Review of *How Can Physics Underlie the Mind: Top-Down Causation in the Human Context* by G.F.R. Ellis

Stephen N. Lyle

February 25, 2018

This book is part of an ongoing rearguard action by many humans against a huge army of straw human beings who believe—and I mean believe, not just hypothesise—that nothing could be relevant to our lives but the doings of elementary particles and associated fields, or strings, or something else that mathematicians have come up with recently, and worse still, that everything humans do is entirely preordained by whatever laws govern those items, whence there is no point trying to do anything with our lives. Ellis' enemies, like the enemies of so many today, would be people who advocate reductionism and determinism. These enemies, if they existed, would be believers of a kind, fundamentalist atheists, deniers of the soul. Perhaps some of them are indeed real flesh and blood, but we need not be concerned with them, for they would clearly be naive and superficial beings, certainly not thinkers at the forefront of the study of human nature today like S. Blackmore [1, 2], D. Dennett [3–6], and S. Pinker [7, 8].

What might give weight to the present book is that the author is himself one of those who propounds the kind of arcane mathematical constructions so dear to the field of fundamental theoretical physics these days. So who better than one of the *cognoscenti* to inform us of their limitations? Looking on the positive side, this is indeed a vast and erudite collection of examples of what Ellis calls top-down causation. There are so many examples, all intended to prove the same thing, that the book has an air of propaganda about it. So if you go along with whatever it is that the author is trying to prove, you may well be delighted with the whole enterprise. But if you feel at the outset that it rings false, you may find that it gets irritating.

The question is whether the things that physics deals with, taken alone, are causally determining for everything we encounter in our world. As the author points out, although later denying it, part of this question remains open, since we do not know whether our theories of microscopic physics are already, or will eventually be, deterministic. Most would say that there is already no hope for them to be deterministic, because they think that quantum theory is intrinsically indeterministic. Ellis is one of these people, although he does briefly consider a deterministic alternative to quantum theory, the de Broglie–Bohm

theory [10, 11], which makes the same predictions as standard quantum theory (whatever that may be). But putting aside the question of whether our present physical theories are already, or will one day become, deterministic, we can still ask whether, insofar as events in the physical world can be said to be caused at all, those causes can be taken to be solely physical ones.

Well, what else could they be? Actually, Ellis views the world as a hierarchy with the subjects of physics at the bottom, then the subjects of chemistry one step up, then the subjects of biology, then the subjects of psychology, and perhaps finally, but that is a big 'finally', the subjects of sociology. So we have strings, branes, quarks, leptons, and all the rest, then atoms and molecules, then big molecules organised into cells and the like, then cells organised into living systems, then the minds of animals like ourselves, then social groups of human beings with their cultures. And the question is: how did all this come about? Ellis' answer is apparently that, for certain things to come about on one level, they need help from a higher level, and that this higher level may not be one that could spring from below.

Let us get to the heart of the matter (Sect. 1.3.4.):

> [...] there are also some effective higher level variables that are not coarse-grained or otherwise emergent from lower level variables: they represent irreducible high level properties and relations. The top-down influence of such variables is a key aspect of the argument of this book. Those variables include mental features like emotions and feelings, abstract entities like theories and plans, and socially determined effects like the value of money.

Human minds were absolutely necessary for aircraft to come about on Earth, or so the argument goes. And who could dispute that? We know that all aircraft we see were made by humans according to a human plan. We also know that the chances of ordinary matter spontaneously assembling itself from the ground into a viable aircraft are negligible. Take away the human mind, and you take away the aircraft. So the mind is a cause and as such must get into our ontology of what there is out there (Sect. 1.3.4):

> [These] higher level features are demonstrably causally efficacious, but cannot be regarded either as coarse-grainings of lower level variables, or as inevitable outcomes of blindly working lower level forces. They are essentially higher level variables.

The mind is an irreducible thing, and a real thing now, in our ontology (Sect. 1.3.5):

> We must recognise the existence of any kind of entity that demonstrably has a causal influence on physical systems.

But when the universe began, there were no minds. So how did they get there? This is the big issue and the reader will look hard in this book to find an answer. Put another way, if some things need a higher level in the hierarchy to 'emerge', and if the higher level itself needs an even higher level, where will it all end?

This brings to mind Dennett's contrast between cranes and skyhooks (see [3, pp. 226 ff] and the many references to this in [4]). A skyhook is of course an imaginary thing. We build things from the bottom, using cranes. And what is so appealing about science is precisely that it so often proceeds from the bottom. There are so many examples that one could write a whole book about that, at great risk of its seeming like propaganda. In fact, it is surely the greatest achievement of the whole scientific endeavour that we have been able to *explain* so much without ever invoking magic, without ever insisting that there be some mystery component that must necessarily remain intrinsically inexplicable. Actually, science proceeds by assuming quite the opposite and seeing what it can do. And every time, it moves forward, encompassing more and more phenomena. But to do this it also has to burn through barriers with the universal acid of Darwin's theory, in the form of the Darwinian algorithm [4, pp. 50, 63] (see below).

