

Reviews

Faith in Theory and Practice: Essays on Justifying Religious Belief. Edited by ELIZABETH RADCLIFFE and CAROL WHITE. Chicago and LaSalle, Ill.: Open Court, 1993. 235 pages. \$34.95; \$17.95 (paper).

As is well known, the philosophy of religion is an exciting discipline in which to work these days. There is much creative and thought-provoking research being done, perhaps especially so in religious epistemology. This volume—a collection of papers originally presented at a conference at Santa Clara University in 1991—gives further evidence of this creativity and vigor. While some essays address traditional topics, for example, Hume's arguments against miracles, many explore new ground, for example, the role of intellectual virtue in religious knowledge. Indeed, as the editors indicate in the introduction, one of the distinctive features of the volume is the claim that "justification may be quite different from finding support for propositions"; it is, rather, "situated in the quality of the life of the believer and manifested in her understanding of herself and her relationships to others" (p. ix). In short, many essays focus not on religious beliefs but on religious life—not on acts but on persons—and thus implicitly critique the dominant duty-based approach to the justification of belief, in which one is justified in holding some belief only if one has fulfilled certain epistemic obligations or at least not violated certain duties. So this volume in some ways represents a challenge to business as usual in religious epistemology.

The essay which most clearly challenges the status quo is Linda Zagzebski's stimulating "Intellectual Virtue in Religious Epistemology." Zagzebski argues, persuasively in my judgment, that an epistemological theory modeled on virtue theory in ethics is superior to one fashioned after deontological or consequentialist ethics. In other words, a normative model which focuses on intellectual virtues—like insight, care, open-mindedness, perseverance—is more adequate than the usual theories which speak only of epistemic duties, obligations, norms, and the like. For example, she contends that a virtue-based theory has greater richness, does not assume that justification is strictly rule governed, and is better able to integrate beliefs with dispositions, desires, and feelings. One important implication of Zagzebski's proposal is that "the Christian emphasis on living a life in imitation of saintly persons would extend to imitating the cognitive activities of persons of intellectual virtue" (p. 184). Like Aristotle, Zagzebski rightly insists that excellence in knowing is more than fulfilling certain duties; it involves, rather, the development over time of habitual dispositions to think in ways which reflect recognized standards. And the acquisition of such intellectual excellences—as with excellences of character or moral virtues—requires imitation. In sum, though certain unanswered questions remain, Zagzebski outlines a

promising alternative to understanding the nature of justification in general and the justification of religious belief in particular.

This view of religious faith as more than merely a set of beliefs is also evident in essays by Joshua Golding and William Lad Sessions, among others. In "On the Rationality of Being Religious" Golding argues that under certain conditions it is rational to be a religious person. Note that the claim is not the typical one that it is rational (or irrational) to hold a certain belief. Like Zagzebski, Golding focuses on persons rather than propositions. He is more interested in the question of whether it is rational to pursue "a way of life that is oriented towards attaining the best relationship with God" (p. 91). He sets forth five necessary conditions and then argues that each constraint can be met and thus that it is rational to be a "God-oriented person." So also in his illuminating essay "The Certainty of Faith" Sessions carefully distinguishes between various meanings of faith and of certainty, construing faith primarily as "a kind of personal relationship" involving not only intellectual assent but also trust and commitment (pp. 75–76). He also very helpfully differentiates between epistemic certainty and clarity, ultimacy, and psychological certainty as well as between various forms of epistemic certainty, for example, truth-certainty, justification-certainty, belief-certainty, and proposition-certainty. Given these distinctions, Sessions shows precisely how "various kinds of certainty are required, prohibited, and permitted" by faith (p. 75). His analysis is quite helpful, if complex, and shows at a minimum how greater precision is needed in the usual discussions of faith and reason.

Space does not permit extensive treatment of each of the essays (and not all deserve equal treatment), but the pieces by William Alston and Alvin Plantinga—two leading epistemologists and philosophers of religion—require comment. Alston's paper, entitled "The Fulfillment of Promises as Evidence for Religious Belief," constitutes an argument for the claim that the fulfillment of divine promises in one's life "provides one with evidence for the truth of the religious belief system within which the claim of those promises was generated" (p. 7). More exactly, Alston argues that "the fulfillment of alleged divine promises provides *some* evidence for the truth of the Christian scheme" (p. 30). That is, certain empirical evidence, for example, spiritual fulfillment, combined with other considerations, such as arguments for the existence of God and direct experience of God's presence, "make up a larger set of reasons that may suffice to make Christian belief reasonable, or even to provide a conclusive case for its truth" (p. 30). Alston's project, in other words, is a form of apologetics—an attempt to render the Christian faith credible. But while he addresses and rebuts various objections, Alston gives inadequate attention to the challenge of religious diversity. As in his *Perceiving God* (Ithaca, N. Y.: Cornell Univ. Press, 1993), here in this essay he dispenses all too quickly with what in fact is a more difficult objection, namely, the claim that fulfillments of divine promises in different, and in part incompatible, religions destroys the evidential force of such fulfillments. This challenge must be taken more seriously.

