

Dennett to Papineau, July 3<sup>rd</sup>, 2017

Dear David,

First, let me thank you for writing such an insightful and candid review (TLS, June 30) You make no bones about the parts of the book that you find incredible, but you do so in a way that will, I think, encourage readers to see for themselves what they make of it. You even offer a “not entirely implausible” (nice phrase) diagnosis of what could have led me to such extravagantly counterintuitive views: my Rylean upbringing. I am happy to acknowledge my deep debt to Ryle, a philosopher whose importance is now beginning to be appreciated anew, but your diagnosis systematically overlooks two of the main themes of the book—gradualism and illusionism, to give them oversimplifying names--a blind spot perfectly captured in one of your closing sentences:

He is constitutionally disinclined to give credit to the powers of the mind, even when it is due.

This is ambiguous, is it not? If you mean I am on a constant vigil to expose unacknowledged reliance on magical (“real magic”) powers of the mind, then yes, you are right. See my essay on illusionism (JCS, 2016). It has been my work for fifty years to provide alternative, naturalistic alternatives to all these chimeras; it is not that consciousness doesn’t exist but that it isn’t what you probably think it is. (Qualia are a theorist’s illusion but the subjectivity of experience is real.) If, on the other hand, you mean that I try to deny the amazing power of human minds to plan and explain and introspect and create, then no. On the contrary, the main point of my account of cultural evolution is to show how those very powers arose; cultural change gradually became de-Darwinized, as its products accumulated—thinking tools by the hundreds of thousands—and empowered their users, permitting them for the first time on Earth to comprehend in the strong sense. Why did I put Bach in my “somewhat unfortunate” title? As a shining exemplar of an intelligent designer. You say:

Scarcely less striking are Dennett’s claims about the development of modern culture. Very little of it is deliberately directed, says Dennett. As he sees it, the technological wonders of the modern world have little to do with people understanding how certain means will achieve certain ends.

Nowhere do I say or imply this. I am, after all, trying to explain how we—we human beings—come to be able to anticipate how certain means will achieve certain ends in advance of specific experience, in other words to understand in a strong sense. Reread the first paragraph of the chapter on post-intelligent design (p371). This is something of a hymn to the “powers of the mind,” is it not, and it summarizes the analysis that goes before it. My constitutional disinclination is to let those powers work as skyhooks, instead of trying to show how their dizzying heights could be explained by a cascade of cranes. It is the traditional, complacent reliance on unexplained “powers of the mind” that I am vying against. No dualism, please, and no wonder tissue, until we have exhausted the resources of naturalism.

You note that I claim that “apparently insightful behavior is nearly always performed unthinkingly.” Yes, indeed. There has been apparently insightful, sometimes breathtakingly

cunning, behavior by unthinking bacteria and even viruses by the quintillions for billions of years. Our thinking triumphs, for all their splendor, must account for a vanishing fraction of one percent of all the apparently insightful behavior that has ever occurred. Our kind of reflective thinking is a very recent innovation, and you yourself agree that the zebra “doesn’t self-consciously comprehend” its impressively appropriate escape behavior. I am claiming—and it seems quite obvious to me, not in the least ‘peculiar’—that we must break our habit of *assuming* “thinking” whenever we see cleverness. This is the main message of Darwin’s strange inversion of reasoning, and many people find it an uncomfortable challenge to their preconceptions. As I note,

Ironically, if we were creationists, we could comfortably attribute all the understanding to God and wouldn’t feel so compelled to endow the organisms with it. They could all be God’s marionettes. It was Darwin’s discovery and exposure of the mindless process of natural selection, with its power to generate free-floating rationales, that freed our imaginations to continue reverse engineering all nature’s marvels without feeling an obligation to identify a mind that harbors the reasons we uncover. (p339)

And recall that I explicitly soften my claim in “Comprehension comes in degrees” (pp94-101). Page 100 has a catalogue of examples of animal comprehension and I close the section with “In animals with more complex behaviors, the degree of versatility and variability exhibited can justify attributing a kind of behavioral comprehension to them so long as we don’t make the mistake of thinking of comprehension as some sort of stand-alone talent, a source of competence rather than a manifestation of competence.”