In Sect. 4.7.6, referring then to all the elements that make up an aircraft, Ellis asserts that there is no way they can spontaneously assemble to make an aircraft by any bottom-up process. But that is exactly what must have happened, since at the time of the last scattering surface (LSS) in the early universe, which Ellis so often mentions (see below), there was nothing remotely resembling the higher levels that Ellis claims were necessary to have complexity emerge. It's just that this assembly spontaneously happened through a long and contorted path, of very great complexity, going via human beings and their plans, too complex for a detailed physical account to be of any use to mere mortals. And our main research programme in science today is founded on the fruitful idea that it happened in part because conditions for the Darwinian algorithm to get going arose from the LSS. In this sense then, an aircraft and the plans for it are just part of our extended phenotype, a logically straightforward, although physically ultra-complex consequence of evolution [9].

There is undoubtedly a sense in which the *chances* of ordinary matter spontaneously assembling itself from the ground into a viable aircraft *are* negligible if one just takes some ordinary matter somewhere and waits for something to happen. Actually, in a deterministic world, where there are no chances for anything except what *must* happen, one could do this repeatedly in various places, then sit and wait. In an indeterministic world one could do it once and still talk about chances. But whichever case, the chances of *any* particular configuration are minuscule in a complex world like ours, zero for all practical purposes. And the fact that a world containing an aircraft strikes *us* as a particularly significant configuration is irrelevant here. Even if the world is entirely deterministic, as it may well be, tracing back the chain of causes to the LSS, for example, does not imply that we can read meaning into the LSS, because meaning is something that evolved *with us* (and presumably some other creatures). And nor does it mean that the glorious future of the LSS, at least insofar as we would process what we observe today, had to be designed by a superbeing. The Darwinian algorithm constitutes a very powerful machine.

Since we were just talking about *explanations*, there is a related question that it is useful to ask right away: how could one ever explain the human

mind by talking about quarks and leptons? Even supposing one preferred the hypothesis that chemistry can be explained by physics, biology by chemistry, and so on, such an explanation of the human mind would be long-winded to say the least, hardly practical, and certainly not enlightening in the detail. At best it would have to remain an *explanation in principle*. Presumably few people would disagree with that. But by reference to practical explanations, can we conclude about causality? It still seems feasible that the quarks, leptons, and friends could be jollying away making minds appear in the universe, without the details of that generation constituting what those of us with minds would describe as a useful, that is to say, meaningful explanation. This is a distinction that Ellis does not clearly examine.

Instead he defines cause in such a way that his top-down causes are causes, and defines his ontology, his list of things that must actually be out there, in such a way as to include those causes as real things out there. These definitions, however, should not hide the fact that he is primarily justifying such moves by the need for pragmatic explanations, a need that only straw beings might deny. The question as to whether we really require these 'things' to be causes remains as open as ever, however many examples Ellis may come up with to illustrate this undeniable need.

Indeed, there is much evidence that we do not require things like the human mind to be ultimate or irreducible causes, because we are already beginning to see how minds can evolve from things without minds. And this is good news when we consider the stark fact that, once upon a time, there were no minds, only the kinds of elementary particle soups that cosmologists like Ellis deal in. But to reach this understanding, we have to grasp the power of the Darwinian algorithm. Indeed, the power of Darwin's idea, suitably formulated [4, pp. 50, 63], is precisely that it is an algorithm:

- It is substrate neutral, i.e., it works whenever and wherever certain things are available to set it running.
- It is mindless, i.e., each step and each transition between steps could be carried out by a machine.
- Results are guaranteed.

The mindlessness is the key here (the second point). But we do need certain things to be in place (the first point), and that is, for example, the only remaining (big) question about how come there is life on Earth. There are still lots and lots of details to be filled in, so many we are unlikely to have time to grasp them all, but at least this is a *genuine* research programme.

Compare this with what is proposed in Sect. 8.7.2, entitled *To Be Done*. Ellis makes the following statement:

> Bottom-up emergence cannot by itself lead to the existence of the relevant higher level variables [in the case of biological macromolecules, digital computer programs, thoughts and plans, social neuroscience]

> (for example, physics *per se* cannot lead to the existence of mathematical theorems, or indeed to theories of physics). This emergence can only take place by developmental and learning processes enabled by top-down causation from abstract possibility spaces to the human mind.

Then, under the heading of experimental tests, Ellis says this:

> These considerations are in my opinion conclusive, but are based on understanding and explaining what one already knows. What one likes in an experimental science is a prediction of something new that can then be verified by experiment or observation. What new experiments or observations can we propose that will substantiate or disprove the causal efficacy of higher level variables?

Compare this programme with the one that science has been so successfully implementing for so long now, which looks to parts of things and mechanisms relating them to explain how those things appear to us the way they do. His idea is to look around for 'things', i.e., in one's ontology, that cannot be better understood by examining them in this way. Take, for example, mathematical theorems and theories of physics.

In fact, for Ellis, nothing leads to *the existence* of mathematical theorems and theories of physics: they are irreducible things that must always have been there. In order to disagree with that, we must ask what their explanation would be in a scenario that attempted to get everything, at least from some reasonably well established starting point, by bottom-up explanation. Here is an idea. Let us make a distinction between our theories of physics and the way things work out there. Then it looks clear that the way things work out there is not necessarily ever going to be explained just by our discovery of physical theories that allow us to make good predictions about them. After all, a physical theory is probably a very human thing, imbued with all that is human about our explanations, that is, imbued with the way we have evolved to understand things through our ancestors' interactions with those things or with similar things. So take away all humans and there is a good chance that we will take away all physical theories.

