

5.2 REFLECTIONS ON DAVID ROSENTHAL

Daniel C. Dennett

This masterful essay, a model of clarity, objectivity, and constructive thinking, exposes a tension in my discussions of *seeming*. The preamble section sets out the all too familiar conflict between first-person and third-person views in ways that were illuminating to me—someone who has been thinking about these very points for half a century and more. In particular, it provides the background for Rosenthal’s explanation of how I have used the intentional stance to build an account of consciousness from the third-person perspective that attempts to do justice to the richness of our first-person intuitions while at the same time showing why our intuition that we have incorrigible access to the contents of our minds is compelling but mistaken. He points out that whereas I have echoed Nagel’s insistence that there is no distinction in consciousness between seeming and reality, my reasoning is not Nagel’s: where Nagel indulges in mystery, I attempt to account for this with first-person operationalism (FPO), which has a decidedly third-person, naturalistic basis. In my account, the vaunted incorrigibility of the first-person point of view turns out to be (trivially) *constituted by* our convictions about how it seems, in much the same way that a center of gravity is incorrigibly located in any object at the point which, if supported, would permit the object to remain in equilibrium in any position. Don’t ask how we know this remarkable juxtaposition; it’s in effect true by definition. “We aren’t wrong about our conscious experiences because when it comes to conscious experiences there is nothing beyond our beliefs to be wrong about” (p. 144).

Rosenthal’s reconstruction of my argument about Orwellian and Stalinesque phenomena, and the question of whether there is always a fact of the matter is deft and fair, and it leads him to an observation about the intentional stance that has gone all but unmarked: the

intentional stance doesn't in itself distinguish between conscious beliefs and unconscious beliefs:

Your observations can reveal the states I'm objectively in, whereas my beliefs about what conscious states I'm in is the last word only about how my subjective stream of consciousness appears to me. How my stream of consciousness appears to me is itself objective; it appears to me in one way and not another, and my beliefs about that are the last word about how it appears to me. But that objective matter is different from the objective matter your observations of real patterns reveal, namely, beliefs and desires I have independently of how my mental life appears to me. (p. 148)

How, then, can I “emphatically reject” (as I did) a distinction between how it seems to me and how it seems to seem to me? Rosenthal has in fact uncovered a large and embarrassing contradiction in my 1991 position, which I unwittingly papered over in my campaign to shake people's allegiance to *qualia*. Unpacking this confusion of mine yields some nice surprises—at least to me.

First, let me grant Rosenthal his central point: my denial that there is a distinction between seeming and seeming to seem is belied by phenomena such as “subliminal perception and other nonconscious mental states” (p. 151). Among such phenomena are standard cases of self-deception, in which, for instance, I sure seem to the onlookers to distrust my neighbor while seeming to myself, as I avow sincerely, to trust him with my life. Others are less common cases of more symmetrical confusion. In fact, I presented a particularly vivid case of this phenomenon in *Consciousness Explained* (1991; hereafter, CE) without drawing attention to—without noticing—this conflict: my misadventure as a baseball umpire calling a crucial play at first base (p. 248). How could I not see that this was a case of seeming to seem, fitting Rosenthal's analysis perfectly?

It has taken me some considerable reflection to solve this mystery of self-interpretation. My blindness grew out of my campaign to deny a similar but logically independent presupposition that is also common among the lovers of *qualia*: the compelling conviction that *really seeming* to see a purple cow somehow involves purple *qualia* as *real* properties of my experience, in contrast with, say, a robot seeming to see a red, white and blue American flag when confronted with a complementary color afterimage. A robot equipped with a good model of human color vision might superficially seem to see a red white and blue flag, according to this popular view, but this wouldn't be real seeming, but at best some “merely behavioristic” kind of seeming.

Rosenthal's version of the distinction has nothing to do with *qualia*. It is about how there can be a conflict between what I seem (to other interpreters) to believe,

etc., and what I seem (to myself) to believe; when such a conflict arises, I can seem to judge that the runner is out (just look at my vigorous hand signal) while also seeming (or seeming to seem) to judge the runner to be safe (just listen to what I say). It is sometimes said that this sense of “seem” is a mere indicator of epistemic modesty (“Is that a deer running through the woods, or does it just seem to me to be one?”), not the assertion about any *real* seeming (“Oh wow, I seem to see an undulating checkerboard with Day-Glo green and silver squares!”). It is this idea of real seeming over and above the non-committal merely epistemic seeming that I still wish to deny. (See Andy Clark’s chapter, 7.1, and my Reflections on him, 7.2, for some new slants.)

