
How our Belief in Qualia Evolved, and Why We Care so much

A Reply to David H. Baßler

[Daniel C. Dennett](#)

David Baßler’s commentary identifies five unasked questions in my work, and provides excellent answers to them. His explanation of the gradual evolution of higher-order intentionality via a Bayesian account leads to an explanation of the persistence of our deluded belief in qualia.

Keywords

Belief in belief | Dispositions | Intentional systems | Qualia

Author

[Daniel C. Dennett](#)

daniel.dennett@tufts.edu

Tufts University

Medford, MA, U.S.A.

Commentator

[David H. Baßler](#)

davidhbassler@gmail.com

Johannes Gutenberg-Universität

Mainz, Germany

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität

Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University

Melbourne, Australia

David Baßler’s commentary is a model of constructive criticism, not only pointing to weaknesses but offering persuasive repairs. I have just two points of minor correction to offer before turning to my understanding of his interesting proposals for extensions to my view, which I am inclined to adopt.

First, then, the quibbles. I am happy to see him endorsing my frequent tactic of asking not how to explain x but rather asking how to explain why we believe in x in the first place, but I think that this is a procrustean bed on which to stretch my concept of intentional sys-

tems. In [Dennett \(1971\)](#) I was indeed offering an account of intentionality that was *demoting*, in that intentionality was not seen as a feature that sundered the universe into the mental and physical (as Brentano and others had claimed), but I don’t like to think of it as dismissing intentionality as a real phenomenon—though of course many have interpreted me that way. [Dennett \(1991\)](#) tried to correct that misconstrual, showing that the phenomena of intentionality are real in their own way—any beings that don’t discover these patterns are missing something important in the world. That aside, I

love the use he makes of Hume on miracles to introduce his treatment of our minds as witnesses, just not very good witnesses; their testimony can be explained in ways that do not grant the truth of some of their most cherished claims. As he puts it, the assumption of phenomenal consciousness “is deeply misleading because it makes us look for the wrong things, namely, the objects our judgments are about, rather than the causes of these judgments, which are nothing like these objects” (Baßler [this collection](#), p. 2).

My other quibble is a similar elision I want to resist. He says: “Large parts of *Breaking the Spell* are dedicated to making understandable how ‘belief in belief’ could have evolved over the centuries, beginning long before the appearance of any religion” (Baßler [this collection](#), p. 4). This misidentifies higher order belief, beliefs about beliefs, with belief in belief. The former did indeed evolve gradually over the eons, and I find Baßler’s “just so story” about this gradual process enticing indeed, and will have more to say about it below, but belief in belief is a much younger (and almost always pernicious) phenomenon, which involves the deeply confused judgment that it is morally obligatory to try to get yourself to believe traditional nonsense when you know better. “If you don’t believe in God, you are immoral. Therefore you must strive to believe in God. Belief in God is a good thing to inculcate in our children and in ourselves.” Belief in belief didn’t arrive on the human scene until the proto-religions (which originally had no need for the concept) hit upon this obligation as a way of protecting their hegemony against the lures of competing dogmas. Some proto-religions were blithely ecumenical, adopting the gods and demons of their neighbors’ creeds as just another bit of lore about the big wide world, but this credulity could not long stand in the face of market competition and growing common knowledge about the objective world. Since many—probably most—people in the world now see through at least most of the nonsense, their persistent belief in belief is now a deplorable anachronism, a systematic source of hypocrisy. (A delightful cartoon in a recent *New Yorker* perfectly cap-

tures this folly. Two armies confront each other, flying identical banners; one mounted warrior says “There can be no peace until they renounce their Rabbit God and accept our Duck God.”)

As I say, these are quibbles I have to get off my chest. Now to Baßler’s substantive proposals. He organizes his commentary around five questions he says I haven’t properly asked, and he has answers to all of them. He’s right that these are gaps in my account. (1) Why do we need to monitor our dispositions? (2) How is self-monitoring accomplished? (3) How did this self-monitoring evolve in a *gradual* fashion? (4) Why do we misidentify our dispositions? (5) Why are we so attached to the idea of qualia?

