

**Can evolutionary theory explain the existence of consciousness?** A Review of Humphrey, N. (2010) *Soul Dust: The Magic of Consciousness*. London: Quercus, ISBN 9781849162371

*Reviewed by Max Velmans, Goldsmiths, University of London.*

*Journal of Consciousness Studies*, Volume 18, No.11-12 (2011), pp. 243-254.

**Key words:** evolutionary theory, evolution, consciousness, illusion, Humphrey, Darwin, Dawkins, reproductive fitness, science, reductionism, doctrine, faith, soul

In September 2000 I organised a public debate on the motion “Can evolutionary theory explain the existence of consciousness?” at the Institute of Contemporary Arts in London. In spite of having eminent speakers such as Jeffrey Gray and Stevan Harnad already lined up, we found it very difficult to tempt a prominent evolutionary theorist into defending the motion. For example, the chair of the debate Alex Kacelnick (professor of Behavioural Ecology at Oxford) was unable to persuade his colleague Richard Dawkins to speak. Why not? Because the problem consciousness poses to evolutionary theory is far more difficult than it might seem. Many readers of this journal will be familiar with the issues, which I do not have space to review here. But the nub of the problem is this: Darwinian evolutionary theory is a *functional* theory. Stripped down to its essence, it has only one explanatory mechanism: novel biological forms and functions emerge through random variation of genes, and only persist if in some way they enhance the ability of organisms (or populations of organisms) to propagate their genes. Given this, for evolutionary theory to explain the existence of consciousness, it must show (a) *how* consciousness emerged through random variation in the genome of organisms in which it was previously absent, and (b) how that emergence enhanced the ability of those organisms to propagate their genes. Readers will recognize that (a) presents a “hard problem” much discussed in the consciousness studies literature. But why is there a problem with (b)?

It’s not that folk haven’t tried, all the way back to George Romanes (1885) a friend of Darwin who suggested that the function of consciousness is to deal with novelty and complexity. With the emergence of cognitive psychology in the 1960’s there were literally dozens of proposals that gave consciousness responsibility for every major phase of human information processing associated with novelty, complexity and much else. There were also a number of weighty books on the subject, for example John Crook’s (1980) *The Evolution of Human Consciousness*. Humphrey too contributed to these efforts over many years, arguing for example that consciousness was necessary for humans to develop a theory of mind, required to operate in complex, co-operative groups (Humphrey, 1982, 2002). But, as I noted in Velmans (1991), none of these proposals really bear up to close scrutiny. To begin with, most so-called ‘conscious processing’ is not conscious at all. Consider for example your ‘conscious’ ability to read this sentence. How much of the processing involved in reading is conscious? None of it! You may visually experience print on paper, phonemic imagery (inner speech) associated with the spoken versions of the words, and a feeling of having understood the meaning of the sentence (or not). But the processes responsible for recognising and decoding visually presented text which *result* in these experiences are *preconscious* —so by the time you have these experiences, the work has already been done. There is also a fundamental philosophical problem: information processing accounts of

mental functioning, like evolutionary accounts, are third-person accounts which require no appeal to first-person experiences for completeness. And that became very evident from what actually happened (and still happens) in the development of cognitive psychological theory. Third-person, cognitive models of how human minds process novel or complex stimuli (in speaking, reading, problem solving and other so-called conscious tasks) require only information processing of a kind that could be programmed into a computer. If a model of how a system might perform such a task is a good one, the computer should work, leaving any conscious experience that might or might not accompany such functioning nothing to do that isn't already explained by the information processing built into the machine. And the same applies to brains: once one has adequately explained their functioning in terms of information processing or in terms of the neurophysiology that embodies it, there is nothing left, in third-person functional terms, for first-person experiences to do—a problem recognised at the dawn of cognitive psychology by the psychologist George Miller (1962) in his book *The Science of Mental Life*, which the philosopher Saul Kripke later called the problem of “overdetermination”.

