Forthcoming in M. Jeeves, The Emergence of Personhood, Eerdmans 2016

"The Emergence of Personhood: Reflections on *The Game of Life*"

Timothy O'Connor, Indiana University

Abstract

The correlated terms "emergence" and "reduction" are used in several ways in discussions of human personhood, engendering confusion or talking at cross-purposes. I try to bring clarity to this discussion by reflecting on John Conway's cellular automaton *The Game of Life* and simple variations on it. We may think of such variants as toy models of our own world that, owing to their simplicity, enable us to see quite clearly, in general terms, two importantly distinct ways ("weak" and "strong") in which organized macroscopic phenomena might emerge from underlying microphysical processes. I will then consider the evidentiary standing for strong emergence associated with aspects of human life in the actual world and argue that this has profound implications regarding the extent of (dis)continuity of human beings with the rest of nature and of our moral and/or theological significance within it.

Human persons are the most profound point of convergence of the world's basic forms of complexity: physico-chemical, biological, psychological, informational, social. We appear to be unique among living species in having the ability to engage in long-range individual and collective planning, or at least planning behavior that exceeds rudimentary foraging and storage behavior and is thought-through. Standing out more dramatically are our abilities to partly control and manipulate the very biological and environmental conditions of our continued survival and to alter dramatically the circumstances of our day-to-day habitat. Furthermore, we think and act within complex moral and religious narratives, apart from which much of what we do would make no sense. A key question raised by the essays in this volume is whether these and other remarkable abilities and tendencies signal a discontinuity in the evolutionary processes that gave rise to us—and if so, to what extent, and in what form.

Each of the many sciences that consider the phenomena of human life do so with different methodologies and from different 'altitudes'. Whether and how they mesh so as to provide a picture of a unitary whole is itself a substantive, open question. Some of the sciences have at best advanced to a mid-game stage. It might well be that we are not currently in position to see which of them merely offer useful models rather than more-or-less accurate maps of their domains.

A good place to consider how theories from different domains interrelate is in their characterization of organized complexity: phenomena that exhibit strikingly different sorts of robust patterns than those seen among their constituent elements when not so

organized. Organized complexity is seen at a number of levels represented in living human beings. Theorists have described each of these patterned phenomena as exhibiting "emergent" (or sometimes "top down") causation or control. It is tempting for specialists in one area to develop an abstract model of their favored phenomena and then to suggest that the model holds the key to thinking about naturally emergent phenomena generally. But it is rash to suppose at the outset that there is a *single* way to accurately and usefully apply a notion of emergence to all cases of interest. Maybe emergence is more 'dramatic' or fundamentally significant in some cases of patterned complexity than in others.

In what follows, I want to try to isolate two distinct senses of emergence ("weak" and "strong"). The two notions are not orthogonal but rather of differing logical 'strengths'. Furthermore, they signal a basic division, on each side of which one may conceive variant forms, possibly such that one may think in terms of degrees of robustness. The contemporary sciences are very complicated and nuanced. This is good for the advancement of knowledge, since the world itself is complicated and nuanced, but it can make it difficult to see the fundamental issue I wish to raise. So I will invite us to consider a kind of toy world whose dynamics are quite transparent. It will be easy to isolate my two senses of emergence in relation to this kind of world and to grasp what kind of evidence would enable one to decide conclusively whether both notions or merely the weaker one has application. I will then consider the evidentiary standing for strong emergence associated with human life in the actual world and argue that this has profound implications regarding the extent of (dis)continuity of human beings with the rest of nature and of our moral and/or theological significance within it.

I The Game of Life and Two Forms of Emergence

My toy world has conveniently been supplied to us by the mathematician John Conway, in the form of his cellular automaton, the 'Game of Life'. (Versions of the *Game of Life* are readily available online.[1]) *Life* is a dynamic and spatially and temporally discrete two-dimensional infinite grid. One sets an arbitrary initial state by assigning one of two basic properties, *live* or *dead*, to each of the square cells. Each subsequent state of the grid is wholly determined by applying the following three rules to every cell (which has eight 'neighbors in every direction, including diagonally):

Birth: A dead cell with exactly three live neighbors becomes a live cell.

Survival: A live cell with two or three live neighbors stays alive.

Death: In all other cases, a cell dies or remains dead.

