

Putting Zombies to Rest: The Role of Dynamics in Reduction

PETER BOKULICH

Massachusetts Institute of Technology

I argue that property dualism is not supported by the purported logical possibility of qualitative zombies. Chalmers's analysis of the logical supervenience of ordinary macroscopic facts on microphysical facts fails to account properly for causal properties. His arguments rely too heavily on kinematic facts and thereby obscure the dynamical facts at the macroscopic and microscopic levels. A proper analysis of the relation between causal and dynamical properties at different levels reveals that we can only imagine qualitative zombies if we beg the question against qualia being physical.

The strongest argument for the non-physical nature of consciousness rests on the conceivability of phenomenal zombies: creatures who are physically identical to us, but who lack qualitative states, or "qualia." This line of argument has been advanced most prominently by David Chalmers (1996, and, e.g., Chalmers and Jackson 2001). Here I offer a diagnosis of the intuitions that lead to the belief that zombies are conceivable, and I argue that these intuitions cannot do the work required of them. This paper can fruitfully be viewed as a how-to guide for finding zombies unimaginable.

Chalmers's argument for dualism, in brief, is that physicalism requires the logical supervenience of macroscopic facts on microphysical facts. He argues, first, that microphysical truths entail *a priori* all physical truths about macroscopic entities (the configuration of body parts, etc.) and, second, that these macroscopic physical truths in turn imply all ordinary macroscopic truths—with the sole exception of truths about phenomenal states.¹

¹ Chalmers believes that indexical and negative truths will also fail to supervene logically on the physical truths. We shall leave aside these finer points except when they are

Here I accept Chalmers's test of logical supervenience on microphysical facts (or *a priori* entailment by microphysical truths) as a necessary and sufficient condition for physicalism. I argue, however, that in applying this test Chalmers neglects the class of facts (or truths) that is most important for the question of reduction: the class of dynamical or causal truths. Once the significance of this class of facts is established, we shall see that, when imagining a physically identical twin, we are required to imagine the cause of my twin's behavior to be identical to the cause of my own behavior. Thus if I believe that qualia explain my own reactions, I must assume that qualia also explain my twin's reactions. The argument for epiphenomenalist dualism is therefore guilty of begging the question: the only way one can motivate the failure of supervenience of qualitative facts on physical facts is by *presupposing* that qualia are nonphysical and have no causal effect on physical properties.

I. Vitalist Reduction?

The criterion that Chalmers invokes for a successful reduction is that of logical supervenience or, equivalently for our purposes, *a priori* entailment. Chalmers claims that all macroscopic facts logically supervene on microphysical facts, the key exception being facts about phenomenal states (and perhaps causal facts, a point we shall return to shortly). If, for example, we fix all the facts (past, present, and future) about the positions and velocities of all the particles that make up a fish, its offspring, its environment, etc. then there is no flexibility for any variation in the biological facts concerning that fish. However, it is still an open question whether or not the fish would have conscious experiences; either possibility (the existence or

relevant to the argument at hand.

absence of consciousness) is logically compatible with the physical facts: "When God created the world, after ensuring that the physical facts held, *he had more work to do*. He had to ensure that the facts about consciousness held. . . . The world might have lacked experience, or it might have contained different experiences, even if all the physical facts had been the same" (Chalmers 1996, p. 124). Thus biological, sociological, and meteorological facts reduce to physical facts, but facts about consciousness do not.

To evaluate Chalmers's account of reduction, let us consider a world in which physicalism is clearly false. Consider, for example, a world in which the vitalist account of biological processes is correct: bodies are still composed of physical particles, but mere physical processes cannot account for the complex self-organization of living beings. A proper account of biological processes (such as growth, reproduction, and purposive behavior) requires reference to a non-physical force, an *elan vital*. One could either say that the physical laws in this world are *violated* by biological processes, or that the physical laws have a *limited* domain of applicability that does not extend into biology. Which characterization we choose is unimportant; the important point is that the world contains irreducible biological laws: biological processes in this world cannot be *explained* by physics.

How would this world fare using Chalmers's test for reduction? Do the biological facts in this world logically supervene on the physical facts? Following Chalmers, we fix all the facts—past, present, and future—about the positions, velocities, etc. of all the particles in the world. In doing this, however, we thereby fix the behavior of all the body parts of living creatures. Consider the particles making up a fish: the facts about their fish-like configuration, about the paths that they travel through the ocean, about their relationship to egg-configured

particles left in stream-beds, etc. will all be held fixed. Thus there seems to be no room for any variation in the biological facts concerning the fish: once God decided on the paths of the particles through time, all the facts about the reproduction, circulation, health, etc. of all biological entities were also decided. Biological facts apparently supervene logically on microphysical facts even in worlds in which vitalism is true. What has gone wrong here?

