

# The Emergence of Self-Awareness

Uriah Kriegel

*Times Literary Supplement* (TLS), March 2007

=====

Douglas R. Hofstadter, *I am a Strange Loop*, 337pp. Basic Books

=====

We often take it for granted that much of our world has been exposed and explained by science. The best remedy for such complacency is to take a closer look at those few remaining mysteries that present the deepest challenges to our understanding of the world. Three in particular come to mind: Why is there something rather than nothing? Why is some of what there is alive? And why is some of what is alive conscious, or self-aware?

We are somewhat in the dark with respect to the first question, though some scientists would tell you that, ultimately, some combination of Big Bang and String Theory will be the answer. Our most impressive position is with respect to the second question, where it is widely believed that the so-called modern synthesis of evolutionary biology and Mendelian genetics explains not only the origin of species but also the origin of life. Oddly, perhaps, it is the third question that presents the gravest challenge to our attempt at an intellectual domestication of the world. It is the one area of science where our understanding is still in what posterity will surely refer to as the pre-historical stage of inquiry. The problem is that we lack any compelling paradigm to guide our research on consciousness and self-awareness. Many take these phenomena to represent science's final frontier.

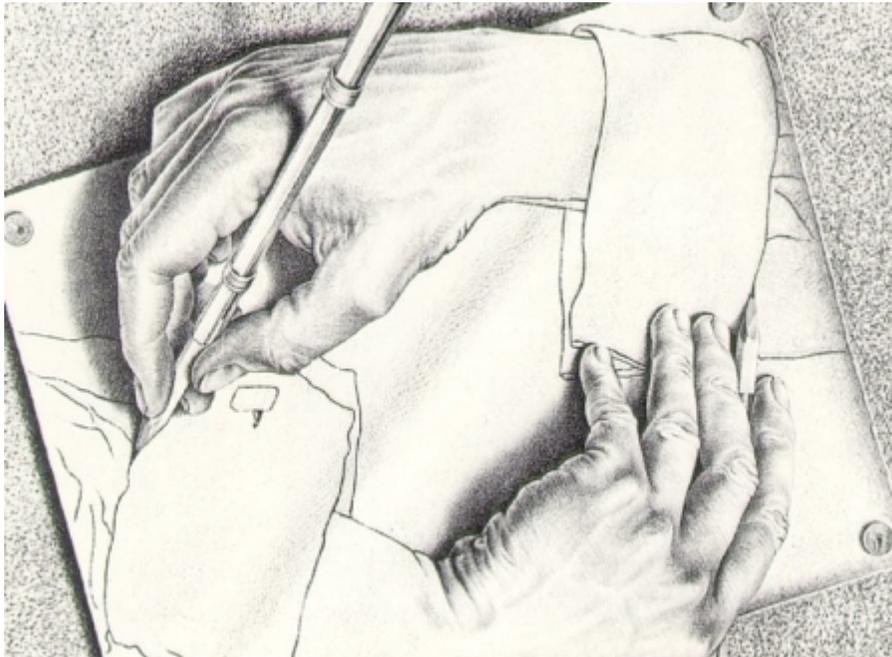
I say "oddly" because our own self-awareness is, in one way, the phenomenon closest to us of all. And yet it has proven extraordinarily elusive. There are probably many reasons for this, but it is sometimes thought that the deep problem is a fundamental incompatibility between the aspired objectivity of scientific inquiry and the essential subjectivity of self-awareness. The *modus operandi* of modern science is to abstract away from our subjective perspective on the world in order to home in on the objective way it is in itself. This might work for most phenomena, but not so much for phenomena that are inextricably tied up with subjectivity, such as self-awareness and consciousness. How is the objective study of subjectivity supposed to proceed? Nobody knows.

It is into this unwelcoming fray that Douglas Hofstadter steps with his new and cutely titled book, *I am a Strange Loop*. Almost thirty years after the publication of his well-loved *Gödel, Escher, Bach*, Hofstadter revisits some of the same themes. The purpose of the new book is to make inroads into the nexus of self, self-awareness, and consciousness by examining self-referential structures in areas as diverse as art and mathematics.