That left many of us scratching our heads. A few of us (including myself) argued that consciousness was simply too important to human life to dismiss as an epiphenomenon, and that rather than try to squeeze it into an information processing box too small to contain it, or a purely functional explanation of human behaviour into which it did not fit, we needed to go back to the drawing board and expand our understanding of the human mind in a way that gives full weight not just to the ways that we function but also to the ways that we experience. Most (including Humphrey) simply redoubled their efforts to squeeze consciousness into the existing conceptual box—and *Soul Dust* is his latest and (according to him) final attempt.

Before I examine his argument, I should say at the outset that this is not really a scholarly book of the kind that one might recommend to students and researchers in the field of consciousness studies. Written in a folksy style and published by Quercus, who focus on fiction books and non-fiction books intended for the general public, Humphrey's book is strong on literary quotes, but weak on the conceptual issues that his theory of consciousness must address—as we will see below. There isn't much reference to empirical research and, apart from frequent references to similarities and differences with Dan Dennett, there is little attempt to place his own views in the context of a long history of attempts to explain the emergence and persistence of consciousness in terms of Darwinian principles. Nor is there very much reference to where insights from others working in the field of consciousness studies converge with and precede his own views. On the contrary, the tone of the book throughout conveys the impression that all the ideas are newly minted and that, in our previous understanding, little has been achieved. For example in his summary of part 1 of the book he writes, “I have tried, in part 1 of this book, to do what no one has done before: to explain how phenomenal consciousness could be an evolved feature of the human mind.” (p69) As I have noted above, many have tried. But did Humphrey, unlike others, finally succeed?

Humphrey's point of departure is the (dubious) claim that consciousness *must be fully discoverable from a third-person perspective*. In spite of its seeming inaccessibility to external scrutiny, the presence of conscious qualia should be observable from the outside,

and it should be possible to discover the full content of consciousness from an external perspective. Why? “The reason is the ultimate one, the hand of natural selection. Since consciousness, as we know it, is a feature of life on earth, we can take it for granted that—like every other specialized feature of living organisms—it has evolved because it confers selective advantage. In one way or another, it must be helping the organism to survive and reproduce. And of course this can happen only if somehow it is changing the way the *organism relates to the outside world.*” (p14)

Humphrey also thinks that there is one truth about consciousness of which we can be *certain*: “We can be sure it did not happen accidentally. It must be the result of natural selection favouring genes that underwrite the specialized neural circuits—whatever they actually are—that sustain the illusion of qualia, giving rise to the magical mystery show for the first person. And *it is axiomatic that this will have happened only if those lucky enough to be spectators of this show have somehow been at an advantage in terms of biological survival compared with their less fortunate cousins.*” (p70—my italics)

We will return to the “illusion of qualia” and the “magical mystery show” below. But before we do, it is worth reflecting on what Humphrey takes for granted in these paragraphs, on which the rest of his argument depends. Note that when Humphrey writes “We can be sure it [consciousness] did not happen accidentally. It must be the result of natural selection favouring [the reproduction of] genes”, he conflates the conditions for how consciousness *came into existence* with the conditions responsible *for its persistence* once it appeared. Although he states here that “we can be sure it did not happen accidentally” he actually argues later in the book that it did—to use his own analogy, Nature was “doodling” (playing out random variations) and it just turned out that this doodle was useful for survival and reproduction. But *how* a random variation in genes might, even in principle, produce conscious qualia remains quite obscure in Humphrey’s account, as we shall see.

Note too, the question-begging nature of this argument. The question is “Can evolutionary theory explain the existence of consciousness?” But for Humphrey that isn’t even a question worth asking. For him, qualia are an illusion and *it is axiomatic* that such illusions could only have persisted if they enhanced biological survival—in which case evolutionary theory *must* be able to explain the existence of consciousness. This axiom also forms the basis for his next questionable assertion—that the presence of qualia and the contents of consciousness should be externally observable. Again, “The reason is the ultimate one, the hand of natural selection”, as qualia could only affect natural selection if they change how the organism relates to the outside world (which should be observable).