The first thing to notice is that there is something quite fishy in this case about holding constant all the physical facts through all of time. If vitalism is true, then the physical particles will behave *differently* when they constitute a living biological entity than they do when they constitute a corpse. The laws of physics do not govern the behavior of the physical particles when they compose a living fish; instead, the *vitalist* laws explain the fish's biological processes. If we help ourselves to the *actual* positions of these particles throughout time (that is, to the particle *trajectories*) then we obscure this all-important fact.

The important question, as far as reduction is concerned, is *why* the fish parts (and/or the particles making up those fish parts) behave the way they do. If the physical *laws* account for all biological processes, then reductionism is true. If we instead need to invoke irreducible laws of biology, then some form of anti-reductionism, such as vitalism, is true. But it is a mistake to tie reduction to the question of whether biological *behavior* is logically fixed by the *behavior* of the constituting physical entities. The relationship between the positions of the fish-parts at a particular time and the position and configuration of the fish at the same time is a trivial relationship that tells us extremely little about whether fish can be reduced to the physical particles they are composed of.

II. Dynamics vs. Kinematics

Another way of characterizing what is amiss in the above analysis of the vitalist world is to notice that we have held fixed the *kinematics* of the situation—that is, how the positions of the particles change with time—at the expense of the *dynamics*. When we speak of the physical dynamics of a system we refer to the *forces* involved in a particular process. A familiar example of such dynamical concepts can be found in Newton's second law, $F=ma$. The acceleration of a body is obviously a purely spatiotemporal fact, the change in the body's velocity, but the *mass* of the body and the *force* acting on the body are irreducibly dynamical. More generally, the dynamical information of physics is encoded in the Hamiltonian, Lagrangian, or action used to describe a system.

There is no reason to suppose that the "dynamics" of biology in our vitalist world will formally resemble the familiar physical dynamics of our own world. The purposive behavior of a fish need not be derivable from a least action principle applied to a function of *elan vital*, for example. Nonetheless, it is clear that biological properties and relations in the vitalist world will be dynamical both in the broad sense of having causal powers and in the sense that they will affect the *physical* properties of entities, properties such as the configuration and mass distribution of these entities. Let us refer to this ability to affect physical properties as "dynamical efficacy"; it is then clear that both in the vitalist world and in our own, biological properties and processes are dynamically efficacious.

Let us now consider the role that dynamics plays in Chalmers's applications of the test for the logical supervenience of higher-level facts on microphysical facts. Chalmers and Jackson (2001) clearly stipulate that in such tests we are to hold fixed *all* the microphysical facts, both

kinematic and dynamical. The complete information about microphysical facts, they tell us, includes complete information about the structure and dynamics of the world at the microphysical level: in particular it includes or implies the complete truth about the spatiotemporal position, velocity, and mass of microphysical entities. This information suffices in turn to imply information about the structure and dynamics of the world at the macroscopic level, at least insofar as this structure and dynamics can be captured in terms of spatiotemporal structure (position, velocity, shape, etc.) and mass distribution. . . . So the information plausibly suffices for at least a geometric characterization—in terms of shape, position, mass, composition, and dynamics—of systems in the macroscopic world. (Chalmers and Jackson, p. 330)

Despite Chalmers and Jackson's explicit inclusion of dynamical facts, they make no use of these facts at either the microphysical and macroscopic level when developing their argument. They are correct in saying that truths about the spatiotemporal structure of the macroscopic world are implied by the microphysical facts, but the only microphysical facts that are doing any *work* here are the spatiotemporal *positions* of the fundamental particles.² Once we help ourselves to the *actual* spatiotemporal positions of the physical particles, the dynamics at play (the laws leading to the *change* in positions over time) become irrelevant.

