

The Argument from Consciousness

J. P. MORELAND

Section One: The Backdrop for Locating Consciousness in a Naturalist Ontology

Consciousness is among the most mystifying features of the cosmos (see Moreland 2008). During the emergence of the mechanical philosophy in the seventeenth century, Leibniz wrote the following as a challenge to mechanistic materialism:

It must be confessed, moreover, that *perception*, and that which depends on it *are inexplicable by mechanical cause*, that is by figures and motions. And supposing there were a machine so constructed as to think, feel and have perception, we could conceive of it as enlarged and yet preserving the same proportions, so that we might enter it as a mill. And this granted, we should only find on visiting it, pieces which push one against another, but never anything by which to explain a perception. This must be sought for, therefore, in the simple substance and not in the composite or in the machine. (Leibniz 1979, p. 536)

And while different bells and whistles have been added to our conception of matter since Leibniz's time, scientific naturalist explanations for the emergence of consciousness are as inadequate today as they were when Leibniz threw down his gauntlet. As Geoffrey Madell opines, "the emergence of consciousness, then is a mystery, and one to which materialism signally fails to provide an answer" (Madell 1988, p. 141).

Not only are adequate naturalistic explanations for irreducible consciousness hard to come by, there is a widespread suspicion, if not explicit acknowledgement that irreducible consciousness provides evidence for theism. Thus, Crispin Wright notes:

A central dilemma in contemporary metaphysics is to find a place for certain anthropocentric subject-matters—for instance, semantic, moral, and psychological—in a world as conceived by modern naturalism: a stance which inflates the concepts and categories deployed by (finished) physical science into a metaphysics of the kind of thing the real world essentially and exhaustively is. On one horn, if we embrace this naturalism, it seems we are committed either to reductionism: that is, to a construal of the reference of, for example, semantic, moral and psychological vocabulary as somehow being within the physical domain—or to disputing that

the discourses in question involve reference to what is real at all. On the other horn, if we reject this naturalism, then we accept that there is more to the world than can be embraced within a physicalist ontology—and so take on a commitment, it can seem, to a kind of eerie supernaturalism. (Wright 2002, p. 401)

Similarly, William Lyons claims that:

[physicalism] seem[s] to be in tune with the scientific materialism of the twentieth century because it [is] a harmonic of the general theme that all there is in the universe is matter and energy and motion and that humans are a product of the evolution of species just as much as buffaloes and beavers are. Evolution is a seamless garment with no holes wherein souls might be inserted from above. (Lyons 1995, p. lv)

Souls being “inserted from above” is a veiled reference to theism’s explanatory power for consciousness: If “souls” exist, they would have to be “inserted from above”, since natural processes by themselves are “seamless.” Some argue that, while certain features of finite mental entities may be inexplicable on a naturalist worldview, they may be explained by theism, thereby furnishing evidence for God’s existence. For some time, mental entities have been recalcitrant facts for naturalists. Indeed, for philosophers who take the issues and options in philosophy of mind to be significantly influenced by empirical considerations, the proliferation of a wild variety of physicalist specifications of a naturalist treatment of mental phenomena may fairly be taken as a sign that naturalism is in a period of Kuhnian paradigm crisis. The argument from consciousness for God’s existence (hereafter, AC) provides a way of dethroning the naturalist hegemony. Moreover, by giving a more adequate analysis of and explanation for mental entities, it provides a way out of the crisis and, together with other lines of evidence, offers materials for a cumulative case argument for theism.¹

For decades, versions of naturalism have multiplied like rabbits, so before we examine AC and its chief rivals, it is important to clarify two factors that constitute the dialectical background for what follows. First, I shall unpack the ideational structure of a version of naturalism that follows most plausibly from taking it as a worldview that claims explanatory, epistemic superiority to its rivals. Second, I shall lay out the central epistemic conditions relevant to assessing the force of AC vis-à-vis naturalism.

While there will be different nuances given to naturalism by different thinkers, it is still possible to give an accurate characterization of a specific form of philosophical naturalism (hereafter, simply naturalism or scientific naturalism) that is currently enjoying widespread acceptance (cf. Rosenberg 1996; Moreland & Craig 2000). And by clarifying the relationship between a naturalist ontology on the one hand, and its epistemology and creation account on the other, a picture will emerge as to what *ought* to constitute that ontology. This picture will allow us to identify a substantial burden of proof for alternative naturalist ontologies that bloat naturalist metaphysical commitments beyond what is justifiable within the constraints that follow from the other two aspects of a naturalist worldview.

1. Graham Oppy offers a brief critique of AC, especially as formulated by John Locke and Richard Swinburne. See *Arguing about Gods* (Oppy 2006, pp. 382–401). Unfortunately, he rejects all-too-briefly cumulative case arguments (Oppy 2006, pp. 5–6) and, thus, in my view does not give them sufficient consideration.

Naturalism is the view that the spatiotemporal universe of entities postulated by our best current (or ideal) theories in the physical sciences, particularly physics, is all there is. It includes (1) a naturalist epistemic attitude (e.g. a rejection of so-called first philosophy); (2) an etiological account of how all entities whatsoever have come to be, constituted by an event-causal story described in natural scientific terms; and (3) a general ontology in which the only entities allowed are ones that bear a relevant similarity to those thought to characterize a completed form of physics. Whether this ontology should also include *sui generis* emergent properties will occupy our attention shortly.

The ordering of these is important. The epistemic attitude justifies the etiology, which together justify the ontological commitment. Also, naturalism requires coherence among these three areas. David Papineau claims that we should set philosophy within science in that philosophical investigation should be conducted within the framework of our best empirical theories. It follows that "... the task of the philosophers is to bring coherence and order to the set of assumptions we use to explain the empirical world" (Papineau 1993, p.3). Thus, there should be coherence among third-person scientific ways of knowing, a physical, evolutionary account of how our sensory/cognitive processes came to be, and an analysis of those processes themselves. Any entities that are taken to exist should bear a relevant similarity to those characterizing our best (or ideal) physical theories, their coming-to-be should be intelligible in light of the naturalist causal story, and they should be knowable by scientific means.

The naturalist epistemic attitude

Scientism constitutes the core of the naturalist epistemology. Wilfrid Sellars said that "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (Sellars 1963, p.173). Contemporary naturalists embrace either weak or strong scientism. According to the former, nonscientific fields are not worthless nor do they offer no intellectual results, but they are vastly inferior to science in their epistemic standing and do not merit full credence. According to the latter, unqualified cognitive value resides in science and in nothing else. Either way, naturalists are extremely skeptical of any claims about reality that are not justified by scientific methodology in the hard sciences.

For example, that methodology is a third-person one that sanctions only entities capable of exhaustive description from a third-person perspective. Skepticism prevails for entities that require the first-person perspective as their basic mode of epistemic access. For such naturalists, the exhaustive or elevated nature of scientific knowledge entails that either the only explanations that count or the ones with superior, unqualified acceptance are those employed in the hard sciences.² At least two philosophical theses elaborate the naturalistic epistemic and methodological constraints for philosophy. First, there is no such thing as first philosophy; rather, there is continuity between philosophy and natural science. Second, scientific theories that are paradigm cases of epistemic/explanatory success, for example, the atomic theory of matter, evolutionary biology, employ combinatorial modes of explanation. Thus, any process that constitutes the Grand Story and any entity in the naturalist ontology should exhibit an ontological structure analyzable in terms that are isomorphic

2. I am assuming here a realist construal of explanation.

with such modes of explanation. Colin McGinn has defended this idea along with what he takes it to entail, viz., the inability of naturalism to explain genuinely unique emergent properties:

Can we gain any deeper insight into what makes the problem of consciousness run against the grain of our thinking? Are our modes of theorizing about the world of the wrong shape to extend to the nature of mind? I think we can discern a characteristic structure possessed by successful scientific theories, a structure that is unsuitable for explaining consciousness. . . .

Perhaps the most basic aspect of thought is the operation of *combination*. This is the way in which we think of complex entities as resulting from the arrangement of simpler parts. There are three aspects to this basic idea: the atoms we start with, the laws we use to combine them, and the resulting complexes. . . . I think it is clear that this mode of understanding is central. . . . [and] our scientific faculty involves representing the world in this combinatorial style. (McGinn 1999, pp.55–6; cf. pp.54–62, 90, 95)

The naturalist Grand Story

Let us call the naturalist creation account “the Grand Story”: All of reality – space, time, and matter – came from the Big Bang and various heavenly bodies developed as the universe expanded. On at least the Earth, some sort of prebiotic soup scenario explains how living things came into being from nonliving chemicals. And the processes of evolution, understood in either neo-Darwinian or punctuated equilibrium terms, gave rise to all the life forms we see including human beings. Thus, all organisms and their parts exist and are what they are because they contributed to (or at least did not hinder) the struggle for reproductive advantage, more specifically, because they contributed to the tasks of feeding, fighting, fleeing, and reproducing.

The Grand Story has three key features. First, at its core are two theories that result from combinatorial modes of explanation: the atomic theory of matter and evolutionary theory. If we take John Searle to be representative of naturalists here, this means that causal explanations, specifically, bottom-up but not top-down causal explanations, are central to the (alleged) explanatory superiority of the Grand Story (Searle 1994, p.83–93).

Second, it expresses a scientific philosophical monism, according to which everything that exists or happens in the world is susceptible to explanations by natural scientific methods. *Prima facie*, the most consistent way to understand naturalism in this regard is to see it as entailing some version of strong physicalism: everything that exists is fundamentally matter, most likely, elementary “particles” (whether taken as points of potentiality, centers of mass/energy, units of spatially extended stuff/waves, or reduced to/eliminated in favor of fields), organized in various ways according to the laws of nature. By keeping track of these particles and their physical traits, we are keeping track of everything that exists. No nonphysical entities exist, including emergent ones. When naturalists venture away from strong physicalism, however, they still argue that additions to a strong physicalist ontology must be depicted as rooted in, emergent from, and dependent on the physical states and events of the Grand Story.

Third, the Grand Story is constituted by event causality and eschews both irreducible teleology and agent causation (AGC) in which the first relatum of the causal relation is in the category of substance and not event. And the Grand Story is deterministic in two senses:

diachronically, such that the state of the universe at any time t coupled with the laws of nature determine or fix the chances for the state of the universe at subsequent times; and synchronically, such that the features of and changes regarding macroholes are dependent on and determined by microphenomena.

The naturalist ontology

Weak versus strong naturalism

There is a distinction between strong/strict and weak naturalists. Strong naturalists (e.g. Papineau) accept a strict version of physicalism (all individuals, events, states of affairs, properties, relations, and laws are entirely physical) for the natural world, while weak naturalists (e.g. Searle) embrace various emergent entities.

The location problem

A good place to start a more detailed analysis of a naturalist ontology is with what Frank Jackson calls the location problem (Jackson 1998, pp. 1–5). On the basis of the superiority of the naturalist epistemology, naturalists accept a fairly widely accepted physical story about how things came to be (the Grand Story) and what they are. The location problem is the task of locating some entity (e.g. semantic contents) in that story.

For Jackson, the naturalist must either locate a problematic entity in the basic story or eliminate it. Roughly, an entity is located in the basic story just in case it is entailed by that story. Otherwise, the entity must be eliminated. At this point, it is worth recalling that Kim and others have complained that one does not *explain* a phenomenon by labeling it supervenient. Likewise, one might think that someone has not really “located” a puzzling phenomena if all one has done is point out that it necessarily covaries with this or that sort of physical phenomenon. In any case, Jackson provides three examples of location. First, just as density is a different property from mass and volume, it is not an additional feature of reality over and above mass and volume in at least this sense: an account of things in terms of mass and volume implicitly contains, that is, entails the account in terms of density. Second, Jones’s being taller than Smith is not an additional feature of reality besides Jones’s and Smith’s heights because the relational fact is entailed, and in this sense located by the latter.

More importantly, Jackson focuses on the location of macrosolidity. He acknowledges that prior to modern science, there was a widely accepted commonsense notion of macrosolidity, viz., being everywhere dense. However, due to modern science, this notion has been replaced with being impenetrable. Thus, macrosolidity may be located in the basic microstory: given a description of two macro-objects in terms of their atomic parts, lattice structures, and subatomic forces of repulsion, this description entails that one macro-object is impenetrable with respect to the other.

Jackson believes mental properties are troublesome entities for the naturalist to locate, and the naturalist must argue that they globally supervene on the physical. He unpacks this claim with two clarifications. First, he defines a minimal physical duplicate of our world as “a world that (a) is exactly like our world in every physical respect (instantiated property for instantiated property, law for law, relation for relation), and (b) contains nothing else in the sense of nothing more by way of kinds or particulars than it *must* to satisfy (a)”

(Jackson 1998, p. 13). Second, he advocates B*: any world that is a minimal physical duplicate of our world is a psychological duplicate of our world.

The logic of the mereological hierarchy

Jackson correctly grasps the connection between accepting the epistemic superiority of naturalism and deciding between weak and strong naturalism. For Jackson, if naturalism is to have superior explanatory power, this entails strong naturalism. Jackson understands that there are at least three constraints for developing a naturalist ontology and locating entities within it: (a) entities should conform to the naturalist epistemology; (b) entities should conform to the naturalist Grand Story; (c) entities should bear a relevant similarity to those found in chemistry and physics or merely be capable of one-to-one or one-to-many correlation with entities in chemistry or physics or be shown to depend necessarily on entities in chemistry and physics.

Further in the discussion, we will see why these constraints disallow explanations for the existence of emergent properties. Regarding emergent properties, the second disjunct of (c) “solves” the so-called explanatory gap by simply naming the problem and dismissing the need for a naturalist to do any further explanatory work. For many philosophers, including many naturalists, this strategy is inadequate. The second disjunct also suffers from the difficulty of justifying the existence of *sui generis* emergent entities in light of criteria (a) and (b). The third disjunct of (c) suffers from this latter problem and also from difficulties with justifying the claim that emergent entities are “necessitated” by their subvenient physical bases. Defending these claims are central to the desiderata of this chapter. But it may be useful at this stage of reflection to show how (a) and (b) justify the standard mereological hierarchy as the proper naturalist ontology.

Construing the hierarchy in terms of individual entities and properties rather than in terms of concepts or linguistic descriptions, it consists in an ascending level of entities in the category of individual such that for each level above the ground level of elementary microphysics (at which entities have no further physically significant separable parts), wholes at that level are composed of the separable parts at lower levels. Thus, from bottom to top we get microphysical entities, subatomic parts, atoms, molecules, cells, living organisms, and so on. The relationship between individuals at level n and $n + 1$ is the part/whole relation.

Here is a key point about the hierarchy in the category of individual and property (see further discussion): *the “hierarchy” is not really a hierarchy*. There is no ascending anything. Rather, the levels form spatiotemporally wider and wider wholes. So we should think of the “hierarchy” as going out, not up.

Moreover, there are ontological constraints for what sorts of properties a naturalist should include in the hierarchy. As typically presented, the hierarchy entails the causal closure of the basic microphysical level along with the ontological dependence of entities and their activities at supervenient levels on entities and their activities at that basic level. Causal closure and top/down causation are controversial. But acceptance of closure and a rejection of top/down causation are hard for a naturalist to avoid. The basic naturalist argument for causal closure is that if it is rejected, then

you are ipso facto rejecting the in-principle completeability of physics—that is, the possibility of a complete and comprehensive physical theory of all physical phenomena. For you would

be saying that any complete explanatory theory of the physical domain must invoke nonphysical causal agents. . . . It is safe to assume that no serious physicalist could accept such a prospect. (Kim 1998, p. 40)

The completeability of physics is not arbitrary. It follows naturally from the Grand Story, according to which one begins at the Big Bang with a small number of physical entities and explains the origin and behavior of everything else in terms of the laws of physics and new combinations of microphysical entities. The Story itself gives pride of place to microphysical entities and it is bottom/up at its core. The completeability of physics is essential to the explanatory superiority of the Story. The causal closure principle (no physical event has a nonphysical cause) is not arbitrary nor is it an additional postulate naturalists are intellectually free to reject. It follows from the combinatorial mode of causal explanation and the Grand Story's commitment to the sort of micro-macro constitution and determination at the core of the atomic theory of matter, evolutionary biology, and other central theories of how things have come to be. If a naturalist rejects closure, he or she will have to accept *sui generis*, contingent brute facts. In turn, this undermines the claim that a naturalist worldview is superior to rivals because it can explain how all things have come to be.

Besides closure, a related issue for deciding what sorts or properties should populate the hierarchy is the problem of top/down causation. There is severe intellectual pressure from naturalism itself for rejecting top/down causation for genuinely emergent *sui generis* properties. Moreover, the only way to save top/down causation is to reduce it to outside/in causation that occurs with respect to structural wholes at the same level as their parts via causal feedback. I also think that the price to be paid for retaining causal laws in the special sciences is to disallow emergent properties and allow only microphysically based structural properties constituted by microphysical parts, properties, and relationships. If this is right, it follows that an adequate treatment of these desiderata (to preserve "top/down" causation and causal laws in the special sciences) entails that a naturalist ontology constituted by the standard mereological hierarchy can countenance structural wholes in the category of individual and structural supervenient properties in the category of property, but it cannot countenance genuine emergent properties, especially causally active emergent properties. All emergent properties, if such there be, must be epiphenomenal.

An emergent property is a completely unique, new kind of property different from those that characterize its subvenient base. Accordingly, emergent supervenience is the view that the supervenient property is a simple, intrinsically characterizable, novel property different from and not composed of the parts, properties, relations, and events at the subvenient level. We may characterize "novel" as follows:

Property P is a *novel* emergent property of some particular x at level I_n just in case P is an emergent property, x exemplifies P, and there are no determinates P' of the same determinable D as P such that some particular at level $I_{i=1-(n-1)}$ exemplifies P or P'. (Haldane 1996)

A structural property is one that is constituted by the parts, properties, relations, and events at the subvenient level. A structural property is identical to a configurational pattern among the subvenient entities. It is not *sui generis*.

The existence of emergent mental properties presents two problems for naturalism. First, for those who accept a causal criterion of existence, emergent mental properties are

epiphenomenal and, thus, do not exist. One is then faced with a dilemma: accept phenomenal consciousness as emergent and reject causal closure or retain closure and reject phenomenal consciousness on the grounds that it is epiphenomenal. In the subsequent sections, we shall examine versions of naturalism that accept emergent mental properties. These versions must address epiphenomenalism.

Second, it is pretty obvious that mental states are causal factors in our behavior. Knowledge and agency are hard to salvage if this is denied. Indeed, if an analysis of mental states entails epiphenomenalism, this is widely recognized as fodder for a *reductio* against that analysis. Thus, many naturalists think that the only way to save mental causation is to identify it with the physical. Not all naturalists reject top/down causation. But since bottom/up but not top/down causation follows most naturally from (1) the central theories that constitute the Grand Story and (2) the mereological hierarchy with the dependency of lower on higher levels, there is a burden of proof on those naturalists who accept genuine top/down (and not merely outside/in) causation.

Here is one final constraint for a naturalist ontology. If we limit ourselves to macro-properties, an appropriate limitation because consciousness is a macrofeature, then the following principle seems to be *prima facie* justified:

Principle of Naturalist Exemplification (PNE): $(x) Px \rightarrow Ex$

P stands for any property whatever and E stands for the property of being extended. Moreover, x ranges over and only over property instances. Elsewhere I have defended a constituent ontology in which property instances are complex entities, and I shall merely assume this ontology here (see Moreland 2001b). According to this ontology, when some concrete particular e exemplifies a property P , then the-having-of- P -by- e is a property instance that is modally distinct from both P and e . Thus, x is neither identical to P nor e . So understood, property instances are certain sorts of states of affairs and, moreover, if the instantiation of P by e is temporal, then the property instance becomes an event.

Note that P and e are constituents of x . If we focus on paradigm cases that satisfy PNE, it becomes reasonable to hold that the spatial extension of x is grounded in, obtains in virtue of the spatial extension of e . For example, when an apple is red, the-having-of-red-by-the-apple is a property instance spread out through the extended region occupied by the apple. It is in virtue of the apple's extension that the particular instance of red is extended. This may be seen, for example, by noting that it is because the apple has a particular shape that its instance of red has that shape as well.

PNE says that if a property in the naturalist ontology is to be exemplified, then a necessary condition is that both the concrete particular that exemplifies P and the property instance that results have spatial extension.

PNE seems to capture nicely the wide range of properties in macrophysics, chemistry, geology, neuroscience, and so forth. It could be objected that PNE fails because certain entities, for example, some quantum entities or the point particles of Roger Boscovich were unextended and provide counterexamples to PNE. I do not think this objection works. Regarding quantum entities, there are at least eight different empirically equivalent philosophical models of quantum reality and, it is irresponsible to make dogmatic claims about the ontology of the quantum level (Herbert 1987, p. 15–29). And since I have limited PNE to the macrolevel, we may set aside the quantum world for our purposes. Regarding entities such as Boscovich's particles, rather than conclude that they are counterexamples

to PNE, their lack of spatial dimensionality may be taken as a *reductio* against them. Indeed, this is how the history of physics ran. Boscovichian particles fit more easily into a spiritualist ontology (e.g. Berkeley's) than in a straightforward version of materialism, and like action at a distance, they were rejected.

There is a debate about whether individual mental states such as pains and thoughts are extended. I cannot enter that debate here. But on the basis of PNE, if it turns out that mental states are not extended, then PNE banishes them and their constituent properties from a naturalist ontology. In this case, PNE counts against any naturalist ontology that quantifies over emergent mental properties.