But we will be left with the way things work out there. So are we claiming to explain that, or to find a cause for that? It would indeed appear to be the aim of a 'theory of everything'. But then we should ask how our theory of everything could explain its own existence? Could it be that we are outstretching ourselves? Becoming too greedy once again, like the pre-Copernicans? Some things should perhaps be viewed as being forever out of our reach, although one may of course use an ideal to guide a research programme. But at least we can glimpse how physics, i.e., physical processes at some lower level, not necessarily an absolutely bottom level, might lead to humans that can make good theories of physics that can make good predictions about what will happen in various situations. So there is a point of contact with Ellis here: we shouldn't expect to explain why the laws of physics are the way they are, at least not *all* of them, and not why they actually exist. But we may still reject the idea that human

minds would never be brought about by bottom-up causation from a primordial soup, precisely because we make a distinction between causation and practical explanation.

What then of mathematics? In the penultimate quote, Ellis claims that physics *per se* cannot lead to the existence of mathematical theorems. But insofar as mathematical theorems are just what gets instantiated in human brains, or in books that only humans are ever going to use to form such instantiations, we can just as well glimpse in principle how physical processes at some lower level might lead to humans developing neural circuitry that achieves this. Why does there have to be some abstract thing out there? What could it be, after all? Indeed, how could it be abstract and out there? Aren't abstract things just in human minds? Ellis thinks he has the answer when he uses the term 'abstract possibility spaces'. But what are these made of? Where are they? And above all, do we really need them? Once again, we may well suppose that something about the way things are organised out there corresponds pretty well to these representations in our brains, otherwise they would not prove so useful. But why do we need to impute more existence to them than that? Isn't it better to focus on what we can know of them, that is, the representations we make of them?

Hopefully, put like this, we can see the problem with Ellis' dictum (Sect. 8.7.2):

> **Existence of Top-Down Causation.** The basic point is that one demonstrates existence of top-down causation whenever manipulating a higher level variable can be shown to alter lower level variables.

But the idea of manipulating something like Pythagoras' theorem is problematic. Talking about non-Euclidean geometry doesn't help here, even though our representations of the world come out differently when imagined in a pseudo-Riemannian manifold. This is not like altering the temperature setting on a thermostat. If general relativity is right, Pythagoras' theorem will never hold more than approximately anywhere. It is just an idea in human heads and books, and will disappear with us (or all other processes that resemble us sufficiently well to conceive things like that), except insofar as it is an approximation to the way things are organised out there. And it's not because things out there are what we would describe as organised that there has to be an organiser, or indeed any kind of explanation that could come within our grasp, or even a cause for such organisation. Cause and effect are also just the human route to what humans consider to be explanations, although something like cause and effect presumably occurs out there.

Indeed, we may consider that our brains have evolved to understand events on our length and time scales in this *deterministic* way, and this would appear to be part of the genetic recipe for brain development in young humans, precisely because there is something about the way things work out there that makes this a useful predictor. Whether or not the world is somehow intrinsically indeterministic, whatever that may mean, on microscopic scales, as some would say the quantum theory suggests, is neither here nor there on our length

and time scales, where things fit well with the deterministic picture. This is relevant to the notion of free will, discussed below, where I shall claim that it is best viewed as an essentially deterministic cause-and-effect representation of what happens to 'us', that is to say, the 'self', and that the self is also just a convenient representation of the complex physical process of a human being, these representations constituting effective short cuts to understanding the role that such complex processes play in their interactions with their environment. This picture stresses, not just the usefulness of such short cuts, but the fact that they are an integral, evolved part of being human (as expressed so clearly in [15]).

Basically then, in a more modest approach, and indeed a more economical approach, we impute existence to what is out there because we are realists, but we need not claim more understanding of what is out there than our very human representations of what is going on. And we do not need abstract possibility spaces to exist independently. Their rightful place is in our own minds, product of some neural correlate.

Ellis' Sect. 8.7.2, the one entitled *To Be Done*, is a very short section indeed when compared with the tremendous research programmes launched by an intellectual investment in bottom-up causation. And there remains the vexatious question of where it all stops when we work our way up the hierarchy. Ellis' answer here is also disappointing. We don't even find a god—intelligent design is flatly rejected in Sect. 8.1.4. This section is interesting in itself. The discussion centers on a thing called the last scattering surface (LSS) which cosmologists hypothesize, occurring early on in the story of our universe, some 300 000 years after the initial singularity (in our current model), when matter and electromagnetic radiation are surmised to have decoupled. In view of a *reductio ad absurdum*, our guide asks us to assume that everything in the history of the Earth is indeed written into the fluctuations on the LSS:

> We then have to explain how the theory of general relativity, the Mona Lisa, the international banking system, and so on and so forth, could all have been encoded there in the initial data that determined the later universe.

Of course, we must be careful here not to confuse practical explanation, obviously impossible in this case, and cause. For we may still have bottom-up causes leading all the way from the LSS to the Mona Lisa, but whose details in no way provide a practical explanation.

But here is how Ellis deals with this conundrum. He begins Sect. 8.1.4 by saying "forget quantum theory". This is because, even before the further considerations in this section, he adopts the view that the last scattering surface was not a specific thing out there, but one of those vague and indeterminate things that belong to the ontology of most quantum theorists. However, although the 'standard' view of quantum theory might introduce a vague notion of indeterminacy in the actual state of the LSS, we are not forced to accept this fuzzy ontology in quantum theory [10, 11], and even if we did, it wouldn't

change the situation, since nothing is required of the LSS in terms of meaning, thanks to the power of the Darwinian algorithm.