In short, I am not recanting my dismissal of qualia, as philosophers consider them, but I am taking on board Rosenthal’s claim that there are plenty of cases of where the third-person use of the intentional stance conflicts with the first-person use. Indeed, it is in order to handle these possibilities gracefully that heterophenomenology declares subjects to be logically incorrigible about what they *seem* to experience. That is FPO in action, letting subjects constitute their heterophenomenological worlds by fiat in their considered judgments about what it is like to be them, a preliminary tactic that fixes *what is in need of explanation* while leaving wide open the prospect of discovering that subjects are, in spite of the confidence of their convictions, wrong about what is really going on in them. Thus Roger Shepard and Zenon Pylyshyn can agree that their subjects seem to be rotating images—that’s what they all insist on, if you ask them—while disagreeing about what is really going on inside them. (For an early version of my FPO applied to this controversy, see my “Two Approaches to Mental Images” in *Brainstorms*, 1978, where I distinguished the “ β -manifold” of intentional objects from the underlying machinery.)

But now, what of Rosenthal’s further conclusion, that my FPO fails in its purpose? “Dennett’s account does not, after all, provide a successful path from the third-person aspect of mind to its first-person aspect” (p. 153). I am not persuaded of this, since I think I can happily acknowledge that the distinction Rosenthal draws between seeming and seeming to seem is honored by my account after all, but doing so requires me to remove another bit of papering over, which (it seems to me) is also a problem for Rosenthal’s “close cousin” to my view, the latest version of his famous HOT (Higher Order Thought) theory of consciousness: “what it is for a mental state to be conscious consists in one’s being aware of oneself, in the right way, as being in that state; a state is conscious if one has a suitable high-order awareness that one is in that state” (p. 154).

I have always been drawn to some aspects of HOT theory, as I explained in *CE*. In particular, HOT theory provides a sensitive account of one very tricky feature of human consciousness: reportability in language. There is a big difference

between *expressing* a mental state and *reporting* a mental state; the poker player who can't manage to maintain a good—uninformative—poker face involuntarily exhibits facial expressions, hand gestures, postures nicely known as “tells” in the world of poker. Discerning the tells of the other players as tantamount to “reading” their minds, a tremendous advantage to the observant player. In spite of the words “tell” and “read,” this is not verbal communication; tells are instances of negligent self-exposure, not intended speech acts. Tells express mental states without reporting mental states. But when a person does *report* a mental state by *saying* (in words, or via prearranged button press) “the circle on the left seems to be larger,” this intentional action ipso facto *expresses* a mental state, the higher-order belief or thought that the circle on the left seems to be larger, a claim *about* a mental state.

This is all very well when we are considering the heterophenomenology of human consciousness, but what about animal consciousness? In spite of the ubiquitous but misleading practice of psychologists who let themselves speak of their animal subjects “telling” them which state they are in by performing one highly trained action or another (usually an eye blink or button press) this *misattributes* a communicative intention to the subject. Monkeys and rats are not trained to *communicate* with the experimenters (the way human introspectors were trained by the various early schools of introspectionism); they are trained to *inform* the experimenters in much the way a poker player might be patiently and subliminally encouraged by another player to exhibit tells. The animal subjects, unlike human subjects in heterophenomenological experiments, do not have to know what they are doing. But then it follows that the states of mind revealed by such tells are importantly *unlike* the states of mind we humans spend so much time and ingenuity conferring about.

I have long stressed the fact that human consciousness is vastly different from the consciousness of any other species, such as apes, dolphins, and dogs, and this “human exceptionalism” has been met with little favor by my fellow consciousness theorists. Yes, of course, human beings, thanks to language, can do all sorts of things with their consciousness that their language-less cousin species cannot, but still, goes the common complaint, I have pushed my claims into extreme versions that are objectionable, and even offensive. Not wanting to stir up more resistance than necessary to my view, I have on occasion strategically soft-pedaled my claims, allowing animals to be heterophenomenological subjects (of sorts) thanks to their capacity to *inform* experimenters (if not *tell* them), but now, my thinking clarified by Rosenthal's, I want to recant that boundary blurring and re-emphasize the differences, which I think Rosenthal may underestimate as well. “Thoughts are expressible in speech,” he writes (p. 155), but what about the higher-order thoughts of conscious animals? Are they? They are not *expressed* in speech, and

I submit that it is a kind of wishful thinking to fill the minds of our dogs with thoughts of that sophistication. So I express my gratitude to Rosenthal for his clarifying account by paying him back with a challenge: how would he establish that non-speaking animals have higher-order thoughts worthy of the name? Or does he agree with me that the *anchoring* concept of consciousness, human consciousness, is hugely richer than animal consciousness on just this dimension?

