THE EMERGENT SELF

Philip Clayton

Abstract: The debate over the nature of human personhood continues to be dominated by dualist positions on the one side and materialist reductions on the other. But new understandings of emergent complexity over evolutionary time offer a more balanced and nuanced account, one that leaves place for freedom, moral responsibility and personal agency. This account is upwardly open to God and the possibility of continuing existence after death, without however depending on strong knowledge claims that conflict with science.

Keywords: BIOLOGICAL EVOLUTION, DIVINE ACTION, DUALISM VERSUS MATERIALISM, EMERGENCE OF PERSONHOOD, EMERGENT COMPLEXITY, FREEDOM, LIFE AFTER DEATH, MORAL RESPONSIBILITY, PHYSICALIST THEORIES OF THE SELF, SOCIAL LEARNING

Moving on from Dualism and Materialism

Traditional Christian theologians, for whom the orthodox creeds are decisive, have no choice but to affirm a view of the self that is dualistic in at least one sense. There must be an eternal, non-physical soul, which represents the essence of the self. After the death of the body that soul, or self, proceeds to live in eternal beatitude with God – generally reunited with a resurrected body, though some theologians have described it as a purely spiritual state.[1]

1 Note that traditional Christian theology is dualistic only in this very specific sense. Many theologians insist that existence without a body is a deficient state, and Thomas Aquinas defends a hylomorphic anthropology, which requires the union of form and (at least some kind of) body.

 doi:10.3828/mb.2016.12

In the opposing camp, scientific materialists (reductive physicalists) have been constrained by their own version of orthodoxy. Generally within science, explanations are causal and only those entities (objects, organisms) that play a causal role in scientific explanations are taken to be real.[2] Many things (objects, entities, organisms) can emerge and have emerged over the course of cosmic history – quarks, bacteria, wolves and human persons – and to outward appearances these are real things with genuine causal powers. However, if it were the case that matter/energy, as studied in physics, is ultimately the only causal power, then these other emergent objects would enjoy only a derivative reality. The biologist may speak *as if* the bacterium were an agent and did something, but what really happened were changes in a physical system, and in the end it is physics alone that is responsible for describing what physical systems are and do. This is the orthodoxy that sways the scientific materialist. It is not (as is usually claimed) a conclusion demanded by the data, but an underlying theoretical position in whose light all data must be interpreted.

There should be no mystery how this camp will speak about selves and their experience; it is the same answer that we encountered with the bacterium, just a bit more complicated. No one has put this point more clearly than Stephen Weinberg, the Nobel-winning physicist, who seems to think that all explanatory arrows point downward.[3] When you encounter something as strange as a human being who claims to experience the Divine, you look for causes that you can make sense of. The immediate causes are linguistic, social, cultural and historical. All these have somehow been stored in the electrochemical system that is the brain, and biophysics has much that it can tell us about that causal system. Biological evolution produced the brain, which means that we can trace neural structures and functions backwards to their contexts of

2 Wesley Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton: Princeton University Press, 1984).

3 Steven Weinberg comes closest to this position in his *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature* (New York: Pantheon Books, 1992). The phrase 'all explanatory arrows point downward' is often attributed to Weinberg with reference to this book, but interestingly the phrase does not actually appear anywhere in these pages.

evolutionary origin. The further back and the further down we go, the closer we get to physics. When the explanations have become physical explanations, we know we are getting to something solid. Finally – or so we are told – when one can run the explanations backwards to physics, and run the causal story forward to the neurophysiological state that correlates with Jane's experience of God, scientists will be fairly sure that they know what is really going on.

Fortunately, this is not the only way to construe a science-oriented theory of the self. The opposing camps of classical dualism and classical materialism (reductive physicalism) do not cover all the options open to us; indeed, they both appear to be throwbacks to a more metaphysical age, and to build one's theory of the self upon either one of them is not advisable. A third approach, which I represent,[4] is no less excited than the materialists are when scientific work is able to write down simple, fundamental laws that express patterns across a wide range of phenomena. The difference is that this approach – the emergentist hypothesis – emphasises two facts that are inadequately acknowledged by our opponents: evolution produces discontinuities as well as continuities, and the existence of general laws does not complete the explanatory task. Consider the new picture that now unfolds.