But having disregarded the effects of quantum vagaries, he considers that there remain two options for examining this paradox. One is that the Mona Lisa is generated deterministically from random Gaussian fluctuations on the LSS, the second that the fluctuations on the LSS may not be random Gaussian fluctuations, despite the fact that this is the standard view today. In the first situation, he invokes diffusion processes and random collisions between particles to explain why we have difficulty even in showing how the LSS will result in the present observed details of the distribution of dark matter halos. But just because something is difficult means nothing on the purely logical level. It just means that such an explanation of the Mona Lisa would be impractical, but it says nothing about causes, or even partial, partly indeterministic causes. And yet to impress us with the impossibility of his first option, he states that:

> The idea that there is some way that the unique English language on this page should be embodied in [a transfer function] acting on random Gaussian fluctuations is not remotely plausible, for there is no hint of how this could happen in standard discussions of the transfer function.

But there certainly is a hint of how this could happen, even though many, many details may remain: it's the standard research programme of modern science, helped along by the much feared Darwinian algorithm.

What about the second option, the idea that the fluctuations on the LSS are not random Gaussian fluctuations? Ellis interprets this as meaning that:

> [...] these fluctuations must explicitly encode the formulae of the theory of general relativity as written in Einstein's 1915 paper, the words in Darwin's majestic opus on the origin of species, and so on. The issue then is, who or what could have written all this, rather than random Gaussian fluctuations, into the molecules on the LSS precisely so as to get these results? To determine every thought that Maxwell, Einstein, Karl Marx had?

The answer for Ellis is that only a demiurge could have been responsible. This would be intelligent design, something he thinks he is avoiding by the arguments in his book. According to him then:

> [...] genuine emergence took place, leading to intelligent outcomes not uniquely determined by this initial data.

All is confusion here.

The problem with Ellis' reasoning is that he seems to think that, because something meaningful evolves out of the LSS, that meaning already had to be there, even though there was no one for it to mean anything to. But this just doesn't follow. It merely shows that meaning evolves with processes capable of appreciating it, and exploiting it. This is the subject of Dennett's book *Freedom*

Evolves [5]. The hypothesis that all can evolve bottom-up from the LSS doesn't require the LSS to be interpreted as meaning anything. It was surely random as far as we are concerned. It may or may not be that other configurations of the LSS (none of which actually occurred) would have led to a world with human beings, but even that is irrelevant. The point is that all the design that occurred from that point on was mindless. That's what the Darwinian algorithm does for us. No matter how designed things look today, it doesn't mean that there was ever a designer, no more at the time of the LSS than anywhere along the way.

So what of the top levels in Ellis' hierarchical world? We have something about this in Sect. 7.4.1 on goals and purpose, in the chapter on the mind and the brain. Naturally, the mind, the way we understand our cerebral activity, or the way it seems to us, is for us the 'highest' level in the hierarchy of all things that concern us. It certainly looks like the most complex level we have to deal with:

> Given that there is top-down causation, what is the top level? It is not clear that there is a fixed top level in physical terms: there may rather be transient oscillatory bindings of neuronal groups that act as the top for a while. However, in psychological terms there is indeed a top level: it is the level of meaning or *telos*, associated with ethics and aesthetics.

So where does it all end when we go up? Ellis says this:

> Obviously this process faces a problem of infinite recursion: where does the next higher level of selection criteria come from? At some point we have to draw a line and say, this is where I stand: these are my founding principles, this is the purpose in my life. That is where one makes value choices based on one's view of meaning. This may be to help others, to serve God, to save the Earth, to create great art, to maximise one's own welfare, to understand how things function, or just to survive.

This is the only unquestioned appearance of a god in the book, but one is surely enough. Luckily for those of us who would require evidence of such a thing, or even just the utility of such an idea, Ellis allows other ultimate goals like creating great art and saving the Earth.

But once again, we have that nagging feeling that something is wrong here. For what of these goals without minds? They disappear without trace just when our bodies disappear, and gods with them. So if we have to look up in the hierarchy, we find that these highest causes require something made of flesh and blood, hence cells, hence molecules, hence atoms.

We can see what is going on here: all this talk of abstract possibility spaces somehow existing outside our minds, and ultimate goals which seem at first glance to be thoroughly disembodied, or god-given, or part of some god, are just a way to keep human culture, knowledge, and achievement out of the reach

of mere matter. Naturally, the notion of cause and effect is a key item in what we call explanation, but we can have chains of cause and effect that are too long and complex to provide what we consider to be a useful explanation, and usefulness stems also from what meaning has evolved along the way. So no one will deny the relevance of abstract understandings of the world or ultimate goals to the way we live our lives, and we can understand them as causal and existing really, if not materially, if we so wish, if that proves useful (as Hofstadter does in [15, Chap. 3], without falling into the same trap as Ellis here). But there is in fact no difficulty in imagining them as evolving out of mere matter through merely physical processes. Then at least we avoid the difficulty of having to postulate their existence before they could even be appreciated by ourselves, and along with that, we gain a genuine research programme, the one that science as a whole is currently working on.

Nowhere does Ellis mention Karl Popper, who argues at length in favour of an objective mind [12, Chap. 4], and postulates the existence of three worlds in a way that is in some ways reminiscent of Ellis' thesis:

> [...] the world consists of at least three ontologically distinct sub-worlds; or, as I shall say, there are three worlds: the first is the physical world or the world of physical states; the second is the mental world or the world of mental states; and the third is the world of intelligibles, or of *ideas in the objective sense*; it is the world of possible objects of thought: the world of theories in themselves, and their logical relations; of arguments in themselves; and of problem situations in themselves.