Let us use the term "geometrical facts" to refer to all facts about the positions of particles over time, together with facts about the configurations of macroscopic objects. Clearly the geometrical facts, so construed, are only a small subset of all physical facts: the geometrical facts

² We do not need to follow Chalmers and Jackson in specifying a velocity in *addition* to a body's "spatiotemporal position," for spatiotemporal position simply means the position of the body at different times—which, of course, includes all facts about velocity.

leave out all facts about charge, mass, forces, etc. Let us now ask whether the persuasiveness of Chalmers's arguments is in any way altered if, instead of considering *physical* facts as our supervenience base, we simply consider the *geometrical* facts:

Say a wombat has had two children in our world. The ~~physical~~ [geometrical] facts about our world will include facts about the distribution of every particle in the spatiotemporal hunk corresponding to the wombat, and its children, and their environments, and their evolutionary histories. If a world shared those ~~physical~~ [geometrical] facts with ours, but was not a world in which the wombat had two children, what could the difference consist in? Such a world seems quite inconceivable. Once a possible world is fixed to have all these ~~physical~~ [geometrical] facts the same, then the facts about wombathood and parenthood are automatically fixed. The same goes for architectural facts, astronomical facts, behavioral facts, chemical facts, economic facts, meteorological facts, sociological facts and so on. (Chalmers 1996, p. 73)

How can a world in which all particles follow the exact same trajectories as in our world differ with respect to any of these higher-level facts? All the processes in the two worlds are the same: cells divide, money is exchanged, stars explode, and so on. Such a world would *appear* identical to our own because we have stipulated that the kinematic facts in both worlds are identical, and yet it should be clear that fixing the geometrical facts is not a guarantee that we have captured all ordinary macroscopic facts (biological, architectural, etc.) about our world. To see this, we can consider a world that is geometrically identical to our own, but in which the physical facts differ.

III. Hologram Worlds

The relevance of non-geometrical physical properties and facts can be highlighted by recalling that now-familiar dynamical properties, such as mass and force, were only discovered with great difficulty in the history of physics, and only through non-trivial *a posteriori* reasoning. The claim that the motions of all corpuscles are dictated solely by their geometrical properties, with no appeal to additional dynamical properties such as mass, was strongly advocated by some mechanistic philosophers—indeed, Descartes actually claimed it as an *a priori* truth. However, we now know that the properties of everyday entities depend essentially on fundamental dynamical properties such as mass, charge, quantum spin, etc.

Despite the fact that our physics has discarded the simplistic mechanistic picture in which all powers of an entity are dictated by the shape and motions of corpuscles, we philosophers often rely too heavily on visualizable configurations when considering claims of physicalism.³ Upon reflection, however, it should be clear that merely holding fixed the configurations of the objects in our world does not suffice to capture all the physical facts contained therein; we can, for example, imagine a world that is geometrically identical to our own, but which is devoid of dynamical properties. In this world particles have *positions* that change with time, but no mass, no charge—no physical property but spatiotemporal position. As an aid to our intuitions, we

³ Indeed, this paper's own simplistic talk of fundamental micro-physical entities having positions and momenta (velocities and mass) is deeply problematic given that the current best contenders for fundamental physics are *quantum field* theories. The *quantum* aspect of these theories is standardly taken to imply that even in quantum *particle* theory there are no trajectories, and the *field* aspect arguably denies even the existence of particles. Further, the recovery of more classical and local behavior at a higher level is an extremely difficult, and outstanding, problem. For the purposes of this paper we shall set these formidable worries aside, but see Bokulich (2004).

might recall popular characterizations of ghosts that have shapes and move about—but that cannot interact with other physical objects. We could also imagine the holograms contained in various science-fiction narratives: beings that, like ghosts, have a definite configuration and move about, but which are intangible. Of course, in these popular characterizations, both ghosts and holograms are *visible* (and perhaps audible), which seems to imply that they causally interact with light (and perhaps sound). Thus when we imagine our "hologram world" we need to go one step further and imagine entities that have shapes, motions, etc. but no causal powers whatsoever.

We are now to imagine that the objects in this world nevertheless behave in precisely in the same manner that objects in our own world behave—that is, the geometrical facts in the hologram world are identical to the geometrical facts in our world. We can, for example, imagine that all of the particles in this world just happen to follow these trajectories by chance; or we might suppose that the particle motions are dictated by the direct will of a (nonspatial) deity. This latter position was apparently held by Malebranche, who claimed that God is the only true cause of motion, and all the apparent causal interactions between physical objects are merely occasions for God's intervention. In imagining a hologram world, we are clearly imaging a world that is physically distinct from our own, even though all the geometrical facts are by stipulation identical. Hologram-fish are not real fish, and hologram-water is not real (physical) water. Real fish in real water rely on the mass, charge, and other dynamical properties of their constituent particles to behave as they do; hologram-fish in hologram-water do not.⁴

⁴ The so-called "holographic principle" coming out of recent work in quantum gravity does not speak against this claim, for this suggested principle asserts that a world with gravity is (in a specific, technical sense) equivalent to a world of one fewer spatial dimensions without gravity; it does not undermine the central place of dynamics and causation in physics.