It is time to summarize what a naturalist ontology should look like (cf. Moreland 1998b). In the category of individual, if we reject an eliminativist strategy, then all wholes "above" the microphysical level are structural, relational entities constituted by the parts, properties, and relations at the microphysical level. Such wholes stand in a constituent/whole relation to these microphysical entities and are actually wider entities at the basic level. Regarding the category of property, consider the following:

- Emergence₀: new features that can be deduced from base (e.g. fractals)
- Emergence₁: ordinary structural properties (e.g. being water, solidity)
- Emergence_{2a}: *sui generis*, simple, intrinsically characterizable, new kinds of properties relative to base that are also epiphenomenal (e.g. being painful construed epiphenomenally)
- Emergence_{2b}: *sui generis*, simple, intrinsically characterizable, new kinds of properties relative to base with new causal powers construed as passive liabilities (e.g. being painful understood as having top/down causal liabilities)
- Emergence_{2c}: *sui generis*, simple, intrinsically characterizable, new kinds of properties with active power
- Emergence₃: an emergent, suitably unified mental ego with active power

Clearly, Emergence₀ and Emergence₁ fit nicely in the mereological hierarchy and conform to the naturalist epistemology (e.g. combinatorial explanation) and Grand Story. But Emergence_{2a} through Emergence₃ should be disallowed for reasons we have already investigated. It would seem that all a naturalist could do with them is simply to label them as contingent brute facts and assert that they are not a problem for the naturalist. We will look at different attempts to handle some of these sorts of properties in subsequent sections. But we have already examined reasons to be highly suspicious of a naturalist view that accepts one or more of these sorts of properties and also claims that naturalism is explanatorily and epistemically superior to alternative worldviews.

Moreover, there is an increasingly heavy burden of proof on a naturalist ontology as one moves from Emergence_{2a} to Emergence₃. All types of emergence fall prey to previous arguments against emergent entities. Emergence_{2a} requires less justification than stronger forms of emergence because it does not require a rejection of closure. Emergence_{2b} is subject to these arguments and additional difficulties with top/down causation and causal closure. But relative to Emergence_{2c} and Emergence₃, it has the advantage of exhibiting the same sort of causal power – passive liability subject to law – that characterizes causal particulars at the microphysical level.

Emergence_{2c} has all the problems exemplified by Emergence_{2b} and it also suffers from having a unique sort of active causal power different from causal powers of the naturalist ontology besides agent-causal events. Emergence₃ shares difficulties with Emergence_{2c} and

it also suffers from two further facts not easily accommodated in the naturalist ontology if they are taken as irreducible and uneliminable facts about the world: the indexical fact associated with “I” and difficulties with explaining how one can get a sort of primitive, substantial unity in which its various inseparable parts/faculties are internally related to the substantial subject from a mereological aggregate constituted by a structural arrangement of separable parts that stand in external relations to each other and their mereological whole.

Serious metaphysics, simplicity, and emergent properties

Frank Jackson begins his attempt to develop a naturalistic account of the mental by contrasting two very different approaches to metaphysics. The first he calls serious metaphysics. Serious metaphysics is not content to draw up large pluralistic lists of *sui generis* entities. Advocates of serious metaphysics tend to approach the discipline with a prior epistemic commitment of some sort which functions as a criterion of knowledge or justified belief for quantifying over some entity. Thus, naturalist commitment to serious metaphysics usually includes epistemological methodism constituted by the naturalist epistemic attitude. Accordingly, serious metaphysics is primarily *explanatory* and not *descriptive* metaphysics. Thus, advocates seek to account for all entities in terms of a limited number of basic entities and in this way serious metaphysics is inherently reductionistic. For naturalists, these entities will constitute those at the core of the Grand Story: A property/event/object *x* exists iff it is contained within (truth functionally entailed by) the Grand Story.

The second perspective we may call a “shopping-list” approach whose primary goal is a careful description and categorial analysis of reality. Advocates usually employ epistemological particularism, and it is no accident that Roderick Chisholm is the paradigm case of epistemological particularism and shopping-list metaphysics (cf. Chisholm 1989a,b, pp. 162–8).

Jackson correctly claims that the scientific naturalist will prefer serious metaphysics. His naturalist approach to metaphysics expresses a certain form of the principle of simplicity and provides material content for that principle of simplicity most suited for a philosophical naturalist. To see this, let us compare two versions of the principle of simplicity, an epistemic and ontological version, respectively:

Simplicity_E: entities must not be multiplied beyond necessity

Simplicity_O: our ontology/preferred theory about the world should be simple

Of course, there are various ways to state each principle, but these will do for our purposes. Simplicity_E may not be easy to apply (one rival may be simple in one respect and the other in a different respect; one rival may be simpler and the other may be more empirically accurate), but its rationale is fairly straightforward. All things being equal, if a simpler theory does the epistemic job, then the more complicated theory has baggage that serves no important epistemic function. Ontological simplicity is quite different from epistemic simplicity, and some philosophers conflate the two principles. For example, Kim rightly advocates epistemological simplicity for the same reason just mentioned. But he then passes over into ontological simplicity, apparently without noticing the equivocation. After embracing “entities must not be multiplied beyond necessity,” he urges with no justification

or further explanation that “we expect our basic laws to be reasonably simple, and we expect to explain complex phenomena by combining and iteratively applying these simple laws” (Kim 1996, p. 91).

Ontological simplicity does not follow from epistemic simplicity. In fact, it sometimes happens that progress in an area of science entails adopting a more complicated ontology even though both the simpler and more complicated ontologies are epistemically simple. The shift from the simpler ideal gas equation to the more complicated van der Waals equation is a case in point. That said, I believe that the naturalist should adopt both principles of simplicity, and Kim and Jackson give the reason why. Each makes reference to the Grand Story (which, in turn, is justified by the naturalist epistemology), which is inherently reductionistic.

Moreover, if naturalists claim to have epistemic/explanatory superiority over rivals, then their employment of the Grand Story must be done such that entities that cannot be identified with some structural combination of fundamental microphysical entities must be eliminated. Kim and Jackson both understand this, and while Jackson seeks to carry out this way of understanding the location project, Kim has abandoned it in recent months (see Kim 2005, chap. 6; cf. 1998, chap. 4). Still, Kim’s appeal to ontological simplicity ever bit as much as Jackson’s provides a representative naturalist employment of the principle.

And their characterization of it provides a way of transforming the merely formal principle Simplicity_O into a related version with material content. For Kim, we begin with simple, basic laws – and presumably microphysical particulars governed by them – and allow more complex entities into one’s ontology only if they are subject to combinatorial modes of explanation that involve the iterative application of the basic laws. Similarly, Jackson says one should start with the Grand Story and allow entities into one’s ontology only if they are entailed by that ontology (Jackson 1998, pp. 24–7).³ For Jackson, this means accepting only structural entities that are Emergence₀ or Emergence₁. Expressed in terms of the appropriate naturalist material principle of simplicity, we have

Simplicity_{ON}: our ontology or preferred theory about the world should be simple in the sense that it contains the microphysical entities of an ideal physics or entities whose existence can be explained by the naturalist epistemology (e.g. combinatorial modes of explanation) applied to the microphysical entities that constitute the Grand Story.

Simplicity_{ON} would seem to rule out entities that are Emergence₂ or Emergence₃.

A realist view of causation and emergent properties

We have seen reasons for adopting a *prima facie* burden of proof on any naturalist ontology that includes emergent entities. If such entities are accepted, then a naturalist would owe us a causal account of their coming-to-be. In closing this section, it is important to state

3. By “entails” here, Jackson means the ordinary truth-functional connective. Jackson actually thinks physicalism *a priori* entails the psychological and that this is a necessary truth. If physicalism ϕ is true, then, of necessity, the psychological truths ψ follow *a priori*. Jackson employs a version of two-dimensional semantics to defend the claim that instances of $\phi \rightarrow \psi$ are *a priori* necessary. But this is a stronger claim and many naturalists would not follow him in this, so I shall employ the weaker truth-functional version in what follows. I am indebted to Shaun McNaughton for pointing this out to me.

certain constraints on such an account. In the sections to follow, we shall look at naturalist views that seek to conform to or disregard these constraints. But these constraints seem *prima facie* justified because they follow naturally from the naturalist epistemology, Grand Story, and other aspects of the naturalist ontology.

Regarding emergent properties, although some demur, at least five reasons have been proffered for the claim that causal explanations in the natural sciences exhibit a kind of causal necessity, that on a typical realist construal of natural science, physical causal explanations must show – usually by citing a mechanism – why an effect must follow given the relevant causal conditions:

- (1) Causal necessitation unpacks the deepest, core realist notion of causation, namely, causal production: a cause “brings about” or “produces” its effect.
- (2) Causal necessitation fits the paradigm cases of causal explanation central to the core theories that constitute a naturalist worldview and in terms of which it is purported to have explanatory superiority to rival worldviews.
- (3) Causal necessitation provides a way of distinguishing accidental generalizations or coincidences from true causal laws or sequences.
- (4) Causal necessitation grounds the derivation of counterfactuals.
- (5) Causal necessitation clarifies the direction of causality.

Three points of clarification are in order. First, minimally, the sort of modality involved may be taken as physical necessity, a form of necessity that runs throughout possible worlds relevantly physically similar to our actual world (e.g. in having the same physical particulars, properties, relations, and/or laws). Second, strong conceivability is the test that is used to judge causal necessitation (given the lattice structures and so forth of two macro-objects impenetrable with respect to each other, it is strongly inconceivable that one could penetrate the other).

Finally, Principles (3)–(5) have sometimes been offered as additions to a covering law form of explanation to provide an adequate natural scientific causal explanation. Strictly speaking, a covering law “explanation” is just a description of what needs to be explained and not an explanation. But by adding a causal model that underwrites it and that exhibits causal necessitation, the total package provides explanations for both what and why the phenomena are as they are. For brevity’s sake, I will talk as if a covering law explanation is an explanation, but when I speak of a covering law explanation I mean to include in it an underwriting causal model.

In this section, we have examined the limitations on a naturalist ontology that follow from naturalism itself taken as a worldview epistemically/explanatorially superior to its rivals. Let N stand for the truth of naturalism. In the terms of epistemic appraisal proffered by Chisholm, it seems that – (N & Emergence_{2a}) is at least *epistemically in the clear* where a proposition is *epistemically in the clear* provided only that subject S is not more justified in withholding that proposition than in believing it. Alternatively, it is at least *reasonable to disbelieve* (N & Emergence_{2a}) (S is not more justified in withholding that proposition than in disbelieving it) (Chisholm 1977, p. 16).

However, there are additional limits for a naturalist ontology when a plausible rival worldview is brought into the picture. As Timothy O’Connor points out, emergent properties, especially mental properties, must be shown to arise by way of causal necessitation from a microphysical base if we are to “render emergent phenomena naturalistically

explicable” (O’Connor 2000, p. 112). Among his reasons is the idea that if the link between microbase and emergent properties is a contingent one, then the only explanation for the existence and constancy of the link is a theist explanation (O’Connor 2000, pp. 70–1, n. 8). O’Connor claim seems to me to be correct, and to probe this matter further, we turn to an examination of the theistic argument for God’s existence from consciousness.

Section Two: The AC

In this section, I shall clarify and defend this AC by describing three issues in scientific theory acceptance relevant to assessing AC’s force, presenting three forms of AC and offering a brief defense of its premises. I hope to show that an important factor in theory acceptance – scientific or otherwise – is whether or not a specific theory has a rival. If not, then certain epistemic activities, for example, labeling some phenomenon as basic for which only a description and not an explanation is needed, may be quite adequate not to impede the theory in question. But the adequacy of those same activities can change dramatically if there is a sufficient rival. Section one presented reasons for denying emergent mental properties/events that follow solely from naturalism. In this section, we shall discover additional reasons for naturalists to eschew emergent mental entities that follow because of the presence of AC. The combined force of sections one and two place a severe (and increasing) burden of proof on any naturalist who seeks to reconcile the existence of emergent mental entities (from Emergence_{2a} to Emergence₃) with naturalism.

Three issues in scientific theory acceptance

Basicity

While theism and naturalism are broad worldviews and not scientific theories, three issues that inform the adjudication between rival scientific theories are relevant to AC. The first issue involves deciding whether it is appropriate to take some phenomenon as *basic* such that only a description and not an explanation for it is required, or whether that phenomenon should be understood as something to be explained in terms of *more basic* phenomena. Attempts to explain uniform inertial motion are disallowed in Newtonian mechanics because such motion is basic on this view, but an Aristotelian had to explain why a particular body exhibited uniform inertial motion. Thus, what is basic to one theory may be derivative in another.

Naturalness

Issue two is the *naturalness* of a postulated entity in light of the overall theory of which it is a part. The particulars, properties, and relations postulated should be at home with other entities in the theory and, in this sense, be natural for the theory. Some entity (particular thing, process, property, or relation) *e* is natural for a theory *T* just in case either *e* is a central, core entity of *T* or *e* bears a relevant similarity to central, core entities in *e*’s category within *T*. If *e* is in a category such as substance, force, property, event, relation, or cause, *e* should bear a relevant similarity to other entities of *T* in that category. This is a formal definition and the material content given to it will depend on the theory in question. In

section one, I argued that the basic entities constitutive of the Grand Story provide the material content for naturalism.

Moreover, given rivals R and S, the postulation of e in R is *ad hoc* and question-begging against advocates of S if e bears a relevant similarity to the appropriate entities in S, and in this sense is “at home” in S, but fails to bear this similarity to the appropriate entities in R. The notion of “being *ad hoc*” is difficult to specify precisely. It is usually characterized as an inappropriate adjustment of a theory whose sole epistemic justification is to save the theory from falsification. Such an adjustment involves adding a new supposition to a theory not already implied by its other features. In the context of evaluating rivals R and S, the principle just mentioned provides a sufficient condition for the postulation of e to be *ad hoc* and question-begging.

Naturalness provides a criterion for advocates of a theory to claim that their rivals have begged the question or adjusted their theory in an inappropriate, *ad hoc* way. Naturalness is also useful for deciding the merits of accepting R, which depicts phenomenon e as basic, versus embracing S, which takes e to be explainable in more basic terms. If e is natural in S but not in R, it will be difficult for advocates of R to justify the bald assertion that e is basic in R and that all proponents of R need to do is describe e and correlate it with other phenomena in R as opposed to explaining e. Such a claim by advocates of R will be even more problematic if S provides an explanation for e.

Epistemic values

Issue three involves *epistemic values*, normative properties which confer some degree of justification on a theory possessing them. Examples are theories should be simple, descriptively accurate, predicatively successful, fruitful for guiding new research, capable of solving their internal and external conceptual problems, and use certain types of explanations or follow certain methodological rules and not others (e.g. “appeal to efficient and not final causes”). Studies in scientific theory assessment have made it clear that two rivals may solve a problem differently depending on the way each theory depicts the phenomenon to be solved.

It is possible for two rivals to rank the relative merits of epistemic values in different ways or even give the same virtue a different meaning or application. Rivals can differ radically about the nature, application, and relative importance of a particular epistemic value. Thus, in arguing against B, it may be inappropriate for advocates of A to cite its superior comportment with an epistemic value when B’s proponents do not weigh that value as heavily as they do a different one they take to be more central to B. For example, given rivals A and B, if A is simpler than B but B is more descriptively accurate than A, then it may be inappropriate – indeed, question-begging – for advocates of A to cite A’s simplicity as grounds for judging it superior to B. I am not suggesting that rivals are incommensurable. In fact, I believe that seldom, if ever, is this the case. Only on an issue-by-issue basis can one appropriately make judgments about the epistemic impact of the conflict of disparate epistemic values.

The AC

The deductive form of the argument

Theists (e.g. Robert Adams (1992, pp. 225–40) and Richard Swinburne (1979, chap. 9; 1986, p. 183–96)) have advanced a different theistic argument from consciousness. The argument

may be construed as an inference to the best explanation, a Bayesian-style argument, or a straightforward deductive argument in which its premises are alleged to be more reasonable than their denials. Setting the inductive forms aside, AC becomes the following:

- (1) Mental events are genuine nonphysical mental entities that exist.
- (2) Specific mental and physical event types are regularly correlated.
- (3) There is an explanation for these correlations.
- (4) Personal explanation is different from natural scientific explanation.
- (5) The explanation for these correlations is either a personal or natural scientific explanation.
- (6) The explanation is not a natural scientific one.
- (7) Therefore, the explanation is a personal one.
- (8) If the explanation is personal, then it is theistic.
- (9) Therefore, the explanation is theistic.

Overview of deductive premises

In my view, Premises (3) and (6) are the most crucial ones for the success of AC, since they are the premises most likely to come under naturalist attack. Let us set them aside for the moment. We are assuming the truth of Premises (1) and (2). All the naturalist rivals of AC we are considering agree with them.

The main justification for Premise (4) is the difference between libertarian and event causal theories of agency. J. L. Mackie rejected (4), claiming that personal explanation is simply a subclass of event causal explanation. Moreover, divine action in Swinburne's account of personal explanation involves the direct fulfillment of an intention on the part of God. But, argued Mackie, since human action is a type of efficient event causality between the relevant prior mental state, for example, an intending, and a fulfillment which runs through and depends on a number of intermediate events which are part of a complex physical mechanism, there is a disanalogy between human intentional acts in which intentions are fulfilled indirectly and those of a god in which, supposedly, intentions are directly fulfilled. On Mackie's view, this disanalogy makes alleged divine action and the relevant sort of personal explanation mysterious and antecedently improbable. Thus, (4) is false and, even if it is true, it makes theistic personal explanation less, not more probable.

Is Mackie's argument successful against (4)? I do not think so. For one thing, *pace* Mackie, it is not at all clear that libertarian agency and the associated form of personal explanation are not to be preferred as accounts of human action to event-causal accounts. Obviously, we cannot delve into this issue here, but if libertarian agency is correct, then Mackie is wrong in his claim that (4) is false.

Secondly, a defense of (4) may only require a *concept* of libertarian agency and personal explanation, even if we grant an event-causal theory of action for human acts. If we have such a clear conception, then even if human acts do not fall under it, under the right circumstances, it could be argued that a form of explanation clearly available to us is now to be employed. What those circumstances are and whether they obtain are more centrally related to Premises (3) and (6) of AC and not (4). But since Mackie criticized (4) on the grounds that if true it would make theistic explanation antecedently improbable, I want briefly to say something about what could justify the claim that a personal explanation of the libertarian sort should actually be used.

Many have tried to state necessary and sufficient conditions for personal action in event-causal terms with John Bishop's account being the most sophisticated to date. But Bishop admits that our concept of agency is different from and irreducible to event causality and is, in fact, libertarian (Bishop 1989, pp. 58, 72, 69, 95–6, 103–4, 110–1, 126–7, 140–1, 144). For Bishop, the pervasiveness and power of the libertarian conception of agency places the burden of proof on the defender of a causal theory of action. Bishop claims that his own causal theory works only for worlds relevantly similar to ours in being naturalistic worlds. He does not offer an analysis of action true across all possible worlds because he admits that our concept of action is libertarian and there are worlds in which it is satisfied. His justification of this minimal task is a prior assumption of naturalism, but such an assumption is clearly question-begging against AC. So if we have a clear, powerful, and *prima facie* justified libertarian conception of agency, Mackie's point about the mysteriousness and antecedent improbability of anything answering to this concept is seriously overstated.

Granting the nonphysicality of mental states, a causal theory of personal action will boil down to the claim that person P does some act e (raising one's hand to vote) if and only if some event b (the hand going up), which instantiates the type of state intrinsic to e-ing is caused by the appropriate mental state in the appropriate way. Note carefully that, regardless of the details of such an account, it will amount to nothing more than a causal correlation between certain physical states and the relevant mental events. According to Premises (2) and (3) of AC, these correlations need and have an explanation. A causal theory of action will not do for the origin, regularity, and precise nature of these correlations, since these are what constitute a causal theory of action in the first place. If a causal theory of action presupposes mental states, then it will be impotent to explain the existence, regularity, and precise nature of those mental states themselves unless, of course, a divine causal theory of action is used. If this is so, and if we possess a clear concept of libertarian agency and personal explanation, then there is no good reason why a theist cannot use this type of explanation in this case.

However, a defender of (4) could deny a libertarian view of agency and personal explanation. After all, some Christian theists (e.g. certain Calvinists) employ a causal theory for divine action. One could argue that there is some difference between normal physical event causality and a causal theory of personal action. Minimally, the latter utilizes appropriately related mental states as parts of causal chains. Since (4) simply notes that there is a distinguishable difference between personal and natural scientific explanation, the alternative we are now considering may be all that AC needs to rebut Mackie. Bishop claims that for a naturalist causal theory of action must be combined with a strong physicalist theory of mental states (Bishop 1989, pp. 8, 43, 103). But setting this aside, since we are assuming the reality of mental states, Bishop's physicalist rendition of the causal theory of action simply does not apply here and a suitable statement of the nature and role of mental states in a causal theory could be all that is needed to distinguish personal from natural scientific explanation according to (4).

The presence of personal explanation as a unique argument form means that when it comes to explaining emergent properties such as those constitutive of consciousness, one does not need to acquiesce with Samuel Alexander's dictum that such properties are "to be accepted with the natural piety of the investigator." Thus, it is more than curious to find naturalists jump straightaway from the recognition that mental properties are genuinely emergent and incapable of naturalist explanation to the conclusion that we must take them as brute facts.