Plantinga's essay is also a piece of apologetics—not pro the Christian faith but contra evolutionary naturalism. Indeed, in "An Evolutionary Argument against Naturalism" Plantinga argues with typical humor and sophistication that evolutionary naturalism is self-defeating and hence that it is irrational to accept it,

since, in essence, evolution does not guarantee that our cognitive faculties are reliable but only that our behavior is adaptive. In other words, given contemporary evolutionary theory and given the belief that there is no God, it is improbable that human cognitive abilities produce mostly true beliefs; but we are rational to believe that our cognitive capabilities are reliable and produce mostly true beliefs—indeed, the vast majority of us most of the time assume that our cognitive capabilities are reliable; thus “the conjunction of naturalism with evolutionary theory is self-defeating” (p. 61). In contrast, with theism the common assumption that our cognitive faculties are reliable is easily explained by reference to a God who created us in his image. Plantinga’s argument is insightful as well as incisive. In conjunction with his own work in epistemology—see, for example, *Warrant: The Current Debate* (New York: Oxford Univ. Press, 1993), and *Warrant and Proper Function* (New York: Oxford Univ. Press, 1993)—Plantinga has set his critical sights on the theory of evolution and many of the claims assumed to follow from it. Readers of this journal will find Plantinga’s arguments of special interest. As with Alston, whether or not one agrees with him, Plantinga provides invigorating reading.

Other essays—by Richard Creel, Steven Grover, Francis Beckwith, Jess Hobbs, James Sennett, and Michael Brown—deal with related topics in the philosophy of religion, for example, anecdotal evidence, skepticism, reliabilism, and each in its own way raises a number of the same concerns as the articles surveyed above. All of the chapters in this volume are written by analytic philosophers and display the clarity and rigor for which that tradition of philosophy is justly known. But as a result some of the essays are, as alluded to above, quite technical and will be difficult to understand for those unfamiliar with philosophy. However, since many of the essays compare religious beliefs and their justification with scientific theories and their testing, this volume will be of interest to many readers of this journal. All in all, *Faith in Theory and Practice* is an interesting collection of essays in religious epistemology—a collection which gives evidence of some promising new directions in the field.

STEVEN BOUMA-PREDIGER
Assistant Professor of Religion
Hope College
Holland, MI 49422

The Knight's Move—The Relational Logic of Spirit in Theology and Science.

By JAMES E. LODER and W. JIM NEIDHARDT. Colorado Springs, Colo.:
Helmets and Howard, 1992. 365 pages. \$27.95 (paper).

The Knight's Move is an ambitious, worthwhile, unique, and challenging book for those interested in the interaction of science and Christian theology—but it is hardly casual bedtime reading. In fact, this is a book which one must read again (and perhaps again). This book is co-authored by a theologian and a physicist—James E. Loder, a Princeton Seminary professor and philosopher of Christian education, and the late W. Jim Neidhardt, physics professor at the New Jersey Institute of Technology. Both are strong advocates of interdisciplinary dialogue between theology (actually Reformed theology) and modern science, as a goal of

their book is integration and discovery of interrelationships between science and theology. The source of Loder and Neidhardt's principal argument of this book is the theological thought of Søren Kierkegaard. Loder and Neidhardt find Kierkegaard's qualitative dialectic and Absolute Paradox woven into Niels Bohr's quantum principle of complementarity and into important strands of the work of such diverse thinkers as Piaget, Einstein, T. F. Torrance, and to a lesser extent K. Barth, K. Gödel, W. Heisenberg, W. Pannenberg, M. Polanyi, and I. Prigogine.

The principal purpose of the book as summarized in the Epilogue is an attempt to "engage the contemporary cultural fragmentation between theology and science in such a way as to counteract any assumption that each is a universe of discourse closed off from or radically incommensurate with the other . . . [and lead] us into scanning for commonalities and then to the creation of this study of relationality" (p. 307). Earlier in the book the authors state that their main concern is not a critique of culture but rather "an interdisciplinary search for ways, models, and patterns by which we can approach the inherent order of creation and facilitate some reintegration of the fragmented fields of study in our culture" (p. 7).

Chapter 1 introduces the reader to many of the principal ideas on which the authors build their case. One of these concepts is Kierkegaard's qualitative dialectic, which the authors liken to the knight's move (the title of the book) in the game of chess, a move which they see as representing the unpredictable creative act, the leap of insight, or even a leap of faith. Furthermore, Loder and Neidhardt suggest, a crucial element of the knowing act is human participation and the open-endedness of what can be known. The authors suggest that science is fully human and open-ended, and they conclude that science and theology are not incompatible but rather are complementary. Another principal argument of the book depends on the nature of this complementarity, and the authors exert considerable effort in developing this.

It is clear that Loder and Neidhardt take both modern science and theology very seriously and feel that a mutual integration of science and theology will play a crucial role in solving the kinds of problems which they wish to address.

"Spirit" is an additional significant component for these authors as they continue to assemble a strategy to address the problems of fragmentation and dualism. The authors define spirit to mean "*a quality of relationality*, and it is way to conceptualize the dynamic interactive unity by which two disparate things are held together without loss of their diversity" (p. 10). What Loder and Neidhardt propose is a concept of spirit which can apply to both a human "spirit" as well as the third person of the Trinity, the Holy Spirit. In this way the authors are trying to arrive at a generic model of spirit, which they want to be helpful in their interdisciplinary dialogue. For our authors such a model of spirit is an important element in their epistemology.

Near the end of chapter 1 Loder and Neidhardt summarize the foci of their book by stating, "four themes are interwoven through this text: the theology/science dialogue; the central reality of Christ's nature; the relational nature of that and of all reality; and the positive spiritual quality of relationality as it points toward Christ's nature through this dialogue" (p. 13).

Following the introductory chapter 1, the book is divided into three main parts. Section 1 sets forth the problem which the authors want to address, and the fundamental multifaceted components of their proposed solution are presented in some detail. A principal feature of their approach is a focus on an exploration of the concept of spirit, the topic of chapter 2. The authors begin by suggesting that the Holy Spirit is the neglected member of the Trinity and that we must have an understanding of the Holy Spirit in order to understand the human spirit. The disaster which the authors wish to address is humanity's fall into dualistic thinking and the resulting cultural fragmentation, which they assert is at least partially a result of our misunderstanding of Holy Spirit in addition to our failure to retain the integrity of our human spirit as spirit.