This claim, of course, will not be entirely uncontroversial. Anti-realist accounts of causation will claim that all causal—and, presumably, dynamical—facts will be exhausted by the regularities between the events in the world. However, such a position seems extremely unattractive in light of the explanatory power of physical dynamics. The success of theories that posit the existence of forces, charges, mass, etc. gives us strong reason to believe in robust causal facts based on these properties, and we can imagine the absence of such facts in a world that nevertheless retains all the actual regularities of our own world. Further, we can discount such Humean accounts of causation for the purposes of this paper because Chalmers explicitly rejects such positions (1996, p. 151).

Although Chalmers takes causation seriously, and refers to dynamical properties when discussing microphysical facts, he fails to see any substantial relation between ordinary causal facts and microphysical dynamics. The reason for this, I argue, lies in the fact that by holding constant the *actual* positions of physical entities throughout the history of the world, he renders the dynamical facts irrelevant to the thought experiment. We have fixed the behavior of wombats, buildings, and societies without paying any attention to *why* these entities behave the way they do. Does the fish have offspring that resemble it because of the physical dynamics involved, or because of some non-reducible law of biology? We can only address these questions by considering with more care the *causes* that are at work in both our own world, and in other possible worlds that we are considering.

IV. Reducing Causal Powers to Dynamics

Chalmers recognizes that causal facts slip through his analysis of physical reduction

based on logical supervenience. He reports that facts about causation will fail to supervene on the physical facts, "taken as a collection of particular facts about a world's spatiotemporal history" (Chalmers 1996, 86). This much, of course, is true, but it is extremely misleading to limit our attention only to facts about a world's spatiotemporal history: a hologram world could be *spatiotemporally* identical to our own without being *physically* identical. This misplaced focus has apparently led Chalmers into mistakenly believing that the question of causation can be divorced from the question of reduction.

The causal powers of ordinary objects and processes are essential components of the facts that we are trying to reduce to physics. A bowl containing a hologram-fish in hologram-water is different from a bowl containing a physical fish in physical water, even if both bowls are causally isolated from the rest of the world. "Water" that ignites when a lit match is dropped in it is not real water, for it clearly lacks the power to extinguish fire. These causal powers have to be properly fixed even if this water never comes into contact with flame.

Chalmers's stipulation that we hold fixed all the facts about the physical *configuration* of our world would be appropriate if we were considering an ontology of "geometrism"; but once we recognize that the best contender for the ontology of our world is instead *physicalism*, we see that we need to take care to include properly dynamics with our kinematics. This implies that causal powers are not something that can be carved off from the everyday facts that we are trying to explain and treated as an accidental afterthought. The relationship between the dynamics at the physical level and the causal powers at the level of chemistry, biology, sociology, etc. is at the very heart of reduction. The physicalist cannot be indifferent as to whether the dynamics of fundamental physics can account for the ability of copper to conduct electricity, the efficacy of

penicillin in curing infections, or the power of increased economic demand to drive up prices.

To flesh out the facts in the world under consideration, to guarantee that we have actually captured the relevant *physical, chemical, or biological* facts, we need to make sure that we have included the relevant physical dynamics—and that these dynamics imply all relevant higher-level *causal* facts.

The problem with fixing all truths about the past, present, and future configuration of physical entities lies not primarily in the fact that these truths are insufficient for logically determining all macroscopic truths, although this is certainly the case: the real danger lies in the fact that fixing these truths can *obscure* the all-important dynamical relations that lie at the heart of reduction. To see this, consider how, in a vitalist world, one would discover that physical truths do not imply biological truths. A Laplacian demon could have at its disposal all of the facts about the dynamics and the spatiotemporal positions of all fundamental particles. This demon would then have to recognize that the physical dynamics alone cannot account for the actual positions of the particles—the physical dynamics fails to describe accurately biological processes. To recognize this fact, however, the demon would have to consider what macroscopic processes would look like if they were dictated solely by physical laws. It would be a mistake for the demon simply to compare the *actual* spatiotemporal positions of the physical particles with the configurations typical of biological processes. However, Chalmers invites us to make this mistake when imagining whether facts about wombathood supervene on the physical facts. The relevant issue when evaluating whether biology reduces to physics is whether the physical *laws* or *dynamics* can account for biological processes, and this issue cannot be pursued without temporarily *neglecting* some of the actual physical facts about our world.

