Why is animal consciousness controversial? A trialogue

Jonathan Birch

Centre for Philosophy of Natural and Social Science,
London School of Economics and Political Science.
j.birch2@lse.ac.uk

6 January 2025

The idea that many other animals have conscious experiences, including all vertebrates and many invertebrates, is less controversial in science than it once was. I don't imagine it was ever all that controversial in the rest of society. In my experience, people do not seriously doubt the sentience of their pets, be it a dog, cat, iguana, fish, or tarantula. On this issue, scientific opinion seems to be moving into closer alignment with popular sentiment. One bellwether is the New York Declaration on Animal Consciousness (Andrews et al. 2024), a public statement signed by 536 scientists and philosophers that aims to capture this change in attitudes (I was one of the co-organizers).

Yet many skeptics remain, including some high-profile neuroscientists (Rolls 2014; LeDoux 2019; Lau 2024), psychologists (Heyes 2008) and philosophers (Carruthers 2019). Others are ghosts from the past, like René Descartes, who wrote that "language is the only certain sign of thought hidden in a body" (Seris and Voss 1993, p. 179); pioneering evolutionist Thomas Henry Huxley, who argued that, because "states of consciousness can have no sort of relation of causation to the motions of the muscles", questions of animal consciousness "do not lie within the scope of physical science" (Huxley 1874, pp. 365-6); and arch-behaviorist John B. Watson, who wrote that "one can assume either the presence or the absence of consciousness anywhere ... without affecting the problems of behavior by one jot or one tittle" (Watson 1913). Their arguments shaped the scientific culture of their time, and for generations thereafter, and their imprint can still be felt strongly.

It can be tempting to accuse skeptics of motivated reasoning. Descartes, Huxley and Watson were all prominent defenders of vivisection. Some of today's skeptics about fish sentience are financially dependent on the aquaculture industry. But not all skeptics stand to benefit commercially or professionally from their skepticism. There are reasonable skeptics, including those just cited—in fact, I am a reasonable skeptic too. I often have inner disputes in which I try to make the skeptical case as sincerely and honestly as I can in order to assess how strong it really is.

In these disputes, I find myself seeking a middle course between credulousness—attributing consciousness to other animals on the basis of weak evidence—and an excessive skepticism that holds attributions of consciousness to a higher bar than any ordinary scientific claim. In

other words, there is an ongoing trialogue within me. One voice stands up for common sense, denying that the popular belief in animal consciousness requires scientific validation. Another takes the opposite view, claiming that, for principled reasons, no amount of evidence will ever be enough to warrant attributing consciousness. A third voice looks for the middle way, accepting the need for scientific validation while asserting that the demand can be met, provided we do not insist on a level of certainty we would never expect in any other area of science.

I want to put that inner trialogue out in the open. I've framed it as a conversation between three characters: Credulus, who thinks conscious mental states can and should be attributed to other animals without any need for scientific inquiry, Skepticus, a critic with behaviorist leanings, and Moderus, who sees a middle path in the emerging science of animal consciousness.

*****

Credulus: Complex animals—and I include invertebrates like octopuses, crabs, insects and spiders in this, along with vertebrates—are not robots. It's obvious they have feelings. Their feelings surely differ from mine in many ways, because their bodies and their sensory abilities differ. But certain basic experiences—joy, pain, hunger, thirst, tiredness, for example—are widely shared. People have known this for thousands of years. Regarding such claims with a default attitude of skepticism is an aberration of modern Western science. Non-Western cultures have never bought this nonsense, and not even Western societies buy it, once one looks outside the academy.