There are two sides to (5): is personal explanation different from natural scientific explanation and are there other explanations for the facts mentioned in (1) and (2) besides these two? We have already dealt with the first question in conjunction with (4). Regarding question two, I think it is safe to say that, given the current intellectual climate, a personal theistic or a naturalistic explanation would exhaust at least the live, if not the logical, options. It is true that Thomas Nagel suggested that panpsychism may be necessary to explain the mental (Nagel 1986, pp. 49–53). But it is widely recognized that panpsychism has serious problems in its own right – for example, explaining what an incipient or protomental entity is or how the type of unity that appears to characterize the self could emerge from a mere system of parts standing together in various causal and spatiotemporal relations (cf. Moreland 2008, chap. 6). Moreover, panpsychism is arguably less reasonable than theism on other grounds, although I cannot pursue this point here. Further, it is not clear that panpsychism is an *explanation* of the phenomena in question. As Geoffrey Madell notes,

the sense that the mental and the physical are just inexplicably and gratuitously slapped together is hardly allayed by adopting . . . a panpsychist . . . view of the mind, for [it does not] have an explanation to offer as to why or how mental properties cohere with physical. (Madell 1988, p. 3)

For these and other reasons, I shall not consider panpsychism further except as part of Timothy O'Connor's project.

Premise (7) follows from previous steps in the argument and asserts the adequacy of a personal explanation for the facts expressed in (1) and (2). One may reject (7) (or (5)) on the grounds that personal explanation, theistic or otherwise, does not give us any real understanding of an explanandum, especially one like (1) and (2). Sometimes this objection assumes that an explanation must cite a mechanism before it can count as adequate. My response to this problem centers on the difference between libertarian and event causality and their associated forms of explanation.

Advocates of libertarian agency widely employ the following form of personal explanation (that stands in contrast to a covering law model): A personal explanation (divine or otherwise) of some basic result R brought about intentionally by person P where this bringing about of R is a basic action A will cite the intention I of P that R occur and the basic power B that P exercised to bring about R. P, I, and B provide a personal explanation of R: agent P brought about R by exercising power B in order to realize intention I as an irreducibly teleological goal.

By way of application, the adequacy of a personal explanation does not consist in offering a mechanism, but rather, in correctly citing the relevant person, his intentions, the basic power exercised, and in some cases, offering a description of the relevant action plan. Thus, if we have some model of God and His intentions for creating a world suitable for human persons (from revelation or otherwise), we can make reference to God, His intentions for creating a world with persons with mental states regularly correlated with their environment, and the adequacy of His power to bring about the basic results captured in (1) and (2).

Premise (8) seems fairly uncontroversial. Humean style arguments about the type, size, and number of deities involved could be raised, but these issues would be intramural theistic problems of small comfort to someone committed to naturalism (cf. Martin 1990,

p. 220). And if we take live options only, then it seems fair to limit our alternatives in (5) to theistic or naturalistic. If that is acceptable, at least for the purposes of arguing against Searle and other naturalists like him, then (8) should not be objectionable.

In the terms of epistemic appraisal proffered by Chisholm, it seems that, given AC and what we have seen about the naturalist ontology from section one, $\neg(N \& \text{Emergence}_{2a})$ is at least *beyond reasonable doubt* where a proposition is *beyond reasonable doubt* for a subject S means that S is more justified in believing that proposition than in withholding it. Alternatively, given AC, $(N \& \text{Emergence}_{2a})$ is at least *reasonable to disbelieve* (S is more justified in disbelieving that proposition than in withholding it) (Chisholm 1989a, pp. 10–7). However, it would be premature to conclude that this is the correct epistemic appraisal of $(N \& \text{Emergence}_{2a})$. We still need to look at Premises (3) and (6). Rather than doing so directly, I shall examine them in sections three through five in the context of naturalist attempts that, if successful, would defeat (3) and (6).

Preview of sections three through five

We have seen reasons that follow from the nature of naturalism itself and from the presence of AC as a rival for why a naturalist ought to be a strong physicalist. Unfortunately, strong physicalism is a tough sell, and a growing number of philosophers are dissatisfied with it. Perhaps our conclusion that a naturalist ought to be a strong physicalist is premature. Maybe there are adequate naturalist accounts of the mental. In sections three through five, we will look at representative samples of the major strategies employed to provide such an account. I will conclude that none of these solutions is adequate and that AC is to be preferred. If I am right about this, then the existence of finite mental states provides good evidence that God exists. The best thing for a naturalist to do in this case is to opt for a strong form of physicalism.

Section Three: John Searle and Contingent Correlation

The weakest position for a naturalist who accepts emergent mental properties and events is one according to which all the naturalist must do adequately to explain the mental is to establish contingent correlations between physical and mental states and leave it at that. Searle's view is the most prominent attempt to flesh out this approach.

Searle's position

Contingent correlation

Actually, Searle acknowledges that correlations are not enough and an adequate account should include the transformation of correlations into causal relations by showing that the manipulation of the physical state alters the mental state and by providing a mechanism as to how this works. But for three reasons, I believe it is appropriate to take him as an example of a contingent correlation position. First, he takes such correlations to be adequate to justify the superiority of biological naturalism, so they are sufficient conditions for a naturalist account of consciousness. Second, he claims that a causal explanation

of consciousness may be, in principle, beyond our abilities to obtain; even so, biological naturalism remains standing. Third, he argues against the need for a naturalist to meet some necessitation requirement, according to which one can show that the relevant mental state *must* occur given a certain physical state. That leaves us with correlations for which the establishment of counterfactual covariance would be nice but not necessary for biological naturalism to be adequate.

Biological naturalism

Searle says harsh things about the last 50 years or so of work in the philosophy of mind (Searle 1994, chaps. 1 and 2; cf. Burge 1992).⁴ He says that the field has contained numerous assertions that are obviously false and has cycled neurotically through various positions because of the dominance of strong physicalism as the only option for a naturalist. For these naturalists, if one abandons strong physicalism one has rejected a scientific naturalist approach to the mind/body problem and opened himself up to the intrusion of religious concepts and arguments about the mental such as AC.

Searle offers his analysis of the mind as a naturalistic account because, he says, no one in the modern world can deny “. . . the obvious facts of physics—for example, that the world is made up entirely of physical particles in fields of force . . .” (Searle 1994, p. 28). Naturalism is constituted by the atomic theory of matter and evolutionary biology both of which allow for micro-to-micro or micro-to-macro causal explanations, but not macro-to-micro ones. Dualism in any form is widely rejected because it is correctly considered to be inconsistent with the scientific worldview. People educated in the scientific worldview know how the world works, and the existence of God is no longer a serious candidate for truth. But a commitment to naturalism and a concomitant rejection of dualism have blinded people to the point that they feel compelled to reject what is obvious to experience, namely, the obvious nature of consciousness and intentionality.

Searle’s own solution to the mind/body problem is biological naturalism: consciousness, intentionality, and mental states, in general, are emergent biological states and processes that supervene upon a suitably structured, functioning brain. Brain processes cause mental processes that are not reducible to the former. Consciousness is just an ordinary (i.e. physical) feature of the brain and, as such, is merely an ordinary feature of the natural world. Despite the frequent assertions by a number of philosophers that Searle is a property dualist, he denies the charge and seems puzzled by it. However, in my view, Searle is indeed a property dualist and an epiphenomenalist one at that, although he also denies the latter charge as well. To show this, let us consider the charge of property dualism first. Searle’s characterization of neurophysiological and mental states are exactly those of the property dualist who insists that mental and physical properties are to be characterized in a certain way and that they are two, different types of properties. In light of Searle’s descriptions of the mental and physical, it is obvious why most philosophers charge him with property dualism, and the burden of proof is on him to show why he is not.

4. Since the publication of *The Rediscovery of the Mind*, Searle has restated his views on these topics, but he continues to cite this earlier work as his most thorough treatment on the topic from which he has not deviated. See his works *The Mystery of Consciousness* (1995, p. 194) and *Mind* (2004, p. 2). Thus, I will rely on *The Rediscovery of the Mind* in explicating Searle’s views and supplement them when needed.

Searle's response is twofold. First, he seems to think that a property dualist must accept the entire Cartesian metaphysics. Second, he says that dualists accept a false dichotomistic vocabulary in which something is either physical or mental but cannot be both. So biological naturalism is to be distinguished from property dualism in that the former does not include the entire Cartesian apparatus and it rejects this dichotomistic vocabulary. If this is how Searle distinguishes biological naturalism from property dualism, his response is inadequate. For one thing, it is absurd to claim that one must accept the entire Cartesian metaphysics to be a property dualist. Thomas Aquinas was a certain sort of property (and substance) dualist, but obviously, he did not accept the Cartesian apparatus (cf. Moreland 1995; Moreland & Wallace 1995). Swinburne defends Cartesian property and substance dualism without accepting Descartes' entire metaphysical scheme (Swinburne 1986, chap. 8). Moreover, Searle's own view has a dichotomistic vocabulary in which he distinguishes normal physical (e.g. neurophysiological) properties from emergent biological "physical" (i.e. mental) properties. So he has simply replaced one dualism with another one.

But perhaps there is a different and deeper distinction between (at least) Cartesian property dualism and biological naturalism for Searle. For the property dualist, mental and physical properties are so different that it is inconceivable that one could emerge from the other by natural processes. However, for the biological naturalist, biological physical properties are normal physical properties in this sense: they are like solidity, liquidity, or the properties of digestion or other higher-level properties that can emerge by means of natural processes. I do not wish to comment further on this claim here except to say that Searle's employment of it to distinguish biological naturalism from property dualism amounts to nothing more than a mere assertion combined with a few undeveloped examples (e.g. liquidity) that are supposed to be good analogies to emergent mental states. But this assertion is simply question-begging in light of AC and, as I will show later, it amounts to an abandonment of naturalism. At the very least, one should stop and ask why, if Searle's solution to the mind/body problem is at once obvious and not at all problematic for naturalists, a field of philosophy dominated by naturalists for 50 years has missed this obvious solution?

Searle's three reasons why biological naturalism is not a threat to naturalism

Why are there no deep metaphysical implications that follow from Searle's biological naturalism? Why is it that biological naturalism does not represent a rejection of scientific naturalism which, in turn, opens the door for religious concepts about and arguments from the mental? Searle's answer to this question is developed in three steps. First, he cites several examples of emergence (e.g. liquidity) that he takes to be unproblematic for a naturalist and argues by analogy that the emergent properties of consciousness are likewise unproblematic.

Step two is a formulation of two reasons why, appearances to the contrary notwithstanding, consciousness is not a problem for naturalists. First, Searle says that naturalists are troubled by the existence of irreducible mental entities because they are misled into thinking that the following is a coherent question that needs an answer: "How do unconscious bits of matter produce consciousness?" (Searle 1994, p. 55; cf. pp. 32, 56–7). Many "find it difficult, if not impossible to accept the idea that the real world, the world described by physics and chemistry and biology, contains an ineliminably subjective element. How could

such a thing be? How can we possibly get a coherent world picture if the world contains these mysterious conscious entities?" (Searle 1994, p. 95).

For Searle, the question of how matter produces consciousness is simply a question about how the brain works to produce mental states even though individual neurons in the brain are not conscious. This question is easily answered in terms of specific, though largely unknown, neurobiological features of the brain. However, Searle thinks that many are misled into thinking this question is about something deeper and more puzzling. Setting consciousness aside, in all other cases of entities arranged in a part/whole hierarchy of systems, we can picture or image how emergent features arise because these systems and all their features are objective phenomena. Our problem is that we try to image how consciousness could arise from a system of unconscious bits of matter in the same way, but this is not possible because consciousness itself is not imageable and we cannot get at it through a visual metaphor. Once we give up trying to imagine consciousness, any deep puzzlement about the emergence of consciousness, given naturalism, evaporates, and the only question left is one about how the brain produces mental states.

There is another reason Searle offers as to why the emergence of consciousness has no deep metaphysical significance. In standard cases of reduction, for example, heat and color, an ontological reduction (color is nothing but a wavelength) is based on a causal reduction (color is caused by a wavelength). In these cases, we can distinguish the appearance of heat and color from the reality, place the former in consciousness, leave the latter in the objective world, and go on to define the phenomenon itself in terms of its causes. We can do this because our interests are in the reality and not the appearance. The ontological reduction of heat to its causes leaves the appearance of heat the same. However, when it comes to mental states such as pain, even though an ontological reduction cannot be found, there is a similar causal pattern; for example, pain is caused by such and such brain states.

So why do we regard heat as ontologically reducible but not pain? In the case of heat, we are interested in the physical causes and not the subjective appearances, but with pain it is the subjective appearance itself that interests us. If we wanted to, we could reduce pain to such and such physical processes and go on to talk about pain appearances analogous to the heat case. However, in the case of consciousness, the reality is the appearance. Since the point of reductions is to distinguish and separate reality from appearance in order to focus on underlying causes by definitionally identifying the reality with those causes, the point of a reduction for consciousness is missing, since it is the appearance itself that is the reality of interest. Therefore, the irreducibility of consciousness has no deep metaphysical consequences and is simply a result of the pattern of reduction that expresses our pragmatic interests.

In step three, Searle claims that an adequate scientific explanation of mental emergence is a set of very detailed, even lawlike correlations between specific mental and physical states.

Critique

Searle versus Nagel on causal necessitation

Searle rejects an argument by Thomas Nagel which denies that mere correlations amount to a scientific explanation. In terms of AC, Nagel would accept Premise (6) (the explanation is not a natural scientific one) and deny that Searle's correlations count as scientific

explanations. Searle rejects (6) and believes such correlations count as adequate scientific explanations. Nagel claims that in other cases of emergence such as liquidity, a scientific explanation does not just tell us what happens, it explains why liquidity *must* emerge when a collection of water molecules gather under certain circumstances. In this case, scientific explanation offers physical causal necessity: given certain states of affairs, it is causally necessary that liquidity emerge and it is inconceivable that it not supervene. But, argues Nagel, no such necessity and no answer to a why question is given by a mere correlation between mental states and physical states in the brain.

Searle's response to Nagel is threefold. First, he says that some explanations in science do not exhibit the type of causal necessity Nagel requires; for example, the inverse square law is an account of gravity that does not show why bodies have to have gravitational attraction. This response is question-begging against Nagel because the inverse square law is merely a description of what happens and not an explanation of why it happens. Interestingly, Newton himself took the inverse square law to be a mere description of how gravity works but explained the nature of gravity itself (due to his views about action at a distance, the nature of spirit, and the mechanical nature of corpuscularian causation by contact) in terms of the activity of the Spirit of God. The point is not that Newton was right, but that he distinguished a description of gravity from an explanation of what it is and his explanation cannot be rebutted by citing the inverse square law. Rather, one needs a better explanatory model of gravity. So Searle's own example actually works against him.

Moreover, even if we grant that mere covering law explanations are, in fact, explanations in some sense, they are clearly different from explanations that offer a model of why things must take place given the model and its mechanisms. Since the AC assumes the correlations and offers an answer to the why question, Searle's solution here is not really a rival explanation but merely a claim that such correlations are basic, brute facts that just need to be listed. In light of what we have already seen, there are at least two further difficulties with Searle's claim.

First, given AC and the nature of theory adjudication among rivals, it is question-begging and *ad hoc* for Searle to assert that these correlations are basic, since the correlations themselves, along with the entities and properties they relate are natural and bear a relevant similarity to other entities, properties, and relations in theism (e.g. God as spirit who can create and causally interact with matter), but are unnatural given the naturalist epistemology, Grand Story, and ontology. As we saw in section one, self-reflective naturalists understand this. Thus, Terence Horgan says that "in any metaphysical framework that deserves labels like 'materialism', 'naturalism', or 'physicalism', supervenience facts must be explainable rather than being *sui generis*" (Horgan 1993, pp. 313–4). And D. M. Armstrong's admits:

I suppose that if the principles involved [in analyzing the single all-embracing spatio-temporal system which is reality] were completely different from the current principles of physics, in particular if they involved appeal to mental entities, such as purposes, we might then count the analysis as a falsification of Naturalism. But the Naturalist need make no more concession than this. (Armstrong 1978b, p. 262)

Horgan and Armstrong say this precisely because mental entities, the supervenience relation, or a causal correlation between mental and physical entities simply are not natural

given a consistent naturalist paradigm. Nor can they be located in Jackson's sense in the Grand Story. Their reality constitutes a falsification of naturalism for Horgan and Armstrong and, given AC, they provide evidence for theism. It is question-begging and *ad hoc* simply to adjust naturalism as does Searle, given the presence of AC as a rival explanation.

Naturalists have long criticized Cartesian dualism on the grounds that the causal relation it posits is so bizarre and its relata so disparate that the relation is virtually unintelligible. Many Cartesian dualists are theists and have sought to rebut this claim by appealing to the alleged clarity of divine miraculous activity in the natural world as a counterexample. However, the dialectical situation worsens if the Cartesian is a naturalist for she must now try to render interaction intelligible solely in light of the resources of the Grand Story, and that cannot be done if the interaction relation is taken to be a natural entity at home in the naturalist ontology. It clearly does not bear a relevant similarity to other entities in that ontology. However, this problem is not a function of the ontological category of the relata. Specifically, it is not a problem that arises for naturalism only if the relata are in the category of individual. It applies equally to the category of property. This is why this problem is sometimes called "Descartes' Revenge." Thus, Searle's employment of a supervenience relation – causal or otherwise – between the brain and consciousness is a serious difficulty for his biological naturalism, one he does not adequately address.

Second, Swinburne's version of AC points out that a correlation can be either an accidental generalization or a genuine law (which exhibits at least physical necessity), and we distinguish the two in that laws are (but accidental correlations are not) noncircular correlations that fit naturally into theories that (1) are ontologically simple, (2) have broad explanatory power, and (3) fit with background knowledge from other, closely related scientific theories about the world. By "fit," Swinburne means the degree of naturalness of the correlation and entities correlated in light of both the broader theory of which the correlation is a part and background knowledge. Now Searle admits that mental phenomena are absolutely unique compared to all other entities in that they "have a special feature not possessed by other natural phenomena, namely, subjectivity" (Searle 1994, p. 93). Unfortunately, it is precisely this radical uniqueness that makes mental phenomena unnatural for a naturalist worldview and which prevents Searle from distinguishing an accidental correlation from a genuine law of nature regarding mental and physical correlations.

So much, then, for Searle's first response to Nagel. His second response is that the apparent necessity of some scientific causal explanations may just be a function of our finding some explanation so convincing that we cannot conceive of certain phenomena behaving differently. Medievals may have thought modern explanations of the emergence of liquidity mysterious and causally contingent. Similarly, our belief that specific mind/brain correlations are causally contingent may simply be due to our ignorance of the brain.

It is hard to see what is supposed to follow from Searle's point here. Just because one can be mistaken in using conceivability as a test for causal necessity, it does not follow that conceivability is never a good test for it. Only a case-by-case study can, in principle, decide the appropriateness of its employment. Now when it comes to things such as liquidity or solidity, Nagel is right. Precisely because of what we know about matter, we cannot conceive of certain states of affairs obtaining and these properties being absent. That Medievals would not be so convinced is beside the point, since they were ignorant of the relevant atomic theory. If they possessed the correct theory, their intuitions would be as are ours. But when it comes to the mental and physical, they are such different entities, and the

mental is so unnatural given the rest of the naturalist ontology that there is no clearly conceivable necessity about their connection. And this judgment is based, not on what we do not know about the two types of states, but on what we do know.

Moreover, a more detailed correlation in the future will not change the situation one bit. There is no noncircular or non-*ad hoc* way to formulate such a correlation and we will merely be left with a more detailed dictionary of correlations that will leave intact the same type of problem of causal necessity true of less detailed correlations. Our current lack of belief in such a causal necessity is not due to ignorance of more and more details of the very thing that lacks the necessity in the first place. Rather, it is based on a clear understanding of the nature of the mental and physical, an understanding that Searle himself accepts.

This is why it will not do for naturalists to claim that they are not committed to anything ultimately or utterly brute (such as the divine will), just to their being something unexplained at any given time but which can be explained through deeper investigation. No scientific advance in our knowledge of the details of mental/physical correlations will render either the existence of mental entities or their regular correlation with physical ones anything other than utterly brute for the naturalist.

But Searle had another line of defense against Nagel: even if we grant Nagel's point about the lack of causal necessity in the mental/physical case, nothing follows from this. Why? Because in the water and liquidity case, we can picture the relation between the two in such a way that causal necessity is easily a part of that picture. But since consciousness is not picturable, we are not able to imagine the same sort of causal necessity. Yet that does not mean it is not there.

Here Searle simply applies his earlier point that, given naturalism, our puzzlement about the emergence of consciousness from unconscious bits of matter is due to our attempt to picture consciousness. Now it seems to me that this point is just false and egregiously so. I, for one, have no temptation to try to picture consciousness. And other naturalists have put their finger on the real difficulty about the emergence of consciousness. Paul Churchland says:

The important point about the standard evolutionary story is that the human species and all of its features are the wholly physical outcome of a purely physical process. . . . If this is the correct account of our origins, then there seems neither need, nor room, to fit any nonphysical substances or properties into our theoretical account of ourselves. We are creatures of matter. And we should learn to live with that fact. (Churchland 1984, p. 21)

Regarding need, I take it he means that everything we need in order to explain the origin and workings of human beings can be supplied by physicalist causal explanations. Regarding room, entities do not come into existence *ex nihilo* nor do radically different kinds of entities emerge from purely physical components placed in some sort of complex arrangement. What comes from the physical by means of physical processes will also be physical.

Searle is simply wrong about the problem being the imageability of consciousness. The problem here for naturalism is ontological, not epistemological.