Having set forth the problem, Loder and Neidhardt propose their solution in chapter 3 as modeled by the "strange loop" (the Möbius band—the twisted and two surfaced yet one-sided, continuous, and closed loop); they see this strange loop as representing what they call bipolar self-relational differentiated unity. The authors are dealing with paradox, as illustrated both through the pictures of M. C. Escher and the strange loop, and they see the one-sided Möbius closed loop as a picture representing a number of paradoxical relationships in terms of a bipolar differentiated unity. A crucial example of this is what the authors call the paradoxical powers of self-reference: the "precise formulation of the bipolar-relational nature of any knowing act (recognized by Hofstadter) points, via Polanyi, to an irreducible quality of human existence: namely, conscious human existence is inevitably and irrepressibly self-relational. Moreover, the *relationship*, not either polarity alone, is the vital center of human existence" (p. 42).

Loder and Neidhardt remind the readers that it was Polanyi who first suggested the bipolar relational unity of what he called personal knowledge. They further suggest that the Holy Spirit can be seen as God's self-relatedness.

The bipolar/self-relational personal interconnectedness theme is further applied to what the authors refer to as the core of scientific investigation, namely, observations of the quantum mechanical microscopic world. The authors see the bipolar strange loop description of the universe in terms of quantum theory as further justification for using these ideas as a guide in trying to understand ultimate reality.

The next bipolar relational differentiated unity explored is the human spirit and the Holy Spirit. Here the authors state, "The human spirit then finds its true home only when it is in one accord with the Spirit of Christ. There its integrity is preserved and it becomes a human analogy for how the Holy Spirit, who never loses relational integrity, searches the depths of God so as to disclose to us the wisdom hidden for us in the mind of Christ" (p. 48).

The next chapters (4 and 5) are the first attempts at establishing the interpretative and explanatory power of their epistemology. And these attempts are indeed intriguing. Our authors find crucial linkages between two Danes, the eighteenth-century theologian Søren Kierkegaard and the early twentieth-century atomic and quantum physicist Niels Bohr, and the fifth century Chalcedonian Christological formulation. In each case they see an apparent paradox which fits nicely with the strange loop model of the previous chapter.

Kierkegaard's "Absolute Paradox" addresses human nature in terms of a frame

of reference designed to deal with the dichotomies of despairing human nature and its transformation by what Kierkegaard refers to as the leap of faith. Loder and Neidhardt suggest that Kierkegaard's "qualitative dialectic" (which holds thesis and antithesis in tension) is closely related to Bohr's assertion that an explanation in quantum mechanics may entail more than one equally valid conclusion, and that the observer may actually determine which possible outcome will be observed, having "the effect of putting the irreducible, open-ended, self-reflective consciousness of the knower at the center of the knowing act" (p. 70).

Bohr added the key notion of "complementarity of contradictories," where he was forced to explain certain physical phenomena (like the wave-particle duality of light) by using mutually exclusive and seemingly contradictory approaches, both of which were necessary for a complete explanation. Loder and Neidhardt suggest that Bohr was not only aware Kierkegaard's ideas of Absolute Paradox and the qualitative dialectic but also profoundly influenced by them. The third piece of the puzzle is the Christological complementarity of contradictories in the formulation at Chalcedon, where the Second Person of the Trinity was declared to be both fully human and fully God.

The authors see these three examples as situations where the explanatory power of the strange loop pattern discloses a relational model of spirit in human nature, physical nature, and Divine. I find the linking of the three to be rather compelling, but I am not sure that I fully grasp the relationship of each to the particular representation of the strange loop, and I wonder how crucial this model is to the overall thrust of the book.

In Section 2 the authors apply the principles developed in Section 1 to a number of interdisciplinary areas. This section is consistent with their goal of making a contribution to reintegration of what they see as fragmented fields of study in our culture. In chapter 6 Loder and Neidhardt point out Kierkegaard's use of the qualitative dialectic as he criticized the dualism of Hegel. In the following chapters the authors apply the relational epistemology of Kierkegaard in interaction with the epistemologies of Jean Piaget (relationality in human development), Albert Einstein (relationality in scientific discovery), and Thomas Torrance (relationality in theology—the knowledge of God), all of these discussions under the umbrella of spirit in the context of knowing.

In Section 3, the final section of the book, Loder and Neidhardt apply their basic relational model to a number of areas involved in the structure and dynamics of human experience—to intelligence in the act of discovery, to the relational dynamics of imagination in spiritual experience and conversion, to the relationality inherent in the intensification process, and to the trinitarian foundation of human interaction in communal life.

Following the Epilogue are a Glossary of Key Terms and Concepts and an Appendix: Resources for Theology/Science Dialogue. All are quite helpful.

This book covers enormous ground and is an ambitious and unique interdisciplinary attempt to address a number of concerns for end-of-twentieth-century humans. It is especially noteworthy that the authors try to integrate the theological work of Kierkegaard and the scientific work of both Bohr and Einstein, showing rather effectively how science and theology can be integrated and inform one another in helpful ways. It also is noteworthy that Loder and Neidhardt have

given the thought of Thomas Torrance serious consideration, finding that his theological ideas are consonant with their basic thesis. Whether the formulation of Loder and Neidhardt completely solves the problem of cultural dualism is not clear to me, but nonetheless this is a significant book, one that deserves serious consideration by any who are concerned about the relationship of science and theology and about the fragmentation of our culture, which our modern philosophical thought has brought about.