I'm very fond of an essay by J.M. Coetzee called "The Lives of Animals" (Coetzee 1999). It contests the idea that we must await scientific validation before acknowledging the joy and suffering of our fellow creatures—and our shared vulnerability. Other animals are treated appallingly by humans, yet the scientific norm is to suspend judgement about their feelings until certain approved kinds of laboratory evidence have been collected. The experience of living with animals on a day-to-day basis is discounted as insufficiently scientific. Yet the lab-based evidence is never strong enough, it seems. Even unusually intelligent animals like dolphins and chimpanzees fail to supply the *linguistic* evidence demanded for consciousness attributions, so judgement can be suspended forever. We should not play this game. To make our empathy contingent on prior scientific endorsement is a huge mistake.

Skepticus: With respect, I think it's possible to sustain one's faith in common sense right up to the moment you confront the evidence—at which point you realize that even quite sophisticated behaviors can be controlled non-consciously.

The most famous example is "blindsight". Blindsight happens in rare cases, when a patient has an injury to a part of their brain called the primary visual cortex. The injury affects a specific region of their visual field: the blind field, or scotoma. The patient reports having *absolutely no conscious vision* in the blind field. Yet information about objects in the blind

field reaches other parts of the brain and can guide actions. In a famous example, a patient fully blind across the whole visual field according to their own reports could "successfully navigate down the extent of a long corridor in which various barriers were placed . . . skillfully avoiding and turning around the blockages" (de Gelder et al. 2008).

If complex, visually guided behavior can be controlled by mechanisms that operate below the level of conscious experience, we cannot point to such behavior in other animals and say "That's evidence of consciousness!". No, it isn't, sorry. It's evidence of them having mechanisms of action control that may be conscious or unconscious.

To be clear, I'm not the kind of skeptic who says "Whaddya mean by consciousness?" and sees in this a knockdown argument. Most of the terms we use to denote broad areas of inquiry in the sciences of mind and brain (e.g. perception, memory, cognition) lack precise definitions. My point is rather that, on *any* reasonable definition of consciousness—define it how you want—it is *not* there in the blind field of a blindsight patient, and yet guidance of complex behaviors by visual information in the blind field remains. So, when we turn to other animals, the fact we observe complex behavior guided by sensory information (which I grant!) is no evidence that the information is consciously experienced.

And there is a second body of evidence that should make you uncomfortable: the literature on *anthropomorphism*. We are very prone to "see" mentality in systems that don't have it but that mimic human features, like humanoid robots and animated characters. Even an animated character without any body parts at all—a moving triangle, for instance—can evoke very strong projections of mental states. Accordingly, a foundational idea in the science of animal minds is that we need to build in safeguards against our anthropomorphic tendencies. Otherwise, we'll end up attributing conscious minds to unicellular organisms, like amoebae and paramecia, simply because they behave in ways that vaguely resemble the things we do consciously.

Credulus: As well as anthropomorphism, there is "anthropodenial", as Frans de Waal called it: the human tendency to deny mental continuity with other animals (de Waal 1999). There is also "anthropofabulation": our tendency to think that human behaviour and cognition is oh-so-special and sophisticated, refusing to believe that it may largely rely on mechanisms shared with other animals (Buckner 2013). Given that we have all three tendencies, and two of them bias us *against* recognizing mentality in other animals, there's no justification for the scientists' selective obsession with mitigating the risks of anthropomorphism while neglecting these other risks.

Moderus: Maybe it will help to backtrack a little. What entitles me to infer that the *people* around me have conscious minds? It is an "inference to the best explanation" or "abductive" inference, a kind of inference found in all areas of science. If we don't accept these inferences in the case of consciousness, we are applying an unjustified double standard. As long as we do accept this kind of inference, the case for consciousness in other humans is overwhelming.

I have a lot of behavioral data that needs explaining, after all. I am constantly having conversations with other people about my experiences and theirs, about our shared experiences, about patterns of similarities and difference. I ask them how they felt when they heard some good or bad news, whether they are feeling more tired or energetic than they were yesterday, whether that painting evokes the same emotions in them as it does in me.