Over time, stable clusters of various kinds arise and exhibit macro-level patterns of activity and interaction with other cluster types. (Apt names have been given to certain recurring sorts, such as 'oscillator', 'glider', 'puffer', and 'eater'.) Once these clusters appear, their macro-level behavior can be studied in ignorance of the three micro-level rules that underlie them. Interestingly, different sorts of high-level patterns are observed in games with different initial conditions.

If we think of Conway's 2-D grid as a kind of toy world (there are 3-D versions of it as well), the three basic rules and its initial state constitute its fundamental 'physics.' The high-level patterns can be thought of as its 'chemistry' or 'biology.' These high-level

[1] See http://conwaylife.com/.

patterns exhibit what I will term ‘weak’ (or ‘physics-closed’) emergence: they are emergent in the sense that one cannot—in any straightforward way—derive the high-level rules from the fundamental rules alone (though, at least for the case at hand, one could do so in principle, in a very laborious, roundabout way, from the low-level rules plus the initial conditions).

> **Weak Emergence**$=_{def}$ Some aspects of the behavior of a specifiable type of composite, organized system exhibit ongoing or recurrent lawlike patterns that, outside this organizational context, are neither exhibited by the system’s components nor describable in terms of the concepts of the most compact theory that fully describes the component behavior.

The high-level rules of *Life* are only weakly emergent in that they do not in any way alter or supplement the basic dynamics that drive the world’s evolution. The ‘physics’ of a standard *Life* world is causally closed, with each total configuration of the grid at a time t_1 being strictly determined by its state at the previous time t_0 in accordance with the three basic rules.

Someone wishing to emphasize the significance of weak emergence might say that the high-level *theory* that apply under the right conditions to composites such as gliders and eaters is irreducible to the basic cellular theory and constitutes a relatively autonomous *explanatory* system: we can predict/explain macro-level phenomena using only our knowledge of macro-level laws and facts. (Absent the disruption of collisions with merely ‘physical’ clusters, of course: just as in our world, biology is not going to tell you

what will happen next when a meteor strikes a biological niche.) Moreover, for certain purposes, an explanation of a configured state that is at once physical and biological in terms of biology is more illuminating than one in terms of physics. If we not only observe but intervene in a *Life* world by manipulating the value of one higher-level variable and holding other relevant factors constant, we can change the value of another higher-level variable in very predictable ways and in a variety of contexts. This indicates that we have not only high-level explanation, but also high-level *causal* explanation.

However, we should observe that this *theoretical/explanatory irreducibility* in standard *Life* worlds is matched by *physical reduction* in the following sense: First, the state of every stable cluster at any moment is wholly fixed and constituted by the properties of each of its cells. Second, the cluster's evolution over time is fundamentally determined by the basic 'physical' laws. If we think of the fundamental rules in Life as causal, then each macro-level event (such as a glider's changing position over a time interval) has a micro-level *cause*—the micro-level causes of its components' positions and behavior—and thus a micro-level causal explanation, in addition to the macro-level explanation. The micro-level causation determines the macro-level causation, *and not vice versa*. Macro-level events aren't doing anything "extra", over and above what the micro-level events are doing, in order to bring about the positions and behavior of gliders, eaters, and the like. Macro-level events, then, do *not* make a nonredundant causal difference to the behavior of the objects in such a *Life* world. Macro-level objects and events in such worlds are weakly emergent only: theoretically/explanatorily irreducible, ontologically and causally derivative.

Once one gets used to looking at things in this way, physical reductionism can begin to seem inevitable—both in any imaginable variety of *Life* world and in any law-governed physical world, notwithstanding the added complexity of three spatial dimensions, a richer inventory of basic properties, and continuity in space and time. Surely *any* sort of organized complex phenomena must be grounded in comprehensive underlying patterns that hold without exception in every context?

Whatever is true with respect to comprehensive physical reduction in our own world, we can readily describe variations on standard *Life* worlds that enable us see that this *need not* be the case—there exists a coherent form of alternative. Imagine that you are handed a tall stack of (very large!) numbered sheets of graph paper. On them are changing snapshots of a *Life* world (where the shaded squares are 'live' cells). Your job is to figure out the basic transitional rules. After flipping through several pages, you hit upon the Birth, Survival, and Death rules as the most compact ways to capture all the transitions in each cell. Continuing to check subsequent pages to verify that the rules hold without exception, you hit upon a page where the result departs in a small way from what the rules predict. You observe that the divergence is restricted to a complex star-shaped cluster that first appeared on the previous page. Flipping ahead, you observe that as more of these star-shaped clusters appear, their subsequent evolution, too, departs from what the three basic rules predict. Further investigation reveals that the form the divergence takes is identical in each case.