Chalmers specifically eschews the claim that high-level properties supervene on microphysical laws and boundary conditions. He tells us he is not suggesting that high-level facts and laws are entailed by microphysical *laws*, or even by microphysical laws in conjunction with microphysical boundary conditions. That would be a strong claim, and although it might have some plausibility if qualified appropriately, the evidence is not yet in. I am making the much weaker claim that high-level facts are entailed by all the microphysical *facts* (perhaps along with microphysical laws). This enormously comprehensive set includes the facts about the distribution of every last particle and field in every last corner of space-time: from the atoms in Napoleon's hat to the electromagnetic fields in the outer ring of Saturn. Fixing this set of facts leaves very little room for anything else to vary. (Chalmers 1996, 71-72)

However, the macroscopic facts that we are interested go beyond the mere position of Napoleon's hat—we also want to know whether physics can account for the hat's ability to absorb or repel Napoleon's perspiration, for the hat's ability to fetch hundreds of thousands of dollars at auction, etc. Answers to these questions cannot come from merely helping ourselves to the actual positions of hat molecules, perspiration molecules, and money molecules: we need to have the higher-level causal facts grounded in the microphysical dynamics. This requires us to confront the strong claim that Chalmers hoped to avoid. We have to be able to show (in principle) that the physical laws, together with appropriate initial conditions and boundary conditions, can account for both the configurations and causal powers involved in the high-level

facts of interest.⁵

Once we recognize the importance of dynamics for the question of reduction, and the fact that fixing the spatiotemporal positions of entities offers no guarantee that we have the relevant dynamics correct, we can see that Chalmers's suggested method for using logical supervenience as a test for ontological reduction is fraught with peril. The suggestion that we hold *all* physical facts—past, present, and future—fixed when considering possible worlds runs up against the possibility that the kinematic and dynamical facts about physical entities can pull apart in some situations, and in these situations we are often inclined to notice only the easily visualizable spatiotemporal positions of our imagined particles at the expense of the relevant dynamics. Once we recognize the complicated relationships between the facts of interest, we see that we can only secure some of these facts if we are willing to allow some others to vary.

V. The Cost of Stipulating Causal Closure

A philosopher living in the vitalist world, pondering whether biological facts logically supervene on physical facts, would be ill-advised to hold fixed all the spatiotemporal facts about physical particles when considering the possible logical supervenience of biological facts. Doing so would offer the image of a world apparently containing all familiar biological processes, but this way of considering the situation only masks the fact that particles obey one set of dynamics when they constitute a living organism and a different set when they constitute a corpse. To

⁵ If fundamental physics is indeterministic, the situation will be slightly more complicated, for the physical laws and boundary conditions (including initial conditions) will offer only the probabilities of various physical facts obtaining at a later time. Thus we will also need to include information about what physical facts were indeterministically realized.

address properly the question of reduction, we have to ask whether the laws of physics—together with the appropriate boundary conditions—allow us to explain our familiar world, or whether we also need to appeal to fundamental laws of biology, or chemistry, or what have you. Thus when considering "physically identical worlds," we cannot leap immediately to worlds in which *all* physical facts are the same as in our own; rather we need to consider worlds in which (roughly speaking) the *initial* physical facts are identical to our own *and the physical dynamics or laws* are the same as those in our world.

The problem we now face is that it is extremely difficult to know what such a world looks like. Would a world with physical facts identical to our own a million years ago look like ours today if its evolution was *purely* governed by physical dynamics? Would all the facts about wombats be unchanged? Would there still be wombats? Our confidence in an affirmative answer to these questions will depend on our confidence that biological processes can be successfully reduced to physical processes.⁶ The real work will be done by justification for reduction that lies quite apart from direct considerations of logical supervenience of macroscopic facts on *all* microphysical facts.