It could conceivably be that, although I am having experiences, I live in a world of philosophical zombies who are knowingly faking all these reports (Kirk 2023). It could also be that, while they are not knowingly deceiving me, other people use all the same words to refer to computations that feel like nothing to them, and this vast difference in our inner lives never comes to light. A third explanation is that we do, in fact, have conscious experiences of quite similar kinds, allowing us to convey truths to each other about how we feel.

How can we choose between these explanations? The first two posit a metaphysical gulf between me and the people around me for which I have no evidence or explanation. By contrast, the third one explains everything very simply and straightforwardly, meshing well with everything else I believe, positing no weird gulfs, and so it is reasonable for me to believe the third one.

Now let us turn to other animals. What entitles me to infer that they have conscious minds? *Exactly the same type of inference!* Here too I have a lot of behavioral data in need of explanation. For pet owners, the data are especially rich, which may explain why they are so confident. It is not just anthropomorphism.

It's true enough that *linguistic* evidence is absent (though other forms of communication exist). But linguistic evidence, though a good and important kind of evidence in the human case, was never the *only* evidence. When I interact with preverbal children, I'm again in a situation where *by far* the best explanation for the data is that we are both conscious beings. The idea that the child laughs and cries at amusements and threats unconsciously seen, zombie-like, puts us back in "weird gulf" territory. Kids are not unconsciously playing with their toys, driven by unconscious analogues of joy. This is nowhere near as good as the explanation that posits conscious experiences of similar kinds in both of us.

Why is the situation any different when we interact with a dog? The dog's "excitement-like behavior" when chasing a ball is very well explained by positing a feeling of excitement, and very much less well explained by some kind of unconscious analogue of excitement—it's that weird gulf again, where there is no reason to posit such a gulf. Yes, there are fascinating abnormalities like blindsight, but there is no reason to think human behavior is generally motivated by unconscious mechanisms—by "unfelt emotions" and "unfelt moods". It's speculative to suggest there even *are* such states. This is an explanation we create especially for other animals while denying it for ourselves, and that's, in short, a poor explanation.

Skepticus: I don't concede the point on "not just anthropomorphism", but I won't blather on about this. The question of why people feel so confident about their pets is a side issue, whereas the fundamental issue concerns the standards for a justified scientific inference.

I see it like this: there are standards of inference for everyday life, and there are other, *higher* standards that apply in science. The idea of an "inference to the best explanation" appears in both settings, but it's a mistake to assume that the standards are the same in science and everyday life.

In everyday life, we can be fairly relaxed about explanations that posit causes just to explain something that needs explaining. I see oddly-shaped footprints in the snow, so I infer a person with shoes of that shape walked that way, even though I have no other evidence of such a person existing (Lipton 2004). But in science we can't allow that kind of thing or we'll face an explosion of convenient causes. There are strict entry criteria for what counts as an eligible causal explanation of a phenomenon *in science*.

What are the entry criteria? John Herschel's "vera causa" standard has been influential (famously, it influenced Darwin). Herschel argued that, in science, explanations must cite causes whose *existence* and *competence* to produce the phenomenon of interest has been *independently established*. Darwin took this maxim extremely seriously, arguing via a detailed analysis of selective breeding that unconscious forms of selection exist and are competent to produce large changes. Only then could he appeal to this independently established vera causa to explain adaptation and the origin of species.

Here is the basic problem: *conscious mental states don't meet the vera causa standard*. You want to posit them to explain the behavior of other animals—OK, but you haven't independently established their existence and competence to produce that behavior.

And there are deep reasons why no one can do this. Firstly, the very existence of these states is called into doubt by "illusionism" about consciousness, as defended by Daniel Dennett (2017) and Keith Frankish (2016), and in older literature by "eliminative materialism" (Ramsey 2024). Secondly, their competence is called into doubt by the possibility of epiphenomenalism, as defended by Huxley: the idea that conscious experience might be a sideshow with no effect on behaviour, akin to the steam that trails behind a steam train making no difference to the workings of the engine. There are also theories that we might call "near-epiphenomenalist", in that they give consciousness only very small and circumscribed causal roles, e.g. they say that experiences cause reports of experience, but no other behaviours. Some "higher-order" theories come close to epiphenomenalism in this way (Carruthers and Gennaro 2023).