Emergent Complexity

All existing things trace their origins back to a single Big Bang, but cosmic evolution has (for whatever reasons) produced a series of very different kinds of systems. Some years ago Stuart Kaufmann and I argued that the possibility space of biological systems, and hence the dynamics of biological systems, cannot be finitely prestated in terms of physics.[5] Of

4 Philip Clayton, *Mind and Emergence: From Quantum to Consciousness* (Oxford: Oxford University Press, 2004); *The Re-emergence of Emergence: The Emergentist Hypothesis from Science to Religion,* co-edited with Paul Davies (Oxford: Oxford University Press, 2006); *Adventures in the Spirit: God, World, Divine Action* (Minneapolis: Fortress Press, 2008).

5 On Emergence, Agency, and Organization' (with Stuart Kauffman), *Philosophy and Biology* 21 (2006): pp. 501–21.

course, everything that occurs will be consistent with physical laws; and when organisms do evolve, we can describe their physical and chemical properties. But the dynamics of Darwinian evolution are not theoretically reducible; hence physics cannot predict or explain the evolution of species.

These principles lie at the heart of what is now called emergent complexity.[6] Biological agents – organisms – play an irreducible role in biological explanations. Genes transmit hereditary information across generations and influence (sometimes *strongly* influence) the behaviour of organisms. In order to explain the behaviour of organisms in their environments, however, biologists develop and test theories based on the evolutionary goals and tasks of individuals, groups and species. Structures, functions of behaviour and interactions with other organisms play crucial roles in these explanations. As Stuart Kauffman has written, each organism is 'out to make a living' in its environment,[7] and we need to understand it in that context.

At some point social learning – the transmission of acquired information – begins to play a role in these explanations. Later, social learning begins to acquire certain shared patterns that are unique to subgroups within a species. These cultural patterns are certainly present in bird species, and they play an increasingly important role in the study of the more complex mammals. The famous primatologist Frans de Waal once noted in a talk that he could identify different groups of great apes based solely on cultural differences in their behaviour. It follows from the complexity of higher primates and their cultural practices that differences between individuals also become increasingly important. Primatologists begin to speak of 'the emergent self'.[8]

6 Jeanne E. Arnold, ed., *Emergent Complexity: The Evolution of Intermediate Societies* (Ann Arbor, MI: International Monographs in Prehistory, 1996); Charles Lineweaver, Paul C.W. Davies and Michael Ruse, eds., *Complexity and the Arrow of Time* (New York: Cambridge University Press, 2013); Martin A. Nowak and Sarah Coakley, eds., *Evolution, Games, and God* (Cambridge: Harvard University Press, 2013).

7 Stuart Kauffman, *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity* (Oxford: Oxford University Press, 1995).

8 Raymond L. Neubauer, *Evolution and the Emergent Self: The Rise of Complexity and Behavioral Versatility in Nature* (New York: Columbia University Press, 2012).

This is the starting point for philosophical and theological reflection on the nature of human identity. In our species the evolutionary bet was made not on speed or strength, but on cognition. The layering and striation of the human brain, and especially the major investment in the prefrontal cortex, produced the capacity to develop mental images and concepts, to model the probable behaviours of others, to imagine scenarios and then orient one's actions around these imaginary worlds. The details of beliefs and concepts probably play a greater role in explaining the actions of humans than do the 'givens' of biology and physics. Psychologists study individual self-conceptions and idiosyncratic beliefs; sociologists and cultural anthropologists study shared systems of belief and meaning. None of these capacities contradict the natural sciences or break physical laws; indeed, the biological sciences contribute essential pieces to the complete explanatory picture.

Personal Identity

We have achieved an account, I suggest, that ascribes neither too great nor too small a role to scientific explanation. We have seen that physicalism ascribes too much explanatory authority to physics, whereas dualism undercuts the explanatory role that different sciences play as one traces the results of emergent evolution. We can now explore what these conclusions imply for some of the classic philosophical and theological questions about the nature of the self: the nature of the person, personal identity over time, souls, resurrection and the hope for life after death.

Emergentist approaches to human identity rely on natural (naturalistic) explanations as far as they will take us. (At the end we return to the question of what happens when we have taken them as far as they will go.) The argument for an emergence-based theory of the self involves two crucial steps.

(1) Mental representations of the world

Every creature with a brain represents the external world in some way. The *means* are electrochemical states and processes in the central nervous system, but *what* is represented are aspects of the environment

that are salient to the individual person or animal: where food or danger is; memories; information about other individuals. As brains increase in complexity, more and more information becomes available, allowing for more and more complex behaviours among primates.