In short, for Humphrey, that natural selection can explain consciousness is a *doctrinal matter*, a clearly stated *article of faith*, not open to challenge by any empirical evidence—immune for example to extensive evidence that the conscious experience of others cannot be fully discerned from their behaviour and that experiences as such *cannot* be observed from the outside (fundamental limitations of third-person methods that are widely recognised both in psychological science and philosophy of mind).<sup>1</sup> Whatever this doctrinal commitment may be, it is not science.<sup>2</sup>

---

<sup>1</sup> I am not of course denying that if one had precise knowledge of the neural correlates of a given conscious experience one might then be able to predict the existence of that experience from third-person observation

Placing this major caveat aside for the moment, we can still ask whether Humphrey succeeds in his own terms. How for example does he account for the initial emergence of qualia and what such emergence achieves? To make these problems tractable Humphrey first *redefines what qualia really are* in third-person functional terms. So – although he accepts that being phenomenally conscious *seems* to involve what it is like to be something at a given moment, i.e. to experience feelings or what philosophers call qualia, he goes on to assert as a point of departure that “to experience sensations as having these features is to form a *mental representation* to that effect ....Thus “consciousness” (or “being conscious”), as a state of mind, is the cognitive state of entertaining such mental representations” and “Consciousness can change the subject’s life just to the extent that these *representations feed forward to influence what he thinks and does*”. ( P7)

It should be evident that redefining qualia in such cognitive terms once again *begs the question*. Rather than *demonstrating* conscious qualia to be “a cognitive state of entertaining mental representations” he simply asserts it. Those acquainted with the literature will know that such redefinitions of *first-person qualia* in terms of *third-person viewable cognitive representations* (of varying kinds) is a routine manoeuvre made by all functionalist reductionists, dating back to Armstrong (1968). And they all face the problems outlined at the beginning of this review: cognitive representations may or may not be accompanied by conscious experiences, and their functional roles in information processing systems would be the same whether or not they are accompanied by conscious experiences (for many additional problems of functionalist reductionism see review in Velmans, 2000, or 2009 chapters 4 and 5). Humphrey does not attempt to address these problems—although he does try to head off our natural intuition that in reducing conscious qualia that are “like something” to third-person viewable cognitive representations he is trying he is trying to “pull a fast one”. His defence? “Rather, if you make this objection, I would say you have just pulled a fast one yourself. You have fallen for the tempting idea that there is something conscious experience *actually is* that is separate from what the subject *thinks* it is—that is, the *mental representation* that he makes of it. But it is not so. ...I will argue that what I called at the start of this book the inadequate phrase “it’s like something” is not such a bad phrase after all. Because, when it comes to it, for a subject to have a sensory experience that *is like something* really is for him to represent the object of experience *as if it is something* with some very peculiar features. In short, for the subject to have a sensory experience that is like something is just for him to experience it *as what it is like*.” (p19) I will leave readers to evaluate this defence for themselves.

---

of the correlates. However without first-person access to a given experience as well as third-person access to any accompanying neural activity it would not be possible to *discover* those neural correlates, or to know what that experience is like.

<sup>2</sup> As Karl Popper observed, scientific theories can *never* be regarded as *axiomatic*. Even Newtonian Laws of Motion, which had been extensively verified for around 300 years, were shown by anomalies that were better explained by Einstein’s theory of relativity to give a fundamentally incomplete (and in some senses false) view of the natural world. And if that can happen to the best verified of all scientific theories, it could in principle happen to any scientific theory. Consequently, scientific theories are at heart, “best conjectures”. To count as “scientific” they must, at least in principle, be open to falsification by evidence that does not fit the theory.

Humphrey then develops his reinterpretation of “what it is like” to experience something, arguing that if this commonly used description of consciousness means anything, it must mean consciousness is like *something else* as in situations where we say X is like Y. However, in normal usage, the claim that, e.g. seeing “red” is like something doesn’t require one to compare that sensation with some *other* sensation. Consequently, he argues that there can only be one set of circumstances in which one can meaningfully claim that an experience is like something: “it would be when you recognize that Y does not or could not exist as an entity belonging to the ordinary world where you can test things, but might exist in another world with different rules to which you have no direct access..”(p30). In short, Y must be something transcendental—an interpretation which he then points out doesn’t make sense, as we have no way of making the comparison between conscious phenomenology and something transcendental. From this he concludes that consciousness must be *an illusion*, which must have been planted in the human mind by evolutionary forces (see p32). It’s again a simple matter to evaluate this argument for oneself. When you feel that the sting of a bee is “like something”, are you trying to compare that feeling to something *transcendental* (something otherworldly that is quite different to the pain that you feel)? If not, Humphrey’s argument that consciousness is an illusion can’t get off the ground.

