Who's Afraid of Reductionism? The Study of Religion in the Age of Cognitive Science

Edward Slingerland

This paper aims to defend the application of tools and knowledge drawn from the natural sciences to the study of religion from the common charge that such approaches are overly "reductionistic." I will argue that "reductionism" is ultimately an empty term of abuse—any explanation worthy of being called an explanation involves reductionism of *some* sort. Drawing upon the work of Charles Taylor, I will try to explain what "good," non-eliminative reductionism—one that recognizes the reality of complex, emergent human-level structures of meaning—might look like. I will also argue that these human-level structures of meaning should not be seen as possessing special ontological status, but rather must be understood as grounded in the lower levels of meaning studied by the natural sciences, instead of hovering magically above them. Practically speaking, this means that scholars of religion need to start taking seriously discoveries about human cognition being provided by neuro- and cognitive scientists, which have a constraining function to play in the formulation of theories in religious studies. Moreover, adopting a "vertically integrated" approach—grounded in a post-dualist, embodied pragmatist perspective—will help the field of religious studies to get beyond the unhelpful, and intellectually paralyzing, social constructivist dogma that continues to inform most of the work in our field.

Edward Slingerland, Associate Professor of Asian Studies Canada Research Chair in Chinese Thought and Embodied Cognition, The University of British Columbia, British Columbia, Canada. E-mail: edward.slingerland@ubc.ca. Portions of this article are adapted from Slingerland 2008, Ch. 6, used with permission of Cambridge University Press.

Journal of the American Academy of Religion, June 2008, Vol. 76, No. 2, pp. 375–411
doi:10.1093/jaarel/lfn004

Advance Access publication on March 29, 2008

MY ATTENTION WAS RECENTLY caught by an essay in *JAAR* by Russell McCutcheon (2006a) that focused on the issue of reductionism, framed in the context of the always fraught relationship between religious insider and outsider. Against certain scholars of religion, such as Robert Orsi, who are guided by the Wilfred Cantwell Smith (1959) dictum that "no statement of a religion is valid unless it can be acknowledged by that religion's believers" (42),[1] McCutcheon points out that any interesting work of scholarship involves what Wayne Proudfoot calls "explanatory reduction" (741). This is true whether we are talking about Paul Courtright's (1985) psychoanalytic reinterpretation of the Ganesh story, Jeffrey Kripal's (1995) exploration of homoerotic themes in Hinduism (737) or—despite his claims to the contrary—Robert Orsi's (2005) analysis of the deeper, hidden intellectual baggage that drives contemporary American members of the AAR (738), or the God of the Pentecostals as a symbol of the power of the universal sacred (740–741). To reject reductionism tout court would involve, McCutcheon suggests, "the end of the human sciences as we know them" (736).

THE POSTMODERN FUNHOUSE OF MIRRORS

I entirely agree with this sentiment, but found myself at a loss when it came to reconciling it with the thoroughly social constructivist framework within which it was formulated. Elsewhere in the essay, McCutcheon argues that "all acts of signification ... are a translation of one set of claims into a language that is itself no closer than any other to some presumed authentic source of the Nile" (742), and echoes Baudrillard's claim that all we ever have is representations, "each competing for the chance to stand in for a Real that never was present to begin with" (743). The manner in which this sort of social constructivism has become the basic background dogma of our field was brought home to me by the subsequent exchange between McCutcheon (2006b) and Courtright (2006), where both prefaced their comments by approvingly citing the words (the Word?) of J. Z. Smith. McCutcheon declares that, whatever our differences concerning the relative status of religious insider versus outsider, as scholars of religion we all "follow J. Z. Smith and agree that it is history ... all the way down" (2006b: 756).

We *were* taught this by J. Z. Smith, and Clifford Geertz, and Judith Butler, and all of the other theorists assigned to us in graduate school

[1] Cited in McCutcheon (2006a: 722).

theories and methods courses—the idea that "history goes all the way down" is a deeply engrained truism in most of the core humanities departments, with only philosophy and perhaps rogue historians still valiantly trying to pin down the ever-receding "Real" that Baudrillard has informed us never existed. It is easy to understand why we, as scholars of religion, have become stuck in this postmodernist dead end. Impressed by the postmodern critique of Enlightenment ideals, objectivism no longer strikes us as a viable option; left without a theoretical basis for establishing certain knowledge, we appear incapable of extracting ourselves from the social constructivist quagmire, where the only standard for debate appears to be winning the title of more social-constructivist-than-thou. What has always puzzled me, however, is how absolute conviction in the "truth" of social constructivism can coexist with responsible, thoughtful scholarship—such as that evinced by McCutcheon and Courtright—that itself makes no sense in the light of social constructivism. It is hard to understand the Academy's continued insistence on historical and linguistic accuracy, coherence of argumentation, and textual and material evidence if, "in fact," all we are really doing is wandering around aimlessly down an endless hallway of distorting mirrors.

To be sure, for decades there have been humanities theorists arguing that both old-fashioned Enlightenment objectivism and its twin, postmodern constructivism, are predicated on a metaphysical dualism that needs to be transcended if we are to get, in Richard Bernstein's phrase, "beyond objectivism and relativism" (1983). So far, though, most attempts to shake off dualism have generally ended up getting no further than absorbing the natural world into the social. The philosophical hermeneuts and Rortian neo-pragmatists lead us back into a never-ending, hermetically sealed circle of conversation, Bruno Latour's "amodernism" (1993) dissolves into a morass of vague neologisms, and Bourdieu's otherwise promising attempt to focus attention on somatic knowledge is ultimately unable to get beyond the body as nothing more than a passive storehouse for socially constructed *habitus*—a "living memory pad" (1990: 68) or "depository of deferred thought" (69). Traditional humanists continue to accept a world divided between an inert kingdom of dumb objects, governed by deterministic laws, and the realm of the free and unconstrained spirit—a metaphysical divide expressed most clearly in German, where the sciences of mechanistic nature (*Naturwissenschaften*) are distinguished from the sciences of the elusive human *Geist* (*Geisteswissenschaften*). Within the confines of such a divided world, it is difficult to see how we humanists can ever

escape the solipsistic fate of Borges, endlessly and onanistically spinning stories inside of stories.[2]

This piece is motivated in part by the odd spectacle of otherwise quite thoughtful and analytically gifted scholars genuflecting before the altar of social constructivism before proceeding to make arguments that lose all of their bite if we take such constructivism seriously. I agree wholeheartedly with McCutcheon that reductionism is the bread-and-butter of what we do as humanists, but would want to push his defense of "explanatory reductionism" beyond the mind–body divide in order to defend an embodied approach to human culture—one that can actually break us out of the endless cycle of contingent discourses and representations of representations.

EMBODYING THE HUMANITIES

What I am going to be referring to as the "embodied" approach to the study of culture involves a loose collection of scholars who see the human mind and its products as part of the physical world, not as hovering somewhere above it, and who are therefore committed to breaching the *cordon sanitaire* that has traditionally divided the humanities and natural sciences.[3] This embodied approach claims no privileged access to eternal, objective truths, but argues that commonalities of human embodiment in the world can result in a stable body of shared knowledge, verified by proofs based upon common perceptual access. By breaching the mind–body divide—by bringing the human mind back into contact with a rich and meaningful world of things—this approach to the humanities starts from an embodied mind that is always already in touch with the world, as well as a pragmatic model of truth or verification that takes the body and the physical world seriously.

One of the inspirations for the embodied approach is the growing consensus coming out of the cognitive sciences that metaphysical dualism is not only philosophically problematic, but also empirically untenable: that the mind is the body, and the body is permeated through-and-through with mind. Consciousness, under this understanding, is not a mysterious substance distinct from matter, but rather an emergent property of matter put together in a sufficiently complicated way. Furthermore, it suggests that the manner in which

[2] Borges (1999: 324), quoted approvingly in McCutcheon (2006a: 743).

[3] For some representative introductions to the embodied approach, see Arbib (1985), Johnson (1987), Clark (1997), Lakoff and Johnson (1999), Pecher and Zwaan (2005), Gibbs (2006), and Thomson (2007).

we engage in the study of consciousness and its products—that is, the traditional domain of the humanities—should be brought into coordination with the manner in which we study less complex (or differently complex) material structures. This is the insight behind the arguments for an explanational continuum extending equally through the natural and human sciences that have recently and prominently been offered by, for instance, the ethologist E. O. Wilson with his call for "consilience" (1998), the evolutionary psychologists John Tooby and Leda Cosmides with their argument for the need for "vertical integration" (1992), and the neuroscientist and linguist Steven Pinker with his critique of the humanistic dogma of the "Holy Trinity" (the Blank Slate, the Noble Savage, and the Ghost in the Machine) (2002).

When it comes to religious studies, the embodied movement is exemplified by those who would apply the tools of cognitive science to religious phenomena—asking, for instance, about the innate cognitive constraints on religious representation and transmission (Boyer 1994, 2001), the origin and universality of teleological beliefs (Kelemen 1999, 2004), or the reason for disjunctions between religious dogma and everyday cognitive strategies (Slone 2004). Implicit in the cognitive approach to religion is a picture of the human person as an integrated body–mind system following the laws of nature and produced—like all of the other body–mind systems running around in the world—by evolution.