His answers are constructed by taking on, for the sake of argument, my Intentional Systems Theory, and he gets it right, in all regards. Intentional Systems Theory (IST) presupposes, tactically, that any entity treated as an intentional system “is optimally designed to achieve certain goals. If there are divergences from the optimal path, one can, in a lot of cases, correct for this by introducing abstract entities or false beliefs.” IST is, as I say, a competence model that leaves implementation or performance questions unaddressed.¹

Then comes Baßler’s major novelty: the idea of an intermediate competence between mere first-order intentional systems—which have no beliefs about beliefs (their own or others’)—and full-fledged second-or-higher-order intentional systems—which can iterate the belief context. Such entities he calls (what else?) 1.5th order intentional systems (shades of [David Marr’s 1982](#) two-and-a-half-D sketch!). This is proposed to answer his first and second questions with a plausible and in principle testable evolutionary hypothesis. A system with only 1.5th order intentionality “is able to ascribe desires only in a very particular and concrete

¹ In this regard it is strikingly similar to the free energy principle as presented by [Hohwy \(this collection\)](#); both use the assumption of biofunctional optimizing as an interpretive lever to make sense of the myriad complexities of the brain, assigning to the brain a fundamental task of acquiring accurate anticipations of the relevant causes in the organism’s world. I have not yet been able to assess the costs and benefits of these two different ways of thinking of brains as future-producer: both are abstract, both court triviality if misused. This is a good topic for future work.

manner, i.e., actions that the object in question wants to perform with certain particular existing objects, that the system itself [the ascriber] knows about” (Baßler [this collection](#), p. 6). He is wise to choose basic desires (for food, mating opportunities, safety, . . .) as the intentional states ascribed in this precursor mentality, since they are so readily “observable” in the immediate behavior of the object, giving our pioneer mind-reader a quick confirmation that it’s on the right track, a small, gradual step for a Bayesian brain.

Now what selection pressures would favor such systems evolving gradually from mere first-order systems? To the primitive first-order systems, “the behavior of their conspecifics is unexplained noise to them.” But then they make some simple discoveries. When they see an apple tree, they approach it, *and so do their conspecifics*. If they see a predator, they run, as do their kin. “One might indeed say that the desires of the agents are projected onto the world”, Baßler says. Then, in a very substantive footnote that I wish were in the text—his footnotes contain much of value, and should not be passed over!—he adds: “What I mean by ‘project’ is that instead of positing an inner representation . . . whose function is a desire, along with correct beliefs about the current situation, what is posited is an eat-provocative property of the apple itself. Both theoretical strategies allow for the prediction of the same behavior. The crucial difference is that attributing new properties to objects that are already part of the model is a simpler way of extending the model than positing a complex system of internal states to each agent” (Baßler [this collection](#), p. 7, footnote 9). This answers question (3).

He then imagines, plausibly, that these 1.5th-order systems will evolve a system of communication, but this (as I and others have argued) necessarily involves hiding information from others, which involves having an internal cache of self-monitored knowledge one can choose to divulge or not, depending on circumstances. And this in turn—Baßler’s next major innovation—leads them to become “Agents [who] believe in the existence of a special kind of properties: they believe that they approach

apples because they are sweet, cuddle babies because they are cute, laugh about jokes because they are funny.” This primitive concept of causation serves them well, of course, and is just the sort of simplification to expect in a Bayesian brain, answering question (4).

Now for the icing on the cake, Baßler’s answer to question (5) about why we care about qualia. As he notes, “It is not obvious why we do not react as disinterestedly to their denial as we did to the revelation that there is no ether” (Baßler [this collection](#), p. 5). Here is his explanation: science comes along and starts to dismantle the handy manifest image, with all its Gibsonian affordances, and for those creatures capable of understanding science, a new problem arises: something is being taken away from them! All those delectable properties (and the abhorrent properties as well, of course). Philosophers “still see that there is something missing, and since cuteness is not a property of the outside world, they conclude that it must be a property of the agents themselves” (Baßler [this collection](#), p. 8). “We have the zombic hunch because it seems to us that there is something missing and it seems so because our generative models are built on the assumption that there are properties of things out there in the world to which systems like us react in certain ways. . . . We dismiss robots because we know they can only react to measurable properties, which do not seem to us to be the direct cause of our behavior” (*ibid.*).