### **How the brain deceives itself**

Humphrey then builds on these (shaky) foundations to develop an original story of how cognitive representations become the illusions that he thinks they are, beginning for example with the sensation of red: In its preliminary stages, according to him , “...sensation is sentition [something in between sensation, expression and exhibition]—the privatized expressive activity—as monitored by your mind ... [for example] Red light reflected from a tomato arrives at your eyes, and you create an internalized expressive response; you engage in redding. You monitor what you are doing so as to discover what is happening to you. And the representation you form of your own response is the sensation of red. Thus, for you to have the sensation of red means nothing other than for you *to observe your own redding.*” (p48) In short, sensation is initially a form of *self-monitoring*. Humphrey then admits that self-monitoring is not enough to produce phenomenal consciousness—which leads him to propose that these are *nonphenomenal sensations* (p48). According to him, the critical transition to phenomenal sensations only appears after the next step, at the level of a computation that integrates what happens over time, which requires a pattern of brain activity supported by re-entrant feedback loops. He then develops this account, drawing on Douglas Hofstadter’s idea of a strange loop—a version of a feedback loop that forms a strange mathematical object that deceives the brain that has produced it into thinking that it is a real object—rather like the way that the endless staircase illusion might look real but, being impossible, must be an illusion. Humphrey calls this impossible object an “ipsundrum”, and that, he claims, is what consciousness is: “We wanted a theory of how “it is like something” to be conscious of sensations. And now we have one. Consciousness is a magical mystery show that you lay on for yourself. You respond to sensory input by creating, as a personal response, a seemingly otherworld object, an ipsundrum, which you present to yourself in your inner theatre. Watching from the royal box, as it were, you find yourself transported to that other world.” (p39)

The potential importance of re-entrant feedback loops in the support of human conscious experiences is, of course, not a new proposal in consciousness studies, although Humphrey makes no reference to e.g. the work of Edelman or Llinás. Since the time of Hebb's (1949) work on reverberatory circuits it has also been recognised that patterns of neural activity need to persist for some time in order to support learning and, from the late 1960s, that something of this kind is required for the focal attentive processing and working memory normally associated with human consciousness. Within psychological science, the notion that perceptual and other contents of consciousness are *constructed* by the mind/brain is also a very old and well-accepted idea.

But what about the novel aspects of his thesis? What is it about persistence of neural activity over time combined with mental modelling that produces *qualia*? According to Humphrey, *qualia haven't actually been produced*, as conscious experiences, being impossible objects, are not real. Only brain processes of the kind that can be observed from a third-person perspective are real. That phenomenal qualities *seem* to be real is just a trick (a bit of magic theatre) played by the brain on itself!

In this, Humphrey moves a long way beyond anything that he is able to demonstrate. Other than showing how drawings of illusory objects might not correspond to real objects, Humphrey gives no account of *how* this trick is done. How can the brain can trick itself into thinking it has experienced *phenomenal qualities*, if these qualities do not, in some sense, exist? What could an "illusion" even mean in a world where "veridical experiences" themselves have no existence? Note that the constructed nature of phenomenal consciousness does not suffice to make it an illusion. While phenomenal models can vary in accuracy, they are conventionally thought to represent real states of affairs in the world.