You now can predict, upon seeing the emergence of the star shape, what the future world-states will be, using *modified, disjunctive* forms of the original rules. The new Birth rule, e.g., has the form: a dead cell with exactly three live neighbors becomes a live cell, *except* when occurring within the bounds of, or immediately adjacent to, a star cluster, in which case…. Imagine that as star clusters come into contact, new modifications of the original rules are required to fully capture the way they interact. You find that the most compact way to capture the behavior is (i) to assign primitive new properties ('bright', 'golden') to star clusters and then (ii) to describe the precise impact of the emergent properties of the clusters on the 'ordinary' micro-level dynamics via additional laws *that are no less fundamental*—even though they have application only in limited contexts.

The 'world' you have been given to inspect exhibits what I will call "strong emergence":

> **Strong Emergence**$=_{def}$ Some aspects of the behavior of a specifiable type of composite, organized system are such that (i) they cannot be fully described, even at a microphysical level of description, by compact low-level rules making no reference to macroscopic structures that *can* fully describe behavior outside certain situations of organized complexity, and (ii) they can be fully described by rules that apply only in these contexts and that essentially involve reference to macroscopic structural properties.

Worlds manifesting strong emergence are plainly coherent possibilities. But are they objectionably strange or 'magical', ones that ought to offend the sensibilities of a properly empirically-minded theorist? I don't see why—unless one is in the grips of a

simple building block picture of physical reality, treating the hypothesis that everything that happens is wholly fixed by compact micro-regularities without regard to macro-context as some sort of *a priori* truth. It seems rather that there ought to be no presumption one way or another in advanced of a sufficiently thorough investigation of the empirical facts. As the early 20th-century emergentist Samuel Alexander put it, we should come to the world "with the 'natural piety' of the investigator" (1920, 47).

Perhaps you are thinking that a world with strong emergence is somehow fundamentally disunified. That might be a proper conclusion if departures from the simple rules happened willy nilly. But in the scenario I've outlined, the new patterns, once discovered and the conditions of their appearance learned, will become wholly predictable. There are sufficient conditions, involving organized complexity of a certain kind, for their appearance. Given that this is so, to make sense of such novel phenomena, we need only enrich our conception of the properties of the *fundamental* units (the cells), supposing that in addition to *live* and *dead*, whose dispositional profiles are largely captured by the original unmodified rules, they contain a set of dispositions towards cooperatively contributing to just such behavioral discontinuities, a disposition that is merely latent absent the requisite macro-circumstance. Unity within a physical world does not require causal continuity of *behavior*, only continuity of dispositional structure. In worlds that feature strong emergence, the seeds of every strongly emergent macro-property and the behavior it manifests are found within the world's fundamental elements: latent dispositions awaiting only the right context for manifestation.

II Strong Emergence and Consciousness

Armed with the basic distinction between weak and strong emergence, we can ask ourselves whether specific forms of organized complexity in our own world involve weak emergence only, or are instead forms of strong emergence. Surprisingly, there is quite good evidence for strong emergence at a very *low* level of complexity: within the domain of quantum mechanics itself. Well-confirmed theory indicates that some simple particle systems go into 'mixed' states that cannot be decomposed into any mere 'sum' of the states of the individual particles—they are 'relational wholes'. Outside this domain, the issue gets more difficult to adjudicate. For myself, I think it is *very* hard to establish strong emergence *or* its denial in any domain outside the mental. (More on the mental in a moment.)

Conway's standard *Game of Life* beautifully illustrates the consistency of striking and explanatorily irreducible macroscopic behavior with physical reduction, which two conditions constitute none other than our weak emergence. Note that 'top down causation', the term of choice for some theorists, likewise admits of weak and strong varieties. That there are top-down constraints on individual particle or molecule behavior is evident—a particle caught up in the motion of a wheel rolling down a hill; a molecule caught up in the life of a goal-driven organism as it moves through space. But these cases *may* be wholly a matter of low-level entities, so to speak, coming to 'constrain themselves', entering into and collectively maintaining various kinds of stable boundary conditions merely by doing what they always do in their local interactions. (We should also recognize, however, that the simplicity of *Life* worlds can mask the fact that there

can be different degrees of complexity of high-level patterns in the umbrella category of weak, or 'physics-closed' emergence: the organizational complexity of a living animal moving through its habitat is much greater than that of a wheel rolling down a hill. While degree of complexity is *not* a bridge between weak and strong emergence, it is a theoretically significant feature for the study of complex systems and of interest in its own right.)