Hofstadter is the man for the job. His treatment of the issues is approachable and personal, you might even say subjective. His discussion is never overtechnical and his prose never overbearing. He stays close to the surface of real life at all times, even as he discusses matters of the highest level of abstraction, and his book is full of fresh and rich real-life examples that give texture and authenticity to the discussion. In these ways and others, Hofstadter is what philosophers would call “a good phenomenologist”, meaning a good student of lived conscious experience. If we hope to unravel the mysteries of self-awareness, the combination of good phenomenology and a focus on self-referential phenomena is a promising starting point.

Hofstadter’s principal thesis is that we ourselves, qua conscious beings, are “emergent self-referential structures”. The book thus revolves around two main ideas: the idea of an emergent phenomenon and the idea of self-reference, or of a “strange loop” to use Hofstadter’s technical term.

A strange loop is a phenomenon that involves reference to itself. An artwork, a thought, or a sentence may twist back onto itself and self-refer. Thus, the sentence “this very sentence is written in English” is self-referential, because it refers not to any old sentence, but to itself. A more surprising example explored by Hofstadter (which does not employ the demonstrative expression “this very”) is the sentence ““preceded by itself in quote marks yields a full sentence’ preceded by itself in quote marks yields a full sentence.” (Think about that for a moment.) Many other self-referential phenomena are discussed throughout the book, including self-videotaping videos, self-proving mathematical proofs, Escher’s self-referential paintings, etc.



Different systems may exhibit different degrees of self-referential sophistication, and for Hofstadter, the more sophisticated a system's self-referential capabilities, the more soulful it is – the more robust its selfhood, its existence as an “I”. A snail probably has no conception of itself whatever, and to that extent is soulless. A dog has some conception: it knows that its paw is its own. But the dog's self-conception is very limited. For example, studies show that dogs do not recognize themselves in the mirror. In these studies, a mark is painted on the animal's forehead, and when a mirror is brought in, it is observed whether the animal makes any attempt to wipe the mark off. The number of animals who pass the “mark test”, as psychologists call it, is surprisingly small: the chimpanzee, the orangutan, the bottlenose dolphin, and the Asian elephant are the only ones on record. Even gorillas, baboons, and African elephants fail, as do humans younger than eighteen months. The five self-recognizers (including ourselves) would thus constitute, by Hofstadter's light, a soulful elite within the animal kingdom. At the same time, it is clear that there is a kind of self-awareness that is even more sophisticated and more elusive than the recognition of one's embodied self in a mirror, and there is probably a kind of self-awareness that only adult humans exhibit, and which represents the pinnacle of soulfulness.

One of the deepest questions in this area is how any soulfulness can exist in a world of sub-atomic particles buzzing about in mostly empty space. The ancients believed that the self emerged from the activity of the heart. It is clear today that the seat of selfhood and self-awareness is rather in the brain. A pressing question arises, however: how can something as majestic as self-awareness emerge from the thoughtless activity of millions of nervous cells vibrating inside the darkness of the skull? Some of the most fascinating insights in this book pertain to this issue (avid Hofstadter readers will recognize them from previous writings).

As an example of an emergent phenomenon, Hofstadter tells of the time he tried to take out a wodge of old envelopes from a box in his drawer, and could swear he felt a marble nestled among them. It turned out that the uncanny appearance of a marble was produced by the extra layers around the envelopes' V. This marble appearance is for Hofstadter a paradigmatic emergent phenomenon, arising as it does from wholly unrelated underlying elements. In this case, we are inclined to dismiss the marble appearance as illusory, because of how accidental and idiosyncratic it is. But according to Hofstadter, when an appearance is produced reliably and persistently, under many different conditions and for many different observers, we start taking it more seriously. It becomes “more real” to us. Ultimately, he proposes, our self is an emergent appearance of this sort. In fact, it is the realest emergent object in our inner world.

The issue of emergence arises with special acuity for self-awareness, but it applies already to awareness of things other than oneself. At some level, we know that all that is really going on in our head is the propagation of electrical impulses among nervous cells. But, of course, this is not how we experience our mental life. We experience it as involving the continuous representation of the world. Our ideas, hopes, and desires all employ symbols that represent how the world is and how it could be. The question of

emergence arises already at this point: how can we reconcile our lived, experienced conception of ourselves with the conception of ourselves portrayed by modern science?