Searle versus McGinn on causal necessitation

Searle has one final line of defense against those who place a necessitation requirement on an adequate naturalist explanation for "emergent" properties. Searle seeks to rebut an

argument by Collin McGinn to the effect that such a necessitation requirement is both essential for and unavailable to a strictly naturalist account of consciousness (Searle 1994, pp. 104–5). We will investigate the details of McGinn’s position in section five, but for present purposes, Searle focuses on the following aspects of McGinn’s position: consciousness is a kind of “stuff” that is known by introspection, things known by introspection are nonspatial, an adequate solution to the mind/body problem requires understanding the “link” between matter and consciousness, but given our noetic limitations, it is in principle beyond our ability to know that link and, therefore, there is no naturalist account of consciousness. Searle rebuts McGinn on the grounds that (1) consciousness is a property not a stuff; (2) introspection is a confused notion and should be abandoned; given (1) and (2), there is no reason to deny that consciousness is spatial; and moreover (3) there is no link between consciousness and the brain anymore than there is a link between liquidity and H₂O.

Setting aside until section five the issue of whether or not Searle has adequately rebutted McGinn’s particular formulation of this argument, the more important point is whether or not Searle has rebutted this form of argument if it is stated in more plausible dualist terms. This is a fair approach to Searle’s rebuttal because he explicitly takes McGinn’s premises to represent broad Cartesian-style commitments (except for McGinn’s claim that the link is *in principle* unknowable) and his own rebuttal to be successful against Cartesian dualism in general. Given this broader context, I believe Searle’s rebuttal fails. Consider Premise (1). I do not know of a single property dualist (Cartesian or otherwise) who would take mental properties to be a sort of stuff that, for example, should be referred to by mass terms. Even with respect to mental substances, a framework of stuff is not usually employed. To be sure, some Cartesian dualists may believe in soul stuff, but most substance dualists, including me, employ a substance/attribute ontology to characterize a mental substance as an individuated mental essence; they do not use a separable part/whole framework or the notion of stuff. So Searle is guilty of arguing against a straw man in (1).

What about (2)? Searle’s argument against introspection is as follows:

- (1)₁ If the standard model is true, then there is a distinction (presumably, not a distinction of reason) between the thing seen and the seeing of it.
- (2)₁ The standard model is true.
- (3)₁ Therefore, there is a distinction between the thing seen and the seeing of it.
- (4)₁ If introspection occurs, then there is no distinction between the thing seen and the seeing of it.
- (5)₁ Therefore, introspection does not occur.

There are at least two problems with this argument and they involve (2)₁ and (4)₁. Let us begin with (4)₁. Searle gives no good reason to accept it and, in fact, there are sufficient reasons to reject it. Let us assume as is standardly granted that in introspection, we have a second-order mental state directed upon a first-order mental state. For example, in introspection the self – whatever it is – is directly aware of a sensation of red or a feeling of pain by directing a second-order mental state onto a first-order one. This is a perfectly intelligible account of introspection and it provides the distinction required to reject (4)₁.

If someone rejects this model of introspection, then one can still rebut Searle’s argument by rejecting (2)₁. That is, one can grant (1)₁ for the sake of argument and deny that it applies to introspection on the grounds that it begs the question. After all, why apply the standard

model to introspective acts? Recall that in section two I claimed that at least a certain range of mental states relevant to introspection are self-presenting properties. And according to a standard characterization of them, a self-presenting property presents to a subject the intentional object of that property (e.g. an apple's surface) and the self-presenting property itself (being-an-appearing-of-red). Such properties present other things to a subject intermediately by means of them, and they present themselves to a subject directly simply in virtue of the fact that he has them. Introspective awareness of being-an-appearing-of-red could be understood as the exemplification of a self-presenting property.

In this case, introspection provides a counterexample to the standard model. And while Searle does not mention the self, I see no reason why one cannot be directly aware of oneself. On a certain understanding of intentionality, according to which it is a monadic property, when one is aware of oneself (as opposed to a mental state one has), in direct self-awareness, one simply directs one's intentionality onto oneself and the subject and object of awareness stand in the identity relation to each other. Nothing Searle says comes close to undermining such an understanding of self-awareness.

Searle similarly attacks a spatial metaphor associated with "privileged access" that he alleges to go proxy for introspection: when I spatially enter something, there is a distinction among me, the act of entering, and the thing entered. No such distinction obtains in alleged acts of "private access" and, thus, "private access" should be rejected. The appropriate rebuttal analogously follows the lines of response given to the argument against introspection.

This brings us to (3). As we shall see further in the discussion, liquidity is a bad analogy with conscious properties. Liquidity may be understood as the property of flowing freely, which, in turn, may be characterized in terms of friction, flexibility of bonding angles, degree of spatial compactness, and so forth. In short, liquidity is a structural property and, as such, liquidity constitutively supervenes "upon" a collection of water molecules. There is no causal relation here. Liquidity just is a feature of nonrigid motion constituted by a subvenient base. Thus, it is plausible to deny a "link" between liquidity and a swarm of water molecules. But Searle is clear that conscious properties are simple, *sui generis* emergent properties and, as such, are causally supervenient on the brain. In this case, there is indeed a causal "link" between the brain and consciousness and Searle's analogy employed in (3) is a failure, even in terms of his own views.

I conclude, therefore, that Searle has not succeeded in undermining Nagel: Premise (6) of AC (the explanation is not a natural scientific one) is correct and Searle's correlations are not examples of scientific explanation which count against (6). But what about Premise (3) (there is an explanation for these correlations)? Why is it not reasonable to take mental entities and their regular correlations with physical entities to be utterly brute natural facts for which there is no explanation? The answer is provided by the arguments just mentioned about why Searle's correlations are not really scientific explanations. Mental entities are not natural or at home in the naturalist epistemology, etiology, and ontology. Given theism and AC as a rival explanatory paradigm, and given the fact that mental entities and correlations are natural for theism, it is question-begging and *ad hoc* simply to announce that these entities and correlations are natural entities.

Searle could reply that biological naturalism is not question-begging because we already have reason to believe that naturalism is superior to theism prior to our study of the nature of the mental. The only support Searle gives for this claim, apart from a few sociological musings about what it means to be a modern person, is that it is an obvious fact of physics that the world consists *entirely* of physical particles moving in fields of force. It should be

clear, however, that this claim is itself question-begging and clearly false. When there is a statement in a physics text about the world in its entirety, it is important to note that this is not a statement of physics. It is a philosophical assertion that does not express any obvious fact of physics. Moreover, it is a question-begging assertion by naturalists prior to a consideration of the evidence and arguments for theism, including AC. If Searle denies this, then he should inform advocates of AC of exactly what obvious fact of physics they deny in their employment of the argument.

Most naturalists have seen this and have opted for strong physicalism in order to avoid abandoning naturalism and legitimizing the introduction of religious concepts and explanations into the picture. It may be “neurotic” to deny consciousness, as Searle points out. But it is far from “neurotic” to be driven to do so in terms of a prior commitment to naturalism, and AC makes clear why this is the case.

Mackie on Locke and thinking matter

But perhaps there is a naturalist rejoinder at this point in the form of a *tu quoque* against theists and AC. J. L. Mackie advanced just such an argument (Mackie 1982, pp. 120–1; Williams 1996; cf. Moreland 1998a, 2000, 2001a). According to Mackie, theists such as John Locke admitted that God could superadd consciousness to systems of matter fitly disposed and, therefore, as a result of divine intervention, matter may give rise to consciousness after all. Thus, Locke leaves open the possibility that a mere material being might be conscious given theism. Mackie then asks this question: “But if some material structures could be conscious, how can we know *a priori* that material structures cannot *of themselves give rise to* consciousness?” (Mackie 1982, p. 121). He concludes that this Lockean admission opens the door for the naturalist to assert the emergence of consciousness from fitly disposed matter as a brute fact.

In my view, Mackie’s argument carries no force against AC because a main part of AC consists in the recognition that mental/physical correlations exist, they are not explicable within the constraints of scientific naturalism, and they require a personal theistic explanation if they are to be explained at all. In this sense, the idea that, in one way or another, God could “superadd” thinking or other mental states to matter is required for AC to go through.

However, as I have tried to show, it does not follow from this “Lockean admission” that it is a brute, naturalistic fact that material structures of themselves can give rise to consciousness or that adequate naturalistic explanations can be given for this. Indeed, Locke himself constructed detailed arguments to show that mental states such as thinkings are not within the natural powers of matter nor could they arise from material structures without an original mind to create and attach those mental states to matter (Locke 1959, pp. 313–9). Locke’s view that God could superadd thinking to a material substance just as easily as to a spiritual substance was a conclusion he drew from the omnipotence of God along with the claim that “thinking matter” is not a contradiction and, thus, possible for God to bring about.

I am not defending Locke’s way of arguing that God could superadd thinking to matter. In fact, I do not think it is correct as he formulated it but, clearly, Locke would not have believed that Mackie’s naturalistic conclusion can justifiably be drawn from his own (Locke’s) admission of the possibility of divine omnipotence adding a faculty of thought to a material structure.

Mackie cannot simply assert that material structures have the power to give rise to consciousness and also claim to be operating with a naturalistic depiction of matter. According to David Papineau, matter with emergent mental potentiality is not the sort of matter countenanced by naturalists. This is why when Papineau attempts to characterize the physical in terms of a future ideal physics, he places clear boundaries on the types of changes allowed by naturalism for developments in physical theory. According to Papineau, the naturalist will admit that future physics may change some features of what we believe about matter, but in light of a naturalist commitment and the past few hundred years of development in physics, future physics will not need to be supplemented by psychological or mental categories (Papineau 1993, pp. 29–32).

Given theism, we cannot say *a priori* just what capacities or states God will correlate with specific physical states. But given naturalism, and the commitment to the role of physics in naturalism, along with a view of the physical that is required by physics, we can say that mental potentiality is just not part of matter. Thus, it is question-begging and *ad hoc* against AC for Mackie to adjust naturalism to allow that material structures of themselves can give rise to consciousness.

Consciousness, liquidity, solidity, and digestion

There is one final issue in Searle's defense of biological naturalism that needs to be addressed, viz., his claim that the emergence of consciousness fits a broad pattern of emergence, for example, cases of liquidity, solidity, digestion, and, therefore, since the latter present no problem for naturalism, neither does the former. I offer three responses. First, if we take liquidity or solidity to be the degree of rigidity, flexibility, or viscosity of a collection of particles, then these properties are not good analogies to consciousness because they turn out to be nothing more than group behavior of particles placed in a relatively compressed, stable, ordered structure for solids or a more viscous, less compact arrangement for liquids. So there is no problem about emergence here, since we can easily understand how liquidity and solidity are related to groups of material particles as they are depicted in physical theory.

Second, when we are dealing with genuinely emergent properties that are categorially different from what physical theory takes to characterize subvenient entities, I think that it could be argued that the naturalist has the same difficulty here as with the emergence of consciousness. Recall Searle's point about the pragmatics of reduction: we reduce heat to its causes because we happen to be interested in the objective causes and not the subjective appearances, but in cases of, for example, pain, we are interested in the painful appearance itself, so we do not reduce pain to its causes. In my view, the decision to reduce heat to its causes is not primarily a scientific matter nor is it a matter of our pragmatic interest. I think it has been a function of two things.

First, if we take heat, color, liquidity, or solidity to be identical to the qualia we experience in certain circumstances (e.g. heat is identical to warmth, red is a color not a wavelength, liquidity is wetness), then an ontological puzzle arises analogous to the one about the emergence of mental states: how could warmth emerge in a physical structure as a result of increased atomic agitation? Second, there was a way of avoiding this question in light of a widely held Lockean view of secondary qualities and sense perception. We can locate these secondary qualities in consciousness and identify them as appearances of the real objective phenomena, viz., the objective causes for our experiences of secondary qualities. John Yolton has shown that

during late seventeenth- and early eighteenth- century debates about materialism, immaterialist philosophers (e.g. Ralph Cudworth) regularly argued against the idea that mental entities could emerge from properly structured matter (Yolton 1983, pp. 4–13). A standard rebuttal to this claim was that light and heat were very different from matter but could be generated in material bodies given the right conditions. So mind could likewise emerge. Cudworth and others responded by asserting that light, heat, and other secondary qualities were not in material bodies but were sensations in minds and, thus, the problem does not arise as to how they could arise in a material structure devoid of such qualities prior to the right conditions obtaining. It is clear from this debate at the very beginning of the emergence of modern materialism that one philosophical motive for locating secondary qualities in consciousness was to avoid a straightforward metaphysical problem: *ex nihilo nihil fit*.

If I am right about this, then the ontological puzzle is really the driving force behind what Searle calls normal naturalist cases of emergence. The problem is that these cases are not natural any more than the emergence of consciousness and that is why they were located in consciousness. For example, both secondary qualities such as redness or warmth and painfulness are dissimilar to the properties that constitute an ideal physics. Jaegwon Kim has argued that in Nagel-type reductions, the relevant bridge laws should be taken as biconditionals and not as conditionals because we need materially equivalent correlations between entities (or terms) in the reduced and base theories in order to assert identities between the entities in question (Kim 1996, p. 91). Moreover, says Kim, the identity of reduced and base entities is preferable to mere correlations because the latter raise potentially embarrassing questions as to why such precise correlations arise in the first place.

Kim's point is not confined to mental and physical correlations. All a naturalist can do with them (if we keep these so-called secondary qualities or other categorially distinct emergent qualities in the external world) is to offer a detailed correlation to describe regular relations between physical structures and emergent entities. No amount of knowledge whatever of subvenient entities would take us one inch toward predicting or picturing why these particular entities regularly emerge in such and such circumstances and not others. In discussions of emergence over a century ago, it was precisely their unpredictability from knowledge of subvenient entities that was identified as the hallmark of an emergent property.

In more modern terms, it is the inability to either image or understand why warmth emerges regularly here and not somewhere else, or why it emerges at all given our knowledge of molecular agitation. Note carefully that Searle himself seems to accept picturability as a necessary condition for the acceptance of a claim that one entity emerges from another in the "normal" cases, but picturability is no more available for heat (warmth) emerging from matter than it is for mental states (Searle 1994, pp. 102–3). Nagel's conceivability test applies here just as it does for mental states.

However, even if I am wrong about this, there is a third response that can be given to Searle. There are two features of mental states that make their emergence disanalogous to, say, the properties of digestion. First, mental states are so unique and different from all other entities in the world that it is far more difficult to see how they could emerge from physical states than it is for the so-called normal cases. Second, mental states are quite natural in a theistic world view and have a higher prior probability given theism over against naturalism even if we agree that, say, the emergence of the properties of digestion are equally natural and probable on both world hypotheses.

In my view, these two features of mental states make them more analogous to value properties than to characteristics of digestion. Mackie argued that the supervenience of moral properties would constitute a refutation of naturalism and evidence for theism: “Moral properties constitute so odd a cluster of properties and relations that they are most unlikely to have arisen in the ordinary course of events without an all-powerful god to create them” (Mackie 1982, p. 115; cf. Moreland & Nielsen 1993, chap. 8–10). Presumably, Mackie’s reasons for this claim involve some of the points I have just made earlier: moral properties have the two features that make them natural for theism but unnatural for naturalism. No matter how far future physics advances our understanding of matter, it will not make the emergence of moral properties the least bit more likely, more picturable, or more natural. And the same claim could easily be made for mental properties even if features of digestion are granted equally natural for theism and naturalism.

Searle himself admits that of all the entities in the world, mental states are absolutely unique and radically different from all the others. And as we saw earlier, Armstrong is willing to accept that more ordinary physical or biological properties could emerge when the nervous system reaches a certain level of complexity. But he could not accept the natural emergence of mental states from matter because mental states are of “a quite different nature” from states accepted by naturalists. The jump from physical states to mental states was too far for Armstrong’s naturalism to allow, so he adopted strong physicalism as the only acceptable naturalist solution.

The problem with my third response is that it requires one to weigh the difference between acceptable and unacceptable cases of emergence. But to the degree that mental entities are taken as radically unique from all other physical or biological entities, then to that degree the analogy between the emergence of mental states and other cases of emergence is weakened. And to that degree, the emergence of the mental would be radical as Nagel calls it or unnatural as Adams and Swinburne claim.

After all, naturalists have not spent the last fifty years trying to eliminate or reduce solidity or the properties of digestion like they have mental states. This is because the latter are rightly seen as a threat to naturalism even if the former are not. As B. F. Skinner noted just before his death:

Evolutionary theorists have suggested that ‘conscious intelligence’ is an evolved trait, but they have never shown how a nonphysical variation could arise [in the first place] to be selected by physical contingencies of survival. (Skinner 1990, p. 1207)

Indeed. The constraints on a naturalist ontology discussed in sections one and two place a severe burden of proof on adding emergent mental properties to that ontology, a burden that Searle has singularly failed to meet.

Section Four: Timothy O’Connor and Emergent Necessitation

The vast majority of friends and foes of agent-causal versions of libertarian freedom agree that it is either inconsistent or not plausibly harmonized with a naturalistic view of the world, including a physicalist depiction of particulars taken to populate the naturalist ontology. Thus, naturalist John Bishop claims that:

the idea of a responsible agent, with the 'originative' ability to initiate events in the natural world, does not sit easily with the idea of [an agent as] a natural organism. . . . Our scientific understanding of human behavior seems to be in tension with a presupposition of the ethical stance we adopt toward it. (Bishop 1989, p. 1)

In his excellent and penetrating development of an agent-causal count of freedom, *Persons & Causes*, Timothy O'Connor acknowledges that this is the case: "A great many contemporary philosophers will dismiss [an agent-causal account of freedom] as pointless, since it blatantly contradicts 'the scientific facts'" (O'Connor 2000, p. 108; cf. Moreland 1997; O'Connor 2003). However, O'Connor is actually puzzled by the majority view on this issue, and claims that a robust version of AGC, including his own, may be very plausibly harmonized with the emerging naturalist picture of the world, including a physicalist view of the agent. O'Connor's puzzlement is odd in light of the considerations we noted in sections one and two. In any case, for O'Connor, agent-causal power is an emergent property. To support this claim, O'Connor defends what I shall call the Harmony Thesis: the emergence of agent-causal power may be plausibly located within a widely accepted naturalist ontology, including a physicalist depiction of the agent.

To explain why I think O'Connor has failed to substantiate this claim, I shall describe features of his model and offer three lines of criticism. First, I will expose problems in O'Connor's description of the agent. Second, I will show why a certain model of causation is crucial for O'Connor's project and argue that, given this model, it is not true that consciousness in general, and active power in particular, are emergent properties. Third, I will try to show that certain epistemic features that characterize O'Connor's own case for AGC, if applied consistently, provide adequate grounds for rejecting the Harmony Thesis. Besides problems intrinsic to O'Connor's view, in light of considerations of sections one and two there is a substantial burden of proof – made precise in those sections and shown to be far from arbitrary – that he must meet to be successful. I believe it will become obvious that he fails to meet this burden.

O'Connor is a Christian theist, not a naturalist. Nevertheless, he is concerned to show that AGC, including active power, may be plausibly located in a widely accepted naturalist ontology, and it this claim that I wish to clarify and dispute.

AGC and the emerging naturalist picture of the world (N)

To assess the Harmony Thesis, it is important to get clear on the central features of O'Connor's understanding of AGC and N that are relevant to our present concerns. According to O'Connor, although it may be difficult to do so, AGC may be reconciled with the *Causal Unity of Nature Thesis*, but not with the *Constitution Thesis* (O'Connor 2000, p. 109):

The Causal Unity of Nature Thesis: Macro-level phenomena arise through entirely natural microphysical causal processes and the existence of macro-level phenomena continues to depend causally on microphysical processes.

The Constitution Thesis: All macro-level phenomena are constituted by micro-level phenomena.

AGC

Regarding AGC, O'Connor claims that the core of every free act is an irreducible causal relation between a person and some appropriate internal event that triggers latter elements of the action. O'Connor holds to a realist view of causation, according to which the essence of causality is causal production or the bringing about of an effect. Active power constitutes a special type of causal event that is intrinsically active, that cannot be caused, even by the agent, and that is intrinsically a case of the agent directly causing/controlling his behavior, or at least, the action trigger. Agent causes bring about immediately executive states of intention to act in various ways.

What kind of agent is required for this account? Such an agent must have "rather special properties in her constitution" (O'Connor 2000, p. 49). To elaborate, entities that exhibit event causation are such that the capacity to generate a particular effect is exercised as a matter of course: given the right circumstances, the cluster of properties that ground the capacity directly give rise to the effect. By contrast, having the properties that subserve an agent-causal capacity does not produce the effect; it enables the agent to do so (O'Connor 2000, pp. xiv, 75). Such an agent is a "not wholly moved mover" (O'Connor 2000, p. 67) and an enduring continuant, but not a different kind of substance radically diverse from physical substances (O'Connor 2000, p. 73). Personal agents are biological entities with irreducible emergent properties, where properties are construed as universals that have essentially their dispositional tendencies (O'Connor 2000, p. 73). Sometimes O'Connor uses substance talk to describe the agent (O'Connor 2000, p. 73). However, he also describes the agent as a "complex system regulated by dynamic processes" (O'Connor 2000, p. 95) with a structured capacity, structured by tendency-conferring states of having reasons to act in specific ways (O'Connor 2000, pp. 97–8).

In various places, O'Connor describes the emergent properties essential to AGC. Only entities with more basic attributes can have free will, viz., volition, understanding, practical judgment and the power to believe the act is within one's power (O'Connor 2000, pp. 45–6). Thus, agent causes must possess conscious awareness (O'Connor 2000, p. 122). An agent must be able to represent to himself possible courses of action and have belief/desire sets relevant to each (O'Connor 2000, p. 72). Moreover, given that intentions are action triggers internal to the agent, an agent must be able to cause directly an event internal to the agent (O'Connor 2000, p. 72). In accounting for the role of reasons in AGC, O'Connor claims that an agent directly causes an action-triggering intention the content of which is that an action of a specific sort be performed for certain reasons the agent had at the time. Thus, an agent must have the potentiality to have intrinsic events that exemplify a twofold internal relation of direct reference and of similar content (O'Connor 2000, pp. xiv, 85–6).