Rosenthal says his view is close to mine, since it accommodates my idea that the consciousness of mental states consists in what is given “to be taken in one way rather than another.” But he denies

the operationalist way that FPO spells out what it is for the given to be taken. Being taken is not a matter of ‘writing it down’ in memory or ‘of the effects of various vehicles of content on subsequent action (and hence, of course, on memory)’ (1991a, p. 132) Rather, it is simply a matter of one’s being aware of oneself as being in the mental state in question. (p. 154)

Simply? And then what happens? Or: and then what *may* happen? The agent who is thus aware of being in that mental state is, in virtue of that awareness, able to hinge almost any action on that mental state. (Debnar & Jacoby, 1994. For discussion see Dennett 2017, *BBB*) But do we have any grounds for generalizing the experimental animal’s telltale blink or button-press to a more general ability to *use what it knows about what mental state it is in*? It is a bit of a stretch to call the animal’s response the expression of a *belief*. If we want to secure some form of higher-order thought, we will have to find many ways of assaying the versatility of this ability. To date, the only research I know that makes an inroad on this issue are the metacognition experiments that apparently show animals can use an assessment of their confidence to choose between a risky high-payoff task and an easy low-payoff task (e.g., Shields, Smith, Guttmanova, & Washburn, 2005).

Let me respond, finally, to two footnotes in Rosenthal’s essay. A casual reading of footnote 8 might conclude that I coined “intuition pump” in (1991), in *CE*, but in fact, it was much earlier, in my 1980 comment on Searle.

And footnote 9:

The idea that competition among many nonmental factors results in some folk-psychological conscious states may encourage another claim of Dennett’s, that the search for a neural correlate of consciousness “is probably a wild goose chase” because “the processes that elevate contents to consciousness are like the processes that elevate [evolutionary] lineage divergences into speciation events” (2009, p. 234). But the analogy with evolution is questionable; the competing factors *may well* [my

italics—DCD] result in states we can taxonomize along standard folk-psychological lines, allowing for a subsequent relatively specific process that results in those states' becoming conscious.

I think this is highly unlikely, but I'll grant it is *possible*; my aim in making the analogy to speciation events was to undermine the opposite modal assumption: that, if consciousness is an entirely physical phenomenon, there *must be* neural correlates of consciousness. Not so. Speciation is, I submit, an entirely unmysterious physical phenomenon which eventually yields striking manifestations—lions and tigers and bears. Oh my—but the time and place of its onset can only be *retrospectively* and roughly estimated; there is no Biological Correlate of Speciation. (If, for instance, you attempt to pin down the necessary and sufficient conditions for speciation by some (relatively arbitrary) proportion of accumulated change in “the” genome of a species, making cross-fertility “impossible,” you will face the prospect that any species slowly going extinct must end its days in an pitiful explosion of tiny speciation events, as shrinking and isolated gene pools briefly include a majority of non-cross-fertile genomes. You cannot “see” speciation in a moment; *as with consciousness*, you always need to ask the question: And then what happens?

Works Cited

- Dennett, D. C. (1978). Two approaches to mental images. In *Brainstorms: Philosophical essays on mind and psychology* (pp. 174–189). Cambridge, MA: MIT Press.
- Dennett, D. C. (1980). The milk of human intentionality. *Behavioral and Brain Sciences*, 3, 428–430.
- Dennett, D. C. (1991). *Consciousness explained*. New York, NY: Little, Brown.
- Dennett, D. C. (2017). *From bacteria to Bach and back*. New York, NY: W. W. Norton.
- Debner J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 304–317.
- Shields, W. E., Smith, J. D., Guttmanova, K., & Washburn, D.A. (2005). Confidence judgments by humans and rhesus monkeys. *Journal of General Psychology*, 132, 165–186.

THE PHILOSOPHY OF DANIEL DENNETT

Edited by Bryce Huebner

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

CIP data is on file at the Library of Congress
ISBN 978-0-19-936751-1

9 8 7 6 5 4 3 2 1

Printed by Sheridan Books, Inc., United States of America