But my modified *Life* scenario shows that we can just as readily envision a stronger form of emergence. Unfortunately, once we are dealing with systems of great complexity relative to the microphysical level, it is unclear what feasible means are or could be available for deciding which form of emergence is being manifested. Obviously, it is a fantasy to suppose that we might some day apply fundamental physical laws directly to complex systems, tracking each of the many billions of physical factors constituting or impinging on them over a specified time interval, and see whether the laws prove incomplete and how. The best hope for making a case for a weak emergentist account of a domain's novel complexity is to use simplifying approximation techniques to develop reductive structural models of the target domain that prove to have some measure of empirical backing.

Some seem to think, wrongly, that strong emergence, while possible, is inherently improbable, given the methods of science and especially the advances of physical and biological science from the beginning of the 20th century (e.g., McLaughlin 1992). This keeps them from giving proper weight to the powerful evidence that we have for strong

emergence (which I will discuss shortly) that is right under our noses, in the form of our own conscious mental lives. I will briefly make a couple remarks intended to challenge these sources of built-up resistance to taking strong emergence seriously.

My first remark is historical. From the mid-19th to the early 20th centuries, there was a robust scientific-philosophical tradition in chemistry and biology, centered in Great Britain, that involved commitment to something like strong emergence in the sense that I have defined. (McLaughlin 1992, O'Connor 1994) (There was some measure of conceptual confusion in these discussions, making precise identification of their operative notion(s) difficult.) The British emergentists had a neatly 'layered' understanding of nature: there are discrete and isolated strata or levels within the most complex systems found in the natural world, and each level features a unified system with its own same-level properties and special laws governing their co-evolution over time. This picture was shown to be a failure by a raft of evidence in 20th C science showing that 'lower-level' processes continuously and directly impinge upon and partly regulate those at higher levels. In particular, specific biochemical processes are now known to regulate intercellular processes which, on a large scale, constitute the functioning of biological organs. Even psychological processes are causally influenced in bottom-up fashion by subtle biochemical matters, such as chemical balance levels and neuronal receptor conditions that are implicated in clinical depression and schizophrenia, respectively.

But this kind of evidence from twentieth-century science can be taken to strongly disconfirm the strong emergentist view only if one conflates it with the particular account

proposed by these earlier thinkers, an account that appears unnecessarily crude in hindsight. It is quite possible that higher-level features of a system have a 'downward' causal influence on the evolving microstructure that sustains them, even as the lower-level processes play a vital dynamical role in the way emergent processes themselves unfold. In other words, we should replace the picture of mostly horizontal, or same-level, causal patterns within each level with one in which there is a complex web of myriad upward and downward causal influences that jointly (and nonredundantly) determine the system's evolution through time. Nothing in the successes of the more fundamental sciences in illuminating a vast array of higher-level phenomena precludes the applicability of this interactive sort of emergentist framework.

My second remark is that there is a tendency among some thinkers trained in the methods of empirical science to conflate what we have reason to believe is *true* with what we have *methodological* reason to take as the most fruitful starting hypothesis. Even if (say) the mental is strongly emergent from the physical, if you wish to understand the details of just *how* mental features emerge and the difference they make to the overall functioning of the organism, you must have a thorough understanding of the purely physical processes from which they emerge. Thus, in practice, you should push a reductionist approach as far as you can, thereby isolating any emergent features there may be, allowing for highly specific descriptions of their character and function. Unlike the philosopher, your typical scientist does not spend much time thinking about what will turn out to be true at the end of scientific inquiry. Rather, she is thinking about what working hypotheses are most useful in advancing current understanding. It is easy but

mistaken to slide from the quite proper methodological stance of looking for reducing explanations to drawing an evidential conclusion.

Each time we observe the world in some way, we are simultaneously aware of something else, viz., our *experience* of that bit of the world. We know our conscious experiences not indirectly (not mediated by theory or even inner representation), but by direct acquaintance. I am presently visually aware of a large purple coffee cup next to my computer. This awareness consists in a certain kind of mental representational state, caused by a very rapid but complex processing of light reflectances off the surface of the cup. By contrast, I am aware of the inner representational state itself simply by undergoing it. My acquaintance with it requires no causal or representational mediation; it is ontologically basic. My subjective experience of the cup consists in a unified package of 'looks' of color, shape, relative size, and orientation (to me). These ways the cup appears to me can change without the cup's undergoing any significant change at all, e.g., by dimming the light. These are changing intrinsic features of the experience, ones that I may not *attend* to at all—ordinarily, my attention is on the *cup*—but they are directly present to me nonetheless.