RICHARD CARLSON
Professor of Physics
University of Redlands
Redlands, CA 92373

The Problem of Consciousness. By COLIN MCGINN. Cambridge, Mass.: Basil Blackwell, 1991. 216 pages. \$19.95 (paper).

Rediscovery of the Mind. By JOHN SEARLE. Cambridge: M.I.T. Press, 1992. 270 pages. \$27.50; \$12.95 (paper).

If there are kinds of entities whose properties cannot be explained unless we infer the existence of a supernatural intelligence, then we will have important evidence for God's existence. Human beings seem to be such entities since at this time there is no adequate naturalistic explanation of thinking and consciousness. Indeed, Aristotle in *De Anima*, a study of the life principle in organisms, and Aquinas in his comment on *De Anima* both argue to the existence of a supernatural intelligence on the basis of their analysis of human thinking.¹

By contrast, most twentieth century Anglo-American philosophers do not take this lack of explanation as having such momentous consequences. Some, like John Searle in *Rediscovery of the Mind*, argue that this lack is temporary. It is just a matter of time before neurobiology gives us a successful explanation of human thinking and consciousness without any reference to nonphysical intelligences. Searle argues for this possibility even while recognizing the irreducibly subjective aspect of consciousness and the causal power of mental states. On the other hand, Colin McGinn in *The Problem of Consciousness*, argues that this lack is necessary given the kind of understanding we humans can achieve. The lack of explanation does not entail the existence of a supernatural being, for his thesis is about the limitations of human knowledge, not the character of consciousness per se. Just as the fact that armadillos cannot solve problems of elementary arithmetic does not entail anything mysterious about arithmetic, so the fact that humans cannot solve the problem of consciousness does not entail anything mysterious about consciousness. Like Searle, McGinn believes that consciousness is caused by some combination of physical and chemical interactions in the brain. We just cannot explain how such interactions issue in all the wonders associated with thinking and consciousness.

Both Searle and McGinn agree that consciousness is not identical with some discrete physical part of a human being. They see it as a system feature. They understand consciousness to be a "causally emergent system feature" which is a

result of the interaction among many physical processes going on in a human being (Searle, pp. 111–12).² Hence, the existence of consciousness in humans does not require any inference to a supernatural being. They both deny that consciousness is a “radically emergent system feature” which has causal powers that are not fully caused by the interactions of the elements of the system. While the interaction of the elements is necessary for the coming to be of the system feature, the interaction is not sufficient to account for such a feature. An intrinsic principal or organization characteristic of that type of whole is responsible for this type of emergent system property. This kind of wholistic property may entail the existence of a supernatural being.

In this article I will argue that some of the beliefs Searle and McGinn use to support the causally emergent characterization of consciousness imply that consciousness is a radically emergent system feature. If I am correct, then the features of consciousness and thinking they point to have consequences for the existence of a supernatural being. Before I can criticize them in this way I must consider their rejection of the very possibility of radically emergent system features.

IN-PRINCIPLE ARGUMENTS AGAINST RADICALLY EMERGENT FEATURES

Searle argues against the possibility of any radically emergent system features, which he calls emergent2 properties. His argument can be expressed as follows:³

1. If there are emergent2 properties, then the principle of transitivity of causation is false.
2. The principle of transitivity is true.
3. There are no emergent2 properties.

I understand the principle of transitivity of causation to be the following claim: If *A*, *B*, and *C* are entities or events and if *A* causes *B* and *B* causes *C*, then *A* causes *C*. There are many causal relations that confirm the truth of this principle. Whenever *A* is an agent stimulating some other instrumental agent to act, the principle is true. For example, if I hit the cue ball and my hitting causes the movement of the cue ball which causes the movement of the eight ball, then my hitting the cue ball caused the movement of the eight ball.

The case for the truth of premise 1 also seems strong. Assume we have an entity composed of a number of elements. Assume also that we know the causal series that results in interaction *C*, which is the cause of an emergent feature *E* of the entity in question. *C* is caused by *B*, which is caused by *A*. Emergent feature *E* is causally responsible for feature *D*. According to the principle of transitivity, *A* is the cause of *D*. But such is not the case if *E* is an emergent2 property, since the causal power of an emergent2 property is not a function of the interrelation of the parts. Hence, if there are emergent2 properties, then the transitivity principle is false.

I will argue that premise 1 is false by showing how it is possible that an emergent2 feature exists without the falsity of the transitivity principle. In the reasoning given above it is assumed that *E*, the emergent feature, is caused by the interaction *C*. But that assumption cannot be the case if *E* is an emergent2 feature. The interaction *C* may well be a necessary condition for *E*, but

it is not sufficient, because the new causal power in *E* is not caused by *C*. There must be something additional to *C* and its antecedents that causes this new causal power.

My argument is based on the following principle: The effect of any given cause cannot exceed the causal power of that effect's cause.⁴ For example, a pot of water on the stove cannot be heated to a temperature above the temperature of the flame that is heating it. This is analogous to the second law of thermodynamics. Causal power is likened to energy. If the causal power is not caused by the interaction of the elements, then there must be some causal force involved other than the interaction. It is this other causal force in addition to the interaction that causes the emergent feature.

This is enough to show that the existence of emergent₂ properties does not entail the falsity of the transitivity principle. If "cause" simply signifies a necessary condition, then it is true that the necessary conditions of the interactions antecedent to the coming-to-be of the emergent feature are also necessary for the emergent feature and the emergent feature's causal power. However, if "cause" signifies a necessary and sufficient condition, then the interactions antecedent to the coming-to-be of the emergent feature are not the cause of that feature. If we include this other causal force in addition to the interactions, then we have necessary and sufficient conditions. In that case, the cause will be both the cause of the emergent feature and the cause of the emergent feature's causal power. I have argued that it is possible for emergent₂ properties to be real without entailing the violation of the transitivity principle.