A few more words about illusionism, which I think you also misconstrue, though I take some of the blame for not making my position clearer in the book. (My JCS essay makes up for that in part, but more needs to be done. And by the way, the tutorial on illusionism that Frankish, Lambert and I ran at ASSC in Beijing a few weeks ago overflowed the room, and was well received by the scientists as well as the philosophers. My position may still be a minority position, but it is gaining adherents as theorists confront the unattractiveness of the alternatives.) At one point you quote a sentence of mine

We won’t have a complete science of consciousness until we can align our manifest-image identifications of mental states . . . with scientific-image identifications of the subpersonal informational structures and events that are causally responsible for generating the details of the user illusion . . .

and confess: “This brought me up short.” You declare that this is just what my “mainstream philosophical opponents would say” and then go on to acknowledge that “In the end, I was rather left wondering what the ‘user-illusion’ fuss was all about.” You suspect that you are missing something, and you are right: my mainstream philosophical opponents suppose that we “have access” to “phenomenal properties” of these subpersonal informational structures and events. This is the great illusion: that we—we “first-person subjects”—have any insight *at all* into the properties of the internal causes of our beliefs about our experience. Richard Power, a retired AI researcher, has put the point succinctly in an email to me:

We understand the concept of representation from external representations, such as pictures, or verbal descriptions. For these representations we can have direct experience of both a representer (e.g., portrait painting) and a representee (e.g., the person painted). Call these the medium and the content. . . . For external representations we can experience both medium and content, oil on canvas as well as people, trees, or whatever. But for internal representations, we do not experience the medium AT ALL. Only the content, along with some contextual features such as the time when the percept or imagining occurred.

People are generally comfortable with the discovery that they have no *direct* knowledge “from the inside” of the properties of the blood-purifying events in their kidneys, or of the properties of the peripheral events in the eyeball and optic nerve that subserve vision, but the idea that this ignorance of internal properties of the relevant events in the brain carries “all the way up” is deeply counterintuitive. We seem to be able to “directly inspect” some of our inner states, our conscious states. After all, we can tell people—including ourselves—all about what is happening in our minds, if not our brains. This ability of ours is perfectly real, but we have no intimate knowledge of how we are enabled to do this. (We often cover our ignorance by saying we have “direct acquaintance” with special properties, *qualia*, that defy scientific analysis.)

Electromagnetic radiation gets transduced at the rods and cones of the retina into neuronal spike trains, and there is no second transduction into a “medium of consciousness”; it is spike trains all the way. That is a fact, not a peculiar assumption. But there seems to be a second transduction; that is an illusion, and it can be explained. We’re working on it.

There is much more to be said, of course, but your review laid out the structure and most of the substance of my book so well that we can get right down to business resolving the issues that are still puzzling.

Best wishes,

Dan

Papineau to Dennett, July 15<sup>th</sup>, 2017

Dear Dan

Thank you for your generous and friendly response to my review. I welcome the opportunity to explore our differences further.

Let me start by applauding your rejection of magical explanations. I could not agree more that we need to be vigilant against crediting minds with supernatural powers, and am very happy to join your crusade to find naturalistic alternatives to such mirages.

But this leaves open the extent to which intelligent thought – real naturalistic intelligence – is important. I still think that you give it too little credit.

One issue here is how much of cultural evolution depends on intelligent design. This probably isn't the place to resolve this. We both agree that design plays some role, and differ only on the degree. (But I was bit surprised that you took exception to my claim that you hold that 'Very little of it is deliberately directed.' On p. 246 you say 'Yes, *some* of the marvels of culture can be attributed to the genius of their inventors, but much less than is commonly imagined, and all of it rests on the foundations of good design built up over millennia by uncomprehending hosts of memes competing with each other for rehearsal time in brains.')

We differ more sharply about animal thought. You contrast the 'cunning behaviour' of bacteria and viruses with 'our kind of reflective thinking'. Well, we can all agree that the bacteria and viruses can't think, and that we humans can. But where does that leave the animals in between (not to mention the many humans – we all know some – who are not much given to reflection)? Your inclination is to set the bar high, and so to exclude animals like zebras from the realm of thought.

But why? Along with nearly all other philosophers and cognitive scientists, I take it that zebras harbor internal cerebral states that represent features of their environment, and which combine with other such states to guide zebra behaviour. Indeed recent work on 'model-based learning' argues that such cognitive processing allows animals like zebras 'to anticipate how certain means will achieve certain ends in advance of specific experience' — which you accept is the mark of understanding 'in a strong sense'. All this seems to count against your insistence that zebras have nothing but 'competence without comprehension'.