My visual field is usually quite complex: right now, it is filled with the appearances of a computer, a table, books and papers, and a window beyond which are leafy trees amidst patches of sunshine and shadows, just for starters. But this complexity is *surface* complexity: it is built up out of elements that are individually quite simple: e.g., the 'look' of a certain hue of green in a certain smallish region of the field. These simpler

elements of conscious experience cannot plausibly be identified with anything physical. The only candidates would be complex neural states, perhaps somewhat distributed, and the intrinsic properties of these states are not introspectively graspable and certainly do not have the structural simplicity that phenomenal greenness has. (Indeed, as neuroscience and (further down) elementary particle physics progresses, the idea of any such empirical identification of an experiential quality with a physical state seems more and more fantastic: the layered structural complexity of any given smallish region of our brains is quite extraordinary!)

What goes for visual experience goes for conscious mental phenomena generally: other kinds of perceptual experiences and conscious thoughts, feelings, moods, beliefs, desires, and intentions. All of these have intrinsic aspects not describable in purely third-personal terms. (Hence, the futility of conveying what the visual experience of greenness is like to a congenitally blind person, even one who is quite informed about the current state of the science of visual perception.)

The foregoing reflections indicate that, while our conscious mental states undoubtedly have complex sustaining causes, the future identification of which is part of the business of neuroscience, the conscious states themselves are distinct: caused by, but not identical to, structured physical states of the brain. If this is so, then the strong emergence of such states follows, given only the prima facie evident fact that these states are causally efficacious in all sorts of familiar ways. (My conscious visual experience of beer in the refrigerator caused my persisting belief that there is beer there, which in the fullness of

time causally contributes to the movement of my legs in a fridge-ward direction.) Taking consciousness seriously means abandoning the ontologically reductionist dream in favor of a strong emergentist understanding on which the appearance of new basic properties within certain neurally complex physical systems must (to again quote Samuel Alexander) "be accepted with the 'natural piety' of the investigator." If so, the existence of such properties, while a fundamental fact, may nevertheless be fruitfully studied and eventually explained in detail in *non*reductive fashion, by spelling out the basic inventory of emergent properties, detailing the precise conditions under which physical systems give rise to them, and isolating the precise behavioral impact their presence has on the system.[2]

Of course, there appears to be a radical, 'illusionist' option: that of *not* taking consciousness seriously. That is, one might entertain the idea that we have a deeply illusory conception of our own conscious experience. For certain philosophers and scientists, reductionist physicalist metaphysics trumps, as we might say, how the appearances appear to be—they trump, that is, our intuitive phenomenal beliefs. But far from being a rigorously empiricist outlook, it puts the cart before the horse and threatens thereby the entire edifice of natural science: For we have greater rational warrant for our

[2] Here I differ from both Brown and Sternberg and Zeman in this volume. The authors of both articles give appropriate emphasis to the complexity of interactive processes in human brain activity. But, seeing no alternative to mind-body dualism and a thorough-going physicalism/materialism, they embrace the second alternative and opt for a weak emergentist account of human conscious states. (Or so I read the substantive account given by Brown and Sternberg, despite their suggestion that their view approximate strong emergence in my sense.)

phenomenal beliefs than for our roughly reliable but fallible beliefs concerning even our physical environment and by extension the physical world. Indeed, our beliefs concerning physical reality are grounded, in significant part, in our experiences. If beliefs concerning the physical world derived from the success of 20th C science (such as the shared material basis of all things, living and unliving; the increasing interconnections of the sciences; and the rapid development of neuroscience) are used to call into question our grasp of our experiences themselves, they would thereby be undermining an important basis of their own justification.[3]

III Strong Emergence and Free Will

A specific and important element of our conscious experience is our experience of *will*. We regularly have experiences as of freely forming an intention to do some action while being aware of alternatives that were also available to us. Put differently, these experiences are as of being a purposely efficacious agent (and thus of the holding of an

[3] One can choose a more moderate stance that is deflationary, rather than eliminativist, regarding our awareness of the properties of our conscious states, a stance that lessens the severity of the epistemic problem I raise here. Pereboom 2011 explores the idea that while we are reliably and noninferentially aware of the intrinsic properties of our conscious states, our awareness of them misrepresents them as being structurally simple, when in fact they are quite complex. I believe that there are deep problems associated with this view, too, though I cannot develop my argument for this claim here.

inner causal relation between ourselves and our states of intention or choice).[4] If these experiences are veridical, then it cannot be the case that our actions are wholly determined by impersonal micro-physical causes. If we have reason to believe these experiences are veridical, we have another source of strong emergence.