Four relevant aspects of naturalism

There are four aspects of N relevant to our discussion. First, O'Connor accepts the mereological hierarchy: physics is the basic level of reality and, in the category of individual, all wholes above the fundamental level are systems constituted by parts at lower levels. On this view, the world is fundamentally event causal in nature (O'Connor 2000, p. 107). This seems to mean two things: (1) all strictly physical entities exhibit event causality; and (2) all macrowholes with or without emergent properties exhibit event causality except for

libertarian agents. O'Connor's view of the hierarchy is fairly standard, but it does have an aspect that would be considered controversial among those who accept N, namely, O'Connor rejects the causal closure of the physical (2000, p. 79).

Second, all particulars are physical objects. When discussing N, O'Connor calls the agent "a macrophysical object or system" (2000, pp. 95, 109, 111, 118), and a physical substance (2000, p. 73). According to O'Connor, N requires substance monism (2000, p. 121).

Third, there are genuinely emergent properties (cf. O'Connor 1994). For O'Connor, an emergent property has three important traits: It is (1) a simple, intrinsically characterizable, new kind of property qualitatively different from and not composed of subvenient parts, properties, relations; (2) a property which has its own ontologically basic type of causal influence; and (3) a property which is necessitated by and causally grounded in its base (O'Connor 2000, pp. 70, 110–5, 117–8). Trait (3) requires further elaboration. According to O'Connor, the causal powers of properties are essential aspects of those properties and, thus, belong to properties with an absolute, metaphysical necessity. The causal potentialities of a property are part of what constitutes the property's identity (O'Connor 2000, pp. 70–1, 117–8). It is in this sense, that in the right circumstances, a subvenient property necessitates an emergent property. Thus, properties constitutive of consciousness, including the property of active power, are emergent (O'Connor 2000, pp. 115–23).

Finally, O'Connor embraces Causal Unity but rejects the Constitution Thesis (O'Connor 2000, pp. 108–10). While recognizing that most naturalists take N to require both, he claims that only the former is required. The Constitution Thesis allows only structural macro-properties. In rejecting it, O'Connor accepts emergent properties. And by accepting the Causal Unity Thesis, he believes that he can harmonize AGC with N.

Problems with O'Connor's description of the agent

There are two problems with O'Connor's agent: it is no mere physical particular, and O'Connor cannot justify naturalism over panpsychism as the appropriate ontological framework for locating the agent. Let us consider these in order.

O'Connor's agent is not a purely physical particular

When he speaks of the self *qua* agent, it is essentially mental in nature. When O'Connor describes the self from the perspective of N, he talks as though it were a physical object (O'Connor 2001, p. 51). Galen Strawson claims that a necessary condition for free agency is that one have a concept of oneself as single just *qua* mental, quite independently of whether one also has a concept of oneself as an indissolubly psychophysical thing:

In some very strong and straightforward sense, we intuitively require that there be a *mental subject* in the case of any free agent, a mental subject that is in some way or other properly distinguishable from all its particular thoughts . . . ; a mental subject that is moreover present to itself as such in some way. Whether or not there can correctly be said to *be* such a thing, we require at the very least that any free agent's thought or experience be such that it is overwhelmingly natural for us (and for it) to talk in terms of such a subject. . . . (Strawson 1986, pp. 161–2; cf. 146–69, 323–9)

All that follows from this, says Strawson, is that the concept of the self as a mental particular is a necessary condition for taking the self to be a free agent, not that there actually

are mental substances. O'Connor's description of the agent seems to present it as a subject essentially characterized by a range of mental properties necessary for agency. Since O'Connor offers a characterization of agents themselves, and not simply an analysis of our common sense concept of agents, O'Connor's agent appears to be a mental particular, an essentially mental particular *qua* agent cause.

It is not clear how he can hold that the agent self is a physical substance necessarily characterized by emergent mental properties. If the agent self is essentially mental, and if we recognize that a particular's actual and potential properties are both relevant for characterizing the kind of entity the particular is, then the agent self would seem to be essentially a mental/physical particular, and not simply a physical particular with emergent mental properties attached to it. When John Locke argued that thinking matter was possible, some of his critics (Edward Stillingfleet, S. G. Gerdil, Malcolm Flemyng) responded by pointing out that a "material" substance whose essence was constituted in part by mental potentialities was no longer simply a "material" substance (Yolton 1983). I believe O'Connor's agent is subject to the same criticism.

Perhaps in response to arguments such as these, O'Connor has developed his view of the agent beyond what appeared in *Persons & Causes* and now advances the idea that persons are material substances in a qualified sense (O'Connor & Jacobs 2003; O'Connor & Wong 2005). Working within a framework of immanent universals, O'Connor uses these descriptors for the person-as-agent: a biological organism with emergent properties (in his three senses, including top/down active power) that are as basic as the negative charge of an electron; a three-dimensional continuant with a mental life grounded in its physical nature; a cluster of immanent universals with its own unique particularity not reducible to that of the mereological aggregate from which it arises; an emergent biological organism with a new thisness; a new composite that exhibits an objective substantial unity. These descriptors express O'Connor's desire to steer a *via media* between a mere ordered mereological aggregate on the one hand and a view such as William Hasker's, according to which a brand new emergent mental whole exists and is in no way composed of subvenient entities (Hasker 1999).

O'Connor claims that the standard mereological aggregate is inadequate to ground an enduring continuant, one that is needed to satisfy the requirements for a responsible libertarian causal agent. He also rejects a Haskerian view on the grounds that only a theistic solution along the lines of AC could account for how a complex physical system could give rise all in one go to a brand-new emergent mental entity.⁵ O'Connor wants to avoid universalism regarding composite objects, so he specifies conditions, under which a new emergent individual arises, and he offers an ontological account of how such an individual could arise in the first place. Regarding the former, emergent properties are the best candidates for emergent individuals (and the only clear evidence we have for such properties is consciousness). All other macrowholes are merely mereological aggregates. So in the category of individual, O'Connor's ontology includes atomic simples, mereological aggregates, emergent biological organisms (and as a Christian theist, at least one purely spiritual substance – God).

5. O'Connor rejects this move because it suffers from the causal pairing problem for which most plausible solution to that problem – singular causation – is bogus. O'Connor seems unfamiliar with Thomistic solutions. See Moreland and Wallace (1995).

When it comes to offering an account of all this, O'Connor is not clear about his task, and it is sometimes hard to tell which of these two questions he is answering: (1) How are we to explain ontologically how emergent individuals could come about? (2) When should we judge that an emergent individual has come about? Questions (1) and (2) are ontological and epistemological, respectively, and I shall take (1) to be O'Connor's focus. So understood, he claims that subvenient entities are always trying to bring about the emergent individual, but it is only when a certain threshold level of complexity is reached that conditions are right for that base to cause the emergent individual to come into being. When emergent mental properties appear, they constitute holistic mental states – perhaps enduring baseline mental states – and these, in turn, confer on persons their substantial unity as thinking biological substances, presumably by bringing about through top/down causation a new particularity over and above that of the series of subvenient mereological aggregates that are in a constant state of flux. This “composition-conferred-by-holism” view produces an emergent individual that is somehow composed by its parts yet has a new thisness all its own.

Why should we believe any of this? First, according to O'Connor first-person direct awareness justifies the view that consciousness is emergent in his three senses and this justification overrides any *a posteriori* ascriptions of microstructure to conscious states. All empirical knowledge, he tells us, presupposes this knowledge. Second, we should limit our account to the constraints provided by the naturalist mereological hierarchy and the grounds we have for accepting it, we should avoid a theistic explanation of emergent individuals, and on the basis of theoretical simplicity, we should adopt a view of the emergent individual that does two things: grounds endurance and agency beyond the flux of change in a mere ordered aggregate and is as close to the mereological aggregate as possible in order to fit the naturalist viewpoint.

What should we make of O'Connor's modified view? I believe the objections raised against his earlier position apply with equal or greater force to the modified view. For example, it is still not clear how a particular with basic mental potentialities is a physical object. To his credit, O'Connor seems to recognize this and, thus, he calls persons material substances “in a qualified form.” Moreover, O'Connor's new view is more clearly a version of panpsychism, and it is far from clear that this is a legitimate specification of positive naturalism. For example, when he claims that consciousness is just as basic as negative charge, this claim is closer to theism than to naturalism and it will be a hard pill for naturalists to swallow. This view also renders impossible a strict naturalist explanation of emergence as, for example, in the Causal Unity Thesis. Instead, mental potentialities and their causal interaction with physical conditions are required, and this is a long way from (positive) naturalism.

Besides retaining difficulties from the earlier position, the modified view suffers from some new problems not present in the older version. I mention two. First, there are deep metaphysical problems with O'Connor's emergent individuals. For one thing, the framework of immanent universals renders unintelligible the claim that the emergent individual has its own thisness while at the same time being constituted by the relevant mereological complex. The framework of immanent universals depicts property instances as states of affairs (the so-called thick particular) – in the case of O'Connor's persons, states of affairs that are substantial continuants – with three constituents: the universal, the nexus of exemplification, and an individuator (the thin particular, in my view, a bare particular). Whatever conditions ground the exemplification of the universal are external to (not constituents

of) the instance itself. And since the person can endure even though the mereological aggregate is in constant flux, it would seem that the aggregate is accidental to the continuant. To the degree that his emergent individuals provide what is needed (e.g. being enduring continuants), they look strangely like Hasker's emergent mental ego rather than some via media.

Moreover, there just is no baseline conscious state that is constant throughout a person's life and apt for grounding endurance. The property of being conscious cannot provide such a baseline because it is both a universal and a second-order property of mental properties (being a sensation) that comes-to-be and ceases-to-be exemplified when first-order states come and go. Our mental lives team with flux as does the "underlying" aggregate. There seems to be no account of the individual that grounds its endurance unless we treat the individual as a state of affairs constituted by a mental essence, exemplification, and particularity with the aggregate its cause but outside the being of its effect. But, again, this is Hasker's view, not O'Connor's. Finally, in criticizing Hasker, he claims that unless one appeals to a theistic explanation, one cannot explain how a complex physical system could give rise, all in one go, to unique emergent whole. As an advocate of AC, I am cheered by this admission. Unfortunately, this argument has been repeatedly raised against emergent properties themselves.

Second, I find O'Connor's composition-conferred-by-holism to be deeply troubling. He apparently accepts the dictum that "thought implies a thinker," or more generally, that consciousness requires a particular to possess it. So far so good. But it seems to me that this is so because the bearer of consciousness is more basic ontologically than the mental properties it exemplifies or the mental states that obtain within it. But O'Connor's view has this backward. If I understand him correctly, when the mereological aggregate reaches the proper threshold, emergent consciousness arises and this, in turn, causes the conscious individual to come into existence via top/down conferral (by generating a new thisness). Thus, thinkings cause thinkers, but it seems to me that something like the converse is true – the dependence goes the other way.

O'Connor also claims that emergent states are caused by temporally prior subvenient states and, thus, emergence is diachronic and not synchronic (see O'Connor & Wong 2005). Thus, the following scenario seems to arise: at t_1 subvenient conditions cause emergent conscious state C_1 to obtain at t_2 which, in turn brings about emergent individual I_1 at t_3 . Two things seem to follow. First, the very first mental state in one's life (C_1) seems clearly ownerless, since at t_2 there is no individual to possess it.

Second, beyond the very first conscious state, the following would seem to hold: for all C_{N+1} (for N greater than zero) at t_{N+2} , the individual I_{N+1} conferred by and, thus, ontologically tied to C_{N+1} exists at t_{N+2} . I see no further relevant ontological relationship between a conscious state and an emergent individual other than the conferral relation. If this is correct, then it is hard to see how a continuing "self" can exist, since there just is no single, ongoing "baseline mental state" throughout one's life. Since conscious states are in flux, so are the instantaneous individuals upon whom they confer existence. In this case, for any time t greater than one, there may be an emergent individual that exists while a particular conscious state obtains, but it is the wrong one. In general, each emergent individual at a time is ontologically associated with a mental state that obtained instantaneously earlier and, thus, is ownerless.

Additionally, his modified view is even less compatible with naturalism than his earlier view. In light of the ontology-constraining factors surfaced in section one and their

associated graded burden of proof on any ontology that goes beyond them, O'Connor exceeds those factors (e.g. the mental is as basic as negative charge, the emergence of active power and a new individual, neither emergent entity satisfies the "entry by entailment" condition, top/down causation, epistemic authority given to first-person introspection that trumps *a posteriori* considerations) and fails to meet the burden of proof required for his position to be a plausible version of naturalism. Moreover, given the presence of AC which O'Connor himself acknowledges, his dismissive attitude toward theistic explanation begs the question at several points and fails to take into account adequately the epistemic impact of AC for his project.

O'Connor and panpsychism

So much for O'Connor's depiction of the agent. Here is the second difficulty with his account: As McGinn points out, in the contemporary setting, a "material" substance such as O'Connor's would properly be characterized according to weak panpsychism (McGinn 1999, pp. 95–101). The vast majority of naturalists take panpsychism to be a rival to a naturalist understanding of matter and not a permissible version of N. We will examine this issue in more detail in this section. For present purposes, recall that according to N, the fundamental level of reality is strictly physical and emergent entities "up" the hierarchy depend for their existence, or at least instantiation, on strictly microphysical entities. However, according to panpsychism, mental properties (either potential or actual properties) are fundamental and *sui generis*, and this conflicts with the naturalist hierarchy according to which the fundamental level is strictly physical.

O'Connor can simply disagree here that panpsychism is a rival to naturalism. He acknowledges that his view implies that "the presence of agent-causal capacities in select complex entities has always been among the potentialities of the world's primordial building blocks . . ." (O'Connor 2001, p. 58). Elsewhere he argues that "[t]he basic properties and relations of our world will be those properties whose instantiation does not even partly consist in the instantiation of distinct properties by the entity or its parts. *It is the thesis of emergentism that some basic properties are had by composite individuals*" (O'Connor & Wong 2005, p. 665; italics in the original). Again, "[e]mergent features are as basic as electric charge now appears to be, just more restricted in the circumstances of their manifestation" (O'Connor & Jacobs 2003, pp. 541). I suspect that these are hard sayings for most naturalists. In order for O'Connor to justify the claim that this assertion is a permissible version of N, two things seem to be required.

First, he must show that the emergence of active power is causally necessitated by the relevant physical base. This is a necessary condition for him to show that the actual emergence of active power is consistent with the Causal Unity Thesis. Further in the discussion, I argue that O'Connor fails in this regard. Even if he successfully shows that strictly natural microphysical entities are necessary causal conditions for the emergence of active power, this would not show that his view is an appropriate revision of N and not an abandonment of N in favor of panpsychism as a rival framework because O'Connor's view requires abandonment of the Causal Unity Thesis. Recall that this thesis states that macrolevel phenomena arise through and continue to depend on *entirely* natural microphysical causal processes. On O'Connor's treatment of emergent active power (and consciousness in general), emergence depends on the actualization of nonphysical mental potentialities which are not themselves "natural microphysical properties," even if strictly

natural microphysical entities are necessary causal conditions for such emergence. Second, he could argue that we have prephilosophical intuitions for taking mental properties in general, and active power in particular, to be emergent properties in his specific sense of emergence. I shall consider these moves in what follows.

*The Harmony Thesis, mental properties,
and the causal grounding condition*

Emergent necessitation and contingency

For two reasons, to justify the Harmony Thesis, O'Connor needs the "necessitation" of emergent active power by the subvenient base. The best way to clarify "necessitation" is to characterize it in the context of presenting the first reason. To get at that reason, it will be useful to begin by reviewing insights from Frank Jackson that were presented in section one (Jackson 1998).

According to Jackson, advocates of N should take naturalism to be a piece of serious metaphysics because in so doing, they pattern the epistemic justification of N on that of good scientific theories, and they provide grounds for preferring N to its rivals on the basis of N's superior explanatory power. Accordingly, one must face the location problem: the task of finding a place for some entity (e.g. agency) in the Grand Story. The mereological hierarchy results from serious metaphysics. For Jackson, some entity is located iff it is entailed by the basic account. Thus, if Φ is true of the actual world and all of its minimal physical duplicates told in purely physical terms, and Ψ is the corresponding true description of the actual world told in psychological terms, then Φ entails Ψ . In this way, the physical may be said to "necessitate" the psychological. It is important to keep this framework in mind for what follows.

Although he does not mention it explicitly, O'Connor seems concerned to take N as an expression of serious metaphysics, and he understands this to require the location of emergent properties, including mental properties such as active power, in terms of the understanding of "necessitation" just mentioned. Since he is concerned to show that those who accept N are not thereby given adequate grounds for rejecting AGC, O'Connor must be assuming that AGC may be adequately located in N and, moreover, that AGC does not provide evidence for a rival to N, say theism, along with substance dualism as a component of theism. As O'Connor admits, many – perhaps most – have seen AGC as evidence against N and reject the Harmony Thesis. Thus, O'Connor argues that if one is going to have a scientific understanding of an emergent property, one cannot merely accept a property as emergent without explaining its existence. Rather, one must require that an emergent property be causally grounded in its base properties if it is to be naturalistically explicable (O'Connor 2000, pp. 111–2).

O'Connor also claims that if an emergent property is contingently linked to the base properties causing it to emerge, then apart from an appeal to God's contingent choice that things be so and to God's stable intention that they continue to be so, there will be no explanation for the link itself or its constancy (O'Connor 2000, pp. 70–1). In short, if the link is contingent, the Harmony Thesis is false and AGC provides evidence for theism, and there is less need to preserve physicalism in the category of individual.

The second reason why O'Connor needs the "necessitation" of emergent active power by the subvenient base involves O'Connor's view of causation: The causal powers of

properties are essential aspects of those properties and, thus, belong to properties with an absolute, metaphysical necessity. The causal potentialities of a property are part of what constitutes the property's identity (O'Connor 2000, pp. 70–1, 117–8). O'Connor's realist view of causation – event and agent – entails that a cause produces or brings about its effect in virtue of the properties of the cause, and properties are universals that have essentially their causal powers (O'Connor 2000, p. 73). Since most philosophers identify the supervenience relation with the causal relation in the case of emergent properties, it is in this causal sense that in the right circumstances, the instantiation of a subvenient property necessitates the instantiation of its associated emergent property.

Since an emergent property is the actualization of causal potentialities in the right circumstances, the emergent property seems to be a part of its causal property's identity as well. Thus, an emergent property seems to require its base property to exist. In an earlier account, O'Connor accepted this robust claim about emergent properties because he took an emergent property to be an expression of the very nature of the subvenient base causing it. However, in *Persons & Causes* he says that the notion that an emergent property could not exist without its subvenient base is “possibly gratuitous” (O'Connor 2000, p. 112). His concession seems to result from his desire to offer as minimalist an account of emergence as possible to increase its chances of being accepted by critics and, thus, he leaves open the sort of modality (metaphysical, nomological) required for a minimalist account of emergence. But O'Connor himself continues to accept the more robust account of causality, and this would seem to require that he also continue to accept the stronger notion of emergence.⁶

Unfortunately, while the Harmony Thesis requires the relevant physical circumstances to necessitate emergent mental properties, including active power, the link between mental properties and the relevant physical circumstances seems utterly contingent. Grounded in strong conceivability, thought experiments that provide strong justification for this claim proliferate throughout the literature. For example, inverted qualia and Chinese Room scenarios seem to be coherent and entirely possible. No strictly physical proposition of N employing solely physical terms for particulars, properties, relations, or laws renders these thought experiments broadly logically impossible, even in worlds that resemble ours in every physical respect.

Again, different forms of the well-known knowledge argument seem to be quite plausible. Since O'Connor himself accepts a property dualist interpretation of the argument, given this interpretation, no knowledge whatever of merely physical facts gives one any information at all about the presence, absence, or nature of mental facts. If this is so, it is difficult to see how one could justify the claim that Φ entails Ψ . No amount of information about the former entails anything at all about the latter. Φ is consistent with our world and with inverted qualia and zombie worlds that are minimal physical duplicates of our world. The physical/mental link seems contingent indeed.

Further, the modal argument for substance dualism seems plausible. If so, then at least certain versions of the argument imply that physical entities are not necessary for the instantiation of mental properties. Indeed, theism itself presents (at least) one case in which

6. If I am correct about this, then O'Connor cannot simply argue that the emergence of active power is merely metaphysically compossible with N. Rather, the existence of active power would seem to require N. Thus, the existence of substance dualism as a rival position is a crucial aspect of evaluating the Harmony Thesis, since the presence of substance dualism as a coherent rival counts against this stronger claim.

active power is not dependent upon a physical base. Surely, the existence of God or angels with libertarian power is metaphysically possible, and if so, it is just not clear why the property of active power is causally tied to a physical base.

These thought experiments have been around a long time and there is no sign that they are going away. They provide evidence against the necessitation claim that is central to the Harmony Thesis. As far as I know, O'Connor does not consider the force of the modal argument. I wonder how he would handle cases in which the agent cause is a pure spirit (God). If he says that the presence of the relevant physical base necessitates the emergence of active power but that the latter could obtain without the former, then this would amount to a denial that an emergent property is an essential aspect of the subvenient property whose potentialities actualize it.

Given a functionalist analysis of mental kinds, it may be that a type of mental state could be "realized" in spirits and brains and this fact is consistent with certain brain states in certain circumstances necessitating the realization of a mental state by being sufficient for such a realization. But this admission would not provide O'Connor with a rejoinder to my argument from the instantiation of active power in spirits. Given that active power is a simply, intrinsically characterized property that is instantiated, and not a structural property that is realized, O'Connor depicts active power as a disposition of its metaphysical base as a matter of metaphysical necessity, and it is hard to see how this disposition could be actualized without its categorical base. Further, most naturalists do not cash out emergent supervenience merely as the logical sufficiency of the subvenient base. They spell out emergence in terms of two other principles which, together with logical sufficiency, constitute minimal physicalism:

- (1) The anti-Cartesian principle: there can be no purely mental beings (e.g. substantial souls) because nothing can have a mental property without having a physical property as well.
- (2) Mind-body dependence: what mental properties an entity has depend on and are determined by its physical properties (cf. Kim 1996, pp. 9–13).