Although "embodiment" might have a vaguely appealing sound to most humanists, there are few things that more effectively incite the ire of religion scholars, anthropologists, literature, and art scholars than the word *evolution*—let alone the terms "sociobiology" or "evolutionary psychology." In religious studies, there has been extremely strong resistance to those who advocate applying evolutionary analyses or the tools of cognitive science to the human phenomenon of religion. In order to justify not engaging with the cognitive sciences, religion scholars often pull out the "we live in a post-Kuhnian world" canard about how reliable scientific claims are not possible—when, in fact, there have been a wealth of pragmatist defenses of empirical inquiry that respond to Kuhn and provide quite cogent post-objectivist models of science.[4] The most common rallying point against cognitive approaches to culture, however, is the charge of "reductionism." This is what I would like to focus on here, because our fear of reductionism gets to the heart of the enduring appeal of dualism and our resistance to genuinely embracing a nondualist model of the person. Below I would like to argue that we really *do*

[4] The reader is referred to Haack (2003), Hacking (1983, 2000), and Laudan (1996) in particular.

need to embrace such a model, and that we can do so without having to abandon the intuitions that make vertical integration seem so threatening. This will, though, involve confronting the genuine strangeness and distastefulness of the physicalist model of the self—something that is too often smoothed over by advocates of the embodied approach. It will also require countering some of the more hegemonic rumblings one *does* in fact hear from those with a natural scientific bent, as well as formulating a more nuanced picture of what counts as "real" for creatures like us than one usually finds in the writings of cognitive scientists.

TAKING THE BODY (REALLY) SERIOUSLY

The manner in which the embodied approach can break us out of the dualism that continues to underwrite relativist epistemological claims is, in my mind, best illustrated by a hilarious satire by Marty Smith (1987/1993) called "The Jean-Paul Sartre Cookbook" that has for years been spreading through the internet in various iterations. It purports to be the lost diaries of "a young Sartre obsessed not with the void, but with food," and determined to write "a cookbook that will put to rest all notions of flavor forever." Some representative entries:

> October 6
> I have realized that the traditional omelet form (eggs and cheese) is bourgeois. Today I tried making one out of cigarette, some coffee, and four tiny stones. I fed it to Malraux, who puked. I am encouraged, but my journey is still long.
>
> October 10
> I find myself trying ever more radical interpretations of traditional dishes, in an effort to somehow express the void I feel so acutely. Today I tried this recipe:
> Tuna Casserole
> Ingredients: 1 large casserole dish
> Place the casserole dish in a cold oven. Place a chair facing the oven and sit in it forever. Think about how hungry you are. When night falls, do not turn on the light.
> While a void is expressed in this recipe, I am struck by its inapplicability to the bourgeois lifestyle. How can the eater recognize that the food denied him is a tuna casserole and not some other dish? I am becoming more and more frustrated.

Although the target of this satire is the "individualistic constructivism" of French existentialism rather than poststructuralist theory, the basic

critique of social constructivism is the same. In a certain sense, of course, the satire is a cheap shot: neither postmodernism nor existentialism would deny human physical commonalities. What they *do* deny, though, is the existence of human commonalities at the level of meaning—human bodies as inert physical objects may be subject to a common set of laws, but this has little to do with the lived world of human significance. It is this latter world that is culturally constructed (or, for the existentialists, created by the individual *ex nihilo*), and despite vague animal preferences for cereal over cardboard or cherries over stones, it is this constructed world of human mediated experience that is all that we are really in touch with.

There are powerful theoretical and empirical reasons for thinking that this view of human cognition is wildly incorrect.[5] French existentialists in their dark Parisian cafés drank espresso with sugar rather than, say, dog urine, because of evolved and universal human preferences for stimulants and sugar, and these physical preferences are not different in kind from our preferences for light over darkness, strength over weakness, or truth over falsity. The humor-producing tension of the Sartre satire arises from the conflict between the existentialist assertion of a universe without meaning and the obvious truths of everyday human life: certain things taste good, certain things look good, certain actions make sense, and this ineluctable horizon of significance cannot be erased by a sea of black coffee or a mountain of Galoises.

This is not to deny the power and poetry of the existentialist position—one would have to be dead not to be moved by the quietly courageous and resolutely lucid stance of Camus's *homme absurde* as portrayed in the *Myth of Sisyphus* or *The Plague* (Camus 1942, 1947). But Camus's gift as a writer and rhetorician is what in fact invalidates his basic philosophical point, because—despite his claim that he rejects any "scale of values" (1947: 86)—the very power of his ideal is derived from predetermined and universal human values: being awake is better than being asleep; being clear is better than being muddled; being strong and courageous is better than being weak and cowardly.[6] Camus's creativity consists in his effort to recruit these universal

[5] See Barkow *et al.* (1992), Buss (2005), Carruthers *et al.* (2005, 2006), Gibbs (2006), Hirschfeld and Gelman (1994), Pecher and Zwaan (2005), and Pinker (2002).

[6] Cf. Charles Taylor's comment regarding Sartre that "self-choice as an ideal makes sense only because some *issues* are more significant that others. I couldn't claim to be a self-chooser, and deploy a whole Nietzschean vocabulary of self-making, just because I choose steak and fries over poutine for lunch. Which issues are significant, *I* do not determine. If I did, no issue would be significant ... To shut out demands emanating from beyond the self is precisely to suppress the conditions of significance, and hence to court trivialization." (1992: 39–40).

normative reactions and map them in a quite novel manner: lucidity consists in knowing nothing for certain, and courage consists in rejecting those transcendent truths which once were perceived as requiring strength to defend against unbelief. The mappings are new, but the sources are probably as old as *homo erectus*. Similarly, despite postmodernist social constructivist posturing, the motivations and goings-on at any given annual AAR or MLA meeting would, with a little bit of background explanation, be perfectly comprehensible to Pleistocene hunter-gatherers: friendship, intellectual curiosity, coalition recruitment, exchange of adaptive information (including a heavy dose of social gossip), and an overall direct or indirect goal of achieving security, prestige, power, and sexual access.[7]

Despite Camus's anguished claims, there is no absurd gap between our need for transparent certainty and a dense world devoid of meaning. The world *is* reasonable—not in the sort of transcendent, absolute sense that Camus rightly dismisses as wishful consolation, but in an eminently embodied, anthropocentric sense. The process of evolution ensures that there is a tight fit between our values and desires and the structure of the world in which we have developed. No appeal to eternal verities is required to assure us that a cigarette and stone omelet would make even Malraux puke. Of course, human beings are apparently unique among animals in possessing the cognitive fluidity and cultural technology to effect some radical changes in what gives us pleasure, what we find worth pursuing, and what we deem as meaningful. But all of this cognitive and cultural innovation is grounded in—and remains ultimately constrained by—the structure of our body-minds.

THE PROBLEM WITH EMBODIMENT: DARWIN'S DANGEROUS IDEA

So far, so good—at least for many of us who are tired of spinning our rhetorical wheels, and would like to move beyond the intellectual quagmire of strong postmodern relativism. It is important, though, to fully confront the implications of the embodied model of the human mind. To wit: the human mind is coterminous with the human body (especially the brain), and this body-brain is no more than a very complex physical thing, a product of millions of years of evolution. Human thought is not a ghostly, disembodied process, but rather a

[7] A point made with grace, sympathy, and humor by the novelist David Lodge (1975, 1984, 1989).

series of body–brain states—a series of physical configurations of matter—each causing the next in accordance with the deterministic laws that govern the interactions of physical objects.[8] Ideas, as physical states of matter, can interact with one another, blend with one another, and transform themselves in predictable ways, but there is no super-physical soul or self, outside of the chain of physical causation, controlling or overseeing the process. This means that our thoughts and behavior are, at least in principle, as predetermined and predictable as any other physical process. It also means that the self as we ordinarily understand it—as a disembodied something, soul or spirit or mind, caused by nothing other than itself—is nothing more than an illusion created by the workings of our embodied brain. This picture of the human mind/self/soul, the inevitable conclusion of "Darwin's dangerous idea" (Dennett 1995), is summed up vividly and succinctly in a quotation from the Italian philosopher Giulio Giorello: "Yes, we have a soul, but it is made up of many tiny robots."[9]

The late Francis Crick (1994), who spent the last part of his career exploring the neuroscience of consciousness, called this idea that "*all* aspects of the brain's behavior are due to the activities of neurons" (259)—that is, that consciousness can ultimately be reduced to a physical chain of firing neurons—the "astonishing hypothesis." It is, in fact, more than astonishing: the physicalist view of the human self and the human mind is alien and profoundly disturbing. However odd, this thoroughly materialist view of the self must be grappled with because it is difficult to see what choice we have once we take the decisive step of giving up our belief in a Cartesian ghost in the machine—of believing, to put a finer point on it, in magic. Unless we are prepared to invoke

[8] Those uncomfortable with the idea of the brain as a deterministic, physical system have pinned a great deal of hope on the idea that the brain might be a quantum system, and thus not subject to the constraints of classical physics. Roger Penrose (1989), for instance, has famously argued that quantum indeterminacy is the locus of human free will. There are two fatal problems with this strategy. First of all, there is the inconvenient fact that the desired quality of indeterminacy is present only at the quantum level: once we get up into levels that are humanly relevant, such as that of neurons or hormones, Laplacian determinism re-exerts its iron grip (see Tegmark 2000 for experimental data confirming that the brain seems to be functioning as a classical rather than quantum system). A more basic and fatal problem, however, is that indeterminacy is nothing more than randomness, which is not really what defenders of strong free will are after (Searle 2004: 24–25). The human conception of free will requires that this will be determined by *something*—reasons, desires, spontaneous impulses, etc. Free will as utter randomness is as horrific a concept at a human level as the deterministic absence of free will. See Dennett (1995), Flanagan (2002), and Searle (2004) for more detailed, cogent critiques of the quantum mechanics/free will argument.