To establish the existence and competence of conscious mental states to explain observations of animal behavior, then, you'd first have to refute illusionism, epiphenomenalism, and near-epiphenomenalism, and *you can't do that*—these are all live options. As long as that's the case, conscious mental states will never meet the vera causa standard.

Moderus: I'm starting to think a significant disagreement between us concerns the relevance of first-person evidence. In my view, first-person evidence can establish the existence and competence of conscious mental states.

Indeed, I'd say first-person evidence rules out a strong form of illusionism that denies there is *anything* it feels like to be me. It leaves room for weaker forms that deny some specific theoretical posit (like "qualia" or "phenomenal properties", understood in some theoretically-loaded way), but that's alright—this milder variant is no threat to conscious mental states' being *vera causae*.

More contentiously, I think first-person evidence establishes the *centrality* of conscious experience to human mental lives, including its causal power, especially in the areas of learning and decision-making. The "steam" view of conscious experience just isn't viable. I know that I went to the Musée D'Orsay this year because I *enjoyed* a visit last year. It wasn't some unconscious process occurring simultaneously with my enjoyment that influenced my decision—it was the enjoyment. One can debate the coherence of epiphenomenalism in the seminar room, but there's no reason to take it seriously as a viable theory of consciousness, because first-person evidence destroys it.

More than this, first-person evidence renders even a "near-epiphenomenalist" view wildly implausible. Conscious experience is not peripheral to my mental life, barely doing anything—it's at the heart of how I learn about the world, how I value possible futures, and how I make big-picture, strategic decisions about the shape of my life.

That first-person insight is our way in to studying consciousness scientifically in other animals. We can study their learning and their decision-making, and we can look for the neurobiological and cognitive mechanisms that sit at the center of these processes and compare them across species (Birch et al. 2020).

Skepticus: Right. So, your position is that first-person evidence establishes both the existence of conscious experience and its importance to learning and decision-making, so that we don't need independent, third-person validation of these things before we start positing conscious states to explain the behavior of other animals. Got it.

You can guess what I will say in reply—that a ban on "first-person evidence" is another very important scientific norm. This was the grain of truth in behaviorism. Watson saw how the psychology of his time overly relied on introspection—saw how it was putting an air-brake on progress—and he offered an alternative vision in which first-person data had no role.

I see Dennett, in *Consciousness Explained* (1991), as trying to do for consciousness science what Watson did for psychology. He saw two futures for the field. In one, first-person evidence is taken seriously as evidence (as it is by you, it seems) and the field goes nowhere. In the second, first-person evidence is ruled inadmissible at the outset, and we take people's

public reports of conscious experience not at face value (as accurate descriptions of their inner lives) but rather as our *explanatory target*, using the standard methods of cognitive neuroscience to investigate the processes that lead people to make these reports—an approach he called "heterophenomenology".

If we take Dennett's path, we have no right to go around positing real conscious states to explain animal behavior. At the start of inquiry, the only evidence that these states *even exist or do anything* is inadmissible first-person evidence. It's likely that, by applying the heterophenomenological method, we will ultimately discover real neurobiological mechanisms that can explain a lot more about a person than just their reports of phenomenology (e.g. the "global neuronal workspace" may be a good example of this—Dennett was a fan). At this point, we will be able to turn back to other animals. But then the right question to ask is: *Do other animals have versions of these mechanisms?* Not: *Do they have conscious experiences?* The latter question never arises. By the time we know enough about humans to justify turning our attention to other animals, we also know enough to replace "conscious experience" with more legitimate neuroscientific categories.