Naturalists employ (1) and (2) in their analysis of emergence precisely because they want to ensure that emergent properties are located in the naturalist ontology by guaranteeing that such properties require, depend on, and are causally determined by their entirely physical subvenient bases. If most naturalists are correct about this requirement for locating an emergent property in the ontology of N, then the actuality, or even the metaphysical possibility of the instantiation of active power in a pure spirit is a problem for the Harmony Thesis. It is one thing to reject the existence of God and angels. It is another thing altogether to claim that God or angels are metaphysical impossibilities, even if the modal status of such a claim is limited to possible worlds with the same physical particulars, properties, relations, and laws as the actual world.

O'Connor does address the knowledge argument and inverted qualia thought experiments. Regarding the former, he opts for a dualist interpretation of the argument and claims that two features of many mental phenomena are emergent properties causally necessitated by the appropriate physical bases: the phenomenal feature and subjectivity which he interprets as the fact that one can come into contact with a conscious property only by having it (O'Connor 2000, p. 116). What about the apparent contingency of the

mental/physical causal link? O'Connor simply denies that all causal necessity must be transparent. He says that there is no good reason to think that when we come to have a scientific understanding of some phenomenon, we will just be able to see that a causal effect had to follow from its cause. In the case of conscious properties, although they are necessitated by their causal bases, we just cannot see the necessity of the causal connection. Regarding inverted qualia, O'Connor adopts the same dismissive strategy, claiming that inverted qualia thought experiments "implausibly drive a wedge between a phenomenal property's qualitative features and its causal role" (O'Connor 2000, p. 120).

O'Connor's rejoinder to these arguments sounds very much like a denial that there is a problem, but the intuitions of contingency that lie behind the various dualist arguments in focus are rooted deep within our prephilosophical intuitions, and surely, there is a burden of proof on O'Connor that is not met by his dismissive strategy.

Four arguments against consciousness as emergent

Four additional considerations cumulatively undercut O'Connor's claim that conscious properties are emergent. First, O'Connor himself admits that "there are no widely accepted working theories that are committed to the existence of emergent properties . . ." (O'Connor 2000, p. 114), and "there is a lack of hard evidence in favor of emergence in areas that are well understood . . ." (O'Connor 2000, p. 115). He does not find this particularly troubling, however, because he believes that our scientific knowledge is so incomplete that the absence of emergent properties is far from empirically established. But the burden of proof lies in the other direction, and the proper conclusion to draw is that, currently, "the hypothesis of emergence" is yet to be justified.

Second, it is false to claim that "there is convincing evidence" (O'Connor 2000, p. 116) that mental properties are emergent. For three reasons, it is difficult and may be impossible to justify their emergence empirically. (1) The emergent hypothesis and substance dualism are empirically equivalent models and no empirical evidence counts in favor of one over the other. (2) To correlate mental and physical properties as a first step toward justifying emergence, one of the two correlates is not available for empirical inspection, and this makes straightforward empirical justification of emergence more difficult. (3) It is only in the case of fairly simple mental states (e.g. specific sorts of pains) that we have any hard evidence of specific mental/physical correlations. There is no evidence whatever that complex mental properties, such as the property thinking-about-the-history-of-skepticism, are correlated with specific base physical properties, much less emergent on them. Part of the problem here is the difficulty of providing criteria for individuating complex mental states in an empirically testable way, a problem that O'Connor himself acknowledges (O'Connor 2000, p. 118). On a fine-grained theory of properties, this may be an impossible task, not just a difficult one. Thus, many strong physicalists adopt a course-grained view of mental properties as a response to inverted qualia arguments, but this move requires that mental properties be identified with functional roles, and it is not available to O'Connor.

Third, even if mental properties are, in some sense, emergent, that does not entail that they are emergent in O'Connor's sense. Recall that for O'Connor, emergent properties have these three features: (1) they are simple, intrinsically characterizable, new kinds of properties; (2) they have their own ontologically basic type of causal influence; and (3) they are causally necessitated by their subvenient physical base.

Roughly, the first two features correspond to what John Searle calls Emergence₁ and Emergence₂, respectively (Searle 1994, pp. 111–2). Now Searle is typical of those naturalists who accept emergent properties as merely emergence₁ and not Emergence₂. Since we have seen reasons for this in section one, I will not rehearse them here. But one point needs to be emphasized. O'Connor claims that mental properties are the best examples of emergent properties, since they exhibit subjectivity and a phenomenal nature, and he claims that we have “direct evidence” of emergence in the case of consciousness (O'Connor 2000, p. 114).

I agree that we have direct access to and introspective knowledge by acquaintance of our own mental states, but naturalists such as Searle claim that this “direct evidence” merely justifies conscious properties as Emergence₁ and not Emergence₂. As I will argue next, the sort of introspective evidence that might be cited to support the claim that some mental properties, especially active power, have their own causal powers also supports substance dualism and, thus, that evidence provides a defeater for the claim that mental properties are emergent. At the very least, this additional introspective evidence goes beyond the sort of direct evidence O'Connor cites to justify consciousness as Emergence₂. At best, it merely justifies them as Emergence₁.

However, even if this “direct evidence” justifies taking active power to be emergent in the first two senses, it utterly fails to justify the third sense. The vast majority of people agree that in introspection they are completely unaware of anything physical. They have no introspective acquaintance with their brain or any other strictly physical object, or with any subvenient physical properties. When philosophers argue that consciousness is a set of emergent properties, they do not appeal to first-person introspection to justify the claim. No inspection of the brain or any other candidate for the subvenient physical base from either a first- or third-person perspective provides “direct evidence” for treating any conscious property as emergent in sense three.

This is an important conclusion that O'Connor apparently fails to see. In a publication subsequent to *Persons & Causes*, O'Connor acknowledges that:

[t]he emergentist can and should allow that there is an epistemological presumption against emergentist hypotheses for systems of currently-untested complexity levels *absent special reason to suspect that they are different from run of the mill cases*. (O'Connor & Wong 2005, p. 674)

But right after this concession, O'Connor attempts to refute a claim by Brian McLaughlin to the effect that, while emergence is a coherent concept, it is enormously implausible that there are any such properties, and least for those with ostensible scientific sobriety.

O'Connor's response consists in two claims: (1) a person's experiences and other conscious mental states are *sui generis* simple emergent properties and (2) claim (1) is defeasibly justified by direct first-person awareness of conscious states with an epistemic strength that precludes the *a posteriori* ascription to them of hidden microstructure hidden to introspection. But O'Connor is simply mistaken about this. Direct first-person awareness completely fails to provide any justification whatsoever for his third characterization of emergent properties, and this is the sense he needs to justify conscious properties as emergent in the sense needed for his Harmony Thesis.

Finally, given O'Connor's employment of “direct evidence” to justify the claim that conscious properties are emergent ones, the epistemic grounds for this claim derive from

first-person introspection and not from empirical research. As we have just seen, O'Connor insists on this. Given that this evidence provides accurate information about the intrinsic nature of mental properties (his sense one of emergence), and given that we have a fairly good idea of the nature of physical properties, most have seen their connection to be contingent, and that is why naturalists have had such a hard time "locating" them in light of the necessitation condition discussed earlier.

The contingency of the link between mental/physical properties stands in stark contrast to paradigm cases of located macroproperties. Jackson cites macrosolidity, understood as impenetrability, as something easily construed as necessitated by subvenient base traits (e.g. intermolecular forces, lattice structures) (Jackson 1998, pp. 3–4). Jackson also points out that the prescientific notion of macrosolidity as being everywhere dense has been rejected by those who accept N. The reason for this rejection is clear. If real, the latter notion of solidity would be a macroproperty only contingently connected to its microphysical base and, thus, it would not be located in N.

Irreducible mental properties are like the prescientific notion of solidity. Since they cannot be located, our dualistic prescientific conception of them must be revised according to some strict physicalist strategy. If mental properties are emergent, they fail to resemble paradigm cases of located macroproperties (e.g. solidity as impenetrability), and O'Connor has failed to provide an adequate justification for assimilating them to the paradigm cases. Interestingly, he acknowledges that:

[r]eductionism nowadays is much disparaged. Yet by our lights, the most plausible variety of physicalism is reductionist, as it does not require one to make dubious moves in the underlying metaphysics of physical properties. (O'Connor & Wong 2005, p. 661)

It is no accident that strong physicalism is (and ought to be) the ontology of naturalism precisely because it does not require such dubious moves. For self-reflective positive naturalists, the Constitution Thesis is an essential component that fits naturalism like a hand in a glove.

AGC, the Harmony Thesis, and the epistemic features of O'Connor's case

In contending for his views, O'Connor makes implicit or explicit reference to certain epistemic features of his case both for AGC and the Harmony Thesis. I shall focus on two of these features and argue that, if applied consistently, they place a burden of proof on O'Connor's defense of the Harmony Thesis – specifically, the harmony of AGC and a physical agent – that he has failed to meet: the role of prephilosophical intuitions in his case, and his view of the nature of prephilosophical intuitions about mental properties.

O'Connor and the role of prephilosophical intuitions

In arguing for AGC, O'Connor accepts two important epistemic requirements: (i) one's view of agency should be guided by and justified in light of prephilosophical, commonsense intuitions, which place a burden of proof on views that abandon them; (ii) these intuitions justify beliefs about the nature of human action itself, and not merely about our concept of human action (O'Connor 2000, pp. xii–xiii, 3–5, 42). O'Connor uses these intuitions to

place a burden of proof on compatibilists and on critics of the Harmony Thesis. Thus, his task in both areas of debate is to rebut and not refute his interlocutors. Applied to agency, O'Connor claims that incompatibilism is *prima facie* justified by these intuitions, they ground a modal argument for incompatibilism, and compatibilists fail to overturn the argument based on these *prima facie* justified intuitions. Applied to the Harmony Thesis, given N and the prephilosophical intuitive justification of AGC, O'Connor says that the burden is on those who reject the Harmony Thesis and accept the Constitution Thesis. Since the latter is neither entailed by the Causal Unity Thesis nor empirically established, then we are not required to accept it. Failure to meet this burden, coupled with positive grounds for emergent properties (see further discussion), means that there is no good reason to reject the Harmony Thesis.

How does one know one's prephilosophical intuitions have sufficient justification to do the work required by O'Connor's case? There are two features of such intuitions. First, they should be held widely and deeply by normal folk with no ideological axe to grind. Throughout the literature, friends, and foes of incompatibilism acknowledge that it enjoys this sort of intuitive support, and O'Connor makes explicit use of this fact in his case (O'Connor 2000, pp. 4–5; cf. Kane 1996, p. 4; Foster 2001, p. 267). Second, both sides of a dispute employ concepts derived from or based on those intuitions. John Bishop is typical of many compatibilists when he explicitly employs a libertarian concept of agency to develop his own compatibilist model that falls under that concept "closely enough" to be adequate (Bishop 1989, pp. 58, 69, 72, 95–7, 103–4, 114, 120, 126–7, 140–4, 177–80). Bishop allows a libertarian conception of agency to guide the development of his own account, and to be the legitimate source both of counter arguments in the form of thought experiments and of the sense of adequacy for his responses to those counterarguments. Libertarian intuitions seem pervasive in debates about agency.

Both characteristics seem present for intuitions on behalf of substance dualism and against physicalist views of the self. Friends and foes of dualism admit that it is the commonsense view, and the vast majority of people throughout history have been dualists about the self in one form or another. Jaegwon Kim acknowledges that:

We commonly think that we, as persons, have a mental and bodily dimension. . . . Something like this dualism of personhood, I believe, is common lore shared across most cultures and religious traditions . . . (Kim 2001, p. 30)

Along similar lines, Frank Jackson says that ". . . our folk conception of personal identity is Cartesian in character . . ." (Jackson 1998, p. 45).

Prephilosophical intuitions in support of a substantial, immaterial self are widely and deeply held, and they ground the modal argument for substance dualism. These intuitions seem expressed in the concepts and arguments used by dualists and physicalists. The intelligibility of near-death experiences, arguments from the unity of one's conscious field, thought experiments about personal identity to the effect that the person is merely contingently related to his body or psychological traits, and responses to these thought experiments (e.g. various causal chain analyses of personal identity) seem to employ a substantial, immaterial conception of the self.

O'Connor could respond that in the case of substance dualism, grounds for N justify a rejection of these prephilosophical intuitions, but in light of his own employment of similar prephilosophical intuitions for AGC and the Harmony Thesis, this response seems

arbitrary. After all, most naturalists employ N to justify a rejection of the intuitions in support of AGC, a fact that O'Connor acknowledges. Most naturalists agree that prephilosophical intuitions are on the side of AGC and substance dualism, but they adopt a consistent attitude – rejection – toward both sets of intuitions. While strictly consistent with the grounds for N, most naturalists believe that AGC and substance dualism are not as plausible as compatibilism (or noncausal versions of incompatibilism) and physicalism in light of those grounds.

Moreover, just as the Causal Unity Thesis fails to entail the Constitution Thesis and the latter has not been empirically established, so the empirical grounds for N fail to entail or empirically establish a physical agent. If O'Connor thinks otherwise, he is invited to cite the empirical evidence that accomplishes this feat. In the absence of such evidence and in light of his own epistemic characterization of the requirements placed on those who would reject the Harmony Thesis, it is hard to see what O'Connor would say to the same claim made by substance dualists about the epistemic status of physicalism, given the presence of prephilosophical intuitions for substance dualism.

O'Connor and the nature of prephilosophical intuitions

In addition to the role of prephilosophical intuitions in O'Connor's case for AGC and the Harmony Thesis, the nature of those intuitions is also of crucial importance. Philosophers differ about the nature of intuitions, for example, some hold that they are merely dispositions to believe certain things. However, the traditional view of intuitions takes them to be cases of first-person direct awareness of a relevant intentional object reported by way of the phenomenological use of "seems" or "appears." O'Connor seem to agree: intuitions in support of AGC are the way things "seem" to people (O'Connor 2000, p. 4); people have "direct evidence" of the nature of conscious properties themselves. Thus, one has direct first-person access to one's own mental states and, indeed, if this is so, such access seems to provide nondoxastic justification for prephilosophical beliefs about/concepts of mental properties, including the nature of active power. He also claims to experience himself directly bringing about the formation of an intention (O'Connor 2000, p. 124). If one accepts this account of intuitions, then one has the resources to explain why certain beliefs are so widely and deeply held.

But the same claim is often made by dualists regarding intuitions about the self. Stewart Goetz has argued that we are directly aware of ourselves and, on this basis, we are justified in believing substance dualism (see Goetz 2001, pp. 89–104). It is on the basis of such first-person self-awareness that people have the prephilosophical dualist beliefs they do, and this is why these beliefs (or, at least, dualist concepts) play such a regulative role in philosophical arguments about personal identity and related topics.

Of course, it is fashionable to claim that people have direct access to their mental states but not to their selves. Since Hume, the major strategy employed to justify this assertion is the claim that people just are never aware of themselves. I believe that dualists have provided adequate responses to this strategy, but that is beside the present point because I do not believe that O'Connor can avail himself of this strategy. To see why, we need to examine his response to an epistemological objection raised against his version of AGC. The objection is that we cannot know whether any events are produced in the manner that AGC postulates, since agent-caused events would be indistinguishable from ones that were essentially random (O'Connor 2000, pp. 123–4).

O'Connor points out that this Humean type objection would be equally telling against his realist version of event causation (event causes produce their effects). The Humean allows direct evidence for the pattern of relations among events, but not of the causal event bringing about its effect. O'Connor says that in some cases we seem to observe directly the causal connectedness between cause and effect. He illustrates this by pointing out that we do not merely observe the movement of the hammer followed by the movement of the nail; rather, we see the hammer's moving the nail.

Now it is not clear how one can directly see the hammer's moving the nail without directly seeing the hammer. Similarly, it is hard to see how one could directly be aware of one's own self producing an intention to act without being directly aware of one's own self. Indeed, O'Connor acknowledges that:

. . . in the deliberate formation of an intention, the coming to be of my intention doesn't seem to me merely to occur at the conclusion of my deliberation; I seem to experience myself directly bringing it about. (O'Connor 2000, p. 124)

This would seem to imply that people are able to be directly aware of their own selves. If so, and given that prephilosophical intuitions are widely acknowledged to be of a substance dualist sort, the very nature of intuitions as first person forms of direct access seems to offer defeasible justified beliefs of a substance dualist sort.

Perhaps O'Connor has other reasons for rejecting the use of first-person direct awareness of the self as grounds for substance dualism. To my knowledge, he has not addressed the topic in writing. If he does, there seem to be two requirements for such response. First, without begging the question, he is going to have to provide sufficient grounds for rejecting first-person awareness of the self and the role such awareness plays in justifying substance dualism in such a way that he does not undermine his own use of first-person awarenesses as a source of justification for AGC. For example, he cannot simply assert that naturalism makes substance dualism implausible, so we must reject the force of this dualist argument, because the same thing is widely said about the epistemic impact of naturalism on the justification of AGC.

Second, he would need to offer an explanation of the origin and justification of the various dualist intuitions that are a part of O'Connor's own characterization of the agent, one I accept. From where did it come and why we should believe it. I believe there is a good answer to these questions – first-person awareness of the self – but these questions would need to be answered in a way that avoids lending support to substance dualism. For example, it seems implausible to suggest that we have first-person awareness of ourselves as physical substances. If we are physical substances, yet we lack first-person awarenesses that this is so and, in fact, seem to have awarenesses that support substance dualism, we would need to know the source of and justification for dualist intuitions that form an essential part of the self *qua* agent.

The fact is that it does not seem to most folks that they are macrolevel objects. On the contrary, it seems to them from the first-person perspective – the perspective upon which O'Connor draws to justify AGC – that they are mental subjects who fail to be aware of exemplifying any physical properties. The issue then becomes whether there is any good reason to think we are physical objects, although we are not aware of being such. As far as I know, O'Connor never gives us any reason to think we are physical objects, and he must provide such an argument. When he does, he runs the danger of bringing forth

considerations of a kind (e.g. from the third-person perspective) that, if persuasive, could also undermine our conviction that we have libertarian freedom. If he simply breaks rank with most people and says that he is, in fact, aware of being a material object by first-person introspection, then this would at best justify locating his view within panpsychism and not within naturalism.

Section Five: Colin McGinn and Mysterian “Naturalism”

Unsatisfied with strong physicalism on the one hand and the various extant naturalist solutions for the origin of consciousness on the other, Colin McGinn has offered the most radical “naturalist” alternative to date (McGinn 1999).⁷ It is so bizarre that it is fair to question whether, even if successful, it is a naturalist position in any meaningful sense of the term. In this section, I shall describe and seek to rebut McGinn’s position.

McGinn’s mysterian “naturalism”

According to McGinn, given the radical difference between mind and matter, due to our epistemic limitations inherited from evolution, there is, *in principle*, no knowable naturalistic solution to the origin of consciousness or its regular correlation with matter that stays within the widely accepted naturalist epistemology and ontology. Nor is there a plausible nonnatural alternative. What is needed is a solution radically different in kind from anything previously offered, one that must meet two conditions: (i) it must be a naturalistic solution; and (ii) it must depict the emergence of consciousness and its correlation with matter as necessary and not contingent facts. More specifically, there must be three kinds of unknowable natural properties that solve the problem. We can unpack McGinn’s position by examining four different aspects of his view.

McGinn and property/event dualism

First, McGinn is committed to property/event dualism. He defines consciousness by giving first-person, introspective, ostensive definitions of particular phenomenal states. He also believes that a fairly simple form of the knowledge argument is conclusive.

McGinn on standard naturalist solutions

He also rejects all other naturalist solutions for many of the reasons mentioned in section one: the uniformity of nature, the inadequacy of Darwinian explanations, the centrality for naturalism and inadequacy of combinatorial modes of explanation along with the bottom/up combinatorial processes constitutive of the Grand Story, the acceptance of a necessitation requirement for an adequate naturalist account.

7. Unless otherwise noted, my description of McGinn’s position is taken from *The Mysterious Flame*. McGinn first thought of his mysterian naturalism in the late 1980s [see his *The Problem of Consciousness* (1991, p. vii; cf. chaps. 1–4)], and his view has remained largely unchanged until the present [see his *Consciousness and its Objects* (2004, reprinted unchanged in 2006, p. 1)].

McGinn on antinaturalist solutions

Third, various antinaturalist solutions must be rejected. He evaluates and rejects three of them: theistic dualism and AC, hyperdualism, and panpsychism. For present purposes, let us examine McGinn's treatment of theistic dualism and AC.

McGinn says that AC is a plausible argument and that there is no plausible naturalist rival outside of his own. But for six reasons, AC is a bad argument. For one thing, if we appeal to a conscious God to explain finite consciousness, we generate a vicious infinite regress for we will have to explain why God Himself is conscious. And if we stop the regress with an unexplainable conscious God, we could just as easily do the same thing by taking finite consciousness as an unexplainable brute fact.