### **Why the brain deceives itself**

And why should the brain deceive itself? According to Humphrey, it can only be to enhance survival and the propagation of genes. However in this book he departs in a major way from his earlier work, and finally abandons the view that consciousness in some way enhances *effective functioning*. Rather, he suggests, "What if its role is not to *enable* you to do something you could not do otherwise but rather to *encourage* you to do something you could not do otherwise: to make you *take an interest* in things that otherwise would not interest you, *to mind* about things that you otherwise would not mind about, or to *set yourself goals* that you otherwise would not set?" (p72) And, "I will not hold back my main conclusion, although I expect I may shock you with how simple it is (after a lifetime of working on the question, I have shocked myself). I think what the natural history reveals is that consciousness—on several levels—makes *life more worth living.*" (P74)

Humphrey then warms to this theme, developing an account of the profundity of conscious experience with extensive references to literature, poetry and philosophy for much of the rest of the book, elaborating on issues that *matter* for people—*joie de vivre*, a sense of being a self with its own metaphysical importance living a life that has meaning and purpose, the sense of living in an enchanted world, our fear of death, and even a belief in an immortal soul and an afterlife. All this, he argues, adds to one's *motivation* for survival and consequently the likelihood of passing on one's genes.

For Humphrey, the notion that consciousness makes life *worth living* is a radical departure indeed, presented as an original insight arrived at after a lifetime of work! I have to confess that having come to a similar conclusion about the centrality of consciousness to a life worth living in Velmans (2000, 2009) (a book Humphrey has studied—personal communication) I wasn't really shocked. But rather than providing support for Humphrey's brand of materialist reductionism, my own conclusion was that this challenges it in an even deeper way, for the simple reason that experienced value, beauty, meaning, purpose and the other joys of consciousness have no obvious place in reductionist third-person science.

The apparent force of Humphrey's argument that a first-person sense of value enhances third-person observable survival and reproduction by *motivating* it is just that—apparent! It is, for example, no different in form to the arguments for any of the other third-person functions that have been suggested for consciousness that Humphrey now rejects. How for example, could one *think* unless one were conscious, how could one *monitor one's environment*, how could one *respond appropriately to complex situations*, and so on? Aren't conscious thinking, monitoring, and appropriate responding equally useful to survival? Unfortunately for all such proposals they fail the tests mentioned at the beginning of this review, and the same applies to conscious motivations. *Viewed from a third-person perspective*, the physical correlates of conscious motivations, just like those of conscious thoughts, can be fully described in physical terms, for example as goal structures in information processing systems, as attractors in complex system landscapes, as activation patterns in limbic system structures and so on. Provided that they are accurately described such physical correlates of first-person motivations already account for third-person functioning, leaving the accompanying conscious experience nothing (in third-person functioning) to do—the problem of “overdetermination” referred to above.

Given these, and many related problems, I concluded that the obvious first-person importance of consciousness combined with its seeming irrelevance to third-person causal explanations presents a “causal paradox”, and that an intuitively satisfying theory of consciousness needs to demonstrate how *both* first-person intuitions about the importance of consciousness *and* the third-person findings of science might be true. For example, by demonstrating how first- and third-person observable aspects of mental life could be *complementary and mutually irreducible*. Much as I would like to present a detailed analysis of this alternative here, it would not be appropriate for this review.<sup>3</sup>

In his apparent celebration of conscious experiences does Humphrey also intend to restore their intuitive importance? Not really. Because enchanting conscious experiences don't fit into his reductionist understanding of the world, he argues that they *must* be “impossible objects”; they can only be illusions created by the brain for the purpose of motivating survival and the reproduction of genes. And, he writes, “...the day may come—even come soon—when science will have revealed the illusion of consciousness for what it is, and any rational person will have no choice but to accept the game is up” (p196). The intention of his book is to hasten the arrival of that day.

---

<sup>3</sup> Interested readers can find such an analysis in Velmans (2000, 2009), or, in a more limited version that focuses specifically on causality, in Velmans (2002) along with 8 commentaries and a reply in the JCS Special Issue on “How could conscious experiences affect brains?”