[9] From Giorello (1997), quoted in Dennett (2003: 1). See Dennett (1995) for a powerful, comprehensive account of the implications of Darwinism for our model of the self.

supernatural belief, it is hard to avoid the conclusion we are "little robots" all the way down.

This absurd idea of a thoroughly mechanistic cosmos—absurd in Camus's sense of the word—gets to the heart of humanistic concerns about reductionism. To begin with, thorough-going physicalism goes against a very basic and profound intuition that there is something special about humans: everything that we value is predicated on the conviction that people are not things. To associate people—or people-level phenomena such as art, literature, and religion—with the realm of things therefore strikes us as fundamentally wrong. Moreover, qua humanists, we are also no doubt irked by the implication that we might be out of a job: what need is there for the *Geisteswissenschaften* if there is no more *Geist*? Are literature and religious studies departments to be absorbed into neuroscience? Getting some clarity about what is valid and not valid about charges of "reductionism" will allow us to see how different levels of explanation can, and must, coexist with one another. It will also give us a sense of how we can, and must, acknowledge the reality of Darwin's dangerous idea, while still continuing to live and work in an environment rich with human meaning.

THE BOGEYMAN OF REDUCTIONISM

To begin with, it is important to realize that any truly interesting explanation of a given phenomenon is interesting precisely *because* it involves reduction of some sort—tracing causation from higher to lower levels or uncovering hidden correlations. We are generally not satisfied with explanations unless they answer the "why" question by means of reduction: by linking the *explanandum* to some deeper, hidden, more basic *explanans*. As McCutcheon argues quite cogently in his essay, this is why the manner in which even traditional humanist scholars go about their work is by its very nature reductionistic. Reduction is what we do as scholars, humanistic, or otherwise, and when someone fails to reduce we rightly dismiss their work as trivial, superficial, or uninformative.

When the deeper principles behind things are poorly understood—that is, when lower levels of causation underlying phenomena we are interested in explaining are not accessible to our prying—we are often forced to invent vague, place-holder entities to stand in for the missing information. Sometimes we are aware that this is what we are doing. For instance, Mendel could reason about the inheritance of traits without knowing how information about them was physically instantiated or transmitted, and Darwin could similarly map out the implications

of natural selection without any clear conception of the substrate of inheritance. In such cases, there is an implicit faith that the lower-level entities and processes will eventually be specified; if not, the theory may have to be abandoned. A discipline can find itself in a dead end, however, when it has postulated vague, place-holder entities without realizing that this is what it is doing—when it takes these unspecified and unknowable entities or faculties to have genuine explanatory force. This is what I see as the essence of Nietzsche's point in a wonderful passage where he mocks Kant's analysis of synthetic *a priori* judgments:

> "How are synthetic judgments *a priori possible*?" Kant asked himself—and what really is his answer? "*By virtue of a faculty*" (*Vermöge eines Vermögens*)[10]—but unfortunately not in five words, but so circumstantially, venerably, and with such a display of German profundity and curlicues that people simply failed to note the comical *niaiserie allemande* involved in such an answer. (1886/1966: 18–19)

Nietzsche compares the "explanation" offered by Kant to the answer of the doctor in Molière's *Tartuffe* to the question of how opium produces sleep: "Because it contains a sleepy faculty, whose nature is to put the senses to sleep." "Such replies belong in comedy," Nietzsche concludes, and so should we.

The force of the argument of cognitive scientists and evolutionary psychologists who are pushing for vertical integration between the humanities and the natural sciences is that the humanities have yet to genuinely free themselves from this sort of "Tartuffery," and continue to rely on impressive-sounding but explanatorily empty entities and faculties. For instance, Tooby and Cosmides note that what they refer to as the "Standard Social Scientific Model"—social constructivism—is satisfied by the explanation that the blank slate of human nature gets filled up by means of "learning," which is about as helpful an explanation as that living things are made of "protoplasm." Just as the mysterious protoplasm that featured so prominently in early biology turned out to consist of a collection of distinct intricate structures with specific functions, so, Tooby and Cosmides argue, will words like "learning," "intelligence," and "rationality" turn out to be blanket terms for what are really a variety of specific, modular, evolved cognitive processes that allow human beings to selectively extract and process adaptively relevant information from the world (122–123).

[10] Lit. "By means of a means."

All of us trained in religious studies were weaned on Geertz, who of course spent most of his career vociferously policing the ontological divide between the *Geisteswissenschaften* (the realm of *Verstehen*, or uniquely human "understanding") and the *Naturwissenschaften* (the realm of *Erklären*, or mechanistic explanation)—or, to use categories that he borrowed from Gilbert Ryle, between "thick description" and "thin description." Ryle had asked us to consider a scenario involving two boys contracting their eyelids, one in an involuntary twitch and the other in a conspiratorial wink, claiming that there exists an "immense but unphotographable difference" between the two (Ryle (1971): 480). This difference, Ryle argues, can only be captured by "thick description," which goes beyond the mere physical to what Geertz refers to as the "semiotic meaning" of the gesture (1973: 6). This semiotic approach to the study of human behavior has the effect of systematically denying any possible substantive role to "thin" bodily or physical processes, thereby protecting the interpretation of culture from the prying eye of science by wrapping it in the mysterious cloud of *Verstehen*.

Although this seems more sophisticated than the answer of the doctor in Molière, it is structurally quite similar. To begin with, it is simply not the case that the difference between a twitch and a wink is "unphotographable"—that is, inaccessible to third-person description. Human beings can generally distinguish twitches from winks instantly and effortlessly, in the same way we tell conscious (and thus false) from spontaneous (and thus sincere) smiles: they involve entirely different neural pathways and sets of muscles, and the resulting slight difference in appearance is something to which our minds are exquisitely sensitive (Ekman 2003). Any complete story of how we tell a wink from a twitch is going to have to involve an account of an entire suite of functionally specialized and physically grounded cognitive mechanisms that have for millennia allowed our ancestors to form coalitions, detect cheaters, and suss out potential enemies. The fact that any of these steps can be selectively knocked out by localized brain damage suggests that the empirical self plays more than just a place-holding role in the process of human understanding. Following Geertz and invoking mysterious words like *Verstehen* to "explain" how we know that a wink is a wink in fact explains nothing, and misses the point that recognition of the "semiotic meaning" of this gesture is grounded in an embodied mind that is amenable to empirical investigation.

Of course, Ryle and Geertz can also be understood as making the more plausible point that the larger meaning of a particular wink—Why is *this person* winking at me? What should I do?—is embedded in a set of long, complex stories, and that for the unpacking

and analysis of these stories we require the higher-level expertise of anthropologists, novelists, and historians. Such humanistic work, however, should not be seen as occurring in an explanatory cloud-cuckoo land, magically hovering above the mundane world of physical causation. Human-level meaning emerges organically out of the workings of the physical world, and we are being "reductive" in a good way when we seek to understand how these lower-level processes allow the higher-level processes to take place.

FROM PHYSICALISM TO THE HUMANITIES: LEVELS OF EXPLANATION

Having sent the bogeyman of reductionism back to its cave, it is now possible to talk about good and bad forms of reductionism—because, of course, it is really "greedy" or "eliminative" reductionism that most humanists have in mind when they bandy about this charge. In order to distinguish productive, explanatory reductionism from crudely eliminative reductionism, it is important to get some clarity about the heuristic and ontological status of entities at various levels of explanation.

Levels of Explanation and Emergent Qualities

Although no evolutionary psychologist or cognitive scientist would purport to be an eliminative reductionist, and all give lip-service to the idea that higher levels of explanation can feature emergent qualities not present at the lower levels, there is a common tendency to nonetheless privilege the material level of explanation: we are "really" just mindless robots or physical systems, no matter how things might appear to us phenomenologically. There are some very good reasons for this privileging of lower levels of explanation. To begin with, the physicalist stance has proven extremely productive, allowing such dramatic technological developments as supercomputers and pharmacological treatments for mental illnesses. Moreover, there is an *a priori* reason for giving precedence to the physical: the structure of the various upper levels of explanation emerges out of and depends upon the lower levels, so the lower levels *are* causally privileged in this way. Molecules form and behave in accordance with more basic principles that govern both inorganic and organic substances, which means that a hypothesis in molecular biology that violates well-established physical chemistry principles is wrong, or else a reason for us to rethink our physical chemistry.