Some have suggested that here a radical yet more targeted illusionism[5] is feasible, on which one denies that our experiences of as of willing to do this or that are in fact *efficacious*. The most careful proponents of the doctrine allow that these experiences as of control pretty reliably correlate with (and are likely caused by) wholly unconscious causal processes that are the real sources of our behavior. (Wegner 2002, among many other recent authors) However, this kind of illusionism, too, has unstable underpinnings. Scientific theories, models, and results are themselves the products of scientific *activity*: of human persons acting in certain coordinated, purposive ways and communicating their activities and results to one another. While the reality of reliably-known, purposive action is not part of the theoretical *content* of (most) scientific theories, nor are those theories explicitly *inferred* from this belief, it is a *pragmatic* assumption of such science, in the

---

[4] This raises the question: are there cases of mis-awareness of conscious willing, cases where we have experience as of willing but there is in fact no willing? How would these be identified as such? Supposing there can be such cases, if grasp of conscious willing is ordinarily a reflexive feature of conscious willing, can mis-awarenesses have intrinsically identical phenomenal characteristics to genuine awarenesses? It would take me too far afield to explore these questions here.

[5] What philosopher Eddy Nahmias (2002) has dubbed “willusionism”.

following sense: if we supposed it to be false, we would thereby have reason to doubt the trustworthiness of the outputs of such activity. It is reasonable to accept the trustworthiness of these outputs only insofar as we take them to have resulted from actions guided by the specific conscious purposes and beliefs that the actors report them to have been. Scientists must know what they are doing and why—to a significant extent—if the ideas that are based on the outputs of such actions are to be seen as more or less well supported. The lesson again is that well-grounded science cannot, in the end, *wholly* separate itself from our own self-understanding as human persons, since that self-understanding is a foundational part of its evidentiary basis.

That said, a more moderate 'willusionism' is not self-undermining. One may allow that we regularly consciously control our activity and are aware of our true purposes in so willing, and yet contend that the regularly accompanying belief that we are choosing *freely* is mistaken. And if we never choose freely, then neither are we morally responsible for our actions and their consequences. But why might one suppose that?

The direct empirical case for no-free-will is distinctly underwhelming. Some authors highlight studies that indicate that we can rather easily be induced to form unwittingly false *ex post facto* beliefs concerning *why* we did what we did (Delgado 1969; Gazzaniga 1994) or *that* we did something that we didn't do (Wegner 2002); others note that we can be influenced in our decisions by hidden factors of which we are unaware, even nonrational physical factors (Brasil-Neto et al.1992; Doris 2002). Taking the direct conclusions of such social and neuroscientific studies at face value, neither is in tension

with the commonsense assumption that we often act freely upon known aims. The first category of studies merely indicates the malleability of our memories; the second the imperfect nature of our own self-awareness, which doesn't threaten one's belief in human freedom and moral responsibility if one allows (as one should) that freedom of choice needn't be all or nothing, it can come in degrees.[6] But the most widely discussed empirical basis for free will skepticism stems from the work pioneered by neuroscientist Benjamin Libet (1985) on the timing of the felt experience of willing, which Libet and others took to indicate that a physical trigger or 'unconscious decision' regularly precedes the conscious awareness of willing. There is not space to delve into the complexities of these studies here. It suffices to say, while they have independent scientific interest, there are deep methodological and substantive problems associated with them that undercut any argument from them to our lacking free will, a conclusion that is accepted by numerous authors of various philosophical opinions and sensibilities. (See Mele 2009 and Clarke 2013 for very clear and thorough critical analyses.)

If direct empirical evidence for the unreality of free will is weak, might there be a good indirect argument? Some authors seem to suppose that free will stands or falls with mind-body dualism, the view that we are, or have as an essential component, simple mental substances causally bound up with our bodies. Since contemporary science indicates that we are instead psychosomatic unities, free will is called into question. Even if our conscious psychological states (and so our conscious choices) are strongly emergent, as I have argued, our choices must be causally determined by the underlying physical states

[6] I discuss these and other cases in O'Connor (2009).

that give rise to them.