Could one respond to the foregoing worries simply by stipulating the causal closure of the physical? In our above example of the vitalist world, troubles arose because physical particles

⁶ The situation will be complicated even further by the fact that the quantum nature of our physical theories suggests that the fundamental dynamics will be *stochastic*, which implies that we cannot require the laws of physics to *force* a world that is at one time physically identical to our own to be identical to our own at all times. While the requirement of determinism is clearly too strong, it seems that it would be too weak if we were merely to require that it be *possible* that the world appear identical to our own over time—for it is possible that a world governed solely by chance could mimic our own. Presumably the proper requirements for reduction will be closely tied to whether we have an adequate scientific *explanation* in terms of our physical laws and initial conditions.

sometimes did things for nonphysical reasons. But we have every reason to believe that in our world the behavior of physical particles depends only on physical factors. If the physical realm is indeed causally closed, then we can rule out scenarios, such as vitalism, that require a violation or limitation of physical laws.

Indeed, one might think that Chalmers already has this claim to causal closure in place, either explicitly or implicitly. For example, if the causal closure of the physical is a fact about our world, then it may be included in, or inferable from, the complete set of physical truths about our world. If we know the physical dynamics and the actual kinematics, then (leaving aside worries about indeterminism) it would in principle be possible to decide whether the physical laws apply universally. If they do, then we need not worry about events being driven by any causal processes aside from physics. Further, Chalmers believes that in order to recover all facts about the macroscopic world, including negative facts, one will have to invoke a "that's all" clause stating that the world contains nothing beyond the physical facts. While Chalmers's motivation for including this clause seems to be limited to securing negative facts (his discussion of causation indicates that he does not intend the clause also to stipulate the causal closure of the physical), once the claim is in place, one might argue that it can serve to establish this closure. If we are committed to physical dynamics, and by stipulation "that's all," then there is arguably no room for non-physical dynamics or laws to limit or violate the physical dynamics.

Assuming the causal closure of the physical—whether by explicit stipulation, by assuming that the physical facts will include a claim that physical laws are inviolable and universal, or by relying on a "that's all" clause—will indeed guarantee that we can rely on our knowledge of the actual kinematics for tests of an *a priori* link between physical truths and higher-level truths.

However, we will need to take careful note of the price we pay for this guarantee. First, the real work here is being done by the stipulation of causal closure—together with our justification for this stipulation; the logical supervenience of all dynamically efficacious properties and processes will follow as a trivial consequence of this stipulation. Second, with this stipulation in place, we are committed to the physicality of *all* dynamically efficacious properties and processes, *even if we have no idea how they could be physically manifested*.

Thus, for example, a 17th century philosopher who stipulates the causal closure of the physical is committed to the claim that biological properties are physical properties, because life, reproduction, etc. obviously affect physical entities. This is true even if she has no idea *how* physics could account for reproduction, purposive behavior, and so on. She now cannot even decide what a physically identical world would look like until she *first* determines what properties and processes are dynamically efficacious. More importantly, in imagining a physically identical world, she now has to imagine that all dynamically efficacious properties and processes take place in that world for the *same reason* that they occur in her world. It is not open to her to imagine that in such a world physical particles behave biologically because of a nonphysical *elan vital*; nor may she suppose that physical causes can serve as a *substitute* for familiar biological causes. The stipulation of causal closure of the physical requires us to hold fixed the *actual* causes of our world when we imagine a physically identical world, which means that all dynamically efficacious properties have to be included in such a world. As we shall see in the next section, this fact makes all the difference when we try to conceive of beings physically identical to us, but lacking consciousness.

VI. Just Imagine, Zombies!

Thus far, I have argued that causal properties, far from being superfluous to the question of physicalism, lie at the very heart of the issue of reduction. Chalmers's treatment of reduction not only fails to do justice to these causal facts, it obscures their ontological status by *stipulating* the facts that are supposed to be *explained* by the causal powers in question. The question we now face is whether Chalmers's mistreatment of the reduction of causal properties infects his argument for the claim that consciousness is nonphysical. We shall see that it does, and that a proper appreciation of the relevance of causal facts leaves the argument for the nonphysicality of qualitative states without support.

Chalmers believes that we can draw an important distinction between psychological states and qualitative states, or qualia. The former, he argues, can be picked out functionally, "that is, in terms of the kinds of stimulation that tend to produce it, the kind of behavior it tends to produce, and the way it interacts with other mental states" (Chalmers 1996, p. 14). Qualia, on the other hand, are not fixed by the functional facts—or even the physical facts—of our world, and therefore, Chalmers argues, consciousness is nonphysical. To establish the failure of supervenience of qualia on physical states, Chalmers coaches us to imagine a zombie twin: a being who is identical to me in every physical respect, but who lacks qualia:

What is going on in my zombie twin? He is physically identical to me. . . . He will certainly be identical to me *functionally*: he will be processing the same sort of information, reacting in a similar way to inputs, with his internal configurations being modified appropriately and with indistinguishable behavior resulting. He will be *psychologically* identical to me It is just that none of this functioning will be

accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a zombie. (Chalmers 1996, p. 95)

We now face two questions. First, is the certainty that Chalmers displays regarding the functional equivalence of the zombie twin warranted, and, second, can we be confident that we have succeeded in imagining a being that is identical to me in *every* physical respect when we imagine a zombie twin?