It would be interesting to know how Ellis would react to Popper's analysis. But Popper's mind–body dualism and his view that mental properties of humans are distinct from physical ones is beginning to creak somewhat. The more we look non-invasively into the human brain when it seems to us to be doing this or that, the more we find neural correlates for everything. Not surprisingly, these are complex and still beyond our comprehension, but we are making progress. And what is more, we are beginning to see that our concepts of the self, consciousness, and free will can be well understood as simplified but remarkably predictive representations of the physical process constituted by each human being [15], representations that would evolve in certain circumstances that could themselves have evolved by purely bottom-up processes.

Here it seems is a much better idea than trying to postulate 'higher levels' that must always have been there if Ellis is right in claiming that they could not have arisen from below. For that only raises more questions about how those 'higher levels' got there. It is tantamount to gods, an *ad hoc* thing that looks suspiciously like a skyhook. What we would like to do, and there is still no reason to think it impossible, despite what Ellis suggests, is to see how the inner, material workings of the brain are able to generate the impressions we make of our self, use them to represent the part our material bodies play in the causal chain of events in their neighbourhood, and provide us with what we consider to be our awareness of what is happening to our bodies, and in

particular what is happening in the brain itself (see [15], which provides a much more adventurous and creative account of the human mind than Ellis). Isn't this a rich and motivating scientific research programme? It is certainly much more challenging than just postulating the existence of a hierarchy whose levels act causally downwards in an *ad hoc* way, by decree as it were, in order to get things to come into being.

Indeed, to say that the abstract notion of a computer program, for instance, must exist on a par with the material stuff around us, and in particular have causal efficacy in the same way as that stuff, is like saying that, if *we humans* didn't have an abstract idea about hydrogen atoms and helium atoms, kept away somewhere in an abstract space that we may be the only ones to know about, then nuclear fusion couldn't take place in stars. And computer algorithms are not the only kind of algorithm Ellis would say really exists in some kind of abstract space. Presumably, the Darwinian algorithm really exists there too, and presumably, without its existing there and being causally operative in some unspecified way from its abstract home down to replicating material items here in the material world, evolution could not have taken place. This provides an odd picture indeed, and yet it seems to be a natural extension of the above idea about computer algorithms. Of course, different people prefer different pictures, but what is the utility of this expensive addition to our ontology?

At the end of Sect. 2.7.1, Ellis continues this discussion as follows:

> Here I use as a criterion that whatever is causally effective in the physical world must certainly exist. If this is not true, we will have to face existence of uncaused entities or events in the physical universe.

But this is just wrong. All the 'entities or events' here, viz., changes that occur within a computer, for instance, are caused by particles and fields. If the same computer spontaneously appeared and something spontaneously pressed the right buttons, the same entities and events would occur and we would have difficulty explaining that the program in abstract space was necessary. Ockham thus removes this. And it is a better picture to just say that this is precisely what has happened, i.e., computers did spontaneously appear, etc. But take away the computers and the humans to press the buttons and the program ceases to exist.

It is clear enough why Ellis has attempted this world view: it is his abhorrence of reductionism and determinism. But what threats do these ideas really pose? And why do they stir up so much trouble for writers like E.O. Wilson and R. Dawkins? As Pinker points out [7, Chap. 6], one reason is that quite absurd beliefs are attributed to anyone who suggests a proper scientific study of human nature. Ellis himself does this in his Sect. 7.7.3, entitled *Crick's Fallacy*. It is worth making two short quotes from the end of Sect. 5.6:

> It seems that the profound nature of the quantum revolution has still not permeated the consciousness of many physicists and biologists, who present the situation as if physics were deterministic all the way down. This is not the case: the bottom level is not deterministic.

There is something slightly disrespectful about this statement. Worse, indeterminism is not yet established for quantum theory. Bohmian mechanics provides a deterministic alternative. On the other hand, this is irrelevant anyway for the reasons discussed earlier. He goes on to say:

> At its base level, the universe is indeterministic, allowing the needed causal slack freeing higher levels from lower level causal determinism. And this quantum indeterminism can affect biological processes. [...] If those levels cannot be reduced to lower levels (interactions of electrons and quarks for example), as assumed by Crick in his book *The Astonishing Hypothesis*, then the principle of irreducible intermediate levels is established. If this was not the case we'd be in deep trouble as we don't in fact know what the bottom level is. We'd have to suspend neuroscience and genetics research until the theory of quantum gravity has been sorted out. In reality, no particular level is privileged in causal terms: they all have causal power.

First of all, Crick may never have claimed that all biological *explanation* could be reduced to lower levels, but like many scientists, he would very likely assume that all *cause* could be reduced to lower levels, even if the full causal chain could not provide a useful practical explanation of biological processes, and this because it is such a fruitful research programme, in contrast to the one proposed by Ellis. Furthermore, the assertion that we must stop doing neuroscience until we have sorted out the theory of quantum gravity is just silly. This is a research programme, not just a philosopher's game. Experiments were done on protons and neutrons before they were done on quarks. Why wait? How could we wait? We would never have found the quark model otherwise.

Returning to Sect. 7.7.3, entitled *Crick's Fallacy*, Ellis asks:

> On what grounds does he treat the particle physicists so badly, denying the full reduction to their level of description? The answer is he believes that nerve cells and their associated molecules have real causal power, because that is the level he deals with and understands.