Although McGinn does not use the language of emergent₁ and emergent₂ properties, he believes that it is logically impossible to understand emergent₂ properties. It may seem odd that McGinn takes this position given that he argues that we cannot explain what it is about the brain that gives us consciousness. His position is consistent because he distinguishes between two kinds of naturalism: effective naturalism, which he denies, and existential naturalism, which he affirms (p. 87). Effective naturalism is the thesis that we should be able to actually specify naturalistic necessary and sufficient conditions for the phenomena in question. Existential naturalism is the claim that every phenomenon that occurs in the system under consideration is caused by the interaction of physical and chemical processes. McGinn believes that it is wrong for us to think that the human powers of theory construction are capable of comprehending everything there is. So effective naturalism is false. While the falsity of effective naturalism does not entail the falsity of existential naturalism, McGinn has little reason to believe that existential naturalism is true. He believes that the truth of existential naturalism is a condition for the intelligibility of any natural phenomenon.⁵ But he gives no support for this belief. He even admits that his belief in existential naturalism is an "article of metaphysical faith" (p. 87).⁶ In short, he gives no reason to reject the possibility of emergent₂ properties. I now turn to considering the implications of Searle's account of consciousness.

SEARLE'S IMPLICIT AFFIRMATION OF EMERGENT₂ PROPERTIES

Searle accepts the irreducibility of the subjective aspect of consciousness. He believes that a conscious organism's experience of pain has a point of view. The

experience is from the viewpoint of the experiencing organism. There is no way that we humans can know what it is like for a cat as a cat to experience pain (Nagel 1979, 165–80). There is good reason to think that there is a reality to such a subjective state, for we know that it is something like what it is to experience pain as a human. Thomas Nagel has argued that if an alien race came to earth they could not know what it is like for a human to be a human (Nagel 1979, 170). But that epistemic limitation would not entail that there is no reality to our subjective conscious experience.

While some have taken this irreducibility to be an emergent² feature of consciousness, Searle does not. He is a causal reductionist who believes that the causal powers of consciousness are entirely explicable in terms of the causal powers of brain processes. He does not claim that every feature of consciousness is identical to some physical property of matter. For example, some objects are solid, and this has causal consequences. Solid objects are impenetrable by other objects; they are resistant to pressure. But these causal powers can be causally explained by the causal powers of vibratory movements of molecules in lattice structures. Searle notes that in the history of science successful causal reductions lead to ontological reductions in which objects of certain types can be shown to consist of objects of other types. In such cases we simply redefine the expression that denotes the reduced phenomena in such a way that the phenomena can now be identified with their causes. For example, colors were once defined in terms of the subjective experience of color perceivers. Today physicists can define “red” as photon emissions of 600 nanometers; instances of the color red simply are identified with photon emissions of 600 nanometers. The subjective experience of red has been cut off from the “real” color. The general principle seems to be that if we achieve a successful causal reduction of a property, then when we siphon off the subjective element of such a property we get an ontological reduction. With consciousness, however, the very thing we are attempting to explain is the subjective element, so we cannot siphon that off and get an ontological reduction. Searle avers that we cannot reduce the subjective, conscious, first-person sensation of pain to the objective, third-person patterns of neuron firings in the thalamus and other regions of a person’s brain. No third-person description will convey the first-person character of pain. But this has no deep ontological consequences for Searle, since this irreducibility is simply a function of our definitional practices (pp. 122–23). The reductionistic practice is not designed to capture first-person features, for these realities are appearances. Searle writes, “Indeed, it is a general feature of such reductions that the phenomenon is defined in terms of the ‘reality’ and not in terms of the ‘appearance.’ But we can’t make that sort of appearance-reality distinction for consciousness because consciousness consists in the appearances themselves. *Where appearance is concerned we cannot make the appearance-reality distinction because the appearance is the reality*” (pp. 121–22; emphasis in original).

So far so good. There just is an irreducibly subjective component of physical reality. There is no reason to classify consciousness as emergent², for although it is not fully explained by the interaction of brain processes, inasmuch as it is an appearance, it does not have causal powers that fail to be explained.

However, Searle assigns consciousness a causal role. Searle argues that consciousness is not a computer program, for no program gives meaning; programs

require only syntactical operations. Consciousness gives meaning. Hence, for Searle, consciousness has causal power.⁷

Now things do not look so good for Searle.⁸ How can the irreducibility of consciousness have no troubling ontological consequences when it has causal power? Searle concedes that consciousness is irreducible due to the irreducibility of the subjective feature of consciousness.⁹ So consciousness is not simply a function of the interactions of the parts of the elements of the brain. This irreducibility was innocuous as long as no causal roles were assigned to consciousness. But when Searle gives consciousness the power to make meanings, he gives it causal power. It seems as if consciousness so described is an emergent² property: it has the causal power of making meanings which have a subjective feature that is not a function of the interaction of the elements that constitute the whole of which it is a feature. Such is my case for Searle's implicit recognition of the reality of an emergent² feature of consciousness.

MCGINN'S IMPLICIT AFFIRMATION OF EMERGENT² PROPERTIES

Like Searle, McGinn affirms the irreducibility of consciousness to physical processes, and he believes that consciousness is a causally emergent feature that is caused by natural processes in the brain. But unlike Searle, McGinn argues that it is logically impossible for humans to understand the nature of consciousness. Intelligences which form all their empirical concepts on the basis of introspection and perception are constitutionally precluded from grasping the nature of consciousness. Human intelligence is such an intelligence.¹⁰ McGinn assumes that human consciousness must be grounded in some naturalistic property *P* that is a property of the human brain. We can never know what that cause is in any determinate way, only that it is wholly naturalistic. We have already seen how little support McGinn has for the claim that this property *P* must be wholly naturalistic. My aim here is to show that certain features he assigns consciousness imply that it is an emergent² property.