I know that you allow that comprehension comes in degrees, and so are happy to credit 'a sort of behavioural comprehension' to animals with complex behaviours. But why so grudging? If a zebra can represent the presence of a lion, and so work out how to respond appropriately, why isn't that comprehension in a strong sense? Not only *is* there a reason for the zebra's action – a lion is nearby – but the zebra *knows* the reason.

I wonder whether your position is influenced by your long-standing attachment to the 'intentional stance'. Ever since your 1971 paper 'Intentional Systems' (though not, I think, in your earlier *Content and Consciousness*), you have distinguished between 'intentional' theorizing, which credits agents with representational understanding, but is neutral about their inner workings, and the 'design stance', which aims to uncover the structure of their cognitive machinery. (In your famous example, you cite a chess computer that 'thinks it should get its queen out early'. This intentional attribution might tell us about the computer's inclinations, but we would be foolish to infer that any such directive is explicitly encoded in the program's design.)

As you know, not all philosophers are happy about this dichotomizing of stances. It drives a wedge between portraying subjects as representers and investigating their inner construction. I don't see why we can't do both simultaneously. Much contemporary work on animal cognition posits internal states that interact causally in various ways, but are also representational, in that

they stand proxy for features of the animal's environment, and guide the animal's behaviour accordingly. Here the attribution of intentional powers and the understanding of design seem to proceed hand-in-hand. Far from requiring distinct stances, they look like a single enterprise.

As far as I can see, your commitment to a distinctive 'intentional stance' does not positively require you to restrict genuine understanding to humans – indeed it seems to leave it entirely open which subjects, if any, are 'really' engaging in intentional thought. But it does create room for you to rule out other animals. Given your disconnect between design information and intentional attributions, no amount of knowledge about the internal workings of animals will entail that they have intentional powers – which then allows you to run your behaviourist line that all animal intelligence resides in practical competence without mental comprehension.

It's a coherent story, but I don't buy it. In my book, the question of which subjects have intentional states depends precisely on the details of their cognitive design, and on this score animals like zebras certainly come out as fully capable of thought. They might not have all the high-falutin' mental powers of our exquisite human selves, but they beat the bacteria and viruses hands down.

On consciousness, you explain that you differ from your mainstream opponents on the issue of whether we 'have access' to 'phenomenal properties'. That's helpful. Where you deny such access, I think that nothing could be more obvious.

Let us first agree some common ground. I take it that we are all materialists here. It's neuronal spike trains all the way, as you put it. So any mental properties of people must be constituted by (possible complex) neuronal arrangements in their brains. I also take it that we are not complete eliminativists about the mind. People really do have pains and visual experiences. So properties like *being in pain* or *seeing something red* must themselves be complex neuronal arrangements.

But that is all that I and my mainstream colleagues mean by 'phenomenal properties' – properties like *being in pain* or *seeing something red*. And it seems obvious, to say the least, that we 'have access' to them. I can know directly whether or not I am in pain or seeing something red, in a way that other people cannot. Why would anybody want to deny that?

Now, it is no part of this mainstream position that introspection gives us access to *all* the features of our conscious states. Some philosophers, like David Chalmers, do suppose this, and conclude that conscious states cannot be material, on the grounds that introspection does not reveal them to be so. Perhaps you too are persuaded by this line and hold that, once we allow introspective access, we will be committed to an (illusory) non-material realm. But the materialist mainstream gets off this bus much earlier. Since pains and visual experiences are brain states, it follows that introspection reveals very little about their nature. It tells us when we have them, but not what they are made of.

You also raise the question of *how* we can tell what is happening in our minds. It's an interesting question, and I know of a number of serious naturalistic answers – no magic – currently under investigation. True, as yet we lack a universally agreed theory of our introspective powers. But

that's no reason to deny we have such powers. After all, nobody is quite sure how homing pigeons manage to navigate back to their lofts, but that doesn't stop the pigeons getting there.

I'm not supposing that introspection is infallible. Since it's not magic, it can always go wrong. Your own work – especially in *Consciousness Explained* – has done much to expose the interesting tangles that introspection can get us into. But these limitations are also no reason to reject introspective powers. You might as well say that pigeons lack any access to the way home, because they occasionally get lost.