Hofstadter tells a wonderful story about the emergence of symbolic thought from neural activity. Imagine a pool table with a million small interacting magnetic marbles (“simms”) on it. These simms careen about the space of the pool table, which he calls the “careenium”. In some circumstances, the simms get magnetized to each other, and may form ball-shaped clusters – “simmballs”. The behaviour of single simms is random, but that of simmballs is not. The simmballs move around inside the careenium depending on what kind of external forces impinge on the careenium’s external walls. Thus the behaviour of simmballs inside the careenium comes to reflect conditions outside it.

Our minds, says Hofstadter, work in just this way. Inside the cranium (careenium) are millions of nervous cells whose behaviour is more or less meaningless. But sometimes large clusters of cells coordinate their behaviour in response to the way the external world impinges on parts of the cranium, such as the retina or the ear drums. When they do, these clusters come to constitute symbols (simmballs), symbols that represent external conditions in a sustained manner that effectively constitutes a rudimentary awareness of the external world. The moral is that although we cannot find anything like symbolic thought or awareness when we look at individual brain cells, if we widen our view and consider slightly more abstract and more spread-out structures and patterns within the brain, we just might.

What is true for awareness of things other than oneself is true also for self-awareness. One special symbol which takes more time to form is the “I” symbol. If the careenium developed a simmball with which to represent its own operations, it would come to be a self-referential system and have an “I”. Our cranium does have a symbol that represents itself, and it is therefore self-aware. Importantly, however, our symbolic representations have a somewhat “coarse grain”, as philosophers say. When we represent an ice cube, for example, we are aware of it simply as a single, homogenous, clear-pinkish cube. We are not aware of the millions of hydrogen and oxygen atoms making it up. Likewise, when we represent ourselves, we are not aware of the millions of neurons inside our brain, but rather of the various symbols that clusters of them make up. That is to say, the cranium is aware of itself precisely as a theatre of ideas, desires, and hopes, not as a container of cerebral molecules buzzing about meaninglessly. And that is why we experience our mental life in those terms, even though ultimately it all rests on the purposeless activities of so many individually insentient nervous cells.

The thesis that conscious selves are emergent self-referential structures strikes me as sending us in exactly the right direction. Nonetheless I find two important problems with it. First, it is slightly disconcerting to discover that, like the marble in the envelope box, I am a mere appearance, albeit a stubborn one. Of course, Hofstadter insists that such stubborn appearances are very real “to us”. But that does not make them real (period). When I hallucinate a lion, but am unaware that I am hallucinating, the lion is very real “to me”. Yet for all that it is entirely unreal. And if we had a whole group of people

hallucinating a lion in concert, the lion would not miraculously assume flesh and blood as a result.

Secondly, self-reference can take place in any number of completely unconscious systems, as Hofstadter's varied examples show. Many inanimate objects, including your office desktop, often perform self-monitoring functions. So self-reference by itself cannot suffice for consciousness and self-awareness. Perhaps what is needed is that not only the system, but also specific states of the system, be self-referential. This proposal, however, requires separate consideration, which Hofstadter does not offer.

That problems should arise with any attempt to tackle science's last frontier is par for the course. The important thing at this pre-historical stage of inquiry is that Hofstadter's book points in the right direction: the phenomenologically responsible exploration of emergent self-referential phenomena. Although this is certainly controversial, my own view is that if a viable paradigm to guide research into self and consciousness is to arise, it would have to be from the neighborhood of ideas explored in this book – and explored so entertainingly, no less. Hofstadter's engaging style, his feather-light prose and his determination to establish genuine communication with the reader add up to high intellectual adventure – and not only for those who, like myself, are already sympathetic to his ideas.

=====

Uriah Kriegel is assistant professor of philosophy (and cognitive science) and associate director of the Center for Consciousness Studies at the University of Arizona, as well as SESQUI fellow at the University of Sydney. His work focuses on consciousness and awareness.