And shortly afterwards:

> The threat to your sense of personal identity and free will posed by Crick's quote is undermined.

Some of us will just not believe that Ellis has really saved us from any terrible threat.

We may say that Crick was stating the great achievement of human scientific endeavour, which is to finally understand in principle how our existence is possible, purely on the basis of what we know as the laws of physics, even if those laws must later evolve with improvements in our scientific knowledge. Of course, he would never for a moment have claimed that effective explanation

had to be in terms of quarks, for example. In fact, nobody is suggesting that we use string theory or quantum chromodynamics to explain why squirrels bury nuts in the autumn. Because there is a difference in claiming that life evolved from inanimate matter, in a way that we are striving to understand, and then evolved through various different species until squirrels came about. And that is understood as very much a bottom-up causal process from times when there was no life, through times when there were no squirrels, until times when there were. The suggestion is certainly not that it is at all practical to explain things like that in terms of individual strings in the early universe.

Instead of the greedy reductionism always criticised by anti-reductionists, and which nobody would ever try to defend, Pinker advocates [7, Chap. 3] the notion of *consilience* as put forward by Tooby and Cosmides [13], then Wilson [14]. It would be nice to quote the whole of the first two pages of [7, Chap. 3], which puts things so tidily. Instead, here is a less elegant summary:

- Newton demolished the long-standing wall between the chaotic sublunary world and the apparently pristine cosmos by postulating a single set of laws that governed the motions of all objects in the universe.
- Lyell then demolished the wall between the creative past and the static present by showing that the Earth was sculpted in the past by geological forces we can still witness today in the form of earthquakes and erosion.
- A whole line of scientific minds then demolished the wall between the living and the non-living by showing that the human body is a machine that runs by hydraulics and other mechanical principles, made not of some wonder tissue [3, Chap. 22], but ordinary compounds obeying the laws of chemistry. And Darwin showed how the diversity of life with its clear signs of design could arise from the physical process of natural selection among replicators, while Watson and Crick showed how replication itself could be understood in physical terms.

Here is the in-principle unification of our understanding of life with our understanding of matter and energy. But it leaves one wall intact in the landscape of knowledge:

- The wall dividing matter from mind, the material from the spiritual, the physical from the mental, biology from culture, nature from society, and the sciences from the social sicences, humanities, and arts.

With new ideas from four frontiers of knowledge, viz., the sciences of mind, brain, genes, and evolution, this wall is now crumbling too. No one should have any doubt about this (see [7] and references therein).

But there has a been a widespread negative reaction to any attempt to build a science of human nature. In his discussion, Pinker puts it like this [7, p. 139]:

> The anxiety about human nature can be boiled down to four fears:
>
> - If people are innately different [if they are not blank slates at birth], oppression and discrimination would be justified.

- If people are innately immoral [who knows what we may find?], hopes to improve the human condition would be futile.
- If people are products of biology [in what looks like it may be a largely deterministic way, at least on the relevant length and time scales], free will would be a myth and we could no longer hold people responsible for their actions.
- If people are products of biology, life would have no higher meaning and purpose.

For those who go along wholeheartedly with Ellis' world view but largely in a spirit of fear about these issues, I would heartily recommend that they read *The Blank Slate* by S. Pinker [7] to put their minds at rest, in fact to see why Ellis is wrong to advocate this new form of obscurantism. The present article is just a review, but here at least is a summary of the arguments Pinker puts forward [7, Chaps. 8–11]. The idea then is that we should fare better if we informed ourselves about the way we humans really are, and why we are like that, using as always the scientific approach of conjecture and refutation, rather than the verificationist, even propagandist, model adopted by Ellis.

Regarding the fear of inequality [7, Chap. 8], suffice it perhaps to cite Hume's guillotine: no matter how convincingly one shows that something is true, it never follows logically that it *ought* to be true. Evolutionary success and goodness are not the same thing. And from a pragmatic standpoint, given that there is always a tradeoff between freedom and equality, the more we know about people's innate advantages and disadvantages, the better we can establish this tradeoff. Most importantly, to see what 'better' means here, consider some of the options:

- Social Darwinism places no value on equality.
- The totalitarian left places no value on freedom.
- The Rawlsian left sacrifices some freedom for equality.
- The libertarian right sacrifices some equality for freedom.

Just considering these options should make us realise that we need not be slaves to a biological understanding of any innate differences between human individuals which may turn out to exist. Our faculty of reason allows us to go beyond this. And we do not need to pull the faculty of reason out of a hat as Ellis would suggest. It is another aspect of our brain that has evolved.

Regarding the fear of imperfectibility [7, Chap. 9]. Three features of human nature can lead to a steady expansion of the circle of human cooperators:

- We have the cognitive wherewithal to figure out how the world works, and this yields a knowhow worth sharing.
- We have language, and this allows the sharing of technology in its broadest sense.

- We have evolved with an emotional repertoire—sympathy, trust, guilt, anger, self-esteem—that impels us to seek new cooperators, maintain relationships, and safeguard them against exploitation.

So as Pinker puts it [7, p. 166]:

> Not only is acknowledging human nature compatible with social and moral progress, but it can help explain the obvious progress that has taken place over millennia.