One final issue. Your quotation from Richard Power rightly points out that we standardly identify phenomenal properties in terms of their 'contents'. We talk of a pain 'in my foot', or a visual experience 'of a red square'. And you infer from this that anybody who takes phenomenal properties seriously must end up trading in 'qualia', mysterious entities that are somehow constituted by these 'appearances', and yet retain their identity even when the appearance is a figment, as with an amputee's phantom pain, or an illusory red square.

I don't take our ordinary ways of talking so seriously. My foot isn't really part of my conscious pain, nor is a red square part of my visual experience. Instead these 'contents' can simply be viewed as useful labels for the intrinsic (and ultimately neural) properties that constitute our conscious experiences. Just as we name famous diamonds after their owners, so we can name phenomenal properties after their normal causes, without supposing in either case that the name points us to an essential feature of its bearer. (I seem to remember learning about this trick from your own article 'Beyond Belief' in the late 1970s.)

In the end, maybe the differences between us are not so great. I agree with you that many people, both within philosophy and without, think of consciousness in materialistically unacceptable ways, and that consciousness so understood is an illusion. But rather than rejecting consciousness altogether, I prefer to clean up our thinking. Immaterialism and eliminativism are not the only alternatives. If we conceive of conscious states in the right way, we can simply identify them with internal neuronal arrangements. Your Rylean heritage might make you uneasy about conceiving of mental states as inner mechanisms. But I invite you to throw off your behaviourist shackles, and join the rest of us in the materialist mainstream.

Dennett To Papineausant, July 17th, 2017

Dear David,

Three quick points: There's a difference between saying, about the cunning design observable in human artifacts, that "very little" is due to human genius, and, as I say, "much less than is commonly imagined," a claim also defended, brilliantly, by Joseph Henrich in *The Secret of our Success* (2015). Most people seem to think that almost all design is intelligent design, and they are simply wrong. I expect you agree with me about that.

And we don't differ as sharply on animal thought as you suggest. I say the zebra has a sort of behavioral comprehension, and you say that a zebra can represent the presence of a lion, "and so

work out how to respond appropriately.” I almost agree, but your verb, “work out” suggests a rather more knowing, reflective response than is warranted, and I wouldn’t be surprised if you agree. In one sense, perhaps, “the zebra *knows* the reason” as you say, but it doesn’t know that it knows this.

On consciousness, our differences are also not so great, almost vanishing. I don’t deny “introspective powers” as you claim; I simply insist that we tend to inflate their strikingly limited products into metaphysical epiphanies. You agree with me that the way most people understand consciousness, it is an illusion, while “if we conceive of conscious states the right way, we can simply identify them with internal neuronal arrangements.” So you and your “mainstream colleagues” are illusionists like me after all. I am happy to learn that there is now so much agreement that there is no problem of “intrinsic” qualia. I do not shun or neglect internal structures and processes. My “behaviorism” is not the behaviorism of either Skinner or Ryle, it is the behaviorism of science. Meteorology is behavioristic in this sense, and so is chemistry, and physics and geology and astronomy. When you achieve a theory that explains all meteorological *behavior*, you get to declare victory; you’ve finished the task, because that’s all there is to explain. Or do you think there is something about consciousness to explain beyond the facts uncoverable by heterophenomenology, the standard method of cognitive science?

Papineau to Dennett, July 24<sup>th</sup>, 2017

Dear Dan,

Here now is my short last word (sorry this has to end somewhere, Dan):

I am glad we agree on so much, and in particular about the neural machinery that enables animals to navigate the world, and about the brain processes that constitute the pains, visual experiences and other mental events that humans are introspectively aware of. Given this, our remaining differences strike me as no more than terminological: should we set the criteria for “comprehension” and “consciousness” so high as to deem animals uncomprehending and human consciousness an illusion?

But this leaves me with a puzzle. Why then does your book go out on a limb and insist on the incomprehension of animals and the illusoriness of consciousness, even against those many philosophers, like myself, who have made it clear that we agree with you on the substantial issues? Why so adamant that this is the only way to put things? It is at this point that I wonder about the lingering influence of Ryle’s doctrine that it’s a category mistake to understand mental terminology as referring to inner cerebral mechanisms. Without some such principle, there seems no route to your view that the agreed science eliminates animal thought and human consciousness.