Establishing the first point seems trivial given Chalmers's exposition, for if the zombie twin is like me in every physical respect, then his body will completely mimic my own: he will walk, talk, read and write exactly as I do. All of his behavior will be indistinguishable from my own, and he will *react* to all stimuli precisely as I do. How then could he differ from me functionally?

It should now be clear, however, that this line of reasoning cannot be trusted, for the same claims could be made of a *geometrically* identical twin—but such a twin is clearly not my functional equivalent. Consider, for example, the image of a baseball game on television: the relations between the image of the bat on my screen and the image of the ball are not functionally identical to those between a real bat and a real ball, despite the fact that the input, output, and internal configurations appear superficially identical in the two cases. Likewise, a hologram-computer in a hologram world is not functionally identical to my own computer—even though a seemingly identical input of keystrokes leads to a seemingly identical output of characters on the screen and on the pages produced by the hologram-printer.

By holding fixed all the facts—past, present, and future—of physical particles, Chalmers guarantees that anything composed of those particles in our imagined world will exhibit the same

behavior as it does in our world. However, this alone is insufficient for establishing the crucial *causal* facts in the two worlds, and thus it is also insufficient for establishing the *functional* equivalence of the two. We can only conclude that my physically identical twin is also functionally equivalent to me if we can rule out the possibility of nonphysical dynamically efficacious properties and relations. If vitalism or substance dualism is true of our world, then my twin will not be functionally identical to me, despite the misleading appearances produced by holding fixed *all* the physical facts.

How, then, are we to establish that all dynamically efficacious properties and relations are ultimately microphysical properties and relations, and that qualia are not included among them? As I have argued above, and will flesh out in more detail in a moment, we seem to have two options: The first is to derive—explicitly, in some reasonable approximation—psychological powers and processes from microphysical dynamics and boundary conditions, and then to demonstrate that qualia cannot be recovered in this same way. The second is to offer a general argument for the claim that all dynamically efficacious properties and relations are reducible to physical properties and relations, and then to argue that qualia are not dynamically efficacious.

The most promising avenue to pursue this second option is built on the assumption that the microphysical level is causally closed. As we saw in the previous section, this allows us to conclude that all dynamically efficacious properties and relations in our world are in fact *physical* properties and relations, and so they will be contained unaltered in any world physically identical to our own. It is important to notice what this stipulation amounts to, however: if we assume the causal closure of the physical, then when we imagine a world physically identical to our own, we must imagine that all behavior in that world is caused in precisely the same way that

is caused in our own world. For this reason we do not even know what to include when constructing our physically identical world unless we *first* decide what facts, properties, and relations are dynamically efficacious.

When I imagine my physically identical twin, I need to ask *why* the twin is behaving as he is. If we are stipulating that our world is causally closed at the level of physics, then I am committed to the claim that all dynamically efficacious casual properties in our world are ultimately derivable from microphysical dynamical properties such mass, charge etc., together with the proper dynamical laws and non-dynamical properties and relations. This means that I am committed to imagining that my physically identical twin behaves as he does for exactly the same reasons that I behave as I do. Thus I need to ask why it is that I say "red" when I see a ripe tomato, and why I exhibit pain behavior when I stub my toe. If we have good reasons to believe that I behave in this way *because* I have certain qualitative experiences, then I am committed to the claim that these qualia are *physical* and need to be included in any physically identical world.

Do we have reasons to believe that qualitative experience is dynamically efficacious? Surely we do. On introspection, it certainly seems that the *reason* I say "red" when gazing at a red tomato and reporting its color lies in the fact that I am having a certain perceptual experience of redness. It is extremely difficult to deny that the *explanation* of my behavior after stubbing my toe is a particularly unpleasant conscious experience that I am having. This evidence is defeasible, of course, but as evidence goes, it is certainly extremely compelling.