First, I will show that McGinn is wrong about the epistemically closed nature of consciousness, for if he is correct about that claim, it may be thought that I cannot infer anything about the character of consciousness from descriptions of its powers. The irreducible character of subjectivity has deeper implications for McGinn than it does for Searle. He uses this feature about subjectivity to mount an argument for the claim that we cannot give a true psychophysical theory of mental phenomena.¹¹ McGinn argues that *P*, the naturalistic property that grounds consciousness, cannot be known by introspection, since no examination of consciousness itself will give us the property of the brain by virtue of which we are conscious. Second, he argues that *P* is not a perceptible property of the brain. He writes,

Basically, I think, it is because the senses are geared to representing a spatial world; they essentially present things in space with spatially defined properties. But it is precisely such properties that seem inherently incapable of resolving the mind-body problem: we cannot link consciousness to the brain in virtue of spatial properties of the brain. There the brain is, an object of perception, laid out in space, containing spatially distributed processes; but consciousness defies explanation in such terms. Consciousness does not

seem made up out of smaller spatial processes; yet perception of the brain seems limited to revealing such processes. (pp. 11–12)

McGinn believes that consciousness is not constituted of smaller spatial processes, because the “what it is like to be” aspect of consciousness is not a reality with spatial properties. This is the irreducibility of subjectivity that Searle affirms. But for McGinn this irreducibility has significant epistemic consequences. However, our inability to perceive *P* is not sufficient to justify epistemic closure, since *P* could be known from an inference in the way that a variety of theoretical objects are known: for example, electrons, quarks, strong and weak forces. McGinn anticipates this objection, for he argues that *P* cannot be gotten at through inference:

4. No phenomenon that fails to be in the data to be explained is a phenomenon that can be explained by inference.
5. Consciousness is a phenomenon that fails to be in the data to be explained.
6. So consciousness is not a phenomenon that can be explained by inference.

Premise 4 seems to be analytically true, for no theory can explain what is not part of the phenomenon to be explained. Of course, one theory can explain another theory, but then that theory is not called a phenomenon. It may seem that McGinn has supported premise 5 with his argument that consciousness is not an observable feature of the brain. But if consciousness is an internal experience, a subjective phenomenon, then consciousness can be part of the data. Recall that McGinn showed only that *P* was not an object of introspection, but consciousness itself is an object of introspection on his account. Hence, McGinn’s case for the epistemic closed nature of consciousness is problematic.

If this criticism of McGinn is correct, then I have no good reason for refraining from drawing inferences about the nature of consciousness. McGinn’s claim that consciousness cannot be explained in virtue of the spatial properties of the brain entails that consciousness and the brain do not have a spatial relationship. I will argue that this concession entails that there is no physical explanation of the relationship between consciousness and the brain. If there is some explanation for this relationship, then there must be a nonphysical explanation of the relationship. Such a consequence means that consciousness is an emergent² property of the brain. The argument can be expressed in this way:¹²

7. If it is possible for there to be a physical explanation of a relationship between *x* and *y*, then *x* and *y* have spatial coordinates.
8. The brain and consciousness do not both have spatial coordinates.
9. So it is not possible for there to be a physical explanation of a relationship between *x* and *y*.

McGinn concedes premise 8. Premise 7 is a conceptual truth. To be physical is to have spatial relations and vice versa. Empirical confirmation of the truth of premise 7 abounds. All explanations of relationships in physics or chemistry are explanations of events or properties or stuffs that have spatial coordinates. Any explanation of a relationship that purports to be physical without spatial coordinates between the things to be explained has stretched the boundaries of the

physical beyond recognition.

McGinn might welcome this conclusion, for he agrees that no explanation can be given of consciousness. But he cannot be so sanguine, for this conclusion precludes only physical explanations. Nonphysical explanations remain a possibility. We saw that McGinn has no argument precluding the possibility of such explanations. Furthermore, the argument above is due not to the limitations of our knowing power but to the different kinds of properties characterizing physical reality and consciousness. The nonspatial character of consciousness conceded by McGinn is the ground for holding that there cannot be a physicalistic explanation of consciousness. That means consciousness cannot be simply a consequence of the interaction of the elements of the human organism. Since consciousness exerts causal power in giving meaning and explaining behavior, McGinn's concession entails that consciousness is an emergent² property.

Of course, this argument has no force for those who do not admit that consciousness cannot be explained by spatial processes in the brain.¹³ For them it is useful to call attention to the ability of human intelligence to know instantiable realities like validity, truth, and equality. These instantiables do not occupy a position in space. If they did, they would no longer be instantiable.¹⁴ Arthur Collins offers a similar argument based on the instantiability of beliefs. He distinguishes two notions of belief: (a) the temporally datable and essentially personal circumstance of someone's believing that *P* and (b) the propositional belief that *P*, in the sense of a belief that can be true or false and can be shared by many (or by none) (Collins 1979, 225–43). As a datable and localized event the first sense of belief could not be shared by others and hence is not the kind of entity that can be true or false. If we can know instantiable realities (and it seems that we do), then we have strong evidence for believing that human thinking is an emergent² feature of the human being.

CONCLUSION AND SUGGESTIONS

Although I have not proven that consciousness or thinking are emergent² properties, I did respond to in-principle arguments against the very possibility of emergent² properties. I have also shown that Searle and McGinn assign properties to consciousness or thinking that imply that thinking and consciousness are emergent² properties.