And so we observe an expanding circle of entities considered worthy of moral consideration and it does not have to be powered by some mysterious drive toward goodness [7, p. 167]. It may come from the interaction between the selfish process of evolution and a law of complex systems. Evolution can indeed lead to greater and greater degrees of cooperation. Replicators can team up, specialise to divide labour, and coordinate their behaviour, because replicators often find themselves in non-zero sum games [7, p. 166]:

> Independent agents repeatedly made their fate hostage to a larger system, not because they are inherently civic-minded, but because they benefited from the division of labour and developed ways of damping conflicts among the agents making up the system.

Put another way [7, pp. 166–7]:

> Agents do better when they team up and specialise in pursuit of their shared interests, as long as they solve the problems of exchanging information and punishing cheats.

In short, morality can evolve from a situation where morality does not exist, is not even relevant, and in a perfectly amoral way. The power of the Darwinian algorithm once again. The evolution of the expanding circle, its ultimate cause, is pragmatic and amoral, but the psychology of the expanding circle, in the sense of its proximate cause, need not be.

So what of the fear of determinism [7, Chap. 10]? The first problem here is that we have a distinct feeling of being in control when we decide to do things. Somehow the cause comes from the self, an entity that remains vague. On the other hand, if the world is not deterministic, so that at least some events are at least partially uncaused, that is not going to help, because free will is all about cause. And fortunately the world can be considered entirely caused on our length and time scales. So whichever world view one adopts, the notion of free will remains problematic. This is why Dennett [5] and others have sought to show that the notion of determinism is irrelevant to the notion of free will, and to explain how any free will worth having can still be compatible with mechanistic and deterministic explanations of human behavior.

Furthermore, as Pinker points out [7, p. 176], anyone hoping that an uncaused soul might rescue personal responsibility is in for a disappointment. The last thing we want in a soul is freedom to do anything it desires. If behaviour

were chosen by an utterly free will, then we really couldn't hold people responsible for their actions. That entity would not be deterred by the threat of punishment, or be ashamed by the prospect of opprobrium, and so on. But if the soul is predictably affected by the prospect of punishment, it is no longer truly free.

In practice, we seek to fine-tune our policy of punishment so that it applies only to people who could have been deterred by it [7, p. 183]. This explains why the usual exemptions from responsibility should not be granted to all, even when we think we can explain what led them to act as they did. The notion of deterrence is crucial here. So why should we discard our lever on the system for inhibition just because we are coming to understand the system for temptation. If we believe we shouldn't, that is enough to hold people responsible for their actions without appealing to a will, a soul, or any other ghost in the machine.

Finally, there are both religious and secular versions of the fear of nihilism [7, Chap. 11], i.e., the idea that biological explanations of the mind may strip our lives of meaning and purpose. Consider first the religious one [7, p. 187]:

> [. . .] if scientists are right that the mind emerged from living matter, we would have to give up the value and dignity of the individual, solidarity and selflessness with regard to our fellow humans, and the higher purpose of realizing these values through the love of God and knowledge of his plans. Nothing would keep us from a life of callous exploitation and cynical self-centeredness.

Pinker goes on to refute the accusation that a materialistic view of the mind is inherently amoral and that religious conceptions are to be favoured as being inherently more humane [7, p. 187]:

> The brain may be a physical system made of ordinary matter, but that matter is organized in such a way as to give rise to a sentient organism with a capacity to feel pleasure and pain. And that in turn sets the stage for the emergence of morality.

Naturally, the alternative to any religious theory of the source of moral values is that evolution has endowed us with a moral sense and we have expanded its circle of application over the course of history through the following:

- Reason: grasping logical interchangeability of our own interests and those of others.
- Knowledge: learning of the advantages of cooperation over the long term.
- Sympathy: having experiences that allow us to feel other people's pain.

Of course, what is hard for many to realise is how an entirely mindless and amoral process like evolution could have endowed us with a moral sense. But this is precisely what we can now explain.

In a simple thought experiment, Pinker puts paid to any theory of morality inspired by the notion of a monotheistic god [7, p. 189]:

> What would be the right thing to do if God had commanded people to be selfish and cruel rather than generous and kind? Those who root their values in religion would have to say that we ought to be selfish and cruel. Those who appeal to a moral sense would say that we ought to reject God's command. This shows [we all hope] that it is our moral sense that deserves priority.

And as Pinker points out, this is not just a logical brainteaser, since the history of religion shows that gods, especially the autocratic monotheistic ones, have indeed commanded people to do all kinds of selfish and cruel things [7, p. 189]:

> The recurrence of evil acts committed in the name of God shows that they are not random perversions.

The rest of this passage deals yet another death blow to morality as delivered by religion. With regard to any threat to morality that might be posed by non-believers, he notes this [7, p. 189]:

> If non-believers thought they could elude the legal system, the opprobrium of their communities, and their own consciences, they would not be deterred by the threat of spending eternity in hell. But they would also not be tempted to massacre thousands of people by the promise of spending eternity in heaven.

And finally [7, p. 190]:

> Would life lose its purpose if we ceased to exist when our brains die? On the contrary, nothing invests life with more meaning than the realization that every moment of sentience is a precious gift. How many fights have been averted, how many friendships renewed, how many hours not squandered, how many gestures of affection offered, because we sometimes remind ourselves that life is short?

We don't need to look far to feel that life is worth living, if it is worth living, and when it isn't, the solace provided by religion is often accompanied by fear of the afterlife, quite apart from the fact that the notion of an afterlife is highly unlikely to have any foundation and many 'believers' are aware of this, so harbour a good measure of doubt.