Given our strong reasons for thinking that qualia are dynamically efficacious, we then need to ask how we are to account for the causal properties that are (*prima facie*) associated with consciousness. Following our earlier analysis, we have two options: First, we can stipulate only

the microphysical dynamics and boundary conditions and see whether this yields the macroscopic facts in question. However, our ignorance obviously rules out this possibility, for not only is it likely to be well beyond our abilities of calculation to recover mental causal powers from the microphysical, but we do not even have in hand the fundamental physical theory, nor an adequate understanding of the dynamics of mental processes. Our second option is to follow Chalmers's suggestion and fix *all* the physical facts—past, present, and future—and to justify this by stipulating the causal closure of the physical. Following this road, however, requires us to hold fixed all facts about all dynamically efficacious properties and relations. Thus we must decide *before specifying our "physically identical world"* whether qualia are dynamically efficacious or not, and we cannot exclude qualitative properties from this world unless we have *independent* reasons for believing that qualia have no physical effects.

It is important to emphasize here that our evidence for the dynamical efficacy of conscious experience is completely independent from all consideration of the conceivability of zombies or the proper way to imagine a world physically identical to our own. On the other hand, arguments *against* the dynamical efficacy of qualia appear to rest completely on the alleged conceivability of zombies (and related scenarios such as inverted spectra). One would never entertain the possibility that my saying "red" when gazing at a red tomato might be caused by something other than my sensation of redness if one weren't seduced into believing that intrinsic causal properties are irrelevant to physical reduction.

This response to the argument for dualism therefore goes beyond the standard objection that epiphenomenalism runs counter to our strong reasons for believing that conscious states have causal impact on the physical world. This is not simply a case of one person's modus

ponens being another's modus tollens. Rather, the argument for dualism based on the logical possibility of phenomenal zombies is guilty of begging the question. The argument for epiphenomenalism requires us to imagine a physically identical world without qualia, but we can only imagine this if we are already committed to epiphenomenalism—that is, if we are committed to the claim that qualia are not dynamically efficacious and so can logically be omitted from a world *dynamically* identical to our own. Given our extremely strong evidence for believing that qualia do affect the physical world (I say "red" because I *experience* redness), we are therefore left without justification for the epiphenomenalist position.

Chalmers has been misled by his over-reliance on the *geometrical* facts of our universe and his willingness to abandon causation in his analysis of reduction. Attention to causal properties reveals that we can only imagine zombies in the way that we imagine "physically identical" worlds that are devoid of properties like mass and charge—the worlds appear the same to the unwary mind's eye because we hold fixed the facts about the spatiotemporal configuration of our world, but the worlds are not *physically* identical because we have not taken care to fix the dynamics essential to the physics

VIII. Conclusions

The diagnosis of the intuition driving the belief that zombies are conceivable is that when we imagine our twin saying "red," we allow ourselves to imagine that the twin's utterance is *caused* by something other than what caused our own utterance. That is, we are holding *fixed* the spatio-temporally characterized events (in the way we do in a hologram world) without also bringing in the relevant *dynamics*. Given that we do not understand well how physical dynamics

leads to the brain activity that gives rise to consciousness, it is very easy for us to misjudge the actual causes at work.

Chalmers has claimed that both consciousness and causation fail to supervene logically on physical facts, and this has led him to speculate that there may be an important metaphysical connection between the two (Chalmers 1996, p. 86). I have argued here that reduction requires higher level causal facts to supervene on microphysical dynamics, and that we therefore have no reason to reject the causal efficacy or physicality of consciousness. I therefore heartily support Chalmers's suggestion that there is an important tie between consciousness and causation, but this tie is not something peculiar to consciousness. Our concepts of water and gasoline are also tied to their causal powers, as are our concepts of viruses and penicillin, fish and fireworks, wildfires and wombats.

It may even be that our concepts of qualitative states are more closely tied to the relevant causal properties than are many of our other concepts. What is the sensation of redness, for example, except the way that our mental processes are affected by things like ripe tomatoes? What is pain, but the way we are affected when, for example, we stub a toe? It is perhaps not surprising that if we set up our test for logical supervenience in a way that allows us to imagine away microphysical dynamics and higher-level causal processes, we can also imagine away conscious states.

References

Bokulich, P. 2004. "Functionalism, Fundamentalism, and Physicalism." (in preparation).

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York:
Oxford University Press.

Chalmers, D. J. and F. Jackson. 2001. "Conceptual Analysis and Reductive Explanation."
Philosophical Review 110:315-360.