But what could account for emergent² features if it is not the interaction of the elements of which the whole is composed? The inability to answer this question is what I think drives neurobiologists and philosophers to reject the reality of emergent² phenomena.

Aquinas and Aristotle may be helpful here (see Aristotle [c. 340 B.C.E.] 1984 and Aquinas [c. 1268] 1994). In their view, what emerges is not just properties but a new substance, a new individual with a set of capacities that differ in kind from the capacities of the old substance or substances from which it comes to be.¹⁵ Substances can be physical in the sense that they can be composed of physical parts. In this way, the explanations provided by biochemistry and neurobiology are necessary to understanding human thinking. But as a whole the substance possesses a principle of organization that is not reducible to the

interactions of discrete parts. As an organizing unity the substance has causal power.¹⁶ In this view, the emergent² features are a function of the substance as an organizing unifying whole. For Aquinas and Aristotle this is true for plant and animal substances as well as for humans.¹⁷ The necessity of higher-level biological explanations that do not reduce to lower-level microbiological explanations confirms the explanatory power of substances.

Substances are causal sources, but they are not themselves uncaused. I have argued that humans have an ability to relate to nonspatial realities. If this relationship requires a nonphysical explanation, then the nature of human substances would seem to require a nonphysical explanation as well.¹⁸ The cause of human nature may turn out to be God, for such a cause must be intelligent and lack spatial location.¹⁹ However, much more would have to be done to argue that this agent is the God of Abraham. But if the emergent² aspect of thinking has this consequence, we can understand why McGinn is so insistent that there *must* be a naturalistic way to account for consciousness and why both he and Searle do not want consciousness to be an emergent² property.

MICHAEL J. DEGNAN
Associate Professor of Philosophy
University of St. Thomas
St. Paul, MN 55105

NOTES

Thanks to Gary Atkinson, John Kronen, Martin Gunderson, Thomas Sullivan, Karen Warren, and Henry West for comments and criticisms on earlier drafts of this paper.

1. It is surprising that McGinn (p. 45) confesses ignorance of the ancient and medieval arguments for God's existence based on the nature of thinking. He writes, "I do not know if anyone has ever tried to exploit consciousness to prove the existence of God, along the lines of the traditional Argument from Design, but in this post-Darwinian era it is an argument with more force than the usual one, through lack of an alternative theory. It is indeed difficult to see how consciousness could have arisen spontaneously from insentient matter; it seems to need an injection from outside the physical realm. Only something of the same kind could bring it about to begin with, it might be thought. However, as will become clear, I do not really think we need to resort to God here. . . ."

2. Searle makes the distinctions between causally emergent system features and radically emergent system features which I use in the text. In addition, he (Searle 1984, 21–22) argues that it is possible that a property or feature of a system may be the effect of the interactions of the elements of that system. He takes it as evident, for example, that liquidity, solidity, and transparency are surface features caused by the behavior of elements at the microlevel as well as features of the very systems in question.

3. The basis for this argument is the following passage from Searle 1992, 112: "I cannot think of anything that is emergent², and it seems unlikely that we will be able to find any features that are emergent², because the existence of any such features would seem to violate even the weakest principle of the transitivity of causation."

4. For a defense of this principle see Connell 1995, 132–34.

5. McGinn writes, "Naturalism about consciousness is not merely an option. It is a condition of understanding. It is a condition of existence" (p. 47).

6. McGinn, p. 87. He appears religious in his denial of any non-natural element. He writes, "Resolutely shunning the supernatural, I think it is undeniable that it must be in virtue of some natural property of the brain that organisms are conscious. There just *has* to be some explanation for how brains subvert minds. . . . And we do not want to acknowledge radical emergence of the conscious with respect to the cerebral: that is too much like accepting miracles *de re*" (p. 6).

7. A causal role of consciousness is again asserted in Searle's (1992) argument that syntax is not intrinsic to physics. Searle argues that no computational state is discovered within the physics of

a system. Anything intrinsic to physics is discovered within the physics of a system. Since any characterization of a process as computational is a characterization of a physical system from outside the system, the identification of a state as computational does not identify any intrinsic feature of the physics of the system. Rather, the characterization of a state as computational is an observer-relative characterization. According to Searle, consciousness assigns the characterizations to systems (see pp. 209–12).

8. In Plato's dialogue, *Phaedo*, Simmias puts forth a view about the soul which is very similar to Searle's about consciousness (see *Phaedo* 92a–93b). Simmias holds that the soul is a harmony of the bodily elements. The soul is not identified with any of the bodily elements but is identified with the harmony of the bodily elements. Socrates criticizes Simmias because his account of the soul renders the soul an effect and not a cause of bodily actions. Yet, Socrates avers, the soul does move the body, it fights off bodily temptations, and so it cannot just be a function of the bodily elements. I criticize Searle in a similar vein. If his view of consciousness is as epiphenomenal as it seems given the trivial consequence that it is not reducible to a physical interaction, then how can he attribute to an appearance the power to generate meanings and interpretations of syntax as well?

9. See Searle 1992, 112, where he writes, "On these definitions consciousness is a causally emergent property of the systems." By contrast, on page 122 he writes, "For our present purposes, we can summarize this point by saying that consciousness is not reducible in the way that other phenomena are reducible, not because the pattern of facts in the real world involves anything special, . . ." These claims seem contradictory. Either consciousness is a causally emergent system property and is not irreducible, or it is irreducible and is not a causally emergent system property.