What then of the secular fear that a biological understanding, or scientific knowledge in general, might drain life of meaning [7, p. 190]? Should we be depressed at the thought that all our thoughts may deep down be selfish? Those who reach this conclusion confuse *ultimate causation*, i.e., why something evolved through natural selection, which is starkly amoral, with *proximate causation*, i.e., how the resulting organism works here and now. This confusion arises to some extent from the language that has been used to express the purely metaphorical 'motives' of genes, which would be entirely selfish if they existed at all. The point is that genes are just 'out to' replicate themselves, but the organisms they design have real motives. Pinker puts it like this [7, p. 190]:

> Sometimes the most selfish thing a gene can do is wire unselfish motives into a human brain—heartfelt, unstinting, deep-in-the-marrow unselfishness. The love of children (who carry one's genes into posterity), a faithful spouse (whose genetic fate is identical to one's own), and friends and allies (who trust you if you're trustworthy) can be bottomless and unimpeachable as far as we humans are concerned (proximate level), even if it is metaphorically self-serving as far as the genes are concerned (ultimate level).

Of course, people do sometimes have ulterior motives, exerting their effects from certain inaccessible quarters of the mind. As Pinker points out, when we combine this observation with the common misconception that genes are a kind of essence of the individual, we might mistakenly conclude that the metaphorical motives of the genes are the deep, unconscious ulterior motives of the person. But this is just wrong.

Let me just give one more quote from Pinker that could usefully be compared with what Ellis has to say in his Sect. 8.7.3 entitled *Where Is Truth?*. This is taken from [7, p. 193]:

> Whatever its ontological status may be, a moral sense is part of the standard equipment of the human mind. *It's the only mind we've got*, and we have no choice but to take its intuitions seriously. If we are so constituted that we cannot help but think in moral terms (at least some of the time and toward some people), then morality is as real *for us* as if it were decreed by the Almighty or written into the cosmos. And so it is with love, truth, and beauty.

A Short Conclusion

What Ellis is combating is the claim that, if we had full data of things at an early stage in the universe and if we just had sufficient computing power, we could predict everything that is happening today. But at the same time, the issue here is not actually *prediction*, or even *explanation*. What we are really concerned about is whether things can come into being through, not necessarily deterministic, bottom-up processes alone, and Ellis puts forward no viable argument against that. Of course, nobody would try to *explain* social processes and animal behaviour in terms of quarks. That would not constitute an explanation, except on an in-principle level, because meaning has evolved in such a way [5] that explanation of higher level processes must be higher level explanation. Top-down 'causation' in the sense that Ellis defines it is of course essential to our *explanations* of things in terms of what is relevant to us, i.e., what is meaningful to us, at our level, but that does not mean that it is an essential part of our in-principle explanation of the world in terms of fundamental physical processes. One thing just doesn't rule out the other.

The reason Ellis needs to appeal to higher level processes is that he thinks meaning needs to be a fundamental part of the ontology of the world, whereas it can be just a derived one, that has evolved [5], like animals, in a world whose

ontology only actually requires elementary particles and fields and the laws of physics. The weak point, in fact, the breaking point, in Ellis' whole argument is his own unwitting assumption that meaning is relevant to cause. Meaning is certainly relevant to explanation on the human level, but that is another thing, a human-based thing that itself derives from the purely physical processes we postulate to have led to our existence.

References

[1] S. Blackmore: *Consciousness: A Very Short Introduction*, Oxford University Press, Oxford (2005)

[2] S. Blackmore: *Consciousness: An Introduction*, Routledge, London, New York (2005)

[3] D.C. Dennett: *Intuition Pumps and Other Tools for Thinking*, Allen Lane, London (2013)

[4] D.C. Dennett: *Darwin's Dangerous Idea*, Penguin, London (1996)

[5] D.C. Dennett: *Freedom Evolves*, Penguin, London (2004)

[6] D.C. Dennett: *Consciousness Explained*, Allen Lane, London (1993). Ellis declares that science has no idea how to tackle the problem of consciousness, despite the claims made in this book by Dennett. Perhaps he never got as far as the first paragraph of the preface. Dennett nowhere claims to have solved the problem of consciousness, only to have put forward some ideas

[7] S. Pinker: *The Blank Slate. The Modern Denial of Human Nature*, Penguin, London (2003)

[8] S. Pinker: *The Better Angels of Our Nature. A History of Violence and Humanity*, Penguin, London (2012)

[9] R. Dawkins: *The Extended Phenotype. The Long Reach of the Gene*, revised edn., Oxford University Press, Oxford (1999)

[10] D. Dürr, S. Teufel: *Bohmian Mechanics. The Physics and Mathematics of Quantum Theory*, Springer-Verlag, Dordrecht, Heidelberg, London, New York (2009)

[11] J. Bricmont: *Making Sense of Quantum Mechanics*, Springer-Verlag, Heidelberg (2016)

[12] K.R. Popper: *Objective Knowledge. An Evolutionary Approach*, Oxford University Press, Oxford (1972)

[13] J. Tooby, L. Cosmides: Psychological foundations of culture. In: J.H. Barkow, L. Cosmides, and J. Tooby, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford University Press, New York (1992)

[14] E.O. Wilson: *Consilience. The Unity of Knowledge*, Knopf, New York (1998)

[15] D. Hofstadter: *I Am a Strange Loop*, Basic Books, New York (2007)