Human brains are not qualitatively but only quantitatively different from, say, bonobo brains. Still, the increased capacities are remarkable. As a species, we create shared cultural worlds that become our 'home' perhaps even more than our biological surroundings. (What does it mean that over the last two hours my twins have been creating a 'civilisation' on a computer screen?) As language-users, we orient our lives around symbolic realities; we are moved to tears, and perhaps to the ultimate sacrifice, by the symbolic overtones of a flag or religious symbol.[9] As conscious individuals, we each inhabit a unique private world of hopes, dreams, fears and impressions, some of which are objectively based and others are completely constructed by the individual and projected onto everything he sees or does.

(2) Personhood, freedom and moral responsibility

Physically, each of us has a body, a brain and whatever mental life that brain and the sum total of its inputs make possible. Our personal identity emerges from these givens and what we do with them. Many philosophical traditions have held that personhood has to be grounded in a unique eternal essence (soul, *ousia, substantia, atman, jiva*). While I cannot prove that such realities do *not* exist, I dispute that they are necessary for explaining persons and their actions in the world.

Recall that, on the nonreductive emergentist view, emergent realities that appear later in evolution are no less real than those (such as quarks, gluons and atoms) that appeared earlier. Personhood is an emergent reality that we associate primarily with humans. We know many individuals who are clearly persons; we treat some other individuals *as persons* because they will someday manifest personal qualities or did so in the past; and we can imagine someday meeting and interacting with

9 Terrance Deacon, *The Symbolic Species: The Evolution of Language and the Brain* (New York: W. W. Norton, 1997).

persons who are not human. We project personal qualities onto our dogs, our dolls, our computers and many other objects, although when we stop to think about it we can usually recognise when we are projecting.

Clearly, persons and communities can decide to treat others as persons (or not). But what is *personal unity*, the unity of the person-as-a-whole over time, and how does it emerge? Steven Knapp and I have argued that personhood is intrinsically connected with human freedom, agency and moral responsibility.[10] All four concepts involve the relationship of an individual to herself and others over time. Together, we have suggested, they are sufficient to account for the emergence of human personhood.

Imagine that an individual makes a certain decision at some time T_1. The emergence of her personhood turns on the manner in which, at some later time T_2, she relates to her decision at T_1. To have personal identity over time is to take personal responsibility for that past action, to acknowledge it as an action that *I* performed, and thereby to define myself as the same person who did that thing. More specifically, if a given act is really to be *my act* and not just an accident that befalls me at a certain moment of time, then what causes me to perform that act cannot be an isolated person-state but must somehow be *the person I am*. When a rational agent decides to do something, part of what she is deciding is the moral status of her future self, the type of person she will become. Insofar as she is a rational as well as a moral agent, she is deciding what kind of self her future self will have to take into account, retrospectively, in its own moments of decision.

Personhood, moral responsibility, rational agency and freedom are thus intrinsically interconnected concepts. *A free act is one in which the person, as she decides whether to perform a morally significant act, is making a decision about, and on behalf of, her abiding personal and moral identity*. Likewise, the person in her future states must be able to choose to accept or reject the moral status that the decision of her past self has, so to speak, projected onto her. The agent exists as a person

10 Here I draw on an argument co-developed with Steven Knapp and presented most fully in Clayton, *In Quest of Freedom: The Emergence of Spirit in the Natural World* (Göttingen: Vandenhoeck and Ruprecht, 2009), chapter 4. Parts of what follows are adapted or excerpted from that presentation.

only insofar as she regards herself as possessing an enduring identity of precisely this kind. To function as a person over time, the actor at T_2 must accept or reject her identity with the action at T_1, and indeed with the actor who carried out the action at that time.

When we respond in this way, we define ourselves as persons both retrospectively and from this point forward. When one does this, she defines her past self as *herself* – as the one who launched the moral/personal project that she now continues and for which she is still responsible, one for which she can be praised or blamed in the present. Likewise, she takes a similar stance toward her future self, defining that self as one who will exist in continuity with the present self and who will thus continue to be responsible for the actions taken now.

It is only in the light of the moral status we have inherited (as it were) from our own past acts that we can choose to alter the kind of person we are by the acts that we now decide to perform. And it is only in virtue of our affirming the freedom in which we performed certain acts in the past that we can affirm our present freedom to define the person that we will henceforth be. An individual constitutes herself as a person when she defines herself as a free, rational, moral agent over the arc of time pointing back into the past and over the arc of time pointing into the future. In short: *to be a person is to be an agent whose action is caused by, or in some sense essentially involves, the identity that this very action brings into being*.