It is equally the case, however, that as we move up the explanatory chain we witness the emergence of what appear to be new entities,

which possess their own novel and unpredictable organizational principles. The field of organic chemistry is based upon principles that emerge at the level of organic molecules, which cannot be fully predicted from the perspective of physical chemistry. Similarly, no amount of intimacy with quantum mechanical principles will allow one to even begin to predict the behavior of macro-level solid objects. As Hilary Putnam (1973) famously observed, there are entire fields of human knowledge, such as geometry, that emerge only once we reach the level of macro objects. The fact that a square peg 15/16″ on a side will fit through a 1″×1″ square hole, but not 1″ diameter circular hole, is a function of the peg's geometric properties; referring to the properties of the molecules that make up the pegs or the materials from which the holes are drilled would be heuristically useless.

This mutual dependence and interaction of levels of explanation is taken for granted in the natural sciences, and is in fact one of the guiding principles driving natural scientific inquiry. The challenge for defenders of true vertical integration is hooking the various levels of explanation in the humanities into their proper—and emergent—place at the top of this causal explanatory chain.

Weak Versus Strong Emergence: Blocking the Move to Mysterianism

We are familiar with how the process of evolution and natural selection has produced more and more complex feeding and fleeing machines, working at many different layers in the food chain, as well as wildly diverse strategies of hunting, mating, parenting, and social organization (Dawkins 1976/2006). Very crude survival machines are built to sense temperature and inorganic nutrient gradients and to adapt their movements and simple feeding behaviors accordingly. More complex ones are then built to take advantage of the work already done by these simple machines in concentrating diffused inorganic nutrients in one valuable package: they are the first predators, and require more complicated sensory and behavioral programming to track down and capture their prey. The prey, in turn, become more complex in response to this pressure, acquiring the ability to detect and evade predators. At a certain point in this process of exponentially increasing complexity, trying to rely on the physical stance—still helpful for dealing with simple rocks and trees and coconuts—was simply no longer fast enough.

It is at this point that it became more efficient for certain particularly complex survival machines to begin viewing other complex survival machines as more than mere objects subject to the laws of physics: to see them as *agents*, propelled by invisible "desires," "fears,"

and "preferences." We cannot directly observe such mental properties, but human beings appear to be so constituted as to inevitably and irresistibly *see* them constantly at work in the world, and this seems to be an extremely helpful heuristic—evolution would not have built it if it were not. As Daniel Dennett observes, for organisms with limited processing ability and time, the set of short-cut assumptions provided by postulating the existence of mental entities—the adoption of what he refers to as the "intentional stance"—provides huge computational leverage. "Predicting that someone will duck if you throw a brick at him is easy from the intentional or folk psychological stance; it is and will always be intractable if you have to trace the photons from brick to eyeball, the neurotransmitters from optic nerve to motor nerve, and so forth" (1995: 237).

This set of beliefs related to mental states has come to be referred to by the broader cognitive scientific community as "theory of mind"[11]—"theory"-like because it goes beyond the available data to postulate the existence of unobservables, causing us to "paint" mental properties onto the physical world. Human beings throughout history and cross-culturally appear to irresistibly see their world as populated by "agents," which unlike objects or plants harbor goals and desires, experience emotions and thoughts, make choices, and are propelled by a special sort of internal causality we can term "intentionality." From a very early age, human beings conceive of intentionality as a distinct kind of causality, and distinguish it from both the billiard-ball contact causation that characterizes "folk physics" and the teleological, "vitalistic" causation proper to all living things.[12] We would be very surprised if something we considered an object hopped up and started moving around on its own, and would be forced to reclassify it as an agent. Similarly, we expect a plant to slowly grow upward toward the light, but a plant that moved in (humanly) real time—that, for instance, walked over to the corner for a drink of water, and then back out in the courtyard to plop itself down under the sun again—would trigger our theory of mind (as well as prompt us to keep our doors locked at night). Under the weak emergence view, the reason we chop up the world like this—into objects, growing things, and agents—is because this division has historically worked: agent-like intentionality and mental concepts

[11] Perhaps the best recent (and quite readable) introduction to theory of mind is Bloom (2004); also see Wellman (1990), Baron-Cohen (1995), Spelke *et al.* (1995), and the essays collected in Carruthers and Smith (1996).

[12] For an introduction to and defense of innate human ontology and modularity theory, see the essays collected in Hirschfeld and Gelman (1994).

emerge as useful ways to think about the world once collections of matter get put together in a certain way.

Although mental concepts such as "wanting" and "believing" are heuristically indispensable, and present themselves irresistibly to the human mind as crucial features of causal explanation, most cognitive scientists believe that there is no reason to see them as composed of genuinely novel stuff. Unless we are to resort to metaphysical dualism, it is hard to see what "wants" or "beliefs" could be *made of*, if they are something other than particular states of neurons and other cells in our bodies. As Michael Arbib puts it, referring to folk psychology concepts or "person-talk," they "are useful for encapsulating meaningful patterns of what our brains can do, but not as describing a distinct reality" (1985: 115). Opposing this "weak emergence" position are the advocates of various forms of "strong" or "ontological" emergence, referred to by Owen Flanagan as the "mysterians" (1992: esp. 8–11). These include, as we would expect, old-fashioned substance dualists, who claim that mind and matter are two independent ontological realms. Descartes is the classic exponent of this position, and—despite the ill repute into which Cartesianism has fallen in recent decades—full-blown substance dualists are still fairly thick on the ground. Those who adhere to traditional religious models of the self are obviously and explicitly dualistic in this sense, but most secular humanist substance dualists probably fall into the category of what Patricia Churchland refers to "boggled skeptics" (1986: 315): it just seems impossible to believe that any amount of physical complexity could produce consciousness. John Locke (1690/1975) expressed the boggled skeptic position quite clearly: "For it is as impossible to conceive that ever bare incogitative Matter should produce a thinking intelligent Being, as that nothing should of itself produce Matter" (623). This is not an unreasonable argument, despite its apparent simplicity: conscious beings seem to be able to do things that completely fly in the face of what we know about the behavior of inert matter. The conclusion that there has to be something else involved is therefore quite hard to avoid.

A more updated version of substance dualism is the so-called "property dualism," which argues that things like human "qualia" are ineffable and possess strongly emergent properties. "Qualia" is a technical-sounding philosophical term for what is, in fact, a quite folksy idea: there is a "what-it-is-like-ness" to my conscious experience that is immediately and exclusively accessible only to myself, and that this special qualitativeness is what would be left out of any third-person description of my experience. Thomas Nagel provided the classic

statement of this position in his famous 1974 essay, "What is it like to be a bat?", where he argued that, essentially, we can never answer that question: we are not bats, and no matter how much third-person descriptive knowledge we accumulate about bat behavior and physiology, we can never have access to the first-person (first-mammal?) qualia of bat consciousness. A similar argument has been famously defended for years by John Searle, who argues that no third-person, purely physicalist account can capture the "original intentionality" or "ontological subjectivity" that is an essential characteristic of human consciousness (1983, 1992, 2004). These ideas about original intentionality or ineffable qualia get to the heart of what makes *Verstehen* seem fundamentally different to us than *Erklären*: we can explain away the behavior of objects, but human-level meaning can only be grasped by the free apprehension of a fellow human spirit.[13]

Despite their intuitive appeal, none of these arguments in favor of body–mind dualism seems sustainable in the face of modern cognitive science. To begin with, the qualia "argument," as intuitively appealing as it is, is essentially an item of faith or bald assertion rather than an argument *per se*.[14] The one place where a committed dualist might make a stand is the "boggled" argument, which until quite recently has been very difficult to refute. The last few decades, however, have seen the development of a crucial bit of evidence tipping things in favor of the physicalist view of consciousness: the development of artificial intelligence, which has finally put to rest the claim that no amount of physical complexity could produce consciousness-like phenomena. As Dennett has argued, we have now built machines, which we know are just machines, that are capable of defeating Grand Masters at chess, passing the Turing Test—i.e., plausibly holding up their end of a free-form conversation—and demonstrating many of the powers that were previously seen as the exclusive province of conscious, intentional agents. Dennett observes that "the sheer existence of computers has provided an existence-proof of undeniable influence: there are mechanisms—brute, unmysterious mechanisms operating according to routinely well-understood physical principles—that have many of the competences heretofore assigned only to minds" (2005: 7).

Despite its apparent empirical untenability, we have to acknowledge that the boggled argument—like faith claims about "qualia" or

[13] This is the same intuition behind Robert Orsi's insistence that people are not "data" (quoted and discussed in McCutcheon 2006a: 721–722).

[14] See, especially Dennett (1991, 1995) and Putnam (1999) for eloquent and convincing critiques of both qualia and Searlian "ontological subjectivity."

"ontological subjectivity"—clearly taps into a deeply ingrained human intuition. Why should we want to block these sorts of moves to ontological emergence if they come so naturally to us, and take so much work to get away from? Contrary to some doctrinaire physicalists, there is nothing about physicalism *per se* that makes it uniquely scientific. If we had an accumulation of a critical mass of replicable evidence for existence of some non-physical, causally efficacious, intention-bearing substance, it would be unscientific *not* to be a dualist—and of course we cannot rule out the possibility that such a point will ever be reached.[15] It is just that no one has come up with a story that would explain how something like a soul could exist in the world as we currently understand it, although we are, *qua* human beings, highly motivated to come up with and to believe such stories. In the absence of an empirically defensible account of dualism, the explanation of reality that best enables us to get a grip on the world does not involve ghosts, souls, miracles, or original intentionality: human beings, like all of the other entities that we know about, appear to be robots all the way down, whether we like that idea or not.