Second, the God hypothesis dignifies consciousness with the word "soul" as an independent thing that uses the body, and thereby generates unanswerable questions that undercut AC: Do rats have souls? Why does God give souls to rats and not worms? Third, theists exaggerate the gap between minds and brains. Mind depends on brain. Why would this be so if mind depends on God? Fourth, the existence of causally powerful substantial souls that are dependent upon brains to which they are contingently connected implies that zombie worlds possible. Now, such a world seems *prima facie* possible, says McGinn, but on further inspection it faces an insurmountable difficulty. It means that consciousness is epiphenomenal and any view that entails epiphenomenalism must be rejected. Epiphenomenalism ensues because if a zombie world is possible it follows that the physical will chug along just the same regardless of whether or not consciousness obtains. Fifth, we do not know how God produces consciousness, so at best AC is a stalemate vis-à-vis naturalism.

Finally, AC gets off the ground only if consciousness is a mystery for which we need an explanation. But, claims McGinn, his account provides a deflationary explanation for why consciousness is a mystery and, in so doing, it becomes obvious that the sort of mystery involved is not of the right kind needed to justify AC.

McGinn's solution

Finally, McGinn offers his own "solution" to the problem. He begins by claiming that while evolutionary processes formed noetic faculties in us apt for doing science, it did not develop faculties capable of doing philosophy. Thus, we have cognitive closure regarding philosophical topics, where an organism has cognitive closure with respect to some domain of knowledge just in case that domain is beyond the organism's faculties to grasp. An area of inquiry in which there is no progress is a good sign of cognitive closure, and philosophy in general, and the mind/body problem in particular are cognitively closed to human faculties due to their limitations that follow from the evolutionary processes that generated them. Thus, the mystery of consciousness would not exist if we did not have the cognitive limitations we do.

What we can do, however, is characterize the kinds of conditions that must be true of any adequate solution: (i) There must be some order underlying the heterogeneous appearances of mind and matter because nature abhors a miracle. (ii) It must be a naturalistic solution. (iii) It must depict the emergence of consciousness and its regular correlation with matter as necessary and not contingent facts. More specifically, there must be three kinds of unknowable natural properties that solve the problem: some general properties

of matter that enter into the production of consciousness when assembled into a brain (thus, all matter has the potentiality to underlie consciousness); some natural property of the brain he calls C^* that unleashes these general properties under the right conditions; just as the brain must have a hidden unknowable structure that allows consciousness to emerge from it, so consciousness must have a hidden unknowable essence that allows it to be embedded in the brain.

There is one final aspect to McGinn's position that provides a naturalistic solution to the apparent nonspatiality of the mental. According to McGinn, ours is a spatial world, yet conscious states have neither spatial extension nor location. This raises a problem: If the brain is spatial but conscious states are not, how could the brain cause consciousness? This seems like a rupture in the natural order. The nonspatiality of consciousness raises serious problems for emergence and causal interaction. McGinn proffers two solutions to this problem. First, he argues that the Big Bang had to have a cause, this cause "operated" in a state of reality temporally prior to the creation of matter and space, and this reality existed in a nonspatial mode. So the cause of the Big Bang was not spatial or material, yet it obeyed some laws in the prior state. At the Big Bang, we have a transformation from nonspatial to spatial reality, and at the appearance of consciousness we have a converse transformation. The nonspatial dimension continued to exist in matter after the Big Bang, lurking behind the scene until brains evolved, at which time this dimension showed itself again.

McGinn's second solution focuses on our concept of space. Typically, we think we are correct to depict space as a three-dimensional manifold containing extended objects. But perhaps this depiction is wrong. Maybe it's not that consciousness is nonspatial; perhaps it is spatial according to the real nature of space that is quite different from the commonsense view. If we define "space" as "whatever is out there as a containing medium of all things," then it may be that the real nature of space allows it to contain consciousness and matter in a natural way. Here the Big Bang was a transformation of space itself and not a transition from non-space to space.

Critique

I do not believe that McGinn's position will be widely accepted and that for good reason. In this section, I will criticize his evaluation of theistic dualism and AC and reserve discussion of McGinn's view of the mystery of consciousness for later.

Theistic dualism and AC

McGinn argues that by appealing to God to explain finite consciousness, one generates a vicious infinite regress and if the regress is stopped with divine consciousness as a brute fact, then one could just as easily stop with finite consciousness. This sort of argument has been around a long time and McGinn appears to be ignorant of what many believe is a long-standing, successful rebuttal to it. Let us consider the first horn of McGinn's dilemma. McGinn seems to think that if we acknowledge there is a problem with cases of finite consciousness that must be solved by appealing to other finite consciousness, then this problem generalizes and applies equally to a conscious God. Unfortunately, McGinn is wrong about this and fails to appreciate what motivates the relevant regress and the sort of regress it is.

For one thing, the infinity of the regress is impossible because it involves traversing an actual infinite and, arguably, that cannot be done. To illustrate, one cannot count from one to \aleph_0 for no matter how far one has counted, he will still have an infinite number of items to count. Such a task can begin, but it cannot be completed. Moreover, trying to count from $-\aleph_0$ to 0 can neither be completed (it involves the same number of tasks as going from one to \aleph_0) nor begun for the following reason: trying to reach any number in the past will itself require an infinite traversal as a preliminary step. Now in a *per se* regress (see further discussion), the transitivity of the relation ordering the regress implies that the dependence among members runs from the earlier to latter members. Thus, such regresses are precisely like traversing from $-\aleph_0$ to 0. Space considerations forbid me to discuss this line of argument further, but in philosophy of religion it is part of what is called the *kalam* cosmological argument. I believe the argument is sound, and I refer the reader to some relevant sources that provide a more thorough evaluation of it than can be done here (cf. Craig & Smith 1993; see also Moreland 2004).

If this is correct, the regress must be finite, and this requires there to be a first member. I shall describe next some necessary conditions that must be satisfied if one is to select an adequate first member. For now, I merely note that it is not an arbitrary decision to stop the regress because it is vicious, indeed.

The first problem with the existence of an infinite regress of the sort McGinn mentions is, as it were, its *length* – it involves traversing an actual infinite series of members. Besides the problem of traversing an actual infinite, there is another problem with the regress that McGinn fails to note: by its very *nature* it is vicious. To see this, let us ask how should “vicious” be characterized here? At least four characterizations have been offered. Roderick Chisholm says that, “One is confronted with a vicious infinite regress when one attempts a task of the following sort: Every step needed to begin the task requires a preliminary step” (Chisholm 1996, p. 53). For example, if the only way to tie together any two things whatever is to connect them with a rope, then one would have to use two ropes to tie the two things to the initial connecting ropes, and use additional ropes to tie them to these subsequent ropes, and so on. According to Chisholm, this is a vicious infinite regress because the task cannot be accomplished.

D. M. Armstrong claims that when a reductive analysis of something contains a covert appeal to the very thing being analyzed, it generates a vicious infinite regress because the analysis does not solve anything, but merely postpones a solution (Armstrong 1978a, pp. 19–21). No advance has been made. He says that this is like a man without funds who writes checks from an empty account to cover his debts, and so on, forever.

Chisholm and Armstrong’s analyses are helpful. But far and away, the most sophisticated treatment of regresses, including vicious ones, was provided by Thomas Aquinas and Duns Scotus. According to Thomas Aquinas a vicious regress is a *per se* regress which exhibits two key features (see Brown 1976): (1) it is not just a list of members, but an ordering of members in the sequence; and (2) the relationship among the members of the series is transitive. If *a* stands in *R* to *b* and *b* in *R* to *c*, then *a* stands in *R* to *c*, and so on. According to Aquinas, if there is no first member in the series that simply has the relevant feature in itself, no other member of the series will have that feature, since each subsequent member can only “pass on” that feature if it first receives it.

Consider a chain of people borrowing a typewriter. Whether or not the chain is vicious depends on one’s view of the correct description of entities at each stage in the chain. Suppose *a* goes to *b* to borrow a typewriter and *b* complies, claiming to have just what *a*

needs. If asked how *b* has a typewriter to loan, he claims to have borrowed it from *c* who, having already borrowed one from *d*, has one to give to *b*. Allegedly, at each stage in the chain, the relevant entity can be described as “a possessor of a typewriter who can loan it to another.” Thus, it is alleged, the regress is not vicious.

But it is incomplete to describe each person as “a possessor of a typewriter who can loan it to another.” Rather, each person is “a possessor of a typewriter who can loan it to another who first had to borrow it from another.” At each stage, the person *qua* lender is such only because he is also a borrower. Thus, given the nature of the series, each stage cannot be adequately described without reference to the earlier stage. Because each member is a borrowing lender, no one will ever get a typewriter unless the regress stops with someone who differs from all the other members of the series in being a lender who just has a typewriter without having to borrow it.

Analogously, because finite conscious beings are contingent, before each such being can give what it has (consciousness) to another, it must first undergo the preliminary step of receiving finite conscious being first. In Armstrong’s terms, each member of the chain exhibits the same problematic feature, namely, being a lender of consciousness who must himself “borrow” consciousness from another. In Aquinas’ terms, the members of the regress *qua* conscious lenders stand in a transitive relationship to the relevant other members in the chain, so without a member who just has consciousness without lending it, there would be no consciousness.

Finally, Duns Scotus offered detailed analyses of various regresses some of which is relevant for present purposes (Cross 2005, pp. 17–28). According to Scotus, there are two very different sorts of ordered sequences involving causal or other sorts of dependence relations: an essentially ordered or *per se* regress and an accidentally ordered or *per accidens* regress. The former are irreflexive (if reflexive, Scotus says one will have self-causation, which is absurd), asymmetrical (if symmetrical, then a member will be both a cause and an effect of the same member in the series), and, most importantly, transitive. In some essentially ordered regresses, an earlier member actually causes a latter member to cause: either *a* causes effects in *b* sufficient for *b* to cause the relevant effect in *c* (*a* effects *b*) or *a* causes *b*’s causing *c* (*a* affects *b*). In various sorts of *per se* dependency chains, the ordering of dependency is (at least) an ordering of necessary dependency conditions from earlier to latter members in the chain.

Scotus identifies three essentially ordered regresses relevant to our discussion: existence, getting the power to operate, and exercising the power to operate. Scotus’ main argument against the infinity of such regresses is crafted to avoid a fallacy of composition (e.g. since each member of the series is dependent, the whole must be dependent). His argument is that there is something in the final effect, the last member of the chain about which we are puzzling and seeking an adequate explanation (existence, causal power, consciousness), that is missing in all the other members precisely as essentially ordered with respect to each other, and that requires a first member which is (1) not a part of the chain and (2) simply has the feature of the final effect in itself without having to get it elsewhere.

But why must we stop with God and not some particular finite conscious being? The decision to stop with God is not arbitrary for this reason. The sort of regress we are considering is one such that in the respect relevant to the ordering of the regress’s members, the stopping place must be unique and different from all others. In the typewriter case, the relevant respect is that each member does not simply have a typewriter; he is himself one who must borrow before he lends. The proper stopping place is with a “first mover” who

simply has a typewriter with no need to borrow one before lending it. Now, each finite conscious being is contingent in two senses: with respect to its existence and with respect to the fact that consciousness was actualized in it. These types of contingency disqualify finite conscious beings from being the proper first mover. Being a necessary being in both senses, God is such a proper First Mover.

This kind of dialectic occurs frequently in philosophy. In agency theory, an advocate of AGC begins with certain concerns about human action and responsibility, opts for AGC, and confronts a problem, viz., what does the agent do to bring about an action? Desiring to avoid a vicious infinite regress, the advocate of AGC concludes that an agent cause is a first cause, a first mover, an entity that may bring about a change without having to change first or be changed to do so. In this sense, agent causes are *sui generis* compared with ordinary event causes in that the latter are changed changers characterized by passive liabilities; and agents, being characterized by *sui generis* active power, cannot be caused to act freely.

In epistemology, foundational beliefs are discovered to be such that they provide justification for nonfoundational beliefs without having to receive their entire justification from their relationship with other beliefs. In one way or another, foundationalists stop the epistemic regress with an epistemic first mover, for example, a nondoxastic self-presenting property. In ontology, discussions of relations and Bradley's famous regress lead to the notion that relations are discovered to be able to relate relata without having themselves to stand in a different relation to those relata. They are unrelatable relaters. AC is an argument form relevantly analogous to these.

McGinn's second critique of theistic dualism and AC is the claim that it uses "soul" to dignify consciousness, and this generates serious difficulties (do rats have souls and, if so, why rats and not worms?). As it stands, this is not much of an argument. For one thing, it is simply false. AC does not quantify over souls in any of its premises, and Premise (1) launches AC on the basis of the existence of consciousness or its lawlike correlations with the brain.

Second, the question "Why do rats have souls and not worms?" is an ambiguous question. If it is the question "Why would God, if He exists, give souls to rats and not worms?" presumably, the answer would be along the lines of why I painted my dining room walls and not the bathroom yellow: I wanted to. What is so problematic about that? If He exists, presumably, God wanted to create certain things and give them certain accidental attributes, and He did not wish to do so for other possible beings He refrained from creating or giving certain accidental attributes. If, instead, the question is about why some things are conscious and others are not, one could say that this is just part of the nature of different things. It is part of the nature of a rat to be conscious and not part of the nature of, say, a tree or rock. Obviously, such an answer involves a commitment to some form of essentialism, but whether or not essentialism is a plausible metaphysical framework is not specifically a theistic concern. This theistic response could employ "nature" in a variety of ways and still be successful.

Finally, focusing on consciousness and not souls, McGinn may be claiming that there is a sort of arbitrariness about theistic dualism such that it entails that at some point, God rather arbitrarily decided to create beings with consciousness and others without it. In response, the sort of "arbitrariness" that seems to underlie this claim is precisely what one would expect if property dualism is true. On a widely accepted dualist understanding of the knowledge of other minds, one starts with first-person acquaintance of one's own mental states and is justified in attributing to other minds whatever mental states are

needed to explain the organism's behavior. Ontologically, an organism either is or is not conscious, it either does or does not have some specific mental state. But epistemologically, as organisms become increasingly disanalogous to humans, one is less and less justified in attributing specific mental states or consciousness itself to the organism. Thus, one is increasingly less justified in such attributions applied to another normal human, a rat, or a worm. As with other cases involving degreed properties (in this case, "being justified to such and such a degree"), sorites-style difficulties surface about drawing precise lines among the relevant ordered entities. However, far from being a problem, this is precisely what one would expect from a dualist perspective and McGinn is mistaken if he thinks otherwise.

McGinn also criticizes theistic dualism and AC on the grounds that, if true, it entails that consciousness depends entirely on God's will but this is not true, since consciousness clearly depends on the brain. Again, McGinn's objection is ambiguous. I can see two interpretations each of which is fairly easy to rebut. His question assumes that if something depends entirely on God, then it will not depend on something else in any sense. But this is a bizarre view of divine providence and God's act of sustaining contingent beings in existence. No matter what the precise theistic formulation of these matters is, theists agree that there is a relevant distinction between primary and secondary causality. For example, just because God created and continually sustains the physical universe and its laws, and is in this sense that upon which they "depend entirely," it hardly follows that lightning does not causally depend on certain antecedent conditions within the cosmos. Various causal relations and dependencies within the created order are consistent with the view that if God had not created and does not continually sustain the universe (or some feature within it), then the universe (or some feature within it) would not exist. Clearly, there is no problem here.

Alternatively, the question may be asking why, if the creation of consciousness is a contingent act, there is a covarying dependence among life forms according to which as brains become less and less complex, consciousness does so as well. Note the sort of question this is. It is a theological question about why God would arrange things in this way. So understood, the question is not a request for a scientific answer or even a distinctively philosophical one. It is a question whose answer requires reference to God's possible intentions and motives for arranging things in this way. As I see it, the question is part of a larger one about why there are bodies in the first place.

What are the adequacy requirements for a theological answer to this question? In my view, we have a situation parallel to the difference between a theodicy and defense regarding the problem of evil for theism. A theodicy aims at providing an account of why God actually permits evil in the world. By contrast, a defense offers no such account but seeks merely to show that atheists have failed to carry their case that evil is inconsistent with the existence of God. A defense seeks to undercut the atheist's argument by providing a possible solution on the grounds that there is a substantial burden of proof on the atheist for which a defense is adequate.

It is hard to see the force of this problem. McGinn would need to give reasons for thinking that the dependency of mind on the brain in the manner specified earlier (and the dependency goes in both directions) is such that there is no reason God would have for creating such a situation. To be successful, McGinn would have to assume that there is no possible reason for God to make things this way. But it is hard to see why this would be the case. The theist could easily hold that God has reasons for doing things this way and even if the

details of those reasons are not available to us, the mere fact that God could easily have them is sufficient to undercut this objection.

Moreover, according to a theology of the body that I favor, God created bodies to provide a source of power for living things so they could act in ways independent of God's own exercise of efficient causal power. Bodies provide power for action in the created world. Further, the more complicated an animal's consciousness is, the more complex and finely tuned the body would need to be to be responsive to the fine-graded mental states in causal interaction with it. Consider a form of consciousness with a complexity sufficient to engage in a variety of quite specific actions associated with precise nuances in thought, believe, emotion, desire, and so forth. On this view, if such a consciousness were causally connected to a material object without the physical complexity needed to register in the physical world the appropriate mental complexity, that mental complexity would be wasted. Such a theology of the body is clearly a possible reason God could have for making things the way he has, and it is sufficient for the purposes of defense required to undercut McGinn's objection.

McGinn's fourth criticism of theistic dualism is that, if true, it entails the possibility of zombie worlds that imply an implausible epiphenomenalism regarding conscious states. But the latter entailment is not the case. One could consistently embrace a form of dualism that entails the possibility of zombie worlds, and also believe that causal interaction between consciousness and matter in the actual world is contingent. From this, it follows that an epiphenomenal world is, indeed, a possible world, but it does not follow that the actual world is an epiphenomenal one. One could go on to unpack "brings about" in "mental state M brings about brain state B" in terms of causal necessitation, viz., "M brings about B in all interactionist worlds relevantly similar to the actual world." All this is clearly consistent with zombie worlds.

I am among those dualists who believe that the causal relation (and any other relevant relations, e.g. the emergent supervenient relation construed in noncausal terms) between consciousness and matter is a contingent one. If God wished, He could have created an epiphenomenal world. Inverted qualia worlds, zombie worlds, the metaphysical possibility of body switches, or disembodied existence are part of the case for the contingency of the relevant mind/matter relations. Since McGinn's objection assumes that dualism entails such contingency, I need not defend it in the present dialectic. Rather, I am arguing that if we grant this contingency and the possibility of both zombie and epiphenomenal worlds, it does not follow that our world is an epiphenomenal one. The dualist will hold that as a matter of contingent fact we live in a world of causal interaction and nothing McGinn says threatens this claim.

McGinn's fifth objection is that the theistic solution does not solve anything because it does not tell us *how* God created consciousness. Without providing such a mechanism, the God hypothesis is vacuous and fails to be an advance over a naturalistic explanation which likewise fails to answer the *how* question.

There are two things to be said in response to this argument. First, McGinn's claim simply fails to understand the logic of personal explanation. I will not repeat here our discussion in section two of the nature of personal explanation. I make one simple point: a personal explanation can be epistemically successful without making any reference to a mechanism or other means by which the hypothesized agent brought about the state of affairs in the explanandum. I can explain the existence and precise nature of a certain arrangement of objects on our dinner table by saying that my wife brought it about so we

could have an Italian dinner with the Isslers. That explanation is informative (I can tell its Italian food we are having, that we are having the Isslers over and not the Duncans, that my wife did this and not my daughter, that natural processes are inadequate). And the adequacy of such a personal explanation is quite independent of whether or not I know exactly how my wife did it.

There are many sciences that involve formulating criteria for inferring intelligent agent causes to explain certain phenomena and for refraining from inferring such causes. And in these sciences, such an inference is usually both epistemically justified and explanatorily significant completely independently of knowledge as to how the agent brought about the phenomena. In forensic science, the Search for Extraterrestrial Intelligence (SETI), psychology, sociology, and archeology, a scientist can know that an intelligent agent is the best explanation of a sequence involving the first 20 prime numbers in a row or that such and such is an intelligently designed artifact used in a culture's religious sacrifices without having so much as a clue as to how the sequence or artifact was made.

Furthermore, an appeal to a particular epistemic value, in this case to the requirement that a necessary condition for successful explanation is that a theory explains how a certain phenomenon was produced, is question-begging against AC and represents a naive understanding of the role various epistemic values play in adjudicating between rival explanations of some phenomenon.

For one thing, two rivals may solve a problem differently depending on the way each theory depicts the phenomenon to be solved. Thus, the epistemic values for assessing one theory may differ substantially from those relevant to its rival. Thus, it is often more complicated to compare rivals than McGinn seems to assume. It is possible for two rivals to rank the relative merits of epistemic virtues in different ways or even give the same virtue a different meaning or application. Rivals can differ radically about the nature, application, and relative importance of a particular epistemic virtue. Thus, it is question-begging to claim that a criterion P set by one hypothesis should be most important for its rival such that if it fails to satisfy P it is explanatorily inferior.

Finally, sometimes one rival will consider a phenomenon basic and not in need of a solution, empirical or otherwise. It may, therefore, disallow questions about how or why that phenomenon occurs and, thus, can hardly be faulted for not being fruitful in suggesting lines of empirical research for mechanisms whose existence is not postulated by the theory. By way of application, a theistic dualism could take God's creation of consciousness and its precise causal correlation with the brain to be a basic action for which there is no further "how" question to be asked. And the theistic dualist can also claim that, given the nature of personal explanation, the epistemic value of citing a mechanism in answer to a "how" question is not as important as other epistemic values. Thus, failure to answer such a question is not a significant issue in light of its own inner logic. But the same cannot be said for naturalism, and given the way physical explanation works, the importance of answering "how" questions by citing a mechanism is, indeed, quite high. Thus, the naturalist's failure to answer this question is a serious one but the same cannot be said for theistic dualism.