Moderus: Thanks for the Dennett exegesis. Look, I think this is a respectable kind of skepticism—a skepticism that rests on (a) tough entry criteria for causes that can feature in scientific explanations, plus (b) a ban on first-person evidence, leading to the conclusion that (c) conscious mental states don't meet the entry criteria. We see a huge divide in consciousness science over point (b), and debates about animal consciousness can be seen as a key flashpoint within that conflict. None of this stops a respectable skeptic from making everyday inferences—and believing, like everyone else, that their pets are conscious.

But in the fundamental dispute over (b), I've changed sides over the years. I think at one time I would have agreed that first-person evidence has no place in science. I now see this as an impoverishing move, a throwing away of valuable evidence. We should be very cautious in trusting the deliverances of introspection, of course, but that doesn't mean they can't provide *initial, tentative, defeasible* evidence regarding many questions.

You're familiar with William James (1896) on "the will to believe" and the need to balance the risk of believing falsehoods against the risk of failing to believe important truths. Scientific norms express a conservative attitude to inductive risk—they traditionally skew strongly towards avoiding the affirmation of falsehoods. If some important truths don't get affirmed as a result, that's no problem. Not my job, says the scientist, to worry about that. My job is to stick to where the evidence is most secure.

But I disagree with these conservative norms. When human attitudes towards other animals are in the balance, I think the approach to inductive risk should be different. We should think about the costs, if many other animals are indeed conscious, of a persistent refusal to acknowledge animal consciousness as real and as a legitimate target of scientific inquiry. Yes, it could be that the emerging science of animal consciousness is chasing chimeras—we do run that risk—but it's also possible, and, I would say, far more likely, that it is chasing

profoundly important truths. The absolutism of a ban on first-person evidence, even as a tentative starting point, regardless of what the cost might be in the currency of foregone knowledge, reflects a misguided approach to weighing inductive risk.

Skepticus: It sounds like we agree on a lot, then. We agree that inferences to animal consciousness involve a deliberate loosening of some longstanding scientific strictures, especially regarding the admissibility of first-person evidence—strictures that were introduced for good reasons and that have held firm for a long time. We agree that this intentional loosening is a somewhat risky bet. You think, for Jamesian reasons, that it's important to take that bet. I think, I guess because I fetishize "hard science", "rigor" and whatnot—and don't see it as my place to second-guess the wider social or political context—that these strictures should stay where they are. Shall we agree to disagree?

Credulus: If I can jump back in for a moment—"agreeing to disagree" is an acceptable resolution when nothing urgent hangs on the dispute. But excessive skepticism has, in my view, done immense harm. It has led to a culture of silence in science around the pain and suffering of animals, both on farms and in labs. Where scientists should be speaking out, and could speak with authority and influence, they instead stay quiet or profess agnosticism. They acquiesce. So no, we can't agree to disagree—the culture of science must change.

Skepticus: It's clear that *our* differences are not entirely about standards of scientific inference. We disagree too about the harms and benefits of a skeptical attitude towards animal minds. You see a tight link between skepticism and the mistreatment of animals at human hands. I think, if people see in the skeptical attitude of scientists a license to treat animals as they wish, they are making a profound mistake. Nothing I have said implies that one should not—when setting animal welfare policies—err on the side of caution.

Moderus: But if you support taking precautions, you should also support gathering the evidence needed to do it well! That means supporting a scientific field in which questions of consciousness and sentience can be broached, and where tentative evidence—not conclusive, but good enough to help us make better decisions—is sought and obtained.

Acknowledgements
This research is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant Number 851145.

References and further reading:
To avoid interrupting the flow I have included few references, so the following list also includes further reading suggestions.