THE LIMITS OF PHYSICALISM: WHY WE WILL ALWAYS BE HUMANISTS

Having hopefully blocked the move to mysterianism or ontological emergentism, I would now like to address in more detail the issue of why these intellectual moves are so compelling to us, as well as what this compulsion *does* reveal about the special status of human-level concepts. I think that John Searle is engaging in a bit of philosophical slight-of-hand when he purports to be a biological materialist, but then continues to insist upon a special ontological status for human subjectivity. Searle is, however, a brilliant philosopher with a quite detailed grasp of the state of the field in the cognitive and neurosciences: why this refusal to relinquish the idea of two distinct ontologies? And why two, we might ask, and not three, or ten? In this section, I would like to

[15] I here take issue with Searle's claim that physicalism functions as a modern religious dogma, accepted "without question" and with "quasi-religious faith" (2004: 48). I agree that many—if not most—physicalists are dogmatists as well, but dogmatism is not intrinsic to the position. If we are comfortable jettisoning Cartesian metaphysics and epistemology, physicalism or naturalism can be defended against the background of a modest, pragmatic conception of truth (see Slingerland 2008, Ch. 5 for a more extended argument). Wesley Robbins (1994) takes such an approach in arguing against Alvin Plantinga's (1993) famous position that naturalism is self-refuting: we do not require analytic certainty in order to conclude that naturalism is our current best explanation of the universe.

explore the intuition that I think motivates the defenders of dualism in all of its various forms: the recognition that human-level reality—reality as seen through the filter of theory of mind—is real for humans, and that it is so deeply entrenched that no third-person description can ever completely dislodge it. In other words, we apparently cannot help at some level seeing a *Geist* in the machine, which means there will always be something importantly different about the *Geisteswissenschaften.*

Why Physicalism does not Matter

Hard-core physicalists such as Dennett are inclined to dismiss positions such as Searle's or Nagel's as a mere statement of religious belief or personal sentiment. Dennett and some other advocates of vertical integration argue that, since intentionality and consciousness are helpful for certain heuristic purposes, but possess no underlying reality, the rigorous study of human affairs will eventually be able to dispense with them entirely.[16] A common analogy drawn by those who feel dualism will soon go the way of bell bottoms and disco balls is the shift in human sensibilities that occurred with the Copernican revolution. Copernicanism presented a view of the solar system that contradicted not only Scriptural authority but the evidence of our senses: the Bible states quite clearly that the sun moves around the earth, and this also happens to accord with our everyday sensory experience. Yet an accumulation of empirical evidence eventually resulted in Copernicanism winning the day—trumping both religion and common sense—and nowadays every educated person takes the heliocentric solar system for granted. Dennett argues that the current physicalism versus dualism controversy is analogous to the early days of Copernicanism: we are resistant to physicalism because it goes against our religious beliefs and our common sense, but the weight of the empirical evidence is on its side. Eventually—after all of the controversy has played itself out—we will learn to accept the materialist account of the self with as much equanimity as the fact that the earth goes around the sun (1995: 19).

A basic problem with Dennett's position, however, is that there is a profound disanalogy between the Copernican revolution and the revolution represented by physicalist models of the mind. The Ptolemaic model of the solar system falls quite naturally out of the functioning of our built-in perceptual systems, but it is not itself part of that system: we do not appear to possess an innate Ptolemaic solar system module.

[16] Cf. Owen Flanagan's comment that since concepts such as the "soul" or "free will" "don't refer to anything real, we are best off without them" (2002: xiii).

Switching to Copernicanism, at least intellectually, requires us to suspend our common sense perceptions, but it does not involve a direct violation of any fundamental, innate human ideas. Physicalism as applied to human minds *does* require such a violation, and this has a very important bearing on how realistic it is to think that we can dispense with mentalistic talk once and for all. Flanagan characterizes dualism as something that has troubled us "for centuries" (2002: 8), but seeing agents as something special goes back for at least as long as people have had theory of mind—perhaps 100,000 years.[17] This is the psychological fact behind the argument forward by Searle and others that consciousness is special: it is inescapably real for us.

We Are Robots Designed Not to Believe that We Are Robots

The idea of human beings as ultimately mindless robots, blindly "designed" by a consortium of genes to propagate themselves, has so much difficulty gaining a foothold in human brains because it dramatically contradicts other factory-issued and firmly entrenched ideas such as the belief in *soul*, *freedom*, *choice*, and *responsibility*—in short, all of the qualities that seem to us to distinguish human beings from mere things. The dualism advocated by Plato and Descartes was not a historical or philosophical accident, but rather a development of an intuition that comes naturally to us, as bearers of theory of mind: agents are different from things. Agents actively think, choose, and move themselves; things can only be passively moved. The locus of an agent's ability to think and choose is the mind, and because of its special powers the mind has to be a fundamentally different sort of entity than the body. Even cultures that did not develop a doctrine of strong mind–body substance dualism—such as the early Chinese—nonetheless believed that there was something special about the mind. As the fourth-century BCE Chinese thinker Mencius put it, what distinguishes the heart-mind (*xin*)—the locus of agency in human beings—from other organs of the body is that it issues commands, whereas the other parts of the body merely follow them (Lau 1970: 168).

It is the mind that is felt to be the locus of human free will, as well as the dignity and responsibility that goes along with such autonomy.

[17] Archaic modern humans have been burying their dead for at least 92,000 years (Bar-Yosef 2006), and elaborate, ritualized burial is as good a litmus test as any of the presence of theory of mind. When an implement breaks, you throw it away, and the remains of living prey are disposed of as quickly and conveniently as possible. Special treatment of the human corpse indicates that a shift has occurred, and the human body is now being viewed as linked to something fundamentally distinct from objects.

Most of us also have a powerful sense, whether we would be willing to defend it or not, that this something special about a person is not identical to the mere collection of their cells: the feeling that the most important part of a person—especially ourselves and the people whom we love—might somehow subsist after death presents itself spontaneously and quite powerfully to human beings, appears to be universal, and takes quite a bit of cognitive work to overcome. In other words, although we are obviously capable of entertaining non-dualist ideas at some abstract level, we seem to have evolved in such a way as to be ultimately invulnerable to the idea of thorough-going materialism.

The cognitive module producing this fundamental intuition is theory of mind, which we humans seem inclined to project onto pretty much *anything* moving in a particular kind of way: geometric shapes in a short animation, for instance, or single dots moving around on a screen appear irresistibly to us to be involved in goal-directed, mentalistic behavior, and for this reason engage our sympathy (Kuhlmeier *et al.*, 2003; Barrett *et al.*, 2005). Human beings, no matter how professionally or intellectually committed to physicalism, seem to feel a constant compulsion to project agency onto the inanimate (Guthrie 1993). We are all familiar with this experience, having to deal daily with stubborn, diabolical computers bent on erasing our data or crotchety old cars that refuse to start. The anthropomorphic drive seems to be universal, and appears quite early in development. Deborah Kelemen (1999, 2004) has documented the widespread projection of invisible or supernatural agency onto the world—what she refers to as "promiscuous teleology"—in children of various ages and education levels, and argues that agent-centered, teleological explanations for phenomena seem to be the human cognitive default position, only gradually, with difficulty, and incompletely dislodged by mechanistic explanation. We are obviously capable of withdrawing our projections when we have to—recognizing that our computer is not really out to get us—but it takes cognitive effort, which suggests that it does not come naturally and is not easily sustainable.

It is thus a mistake to say that we will ever completely dispense with mentalistic concepts, or ever entirely succeed in withdrawing our projections from the world. For instance, it is a rare cognitively intact person who can listen to Nina Simone's rendition of the song "Feelin' Good"[18]—a joyful celebration of "birds flying high," "rivers running

[18] *I Put A Spell On You*, 1965; song written by Anthony Newley and Leslie Bricusse for the 1966 musical *The Roar of the Greasepaint—the Smell of the Crowd.*

free," and "butterflies havin' fun"—without feeling in their bones the emotional and mental contagion that is constantly taking place between human beings and their world. The sight of "reeds driftin' on by" can make us feel calm, and a feeling of calmness can color our perception of the reeds. Rivers really do seem to run free, and the play of butterflies cannot help but seem fun to us—even though, qua scientists, we know at some level that nothing is really going on except water molecules being drawn downward by gravity and some large insects engaged in a random feeding pattern. Most importantly, feeling this kind of resonance between our own concerns and the functioning of the universe makes us feel really, really *good.*

This suggest that our promiscuous teleology and overactive theory of mind play more than a merely accidental and peripheral role in the economy of the human psyche, and are therefore not as dispensable as some might think. As the basis of perceiving meaning in the world, theory of mind would appear to be the foundation of any kind of long-term, large-scale motivation. I can be moved to engage in short-term, limited acts—consuming a cheeseburger when hungry or seeking out sleep when tired—without inquiring into the "meaning" of what I am doing, but the universal and pervasive tendency of human beings to tell and hear stories answering the question *why* suggests that long-term planning and motivation requires such a sense. This feeling that our work or our life has a purpose involves embedding it in an at least implicit narrative, and the agent-centered nature of such narratives suggests that the human ability to remain motivated over the course of long-term, multi-step, delayed-gratification tasks may involve the evolutionary hijacking of reward centers in the brain whose original or proper domain is interpersonal approval and acceptance. In cognitively fluid humans, reward expectancy over long-term tasks may be maintained at least in part by the feeling that some metaphorical conspecific "up there" is watching and approving or disapproving of our actions, or (in its modern iteration) a more diffuse, non-theistic sense that what we are doing "matters"—a conceit that makes no sense unless we project some sort of abstract, metaphorical agency onto the universe. In severe depressives, we may see a breakdown of this system: deeply depressed individuals genuinely *do* seem to perceive the world as unfeeling, mechanistic, and meaningless all the way down. The result is not a feeling of clarity or power, however, but profound behavioral paralysis and overwhelming suicidal tendencies.[19] Evolution is a tinker, and

[19] A chilling literary portrait of such a state is provided by William Styron (1992).

when faced with the task of getting live-in-the-moment social animals to start thinking in more complex and indirect ways about the long term, it simply co-opted a previously existing and very big carrot and stick. Pre-human social animals are powerfully motivated to shape their behavior in such a way as to win the approval and avoid the approbation of their literal social group. The great cognitive innovation that led to us—cognitive fluidity, the ability to project from one domain to another—perhaps also enabled literal social approval and disapproval to be projected onto a much larger scale: not just our immediate tribe, but the cosmos itself.