I accept the philosophically controversial thesis that causal determinism is inconsistent with free will. But the strongly emergentist thesis that the existence and persistence of our capacity to choose is caused (or even—what is not the same—causally *determined*) by organized physical states does not entail that the *outputs* of that capacity is likewise physically causally determined. There is no reason to assume *a priori* that whatever sort of systemic capacity strongly emerges from the collective activity of an organized system of basic physical entities must be similar in kind, in specific respects, to the capacities that sustain it. Basic physical systems are ateleological, but it is an empirical question whether this is true of the capacities that emerge from them in certain organized contexts. Our experience of agency suggests that this is not universally true.

This simple reply will occasion an equally simple retort from a certain kind of 'scientifically-minded' thinker:

> All of your negative arguments may be correct, and yet you have provided us with no positive reason to think that we are free and responsible. The fact that we all *assume* this in practice is worthless. It's a mere prejudice of no more evidential weight than folk physics.

This sort of attitude is widespread among willusionists. I think it rests on a naïve conception of the underpinnings of scientific evidence and of human rationality generally. Descartes' heroic effort to doubt all his beliefs and re-build from scratch

notwithstanding, our beliefs about the world must start somewhere. Specifically, if we are to avoid total skepticism about knowledge in general (something that no one in practice can sustain), we must assume the rough reliability of our basic belief-forming capacities and the truth of our most foundational beliefs about reality.

Our basic cognitive capacities include the senses, memory, and basic forms of deductive and inductive reasoning. All of these are demonstrably fallible, but we learn about their limits and how to use them more critically *only* through using them. Our cognitively basic beliefs include such matters as the belief that the world has a causal or lawlike structure that will continue into the future (thereby permitting induction from past experience); that parts of the material world persist even when we're not observing them; that human bodies we observe are, like ourselves, animated by feeling, thinking, and purposive minds—and that our fundamental socio-moral belief that our seeming experience of freedom of choice is veridical. The reliability/truth of none of these items can be established by non-circular arguments. But that they are not reasonable is scarcely credible. Unlike other things we believe, they do not need to be inferred from evidence to be reasonable. As philosophers put it, these beliefs are properly *basic*, rationally speaking—including our belief that we are free and morally responsible. It is innocent until proven guilty. It is open for future science to make a solid case against this pervasive belief, but in the absence of such a case, and despite our ignorance of the details of how it 'works' in relation to the massive physical information processing of the brain, it is perfectly rational for us to continue to believe it.

IV Strong Emergence from Social Processes?

I end by noting (without taking a stance on) the provocative suggestion developed in quite different ways by Baumeister and Zeman (both this volume) that certain important facets of human nature that we usually think of as wholly intrinsic characteristics may in fact be partly socially determined. Their suggestions are compatible with (and in Baumeister's case, seemingly indicative of) a claim that these characteristics strongly emerge not from purely physical underpinning but in part from external factors, whether a nonphysical realm of meaning (also emphasized by Tattersall this volume) or social features of the human species. I make two points of connection: first, there is a nascent literature in cognitive science developing and exploring the thesis that some large, hierarchically-organized human collectives constitute, for a time, group minds, in the sense that the group possesses and processes informational states that none of the individuals possess. (Hutchins 1995 develops this thesis at length in his analysis of the complex, distributed behavior of several thousand people involved in the maneuvers of a modern naval ship.) How this idea fits with either of my two concepts of emergence is an open question. (Theiner 2011; Theiner and O'Connor 2010) Second, and most pertinent to the present volume, the suggestion of a possible social dimension to emergence connects to Thistleton's and Torrance's (both this volume) *corporate* eschatological understanding of the Christian theological doctrine that human persons are divine image bearers. While this takes us very far afield from scientific concerns, it serves to underscore how correctives to Enlightenment individualism are needed in multiple inquiries into what it means to be human.