Andrews, K. (2020). *How to Study Animal Minds*. Cambridge University Press.
Andrews, K., Birch, J., Sebo, J., and Sims, T. (2024) *Background to the New York Declaration on Animal Consciousness*. nydeclaration.com

Balcombe, J. (2016). *What a Fish Knows: The Inner Lives of Our Underwater Cousins*. Farrar, Straus and Giroux.
Birch, J. (2022). The search for invertebrate consciousness. *Noûs*, 56(1), 133-153.
Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford University Press.
Birch, J., Schnell, A. K. and Clayton, N. S. (2020). Dimensions of animal consciousness. *Trends in Cognitive Sciences*, 24(10), 789-801.
Buckner, C. (2013) Morgan's Canon, meet Hume's Dictum: Avoiding anthropofabulation in cross-species comparisons. *Biology & Philosophy*, *28* (5), 853-871.
Carruthers, P. (2019). *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford University Press.
Carruthers, P. and Gennaro, G. (2023). Higher-order theories of consciousness. *The Stanford Encyclopedia of Philosophy (Fall 2023 Edition)*, E. N. Zalta & U. Nodelman (eds.). https://plato.stanford.edu/archives/fall2023/entries/consciousness-higher/
Coetzee, J. M. (1999) *The Lives of Animals*. Princeton University Press.
de Gelder, B., Tamietto, M., van Boxtel, G., Goebel, R., Sahraie, A., van den Stock, J., Stienen, B. M. C., Weiskrantz, L., and Pegna, A. (2008). Intact navigation skills after bilateral loss of striate cortex. *Current Biology*, 18(24), R1128–R1129.
de Waal, F. B. M. (1999). Anthropomorphism and anthropodenial: Consistency in our thinking about humans and other animals. *Philosophical Topics*, 27(1), 255–280.
Dennett, D. C. (1995). Animal consciousness: What matters and why. *Social Research*, 62(3), 691–710.
Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Company.
Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. Penguin.
Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12), 11-39.
Godfrey-Smith, P. (2016). *Other Minds: The Octopus and the Evolution of Intelligent Life*. William Collins.
Heyes, C. (2008). Beast machines? Questions of animal consciousness. In L. Weiskrantz and M. Davies (eds.), *Frontiers of Consciousness* (pp. 259-274). Oxford University Press.
Huxley, T. H. (1874). On the hypothesis that animals are automata, and its history. *Nature*, 10(253), 362-366. https://doi.org/10.1038/010362a0
James, W. (1896). *The Will to Believe and Other Essays in Popular Philosophy*. Longmans, Green & Co.
Kirk, R. (2023). Zombies. In E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition*)*, E. N. Zalta & U. Nodelman (eds.), URL = https://plato.stanford.edu/archives/fall2023/entries/zombies/
Lau, H. (2024). Premature declarations on animal consciousness hinder progress. *The Transmitter*, 18 November 2024. https://doi.org/10.53053/BCOU1212
LeDoux, J. E. (2019) *The Deep History of Ourselves: The Four-Billion Year History of How We Got Conscious Brains*. Viking.
Lipton, P. (2004) *Inference to the Best Explanation* (2nd edition). Routledge.

Paul, E. S., Sher, S., Tamietto, M., Winkielman, P., and Mendl, M. T. (2020). Towards a comparative science of emotion: affect and consciousness in humans and animals. *Neuroscience & Biobehavioral Reviews*, 108, 749–770.

Pence, C. H. (2018). Sir John F. W. Herschel and Charles Darwin: Nineteenth-century science and its methodology. *Hopos: The Journal of the International Society for the History of Philosophy of Science*, 8(1), 108-140.

Ramsey, W. (2024). Eliminative materialism. In E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2024 Edition), https://plato.stanford.edu/archives/win2024/entries/materialism-eliminative/

Rolls, E. T. (2014). *Emotion and Decision Making Explained*. Oxford University Press.

Sebo, J. (2025). *The Moral Circle*. W. W. Norton & Company.

Séris, J.-P., and Voss, S. (1993). Language and machine in the philosophy of Descartes. In S. Voss (ed.), *Essays on the Philosophy and Science of René Descartes*. Oxford University Press.

Tye, M. (2016). *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* Oxford University Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*, 158–177.

Winkielman, P., and Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science*, 13(3), 120–123.