We will thus apparently always see meaning in our actions—populating our world with "angry" seas, "welcoming" harbors, and other human beings as unique agents worthy of respect and dignity, and distinct from objects in some way that is hard to explain in the absence of soul-talk, but nonetheless very real for us. We will continue to perceive our work, families, and lives as being "meaningful" at some inchoate level, and be strongly motivated to make the appropriate changes whenever we begin to lose this sense.

Qua physicalists, we can acknowledge that this feeling is, in some sense, an illusion. For better or worse, though, we are apparently designed to be irresistibly vulnerable to this illusion. In this respect, Appearance *is* Reality for us human beings. This is where, in fact, we see the limits of a thoroughly "scientific" approach to human culture, and need to finesse a bit our understanding of what counts as a "fact" for beings like us.

HUMAN REALITY IS REAL

Humanists and natural scientists concerned with the issue of levels of explanation and emergent properties have much to learn from the work of Charles Taylor. Taylor has grappled with vertical integration and come away unimpressed, and sees his mission as defending humanism against the reductionistic threat posed specifically by sociobiology, and more generally by the broader "naturalistic" bent of the modern world. We do not have to follow Taylor to his conclusion, which is essentially to reaffirm the Cartesian gulf between the *Geistes-* and *Naturwissenschaften*, in order to feel the power of his basic position. His conception of human-level reality provides us with a nuanced, sophisticated model for understanding the place of the person in the great physicalist chain of causation.

One of Taylor's most important points in the opening chapters of his classic *Sources of the Self* (1989) is that human beings, by their very

nature, can only operate within the context of a normative space defined by a framework of empirically unverifiable beliefs. The Enlightenment conceit that one can dispense with belief or faith entirely, and make one's way through life guided solely by the dictates of objective reason, is nothing more than that: a conceit, itself a type of faith in the power of a mysterious faculty, "reason," to reveal incorrigible truth. In addition to the panoply of "weak evaluations"—such as a preference for chocolate over vanilla ice cream—that we are familiar with, humans are also inevitably moved to assert "strong" or normative evaluations. This latter type of evaluation is based on one or more explicit or implicit ontological claims, and therefore is perceived as having objective force rather than being a merely subjective whim.

All of the classic Enlightenment values that we continue to embrace as modern liberals—the belief in human rights, the valuation of freedom and creativity, and the condemnation of inflicting suffering on innocents—are strong evaluations of this sort, dependent on an implicit set of beliefs about human beings historically derived from Christianity, but reflecting common human normative judgments. Although the Enlightenment *philosophes* began disengaging these beliefs from their explicitly religious context, and we in the last century have more or less completed this process, this does not change their status as beliefs. The "self-evident truths" enshrined in such classic liberal documents as the Declaration of Independence of the United States and the U.N. Universal Declaration of Human Rights are not revealed to us by the objective functioning of our *a priori* reason, but are rather items of faith.

Taylor argues that metaphysically grounded normative reactions such as these are inevitable for human beings. The fact that we cannot coherently account for our own or other's behavior without making reference to metaphysical beliefs, as well as the fact that they irresistibly present themselves to us as objective despite our lack of proof for them, says something important about what it means for a thing to be "real" for human beings. Although values are not part of the world as studied by natural science, the fact that value terms such as "freedom" and "dignity" are "ineradicable in first-person, non-explanatory uses" (57) means that they are, in a non-trivial sense, real. "[Human] reality is, of course, dependent upon us, in the sense that a condition for its existence is our existence," Taylor concedes. "But once granted that we exist, it is no more a subjective projection than what physics deals with" (59). For the peculiar type of animal that we are, moral space is as much a part of reality as physical space, in that we cannot avoid having to orient ourselves with respect to it.

To do something that would no doubt horrify Taylor—that is, reformulate his insights within a naturalistic framework—we can say that our overactive theory of mind causes us to inevitably project intentionality onto the world: we cannot help but see our moral emotions and desires writ large in the cosmos. It would be empirically unjustified to take this projection as "real." Nonetheless, the very inevitability of this projection means that, whatever we may assert qua naturalists, we cannot escape from the lived reality of moral space. As neuroscientists, we might believe that the brain is a deterministic, physical system like everything else in the universe, and recognize that the weight of empirical evidence suggests that free will is a cognitive illusion (Wegner 2002). Nonetheless, no cognitively undamaged human being can help *acting* like and at some level really *feeling* that he or she is free. There may well be individuals who lack this sense, and who can quite easily and thoroughly conceive of themselves and other people in purely instrumental, mechanistic terms, but we label such people "psychopaths," and quite rightly try to identify them and put them away somewhere to protect the rest of us (Blair 1995, 2001). Similarly, from the perspective of evolutionary psychology, I can believe that the love that I feel toward my child and my relatives is an emotion installed in me by my genes in accordance with the principles of kin selection. This does not, however, make my experience of the emotion, nor my sense of its normative reality, any less real to me. Indeed, this is precisely what I would expect from the third-person perspective: the gene-level, ultimate causation would not *work* unless we were thoroughly sincere at the proximate level. The whole purpose of the evolution of social emotions is to make sure that these "false" feelings seem inescapably real to us, and this lived reality will never change unless we turn into completely different types of organisms. Completely extracting ourselves from moral space is as impossible as stopping our visual systems from processing information when we open our eyes, or our stomach from registering displeasure when our blood sugar level drops below a certain point.

THE IMPORTANCE OF PHYSICALISM: WHY PHYSICALISM BOTH DOES AND DOES NOT MATTER

To the extent that human-level reality will always have a hold on us, then, we are entitled to say that physicalism does not matter. This leads Taylor to conclude that the unavoidability of human-level concepts is not merely a phenomenological observation, but rather a clue as to the "transcendental conditions" (32) of "undamaged human personhood"

(26), and thereby a refutation of any sort of third-person, naturalistic account of the humanities. If human reality is indeed real for us, why *not* follow Taylor and say that it is just as real as anything studied by the natural sciences?

In short, because it is not. Or, to put this more accurately, human beings appear to possess an innate empirical prejudice that is so constituted that, once we have *explained* something—that is, reduced a higher-level phenomenon to lower-level causes—the explained thing inevitably loses some of its hold on us.[20] There is an important difference between literally believing that God created the world in seven days and thinking that this is a beautiful story that can mean something to us on Sundays, but must be put aside when we go about our daily work. Evolution is such a relatively new idea, and its message so fundamentally alien to us, that its real implications for our picture of human reality have yet to fully sink in, which is why most liberal intellectuals continue to believe that Darwinism does not seriously threaten traditional religious beliefs or conceptions of the self. It clearly *does*, however, and once we have begun down the physicalist path, we cannot go back to the old certainties. This is not merely because it would be illogical to do so—although it would—but because we just seem to be built in such a way that we want to deal with and picture the world as it "really" is, no matter how unpleasant.

We can get a sense of this human "truth" prejudice—really, a preference for lower-level over higher-level explanation—by thinking about a typical reaction to a science-fiction movie that was very popular some years back called *The Matrix* (1999). For those unfamiliar with the plot, the protagonist, "Neo," begins to uncover puzzling clues that his everyday world is an illusion. He eventually discovers that his body and those of others in his apparently real world of "the Matrix" are, in fact, being maintained in sinister life-support tanks housed in a vast factory. Their brain activity is being farmed as a source of energy by the evil machines who created the Matrix—an elaborate virtual world, projected onto the brains of the bodies in the tanks—in order to fool their prisoners into thinking that they are free. Neo eventually gets in touch with a doughty band of humans who have liberated themselves from the life support tanks and who live crude, uncomfortable, but "free" lives in an underground refuge called (rather heavy-handedly) Zion.

[20] See Preston and Epley (2005) for a very helpful recent study illustrating the human tendency to value lower-level explanations that are presented as explanations for higher-level phenomena.