This rings true to me, and I hadn’t seen this way of accounting for the persistence of the zombic hunch. Baßler proposes that “the reason we, intuitively, do not accept a robot as a subject like ourselves is because we know how the robot does it; we know that it calculates, maybe even in a PP manner—we know that it does not react directly to the properties that seem to exist and that seem to count” ([this collection](#), p. 9). He goes on to list five further features his account provides for. The properties we delusionally persist in “projecting” as qualia are (1) “‘given directly’ to a person”, (2) “irreducible to physical, mechanical phenomena”, (3) “atomic, unstructured”, (4) “important to our lives/beings as humans/persons”, and (5)

“known to every living human being; it is not possible to sincerely deny their existence” (Baßler [this collection](#), p. 9). I particularly like the way that his account explains why (4) is a feature: “These properties seem to be the causes of all our behavior: if one did not feel the painfulness of a pain, one would not scream; if one did not sense the funniness of a joke, one would not laugh, etc. Since the model is still needed for interacting with others, despite theoretical advances in the sciences this felt importance of qualia to our lives is very difficult to overcome” (Baßler [this collection](#), p. 9).

I see that my response consists in large measure of approving quotations from Baßler’s commentary! But that is as it must be; I want to confirm in detail and acknowledge the nice way his proposals dovetail with my account, expanding it into new territory, and helping me see what I have so far only dimly appreciated: just how valuable the new Bayesian insights are.

But let me end with a friendly amendment of my own. Baßler’s interpretation of my view is at one point a simplification, probably just for gracefulness of exposition, and perhaps meant itself as a friendly amendment, but I want to issue a caveat. Baßler takes me to be saying that, for such properties as cuteness and color, “we misidentify dispositions of the organism with properties of another object” ([this collection](#), p. 3) and goes on to have me holding that “This means, under a personal level description, that we believe that there are properties independent of the observer, such as the cuteness of babies, the sweetness of apples, or the blueness of the sky” (*ibid.*, p. 4). I want to put this slightly differently. It is not that there is nothing objective about babies that makes them cute (or of the sky that makes it blue) but just that these objective, observer-independent properties are themselves curiously dispositional: they are, as he notes at one point, what I have called “lovely” properties. They can only be *defined* relative to a target species of observers, such as normally sighted—not “color-blind”—human beings, as contrasted with tetrachromats such as pigeons, for instance. But their existence as properties is trivially objective and observer-independent. Thus rubies were *red* before color

vision evolved on this planet in the sense that if a time machine could take normal human beings back to the early earth, they would find rubies to be red. And some strata exposed by primordial earthquake faults would have been *visible*, to some kinds of eyes and not to others. Probably dinosaur babies were cute, since, as John Horner (1998) has argued, evidence strongly suggests that they were altricial, requiring considerable parental attention, and having the foreshortened skull and facial structure of prototypically cute juvenile animals, including birds. The science-endorsed properties, both external and internal, are so hugely different from what the manifest image makes them out to be, that it is a pickwickian stretch to say that science has discovered “what cuteness is” or “what color is,” but it is also deeply misleading to say that science has discovered that nothing is cute, or colored, after all. And so in a similar vein, I have to contend with how to occupy the awkward middle ground between denying that there are qualia at all, or saying that qualia are something real, but something utterly unlike what most people *think* (and philosophers *say*) qualia are.

1 Conclusion

Baßler has provided me with a plausible and testable extension of my Intentional System Theory with his innovation of a 1.5th-order intentional system, showing in outline how higher-order intentional systems might evolve from their more primitive ancestors. And he has also shown new ways of explaining a point that many people just cannot get their heads around. As my former student Ivan Fox (1989) once put it, “Thrown into a causal gap, a quale will simply fall through it.” See also Fox’s essay, “[Our Knowledge of the Internal World](#)” (1994) and [my commentary](#) on it (1994), which I discovered, on rereading just now, to be groping towards some of the points in Baßler’s commentary. I challenged Ivan Fox to “push further into the engineering and not just revel in the specs” (Dennett 1994, p. 510), and Baßler has done just that.

References

- Baßler, D. H. (2015). Qualia explained away: A commentary on Daniel C. Dennett. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, 68 (4), 87-106. [10.2307/2025382](https://doi.org/10.2307/2025382)
- (1991). Real patterns. *The Journal of Philosophy*, 88 (1), 27-51. [10.2307/2027085](https://doi.org/10.2307/2027085)
- (1994). Get real. *Philosophical Topics*, 22 (1 & 2), 59-106.
- Fox, I. (1994). Our knowledge of the internal world. *Philosophical Topics*, 22 (1 & 2), 59-106.
- Hohwy, J. (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Horner, J. R., Gorman, J., Henderson, D. & Blumer, T. L. (1998). *Maia: A dinosaur grows up*. Bozeman, MT: Museum of the Rockies, Montana State University.
- Marr, D. (1982). *Vision*. New York, NY: Freeman.