10. McGinn notes that our thinking about mathematics and logic shows that we can form concepts in ways other than those based on introspection and perception. But these disciplines are not empirical ones (see p. 17).

11. See McGinn, p. 9, for another argument, which can be expressed as follows:

- a. If we can know the true psychophysical theory, then we are able to understand how the cat's brain generates consciousness. (pr)
- b. If we are able to understand how the cat's brain generates consciousness, then it is possible for us to know the subjective form of the cat's experiences. (pr)
- c. So if we can know the true psychophysical theory, then it is possible for us to know the subjective form of the cat's experiences. (1,2, HS)
- d. But it is not possible for us to know the subjective form of the cat's experience. (pr)
- e. So we cannot know the true psychophysical theory. (3, 4 MT)

It seems to me that premise 1 is true only if "true psychophysical theory" means the theory that explains any consciousness. But for a theory that just explained human consciousness and did not pretend to explain anything more, premise 1 is false.

It is not clear that the argument succeeds against a general theory of consciousness either, for premise 2 is problematic. That premise claims that understanding some naturalistic process implies something that is for us a logical impossibility, knowing the subjective form of the cat's experience. If premise 2 is read as an entailment or some kind of implication, then it is false, for it is possible that the antecedent be true and the consequent false.

12. This argument borrows a key premise from an argument about the consequences of being aware of instantiable universals presented by Thomas Sullivan and Russell Pannier (forthcoming).

13. David Chalmers (1995, 83) proposes that we take consciousness as an irreducible feature of reality and connect it to the basic physical realities like electrons and protons by a series of bridge laws. These laws would relate experience to elements of physical theory. But he assumes that consciousness is spatial, for the analogy he makes with electromagnetic charge and gravitational fields are analogies with spatial realities. If consciousness is an irreducible nonspatial reality, however, his proposal will not succeed. On the other hand, if consciousness is spatial, his proposal has much to recommend.

14. This point about instantiables can be found in Sullivan and Pannier (forthcoming).

15. A substance is not a modification of another entity, like color is the modification of skin or surfaces. The existence of a substance is not dependent upon another as a property is dependent upon a subject. There are other kinds of dependencies that substance does exhibit. For example, substances do depend upon other conditions like oxygen, food, and water. In the history of philosophy many have taken substance to be absolutely independent of everything. In this view, only God is a substance.

16. David Braine (1992) sets out an argument for the claim that neither humans nor animals are complex aggregates, but are in fact substances, unified wholes with intrinsic principles of action.

17. Eleanor Stump (1994) observes that in Searle's definition of emergent I features there is an important ambiguity in what counts as causal interactions of the parts. On the one hand, such a phrase can simply refer to the interactions of the parts independently of their being parts of a whole. But on the other hand, the interactions may refer to the interactions the parts have in the form of the whole. In the latter sense, the configuration or form of the whole is smuggled in. This last sense shows the causal role of substance as an organizing whole. If emergent I includes simply the first sense of causal relation, Searle may be mistaken to take liquidity as an emergent I property. Water may be an Aristotelian substance and not just an aggregate of oxygen and hydrogen.

18. The existence of the new nature requires this additional cause. The coming-to-be of the individual may be perfectly explained by biology.

19. Nothing in this argument requires abandoning the insights of evolutionary biology. The explanatory power of natural selection remains intact. However, it is affirmed that these evolutionary explanations are not fully adequate explanations of human thinking. If God does cause human nature, it is possible that he chose natural selection as the means whereby his own causality would be displayed. Natural selection could be an instrumental cause for the coming-to-be of human nature. Just as Michelangelo used a chisel as the instrument for his creation, so, too, could natural selection have been God's instrument. See Connell 1995, 164–66, for more on accounting for new kinds of organisms.

REFERENCES

- Aquinas, Saint Thomas. [c. 1268] [1951] 1994. *Commentary on Aristotle's De Anima*. Trans. K. Foster and S. Humphries. New Haven: Yale Univ. Press. Reprint. Notre Dame, Ind.: Dumb Ox Press.
- Aristotle. [c. 340 B.C.E.] 1984. *De Anima*. In *The Complete Works of Aristotle*, ed. Jonathan Barnes, trans. J. A. Smith. Princeton, N.J.: Princeton Univ. Press.
- Braine, David. 1992. *The Human Person: Animal and Spirit*. Notre Dame, Ind.: Univ. of Notre Dame Press.
- Chalmers, David. 1995. "The Puzzle of Conscious Experience." *Scientific American* (December): 83.
- Collins, Arthur. 1979. "Could Our Beliefs Be Representations in Our Brains?" *Journal of Philosophy* 76: 225–43.
- Connell, Richard J. 1995. *Nature's Causes*. New York: Peter Lang.
- McGinn, Colin. 1991. *The Problem of Consciousness*. Cambridge, Mass.: Basil Blackwell.
- Nagel, Thomas. 1979. "What Is It Like to Be a Bat?" In *Mortal Questions*, 165–80. Cambridge: Cambridge Univ. Press.
- Plato. [c. 367 B.C.E.] 1981. *Phaedo* 92a–93b. Trans. G. N. A. Grube. Indianapolis, Ind.: Hackett.
- Searle, John. 1984. *Minds, Brains, and Science*. Cambridge: Harvard Univ. Press.
- _____. 1992. *Rediscovery of the Mind*. Cambridge: M.I.T. Press.
- Stump, Eleanor. 1994. "Non-Cartesian Substance: Dualism and Materialism without Reductionism." Paper presented at the conference of the Society of Christian Philosophers, Univ. of Notre Dame, Fall.
- Sullivan, Thomas, and Russel Pannier. Forthcoming. *Universals and the Immateriality of the Intellect*.