One of the more interesting points in the movie is when a cowardly informant is induced to betray the inhabitants of Zion and return to the tanks in exchange for a particularly pleasant illusory life-style: in the virtual world of the Matrix, he is to be a rich and powerful man, with every sensory pleasure one could desire. Most importantly, he will not *remember* that this is all an illusion: his fine steak and excellent red wine will taste just as good as the "real" things would, and the pleasure he will derive from his new virtual life will be—to him, at least—inescapably true and powerfully felt. Especially when compared with the threadbare and uncomfortable life in the bleak underground burrow of Zion, this seems like a pretty good deal: if you do not know the Matrix is not real, what difference does it make? If the steak tastes like steak, why should you care that you are "really" pickled in a tank and being farmed by evil machines? If your memories are to be perfectly erased, why would it matter that you had betrayed your comrades and your former cause?

It probably *would not* matter. The important thing, though, is that we, as human beings, feel that it *would*—we feel anger with this traitor, as well as revulsion at the idea of returning voluntarily to the Matrix. Why? Because, as Aristotle said, we are constituted in such a way as to desire the Good, and the Good for human beings involves being properly situated with regard to what we feel to be the "truth." Promised future rewards that we know to be illusory seem less valuable to us, even if we are assured that they will *seem* real when we get them. The same inchoate instinct that makes life in the Matrix abhorrent to us makes it impossible to continue to embrace, at least in precisely the same way, traditional religious ideals that appear to be in conflict with what we are convinced we now know about the world. And—at least as long as physicalism remains our current best explanation of the world—any religious or philosophical belief based on dualism is going to be in this sort of conflict.

This is where the Copernican analogy *is* helpful. We quite happily live our everyday lives in a Ptolemaic solar system, seeing the sun "rise" and "set" and enjoying the felt stability of the earth under our feet. We acknowledge, though, that this appearance is an illusion, and that the earth is really racing through space at 108,000 km/h around the sun. Why does it matter what is "really" the case, if it makes no difference to the way we see things? It matters because making important, practical decisions based on what is really the case, as opposed to what seems to be the case, works better. Launching satellites or sending off space probes simply would not work very well unless we suspended our intuitive Ptolemaic worldview when engaged in this sort of work.

The same is true of human-level realities. The realization that the body–mind is an integrated system is counter-intuitive, but treatments based upon this insight appear to be massively more effective than dualism-based treatments—pharmaceutical interventions, for instance, have arguably done more for the treatment of mental illness in a few decades than millennia of spiritual interventions, from exorcisms to Freudian analysis. Recognizing that there is no point at which the ghost enters the machine allows us to go ahead with stem cell research, and understanding that personhood is not an all-or-nothing affair helps us get a better grip on what is going on with severe dementia in the elderly. Physicalism matters because it simply works better than dualism, and—once the reality of this superiority is fully grasped—this is an irresistibly powerful argument for creatures like us.

DUAL CONSCIOUSNESS: WALKING THE TWO PATHS

How can physicalism both matter and not matter? We can answer this question by returning to Nietzsche's critique of Kant and seeing how *both* Niezsche and Kant were right. To take moral intuitions as an example, we can follow Nietzsche—somewhat updated and put into the role of an evolutionary psychologist—and see why it is important, unavoidable, and revealing to ask about the adaptive forces that cause us to feel the force of synthetic *a priori* claims, rather than just experiencing them as unquestioned intuitions. Answering the question of origins—uncovering the lower-level, ultimate explanations for our moral intuitions—has important practical implications, but most of all we just simply want to *know*. We also need to follow Kant, however, in recognizing that, no matter what the origins of these intuitions, they are the spontaneous product of a very powerful, built-in faculty, the output of which seem inescapably right to us. This means that, as empirically responsible humanists, we need to pull off the trick of simultaneously seeing the world as Nietzsche and as Kant, holding *both* perspectives in mind and employing each when appropriate. Those who have allowed the "universal acid" (Dennett 1995) of Darwinism to finally breach the mind–body barrier thus end up living with a kind of dual consciousness, cultivating the ability to view human beings simultaneously under two descriptions: as physical systems and as persons. On the one hand, we are convinced that Darwinism is the best account we have for explaining the world around us, and therefore that human beings are merely physical systems. On the other hand, we cannot help but feel the strong pull of human-level truth.

We can mine the world's religious traditions for helpful metaphors for what this kind of dual stance toward the world might be like. Jesus famously advised his followers to be "in the world but not of it" (John 17:14–15), and the figure of the Bodhisattva in the Mahayana Buddhist tradition dwells simultaneously in two realms: that of "ultimate" truth, where there are no distinctions and no suffering and therefore no need for compassion or Buddhism, and that of "conventional" truth, where suffering is real and the Bodhisattva is called upon to exercise finely tuned and deeply felt compassion. My favorite analogy comes from the fourth-century BCE Chinese thinker Zhuangzi, who describes his ideal person as "walking the two paths"—that of the "Heavenly"/Natural (*tian*) and the human. From the Heavenly/Natural perspective, there are no distinctions, no right and wrong, no feelings, no truth. From the human perspective, all of these things are acutely real. The key to moving successfully through the world, Zhuangzi believes, is simultaneously keeping both perspectives in mind, seeing the human "in the light of the Heavenly," and thus seeing through to its contingent nature, while at the same time acting in accordance with the constraints of being a human in the world of humans (Watson 1968: 40–41). This kind of dual consciousness is perhaps what Kant was getting at in a curious passage from the *Groundwork* when he declares that we must "lend" the idea of freedom to rational beings:

> Now I assert that every being who cannot act except under the Idea of freedom is by this alone—from a practical point of view—really free; that is to say, for him all the laws inseparably bound up with freedom are valid just as much as if his will could be pronounced free in itself on grounds valid for theoretical philosophy. And I maintain that to every rational being possessed of a will we must also lend (*leihen*) the Idea of freedom as the only one under which he can act. (115–116)

We have become convinced, qua physicalists, that we are not free, but in our everyday lives, we cannot help acting as if we were free, lest we find ourselves exiled from the Kingdom of Ends—that is, no longer recognizable as undamaged human agents.

CONCLUSION

We should not allow our distaste for physicalist explanations of the human to turn us into reactionaries. The subject of humanist inquiry is not the workings of some Cartesian *Geist* in the machine, but the wonderfully complex set of emergent realities that constitute the lived

human world, in all its cultural and historical diversity. The realization of the thoroughly physical nature of this reality does not condemn us, however, to live forever after in an ugly world of things. For psychologically healthy humans, other humans can never be a existentially grasped as mere things,[21] and our promiscuous projection of teleology onto the world assures that we will continue to find the whole materialist universe a rather beautiful place once it is properly understood. The fact that even the most resolutely physicalist conception of the world cannot help but continue to inspire awe and an implicit sense of meaning in human beings is nicely captured in the character of Henry Perowne, a neurosurgeon and committed materialist, in Ian McEwan's recent novel *Saturday* (2005). Prompted by a poem to imagine being "called in" to create a new religion, Perowne declares that he would base his upon evolution:

> What better creation myth? An unimaginable sweep of time, numberless generations spawning by infinitesimal steps complex living beauty out of inert matter, driven on by the blind furies of random mutation, natural selection and environmental change, with the tragedy of forms continually dying, and lately the wonder of minds emerging and with them morality, love, art, cities—and the unprecedented bonus of this story happening to be demonstrably true. (56)

We have nothing to fear from reductionism, because our innate cognitive mechanisms ensure that the modern scientific model of human beings as essentially very complicated things will not lead to nihilism or despair. As humanists, we are not in fact faced by the stark choice of either a meaningless, mechanistic universe or an endless, nightmarish maze of contingent discourses. It is possible to be an empirically responsible intellectual, and embrace an embodied, vertically integrated approach to one's subject matter, without losing sight of the inescapable human reality of this emergent level of explanation. In the end, acknowledging our inescapable embodiment not only possesses the excellent advantage of being "demonstrably true," but also cannot help but enrich our sense of wonder at the dependent and tragic human condition, in all its felt beauty and nobility.

[21] Of course, what counts as "human" is up for grabs, and the idea that every member of the biological species *Homo sapiens* is "human" is a relatively recent idea—the category has historically tended to encompass only one's own tribe. The recurrent reality of genocide, even in our modern world, serves as a chilling reminder of how quickly and easily groups formerly seen as humans can be reclassified as "things."

REFERENCES

Arbib, Michael 1985 — *In Search of the Person: Philosophical Explorations in Cognitive Science.* Amherst: University of Massachusetts Press.

Barkow, Jerome, Leda Cosmides and John Tooby eds 1992 — *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* New York: Oxford University Press.

Baron-Cohen, Simon 1995 — *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: MIT Press.

Barrett, H. Clark, Peter Todd, Geoffrey Miller and Philip Blythe 2005 — "Accurate Judgments of Intention from Motion Cues Alone: A Cross-cultural Study." *Evolution and Human Behavior* 26: 313–331.

Bar-Yosef, Ofer 2006 — "Human Migrations in Prehistory: The Cultural Record." *World Knowledge Dialogue*, Crans-Montana, Switzerland, September 15.

Bernstein, Richard 1983 — *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis.* Philadelphia: University of Pennsylvania Press.

Blair, R. James 1995 — "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57: 1–29.