### Would you choose reproductive fitness or consciousness?

At an International Symposium on “Where does consciousness come from?” at the *Festival della Scienza*, in Genoa (2010) at which Humphrey and I were both speakers. I presented a thought experiment to a very large audience that could be regarded as a direct test of Humphrey’s convictions. Here it is:

“Imagine that you are twenty-one, in full health but you have no children. Tragically, you catch a fatal illness and have just a few days to live. However the doctors know of two drugs that can save your life, *nocon* and *nokid*. Unfortunately each drug has serious side-effects. If you take *nocon* your life would be saved and your biological and behavioural functioning would be entirely normal, including your ability to have many children. However, you forever, irreversibly lose consciousness. If you take *nokid* your life would be saved and your conscious experience would be entirely normal. Your biological and behavioural functioning would also be normal, with one exception. You forever, irreversibly lose the ability to have children (by natural or any artificial means). Which drug would you choose?” (from Velmans, 2009, p348)

What would you do? In Genoa, I counted five people that chose reproductive fitness. Roughly 1,500 chose consciousness. As I go on to explain in Velmans (2009)

“What makes this little thought experiment interesting is that it directly pits the ability to reproduce (which is absolutely fundamental to evolutionary theory) against the ability to experience. If consciousness is just a means to enhance our reproductive fitness, we should opt to retain this fitness and choose *nocon*. I have tried this thought experiment with many students and they overwhelmingly choose to take *nokid*. Why? Because without the ability to experience anything, life would have no point.” (Ibid)

As I go on to admit, “Accounts of human life or survival in terms of whether it *has a point* fit ill with current, mechanistic accounts of nature. But, I repeat that such mechanistic accounts of how nature appears viewed “from the outside,” simply do not address *what it is like to be* a bit of that nature “from the inside.” We *know* what it is like to be conscious. The delight in being able to experience ourselves and the world in which we live in an indefinitely large number of ways, or the sorrow of losing one’s vision or one’s hearing are *subjectively real*. This reality is not diminished by our inability to explain it in entirely, third-person, inclusive-fitness terms. Our own first-person nature is as much part of the natural world as the functioning of our bodies, and, in the long run, our theories of mind need to accommodate *all* the data. If, after our best efforts, we cannot squeeze what are, in their essence, first-person phenomena into a third-person “box,” so be it. The alternative is to broaden our theories of mind to encompass first-person phenomena.” (Ibid)

Given his ultimate dismissal of all that makes life worth living would such contrary evidence (or any other contrary evidence) shake Humphrey’s beliefs? I doubt it—because once one takes the reduction of first person experience to reproductive fitness to be *axiomatic* one can *always* construct a *post hoc* story to explain contrary evidence away. But the effect of this is to weaken the theory. As Popper noted, a theory that can be stretched to explain



everything, explains nothing, for the simple reason that it *excludes* nothing, making it unfalsifiable and no longer science. Although Humphrey does not consider non-reductive alternatives, the alternative to a reductionist science of consciousness is not nonsense or even non-science, but simply a non-reductionist science of consciousness.

Given the many problems I have mentioned above, I think that only those of a similar reductionist faith will find the arguments in *Soul Dust* in any way persuasive. But his book is aptly named. Humphrey takes the wonders of consciousness, the enchantment of the world, the meaning of life, and the mysteries of the soul, and turns them into dust.

## References

- Armstrong, D.M. (1968) *A Materialist Theory of Mind*, London: Routledge and Kegan Paul.
- Crook, J. H. (1980) *The Evolution of Human Consciousness*. Oxford: Oxford University Press.
- Hebb, D. (1949) *The Organisation of Behavior*. John Wiley and sons
- Humphrey, N. (1982) *Consciousness Regained*. Oxford: Oxford University Press.
- Humphrey, N. (2002) The uses of consciousness. In N. Humphrey (ed.) *The Mind Made Flesh: Consciousness and the Physical World*. Oxford: Oxford University Press.
- Miller, G. (1962) *The science of mental life*. London: Penguin Books
- Romanes, G. J. (1896[1885]) *Mind and Motion (Rede Lecture)*, Reprinted in G. J. Romanes *Mind and Motion and Monism*. London: Longmans, Green & Co.
- Velmans, M. (1991) Is human information processing conscious? *Behavioral and Brain Sciences* 14(4): 651-669.
- Velmans, M (2000) *Understanding Consciousness*. London: Routledge/Psychology Press.
- Velmans, M. (2002) How could conscious experiences affect brains? *Journal of Consciousness Studies*, 9(11), 3-29.
- Velmans, M (2009) *Understanding Consciousness, Edition 2*. London: Routledge/Psychology Press.