As a person, I declare myself responsible for my past, for the past things that 'I' have done. When I choose to perform a certain act, I (consciously or unconsciously) imagine a self, namely, the self who is now choosing the identity implied by the act. The self I imagine as the final cause of my act becomes, in that moment, the self that chooses. Personhood is thus a self-reflexive concept, intimately bound together with freedom and with this temporally extended moral responsibility. At the moment of a morally significant action, a person becomes the very self that she chooses to become in virtue of the act that she is performing.

God, Self and Soul

Many of the ideas that preoccupy humans can be traced back to their biological and physical causes and functions, including some religious ideas. In these cases we rightly give weight to scientific accounts: evolutionary psychology, sociobiology, physical anthropology, the biology of belief, the cognitive science of religion. Durkheim, Marx, Freud and their successors become our guiding authorities.

Although it is often claimed that 'we now know' that all religious ideas are fully explainable in terms of their evolutionary origins and functions, this is false. What history shows us instead is a cyclical pattern: times when scientific theories and data are seen as complementary to meta-scientific reflection, and times when they are seen as *replacing* such reflection. From Richard Braithwaite's early claim that all religious language is merely emotive, to E. O. Wilson's influential book *Consilience*[11] (which basically claims the same thing), to the 'new' atheism (which isn't),[12] we undeniably inhabit one of the incompatibility phases.

The only way to rule out meta-scientific reflection would be to establish that, necessarily, all knowledge is scientific knowledge. But because this is a philosophical claim, establishing it would require relying on philosophical arguments and not merely scientific ones – which would contradict the claim in question.

Assuming, then, that a scientific monopoly over all knowledge cannot be established, the door will always remain open to meaningful reflection beyond science. A God may exist who is the creative Source of all things, the Ground of value and the Lure toward the Good. Experiences of the Divine do not need to be dismissed as illusions that are better explained by the psychological or evolutionary functions of religion. Religious views of the world and the self, of ultimate origins and ultimate ends, will always be part of human speculation and spiritual experience. The emergentist breaks no rules when she affirms these things, and claiming

11 Edward O. Wilson, *Consilience: The Unity of Knowledge* (New York: Alfred A. Knopf, Inc, 1998).

12 Terry Eagleton, *Reason, Faith, and Revolution: Reflections on the God Debate* (New Haven, Yale University Press, 2009).

to know they are false goes a step beyond what we can show. Those who have had certain types of religious experience are further inclined to believe and to speculate.[13]

But note the disanalogy. Beliefs about ultimate reality, such as my belief in God, pertain to a realm that is beyond scientific data. By contrast, we possess massive (and growing) amounts of data on human persons. Here scientific answers are available that simply do not exist when comparing metaphysical theories about ultimate reality. The two cases are epistemically disanalogous, for one ought to proportion her belief to the data.

In both cases, attractive metaphysical possibilities exist. For example, when a person accepts moral responsibility for her past and future states, as described above, perhaps she is actually intuiting her unchanging metaphysical essence – her soul. Perhaps this metaphysical substratum is necessarily eternal and therefore survives the death of the body, providing the basis for eternal presence with God.

Perhaps. But here there is natural knowledge of human personhood to take into account as well. The possibility of souls, resurrection and afterlife cannot be ruled out. Yet inability to prove the negative does not establish the positive. These possibilities remain objects of hope and faith. But the necessity of ongoing dialogue with emerging neuroscience demands of believers rather more humility, an epistemically lighter touch, than has been the wont of theology heretofore.[14]

Philip Clayton is Ingraham Professor of Theology at Claremont School of Theology, CA.

13 For example, Alfred North Whitehead held that all of one's thoughts and feelings are eternally remembered by God, which guarantees an 'objective immortality' of the self within God. I follow Marjorie Suckocki (*The End of Evil: Process Eschatology in Historical Context* [Albany: State University of New York Press, 1988]) in affirming both objective and subjective immortality. Not only the objective contents of thought, but also the subjective qualia and agency that one experiences as constituting her self can be preserved as a distinct locus of agency within the eternality subjectivity of the divine Spirit.

14 I wish to give special thanks to my Research Assistant, Kirianna Florez, for her help in researching this paper and preparing it for publication.