2001 — "Neurocognitive Models of Aggression, the Antisocial Personality Disorders, and Psychopathy." *Journal of Neurology, Neurosurgery, and Psychiatry* 71: 727–731.

1990 — *The Logic of Practice*, trans. Richard Nice. Stanford: Stanford University Press.

Boyer, Pascal 1994 — *The Naturalness of Religious Ideas. A Cognitive Theory of Religion.* Berkeley & Los Angeles: University of California Press.

2001 — *Religion Explained: The Evolutionary Origins of Religious Thought.* New York: Basic Book.

Buss, David ed 2005 — *The Handbook of Evolutionary Psychology.* Hoboken, NJ: John Wiley & Sons.

Camus, Albert 1942 *Le mythe de sisyphe.* Paris: Gallimard.

1947 *La peste.* Paris: Gallimard.

Carruthers, Peter and Peter Smith eds 1996 *Theories of Mind.* Cambridge: Cambridge University Press.

Carruthers, Peter, Stephen Laurence and Stephen Stich eds 2005 *The Innate Mind: Structure & Content.* New York: Oxford University Press.

Carruthers, Peter, Stephen, Laurence and Stephen, Stich 2006 *The Innate Mind: Culture & Cognition.* New York: Oxford University Press.

Churchland, Patricia 1986 *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* Cambridge, MA: Bradford Books/MIT Press.

Clark, Andy 1997 *Being There: Putting Brain, Body and World Together Again.* Cambridge, MA: MIT Press.

Cosmides, Leda and John Tooby 1992 "Cognitive Adaptations for Social Exchange." In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. by Jerome Barkow, Leda Cosmides and John Tooby, 163–228. New York: Oxford University Press.

Courtright, Paul 1985 *Ganesa: Lord of Obstacles, Lord of Beginnings.* New York: Oxford University Press.

2006 "The Self-serving Humility of Disciplining Liberal Humanist Scholars: A Response to Russell McCutcheon." *Journal of the American Academy of Religion* 74: 751–754.

Crick, Francis 1994 *The Astonishing Hypothesis: The Scientific Search for the Soul.* New York: Simon & Schuster.

Dawkins, Richard 1976/2006 *The Selfish Gene.* New York: Oxford University Press.

Dennett, Daniel 1995 *Darwin's Dangerous Idea: Evolution and the Meaning of Life.* New York: Simon & Schuster.

Ekman, Paul 2003 "Darwin, Deception, and Facial Expression." *Annals New York Academy of Science* 1000: 205–221.

Flanagan, Owen 1992 *Consciousness Reconsidered.* Cambridge, MA: MIT Press.

2002 *The Problem of the Soul: Two Visions of Mind and How to Reconcile Them.* New York: Basic Books.

Geertz, Clifford 1973 *The Interpretation of Cultures: Selected Essays.* New York: Basic Books.

Gibbs, Raymond 2006 *Embodiment and Cognitive Science.* Cambridge: Cambridge University Press.

Giorello, Giulio 1997 Sì, abbiamo un anima. Ma è fatta di tanti piccoli robot (interview with Daniel Dennett). *Corriere della Sera (Milan),* April 28.

Guthrie, Stewart 1993 *Faces in the Clouds: A New Theory of Religion.* Oxford: Oxford University Press.

Haack, Susan 2003 *Defending Science Within Reason: Between Scientism and Cynicism.* Amherst, NY: Prometheus Books.

Hacking, Ian 1983 *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* Cambridge: Cambridge University Press.

2000 *The Social Construction of What.* Cambridge, MA: Harvard University Press.

Hirschfeld, Lawrence and Susan Gelman eds 1994 *Mapping the Mind: Domain Specificity in Cognition and Culture.* New York: Cambridge University Press.

Johnson, Mark 1987 *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason.* Chicago: University of Chicago Press.

Kant, Immanuel 1785/1964 *Groundwork of the Metaphysic of Morals,* trans. H. J., Paton New York: Harper Torchbooks.

Kelemen, Deborah 1999 "Function, Goals, and Intention: Children's Teleological Reasoning about Objects." *Trends in Cognitive Sciences* 3: 461–468.

2004 "Are Children 'Intuitive Theists'? Reasoning and Purpose and Design in Nature." *Psychological Science* 15: 295–301.

Kuhlmeier, Valerie, Karen Wynn and Paul Bloom 2003 "Attribution of Dispositional States by 12-Month-old Infants." *Psychological Science* 14:402–408.

Kripal, Jeffrey 1995 *Kali's Child: The Mystical and Erotic in the Life and Teachings of Ramakrishna.* Chicago: University of Chicago Press.

Lakoff, George and Mark Johnson 1999 *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York: Basic Books.

Latour, Bruno 1993 *We Have Never Been Modern*, trans. Catherine, Porter Cambridge, MA: Harvard University Press.

Lau, D. C 1970 *Mencius.* London: Penguin.

Laudan, Larry 1996 *Beyond Positivism and Relativism: Theory, Method, and Evidence.* Boulder, CO: Westview Press.

Locke, John 1690/1975 *An Essay Concerning Human Understanding.* Oxford: Clarendon Press.

Lodge, David 1975 *Changing Places.* London: Penguin Books.

1984 *Small World: An Academic Romance.* New York: Warner Books.

1988 *Nice Work.* New York: Penguin Books.

McCutcheon, Russell 2006a "It's a Lie. There's no Truth to it! It's a sin!: On the Limits of the Humanistic Study of Religion and the Costs of Saving Others from Themselves." *Journal of the American Academy of Religion* 74: 720–750.

2006b "A Response to Courtright." *Journal of the American Academy of Religion* 74: 755–756.

McEwan, Ian 2005 *Saturday.* Toronto: Alfred A Knopf.

Nagel, Thomas 1974 "What is it like to be a Bat?" *Philosophical Review* 83: 435–450.

Nietzsche, Friedrich 1886/1966 *Beyond Good and Evil,* trans. Walter, Kaufmann New York: Vintage.

Orsi, Robert 2005 *Between Heaven and Earth: The Religious Worlds People Make and the Scholars who Study Them.* Princeton: Princeton University Press.

Pecher, Diane and Rolf Zwaan eds 2005 *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking.* Cambridge: Cambridge University Press.

Penrose, Roger 1989 *The Emperor's New Mind.* New York: Oxford University Press.

Pinker, Steven 2002 *The Blank Slate: The Modern Denial of Human Nature.* New York: Viking.

Plantinga, Alvin 1993 *Warrant and Proper Function.* New York: Oxford University Press.

Preston, Jesse and Epley, Nicholas 2005 "Explanations versus Applications: The Explanatory Power of Valuable Beliefs." *Psychological Science* 16: 826–832.

Putnam, Hilary 1973 "Reductionism and the Nature of Psychology." *Cognition* 2: 131–146.

1999 *The Threefold Cord: Mind, Body, and World.* New York: Columbia University Press.

Robbins, J. Wesley 1994 "Is Naturalism Irrational?" *Faith and Philosophy* 11: 255–259.

Ryle, Gilbert 1971 "The Thinking of Thoughts: What is 'le penseur' Doing?" *Collected Papers,* vol. x. London: Hutchinson.

Searle, John 1983 *Intentionality.* Cambridge: Cambridge University Press.

1992 *The Rediscovery of the Mind.* Cambridge, MA: MIT Press.

2004 *Mind: A Brief Introduction.* New York: Oxford University Press.

Slingerland, Edward 2008 *What Science Offers the Humanities: Integrating Body & Culture.* New York: Cambridge University Press.

Slone, D. Jason 2004 *Theological Incorrectness: Why Religious People Believe What they Shouldn't.* New York: Oxford University Press.

Smith, Marty 1987/1993 *Free Agent (March 1987);* reprinted in *Utne Reader* (November/December 1993).

Smith, Wilfred Cantwell 1959 "The Comparative Study of Religion: Whither—and Why?" In *The History of Religions: Essays in Methodology,* ed. by Elidae, Mircea and Kitagawa, Joseph, 31–58. Chicago: University of Chicago Press.

Spelke, Elizabeth, Ann Phillips and Amanda Woodward 1995 "Infants' Knowledge of Object Motion and Human Action." In *Causal Cognition: A Multidisciplinary Debate,* ed. by Dan Sperber, David Premack and Ann Premack 44–78. New York: Oxford University Press.

Styron, William 1992 *Darkness Visible.* New York: Vintage.

Taylor, Charles 1989 *Sources of the Self: The Makings of Modern Identity.* Cambridge, MA: Harvard University Press.

1992 *The Ethics of Authenticity.* Cambridge, MA: Harvard University Press.

Tegmark, Max 2000 "Importance of Quantum Decoherence in Brain Processes." *Physical Review E* 61: 4194–4206.

Thomson, Evan 2007 *Radical Embodiment: The Lived Body in Biology, Human Experience, and the Sciences of the Mind.* Cambridge, MA: Harvard University Press.

Watson, Burton trans 1968 — *The Complete Works of Chuang Tzu.* New York: Columbia University Press.

Wegner, Daniel 2002 — *The Illusion of Conscious Will.* Cambridge, MA: MIT Press.

Wellman, Henry 1990 — *The Child's Theory of Mind.* Cambridge, MA: MIT Press.

Wilson, Edward O. 1998 — *Consilience: The Unity of Knowledge.* New York: Knopf.