Four problems with mysterian "naturalism"

We come to an evaluation of McGinn's own position – mysterian "naturalism." For at least four reasons, it must be judged a failure. First, given McGinn's agnosticism about the

properties that link mind and matter, how can he confidently assert some of their features? How does he know they are nonsensory, prespatial, or spatial in an unknowable way? How can he confidently assert that we are naturally constituted from smoothly meshing materials, as seamless as anything else in nature? How does he know some of these properties underlie all matter? These seem unanswerable.

The only one he proffers is that we must provide a naturalistic solution and all ordinary naturalistic ones either deny consciousness or fail to solve the problem. But given the presence of AC, McGinn's claims are simply question-begging and *ad hoc* according to criteria developed in section two. Indeed, his agnosticism seems to be a convenient way of hiding behind naturalism and avoiding a theistic explanation. Given that theism enjoys a positive degree of justification prior to the problem of consciousness, he should avail himself of the explanatory resources of theism.

In a related fashion, it is sometimes argued, and not without some justification, that attempts to draw a line between what we can and cannot know requires that one must first cross the line to draw it. McGinn comes close to doing the very thing he claims cannot be done. Whether or not one accepts this claim about drawing lines, McGinn's view seems self-refuting. He tells us that we did not evolve with faculties apt for doing philosophy, that when confronted with a lack of progress we should draw the conclusion that we are cognitively closed to the subject matter in question, and so on. Yet McGinn's entire book is a species of philosophical argument, and he explicitly states that his purpose is to develop and defend his viewpoint over against rivals. He also derives philosophical theses (e.g. skeptical theses in areas for which we have cognitive closure) by philosophically studying the history of philosophy, he gives an analysis of the nature of human knowledge, he offers philosophical – not scientific – arguments against positions that rival naturalism. I may be missing something here, but it is hard to avoid the conclusion that McGinn's own project is refuted, or at least undercut by his own views that constitute the core of that very project.

Second, it is not clear that his solution is a version of naturalism. His hypothesized properties cannot be known by employment of the naturalist epistemology, nor are they relevantly similar to the rest of the naturalist ontology. McGinn may appropriately call these “naturalistic” properties in the sense that they are (1) not created by God and (2) are regularly involved in giving rise to consciousness in organisms. However, it is vacuous to call these properties “naturalistic” in the only sense relevant to theistic dualism and AC, namely, as entities whose nature, existence, and activity can be located in a natural ontology and given a naturalistic explanation. Given that naturalism is a worldview that claims superior explanatory power to its rivals, these are bizarre, *sui generis* brute facts on a naturalist view. Indeed, McGinn's ontology is so bizarre that it may be taken as a *reductio* against naturalism if McGinn is correct that no other naturalist solution is available. McGinn's solution is actually closer to an agnostic form of panpsychism than to naturalism, he is clear that panpsychism is a rival to and not a legitimate specification of naturalism.

Third, McGinn does not solve the problem of consciousness, he merely relocates it. Rather than having two radically different entities, he offers us three unknowable properties with radically different aspects, for example, his links contain the potentiality for ordinary spatiality and nonspatiality, for ordinary materiality, and for mentality. Moreover, these aspects of the linking properties are just as contingently related as they seem to be without a linking intermediary. The contingency comes from the nature of mind and matter as naturalists conceive it. It does not remove the contingency to relocate it as two aspects of unknowable intermediaries with both.

Finally, there are difficulties with McGinn's solution to the problem of the nonspatiality of mental states. According to his first option, the Big Bang had to have a cause, this cause "operated" in a state of reality temporally prior to the creation of matter and space, this reality existed in a nonspatial mode, and while the cause of the Big Bang was neither spatial nor material, it still obeyed some laws in the prior state.

There is much in this solution that brings a smile to the theist: the Big Bang had to have a cause, presumably because either events *per se* or those in which something comes-to-be must have causes, the cause is not spatial nor is it material. This cause shares important features with the God of classic theism. At the very least, it is hard to see how the hypothesized state of affairs satisfies the conditions for location in a naturalist ontology specified in section one. The presence of temporality is not sufficient to claim this is a naturalistic state of affairs because on the basis of strong conceivability there are possible worlds in which angels alone exist temporally. As Kant argued, finite consciousness entails temporality, so such worlds are temporal but hardly apt for appropriation by a naturalist.

Nor is the presence of law sufficient. In discussing constituent/whole relations, Edmund Husserl described a host of (*a priori*) laws that he claimed governed the coming-to-be and perishing of various entities, and changes that take place among them (Moreland 2002a). However, these laws are not physical laws of nature. Even if Husserl is wrong, his ontology and many others like it demonstrate that the mere presence of laws that govern change in some purported ontological model is far from sufficient to claim that the model is a naturalistic one. Moreover, it seems reasonable to hold that the nature of a relation is constituted by the nature of its relata – spatial, musical, odor, and logical relations are such because they can relate certain kinds of entities and not others. If this is right, it is hard to see how the laws envisaged by McGinn are natural laws.

Finally, McGinn seems unfamiliar with the *kalam* cosmological argument (see chapter 3). It is safe to say that the argument is sufficiently robust to require inclusion in any discussion of the beginning of the spatiotemporal physical universe. If successful, it justifies the claim that time itself had a beginning that was caused by something that can exist without time. And on the assumption that laws of nature govern temporal processes and, thus, require events to be instantiated, it becomes clear that the cause of the first event was not governed by a law of nature. At the very least, McGinn's speculations regarding his first option are grossly incomplete and, moreover, they open the door for considerations quite favorable to theism.

What about McGinn's second option, that we are wrong to think of space as a three-dimensional manifold containing extended objects? Perhaps the real nature of space is "whatever is out there as a containing medium of all things." If this is correct, then the real nature of space allows it to contain consciousness and matter in a natural way.

I do not have a knockdown argument against this option, but I do find it highly counterintuitive and, in fact, unintelligible. And it may be useful to say why. I begin with an observation about the difference between formal concepts and certain material concepts. In my view, formal concepts are capable of being expressed adequately by way of definite descriptions. To illustrate, the formal concept of a substance is "whatever is an essentially characterized continuant"; the formal concept of justice is "whatever outcome is fair and accords with the maxim 'treat equals equally and unequals unequally.'" Functional concepts are good examples of formal concepts. By contrast, material concepts, at least those defined by ostensive definition, are defined by rigid designation. If we limit ourselves to sense

perceptible entities with which we may be acquainted, then “red,” “sour,” and “middle C” seem to express material concepts.

Now I take the notion of extension to be such a material concept. If I am right, then the only intelligible notion of a spatial dimension is the material concept of “extended one-directional magnitude,” which must be defined ostensively. Along similar lines, “space” is a material concept defined by acquaintance as “extended three-directional magnitude.” I, for one, have no idea what it means to use spatial language to speak of multidimensionality or in the way McGinn does. When a scientist claims that a three-dimensional object can be “*spatially* rotated” into other spatial dimensions, I can give no material content to the claim and, thus, I cannot understand what is being said. Likewise, when McGinn tell us that space is “whatever is *out there* as a *containing medium* of all things,” “out there,” “containing,” and “medium” are either used in the ordinary way characterized earlier, in which case the definition is circular and seems to require ostensive definition to give these terms intelligible content, or else they are used equivocally in which case they are unintelligible, at least to me.

I recognize that physicists talk about a multitude of spatial dimensions. In my view, the scientific notion of an extra dimension of space is a mere mathematical device, a formal definition with no material content that can intelligibly be ascribed to reality, and theories that employ such language should be understood in antirealist terms. When scientists speak of multidimensionality with respect to space, they say things such as the following: there are millions of dimensions of space; there could be an infinitely small volume; mass and space are literally interchangeable; triangles can be identical to circles; that a one-dimensional line (a string) could literally have clockwise vibrations in 10 dimensions of space and counterclockwise vibrations in 26 space dimensions (Ross 1996). I find such language unintelligible, and while the problem may my lack of imagination, I suspect that others may agree with me.

I have argued that McGinn’s position is not as plausible as AC and is not a legitimate version of naturalism. Long ago, Thomas Kuhn taught us that there are certain telltale signs of a paradigm in crisis, among which are the proliferation of epicycles and of rival specifications of the paradigm formulated to preserve that paradigm in the face of stubborn, recalcitrant facts. Especially significant are specifications so bizarre that it is hard to recognize them as specifications of the paradigm. I take McGinn’s mysterian “naturalism” to be an indication that naturalism is in serious crisis with respect to consciousness. Kuhn also taught us that as bizarre and *ad hoc* as some of the specifications may be, if there is no rival paradigm, then an advocate of the degenerative paradigm must simply do the best he or she can with the recalcitrant facts and leave it at that. But if there is a plausible rival, a paradigm shift may well be in order. In my view, McGinn’s position, coupled with theism and AC as a rival, serve as evidence that such a paradigm shift away from naturalism toward theism is past due.

Conclusion

Strong naturalism/physicalism has been in a period of Kuhnian paradigm crisis for a long time, and physicalist epicycles have multiplied like rabbits in the last two decades. Moreover, the various versions of physicalism are in a stagnating period of stalemate. Increasingly, naturalists are turning to emergentist views of consciousness.

The truth is that naturalism has no plausible way to explain the appearance of emergent mental properties in the cosmos. Ned Block confesses that we have no idea how consciousness could have emerged from nonconscious matter: “we have nothing—zilch—worthy of being called a research programme. . . . Researchers are stumped” (Block 1994, p. 211). John Searle says this is a “leading problem in the biological sciences” (Searle 1995, p. 61). Colin McGinn observes that consciousness seems like “a radical novelty in the universe” (McGinn 1999, p. 14); he wonders how our “technicolour” awareness can “arise from soggy grey matter” (McGinn 1991, pp. 10–1). David Papineau wonders *why* consciousness emerges: “to this question physicalists ‘theories of consciousness’ seem to provide no answer” (Papineau 1993, p. 119). Papineau’s solution is to deny the reality of consciousness as a genuinely mental phenomenon (Papineau 1993, pp. 106, 114–8, 120, 121, 126). He correctly sees that strong physicalism is the only real alternative for a naturalist.

If one is a positive naturalist who embraces emergent mental properties, then he or she should admit defeat as Frank Jackson acknowledges:

Our primary concern is with physicalism as a doctrine of the *kind* of world we are in. From this perspective, attribute dualism is not more physicalistically acceptable than is substance dualism. (Jackson 1998, p. 6, n. 5)

Emergence, in particular, is a mere name for a problem to be solved, and it is consistent with substance dualism, double-aspect theory, certain forms of personalism, and epiphenomenalism. This is not a result most naturalists will want to accept.

Jaegwon Kim observes that:

if a whole system of phenomena that are *prima facie* not among basic physical phenomena resists physical explanation, and especially if we don’t even know where or how to begin, it would be time to reexamine one’s physicalist commitments. (Kim 1998, p. 96)

For Kim, emergent mental entities are the paradigm case of such a system of phenomena. Not long ago, Kim’s advised fellow naturalists to simply admit the irreality of the mental and recognize that naturalism exacts a steep price and cannot be had on the cheap (Kim 1998, chap. 4, especially pp. 118–20).⁸ If feigning anesthesia is the price to be paid to retain naturalism, then the price is too high. Fortunately, the theistic argument from consciousness reminds us that it is a price that does not need to be paid.

Further Reading

- Armstrong, D. M. (1999) *The Mind-Body Problem: An Opinionated Introduction*. Boulder, CO: Westview Press.
- Bealer, G. (1996) On the possibility of philosophical knowledge. In J. E. Tomberlin (ed.), *Philosophical Perspectives 10: Metaphysics*, 1–34. Cambridge, MA: Blackwell.

8. Curiously, Kim has become an emergent epiphenomenal dualist regarding phenomenal consciousness. See his *Physicalism or Something Near Enough* (2005). It is likely that his ontology has many brute facts, a curious situation for one who accepts ontological simplicity as a guide for ontology.

- Beauregard, M. and O'Leary, D. (2007) *The Spiritual Brain: A Neuroscientist's Case for the Existence of the Soul*. New York: HarperCollins.
- Beilby, J. (ed.) (2002) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*. Ithaca, NY: Cornell University Press.
- Churchland, P. (1988) *Matter and Consciousness*, rev. edn. Cambridge, MA: MIT Press.
- Clayton, P. (2004) *Mind & Emergence*. Oxford: Oxford University Press.
- Connell, R. (1988) *Substance and Modern Science*. Houston, TX: Center for Thomistic Studies.
- Martin, M. (ed.) (2007) *The Cambridge Companion to Atheism*. Cambridge: Cambridge University Press.
- Menuge, A. (2004) *Agents under Fire*. Lanham, MD: Rowman & Littlefield.
- Moreland, J. P. (2003) The argument from consciousness. In P. Copan and P. Moser (eds.), *Rationality of Theism*, 204–20. London: Routledge.
- Moreland, J. P. (2008) *Consciousness and the Existence of God: A Theistic Argument*. London: Routledge.
- Moreland, J. P. and Rae, S. (2000) *Body & Soul*. Downers Grove, IL: InterVarsity Press.
- Nagel, T. (1997) *The Last Word*. New York: Oxford University Press.
- O'Hear, A. (1997) *Beyond Evolution: Human Nature and the Limits of Evolutionary Explanation*. Oxford: Clarendon.
- Robinson, H. (ed.) (1993) *Objections to Physicalism*. Oxford: Clarendon.
- Skrbina, D. (2005) *Panpsychism in the West*. Cambridge, MA: MIT Press.
- Swinburne, R. (1997) *The Evolution of the Soul*, rev. edn. Oxford: Clarendon Press.
- Swinburne, R. (2004) *The Existence of God*, 2nd edn. Oxford: Clarendon Press.
- Wagner, S. J. and Wagner, R. (1993) *Naturalism: A Critical Appraisal*. Notre Dame, IN: University of Notre Dame Press.

References

- Adams, R. (1992). Flavors, colors, and God. In R. Douglas Geivett and B. Sweetman (eds.), *Contemporary Perspectives on Religious Epistemology*, 225–40. New York: Oxford University Press.
- Armstrong, D. M. (1978a) *Universals & Scientific Realism Vol. I: Nominalism & Realism*. Cambridge: Cambridge University Press.
- Armstrong, D. M. (1978b) Naturalism, materialism, and first philosophy. *Philosophia* 8, 261–76.
- Bishop, J. (1989) *Natural Agency*. Cambridge: Cambridge University Press.
- Block, N. (1994) Consciousness. In S. D. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, 210–9. Malden, MA: Blackwell.
- Brown, P. (1976) Infinite causal regression. In A. Kenny (ed.), *Aquinas: A Collection of Essays*, 214–36. Notre Dame, IN: University of Notre Dame Press.
- Burge, T. (1992) Philosophy of language and mind: 1950–1990. *The Philosophical Review* 101: 1, 3–51.
- Chisholm, R. M. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliff, NJ: Prentice-Hall.
- Chisholm, R. M. (1989a) *Theory of Knowledge*, 3rd edn. Englewood Cliff, NJ: Prentice-Hall.
- Chisholm, R. M. (1989b) *On Metaphysics*. Minneapolis: University of Minnesota Press.
- Chisholm, R. M. (1996) *A Realistic Theory of the Categories*. Cambridge: Cambridge University Press.
- Churchland, P. (1984) *Matter and Consciousness*. Cambridge, MA: MIT Press.
- Craig, W. L. and Smith, Q. (1993) *Theism, Atheism, and Big Bang Cosmology*. Oxford: Clarendon Press.
- Cross, R. (2005) *Duns Scotus on God*. Hants, England: Ashgate Publishing Ltd.
- Foster, J. (2001) *The Immaterial Self*. London: Routledge.
- Goetz, S. (2001) Modal dualism: a critique. In K. Corcoran (ed.), *Soul, Body and Survival*, 89–104. Ithaca, NY: Cornell University Press.

- Haldane, J. (1996) The mystery of emergence. *Proceedings of the Aristotelian Society* 96, 261–7.
- Hasker, W. (1999) *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Herbert, N. (1987) *Quantum Reality*. Garden City, NY: Anchor Press/Doubleday.
- Horgan, T. (1993) Nonreductive materialism and the explanatory autonomy of psychology. In S. J. Wagner and R. Warner (eds.), *Naturalism*, 295–320. Notre Dame, IN: University of Notre Dame Press.
- Jackson, F. (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.
- Kane, R. (1996) *The Significance of Free Will*. New York: Oxford.
- Kim, J. (1996) *Philosophy of Mind*. Boulder, CO: Westview Press.
- Kim, J. (1998) *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kim, J. (2001) Lonely souls: causality and substance dualism. In K. Corcoran (ed.), *Soul, Body and Survival*, 30–43. Ithaca, NY: Cornell University Press.
- Kim, J. (2005) *Physicalism or Something Near Enough*. Princeton, NJ: Princeton University Press.
- Leibniz, G. (1979) Monadology 17. In P. P. Weiner (ed.), *Leibniz Selections*, 533–551. New York: Charles Scribner's Sons.
- Locke, J. (1959) *An Essay concerning Human Understanding*. New York: Dover Publications.
- Lyons, W. (1995) Introduction. In W. Lyons (ed.), *Modern Philosophy of Mind*, xlv–lv. London: Everyman.
- Mackie, J. L. (1982) *The Miracle of Theism*. Oxford: Clarendon Press.
- Madell, G. (1988) *Mind and Materialism*. Edinburgh: Edinburgh University Press.
- Martin, M. (1990) *Atheism: A Philosophical Justification*. Philadelphia: Temple University Press.
- McGinn, C. (1991) *The Problem of Consciousness: Essays toward a Resolution* Oxford: Basil Blackwell.
- McGinn, C. (1999) *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.
- McGinn, C. ([2004] 2006) *Consciousness and Its Objects*. Oxford: Clarendon Press.
- Moreland, J. P. (1995) Humanness, personhood, and the right to die. *Faith and Philosophy* 12, 95–112.
- Moreland, J. P. (1997) Naturalism and libertarian agency. *Philosophy and Theology* 10, 351–81.
- Moreland, J. P. (1998a) Locke's parity thesis about thinking matter: a response to Williams. *Religious Studies* 34, 253–9.
- Moreland, J. P. (1998b) Should a naturalist be a supervenient physicalist? *Metaphilosophy* 29, 35–57.
- Moreland, J. P. (2000) Christian materialism and the parity thesis revisited, *International Philosophical Quarterly* 40, 423–40.
- Moreland, J. P. (2001a) Topic neutrality and the parity thesis: a surrejoinder to Williams. *Religious Studies* 37, 93–101.
- Moreland, J. P. (2001b) *Universals*. Canada: McGill-Queen's University Press.
- Moreland, J. P. (2002a) Naturalism, nominalism, and Husserlian moments. *The Modern Schoolman* 79, 199–216.
- Moreland, J. P. (2002b) Timothy O'Connor and the harmony thesis: a critique. *Metaphysica* 3: 2, 5–40.
- Moreland, J. P. (2004) A response to a Platonistic and set-theoretic objection to the kalam cosmological argument. *Religious Studies* 39, 373–90.
- Moreland, J. P. (2008) *Consciousness and the Existence of God: A Theistic Argument*. London: Routledge.
- Moreland, J. P. and Craig, W. L. (eds.) (2000) *Naturalism: A Critical Analysis*. London: Routledge.
- Moreland, J. P. and Nielsen, K. (1993) *Does God Exist? The Debate between Theists & Atheists*. Buffalo, NY: Prometheus.
- Moreland, J. P. and Wallace, S. (1995) Aquinas vs. Descartes and Locke on the human person and end-of-life ethics. *International Philosophical Quarterly* 35, 319–30.

- Nagel, T. (1986) *The View from Nowhere*. New York: Oxford.
- O'Connor, T. (1994) Emergent properties. *American Philosophical Quarterly* 31, 91–104.
- O'Connor, T. (2000) *Persons & Causes*. New York: Oxford University Press.
- O'Connor, T. (2001) Causality, mind, and free will. In K. Corcoran (ed.), *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, 44–58. Ithaca, NY: Cornell University Press.
- O'Connor, T. and Jacobs, J. D. (2003) Emergent individuals. *The Philosophical Quarterly* 53, 540–55.
- O'Connor, T. and Wong, H. Y. (2005) The metaphysics of emergence. *Nous* 39: 4, 658–78.
- Oppy, G. (2006) *Arguing about Gods*. Cambridge: Cambridge University Press.
- Papineau, D. (1993) *Philosophical Naturalism*. Oxford: Blackwell.
- Rosenberg, A. (1996) A field guide to recent species of naturalism. *British Journal for the Philosophy of Science* 47, 1–29.
- Ross, H. (1996) *Beyond the Cosmos: The Extra-Dimensionality of God*. Colorado Springs, CO: NavPress.
- Searle, J. (1994) *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Searle, J. (1995) The mystery of consciousness: part II. *New York Review of Books* 16, 60–6.
- Searle, J. (2004) *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Sellars, W. (1963) *Science, Perception, and Reality*. London: Routledge & Kegan Paul.
- Skinner, B. F. (1990) Can psychology be a science of mind? *American Psychologist* 45, 1206–10.
- Strawson, G. (1986) *Freedom and Belief*. Oxford: Clarendon Press.
- Swinburne, R. (1979) *The Existence of God*. Oxford: Clarendon.
- Swinburne, R. (1986) *The Evolution of the Soul*. Oxford: Clarendon.
- Williams, C. (1996) Christian materialism and the parity thesis. *International Journal for Philosophy of Religion* 39, 1–14.
- Wright, C. (2002) The conceivability of naturalism. In T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*, 401–39. Oxford: Clarendon.
- Yolton, J. W. (1983) *Thinking Matter: Materialism in Eighteenth-Century Britain*. Minneapolis: University of Minnesota Press.