Conclusion

I began by raising the question whether the distinctive abilities and tendencies of mature human beings indicate a discontinuity in the evolutionary processes that gave rise to us. My way of addressing this question was through the abstract philosophical lenses of a pair of concepts of emergence. If human characteristics and behavior are merely weakly emergent, there is space for discontinuity at the level of psychology and its biological underpinnings: capacities that have neither direct precedent in the terms of evolutionary anthropology nor equivalence in other known species. The still-open question is one of *degree* of dissimilarity to historical hominid precursors or contemporary mammalian intelligent species. However, weak emergence assumes strict continuity at the fundamental physical level. (In the *Game of Life*, fundamental patterns are universal and unchanging, even as they give rise to higher-order novel patterns of many forms within different varieties of stable structure.) In a merely weakly emergent world, there is nothing new under the sun, fundamentally speaking. But I have argued that capacities associated with human and other animal conscious awareness, and with the enhanced conscious feature of subjectivity, or having a point of view as a self, are not only weakly but also strongly emergent. If this is correct, then the historical and comparative psychological discontinuities that prove to be the case entail fundamental discontinuities: deeply novel characteristics on the world stage. To be human is to be made of the same substrate as everything else in the physical world and to be a product of a very long process of incremental biological change. Yet it is also to be a part of the world that transcends its constituting material by becoming aware of oneself as a self, separate from the rest of reality and acting in intentional ways that reflect that awareness. It is also to be

rooted in a community of fellow beings, whose purposes and fortunes are deeply interwoven. What is the relationship of 'we' to 'I' and 'you': is it always merely a conjunction, or are there respects in which we—in temporary, interacting communities or in its widest scope, encompassing all of humanity—constitute an ontological whole, making a fundamental difference as a whole to the way our corner of the world unfolds? This question is scientifically fascinating; it is also one whose answer has important philosophical and theological consequences.

References

Alexander, Samuel 1920. *Space, Time, and Deity. Vol.I-II*. New York: The Humanities Press.

Baumeister, Roy (this volume). "Emergence of Personhood: Lessons from Self and Identity."

Brasil-Neto, J.P., Pascual-Leone, A., Valls-Sole, J., Cohen, L.G. & Hallett, M. 1992. Focal transcranial magnetic stimulation and response bias in a forced choice task, *Journal of Neurology, Neurosurgery, and Psychiatry,* 55, 964–66.

Brown, Warren and Sternberg, Lynn Kerlin Paul (this volume). "Brain Connectivity and the Emergence of Capacities of Personhood: Reflections from Callosal Agenesis and Autism."

Clarke, Peter G.H. 2013. "The Libet Experiment and its implications for conscious will." Faraday Paper 17. The Faraday Institute for Science and Religion.

Delgado, J.M.R. 1969. *Physical control of the mind: Toward a psychocivilized society*. New York: Harper & Row.

Doris, J. 2002. *Lack of character: Personality and moral behavior.* Cambridge: Cambridge University Press

Gazzaniga, M.S. 1994. Consciousness and the cerebral hemispheres. In M.S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1391–99). Cambridge, MA: MIT Press.

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

McLaughlin, Brian 1992. "The Rise and Fall of British Emergentism" In Ansgar Beckermann, Hans Flohr & Jaegwon Kim (eds.), Emergence or Reduction?: Prospects for Nonreductive Physicalism. De Gruyter (1992)

Mele, Alfred 2009. *Effective Intentions*: The Power of Conscious Will, Oxford UP, 2009

Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences,* 8, 529–66.

Nahmias, Eddy 2002. When Consciousness Matters: A Critical Review of Daniel Wegner's the Illusion of Conscious Will. *Philosophical Psychology* 15 (4):527-541.

O'Connor, Timothy 1994. "Emergent Properties," *American Philosophical Quarterly* 31, 1994: 91-104.

O'Connor, Timothy 2009. "Conscious Willing and the Emerging Sciences of Brain and Behavior," in George F. R. Ellis, Nancey Murphy, and Timothy O'Connor, eds., *Downward Causation And The Neurobiology Of Free Will*. New York: Springer Publications, 2009: 173-186.

Pereboom, Derk 2011. Consciousness and the Prospects of Physicalism. Oxford: Oxford University Press.

Tattersall, Ian (this volume). “Human Evolution: Personhood and Emergence.”

Theiner, Georg 2011. *Res Cogitans Extensa: A Philosophical Defense of the Extended Mind Thesis*. Frankfurt: Peter Lang.

Theiner, Georg and O’Connor, Timothy 2010. “Emergence and the Metaphysics of Group Cognition”, in Timothy O’Connor and Antonella Corradini, eds., *Emergence in Science and Philosophy*. Routledge Studies in the Philosophy of Science, Routledge, 2010: 78-117.

Torrance, Alan J (this volume). “Retrieving the Person: Theism and the Scope of Empirical Research.

Thisleton, Anthony (this volume). Xxxxxxxxxxxxx.[TITLE?]

Wegner 2002 Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Zeman, Adam (this volume). “The Origins of Subjectivity.”