

TEN NEGLECTED CLASSICS
OF PHILOSOPHY

Edited by Eric Schliesser

OXFORD
UNIVERSITY PRESS

Jonathan Bennett's *Rationality*

DANIEL DENNETT

Rationality (1964), Jonathan Bennett's first book, was published when he was thirty-four years old, and it exhibits the intensity of a young philosopher who is quite sure he sees a way to cut through a forest of dubious, ideology-ridden, squishy philosophy of mind and set a few things straight. Since the terrain he was scouting and clarifying was the same terrain that I was then embarking on, with similar ambitions, I read his book when it first came out through the dust and smoke of my own earliest efforts to conquer these topics. The result, I confess, was that I simultaneously misread, underestimated, and covertly absorbed much of what he was on about and then proceeded to reinvent some of his wheels in my own work without realizing it until years later. First let me lay out what Bennett took himself to have done in *Rationality*, say why that was such a good idea, and then go on to consider why, nevertheless, his book has not had the influence it might have had. I was not the only explorer of this territory who chose to find other paths to Bennett's destination, but here we all are, and a review of his book may consolidate the gains.

How many philosophers would tell their readers on page 1 that they were about to encounter the results of an "extremely ambitious conceptual inquiry"? Why does Bennett say that? For several reasons, I think. First, he sees his analysis of rationality to be redoing a large part of Kant's project in the *Critique of Pure Reason*—but doing it *right*. Kant thought he was discovering synthetic a priori truths, whereas Bennett sees that he is doing *conceptual analysis*, an altogether different undertaking.

Second—and I can only surmise that this may have been a conscious intention on Bennett's part—he wanted to alert readers to the possibility that his brand of conceptual analysis was not the then fashionable brand—ordinary language philosophy—but something much better, harder, more valuable. Bennett's book appeared in *Studies in Philosophical Psychology*, an unduly prestigious series of little red books with gray-green dust jackets and red lettering that was edited by R. F. Holland and published by Routledge and Kegan Paul. The series' title was something of a misnomer. Although all the monographs in it dealt with some aspect of the mind and, hence, at least tangentially with some aspect of psychology, only a few grappled seriously with any empirical theories or experiments drawn from the field or the lab, and most were explorations of the putative implications of interpretations of the meanings of ordinary language expressions about minds. Ordinary language philosophy—aside from the handful of works of genius (by Wittgenstein, Ryle, and Austin) that launched the movement—was in general extremely *unambitious*, sometimes even comically modest, a lot of precious and informal fussbudgeting about the nuances of meaning that could be eked out of considerations of "what we would say" when confronted with one everyday psychological phenomenon or another. The series might better have been called *Philosophical Studies in Folk Psychology*, but that term was

not yet in use. A few of the volumes achieved a wider influence or at least notoriety—Malcolm’s *Dreaming* and Winch’s *The Idea of a Social Science*, in particular—but in general these essays were for ordinary language philosophers only, and with the dwindling of that school of thought already in the mid-1960s, Bennett had to contend with a problem of misperception. Yes, his book was in that famous series, but no, it was not an informal collection of observations about whether we would say of one bit of human activity or another that it was rational.

Bennett begins by considering two widespread and plausible convictions: human beings are much more intelligent, more rational, than other animals, and the key to their intellectual eminence is language. The first conviction is, he thinks, beyond dispute, and this permits him to *define* “rationality” as “whatever it is that humans possess which marks them off, in respect of intellectual capacity, sharply and importantly from all other known species” (5). The second conviction is not just plausible; it is—shall we say—*congenial* to philosophers, indeed to all thoughtful human beings, but this actually hinders our understanding of it, encouraging oversimplified dogmas, such as

Only human beings can reason.

Reasoning without language is impossible.

Animals don’t even have beliefs (they, unlike us, are mere stimulus-response organisms).

We need to know *how* and *why* “human talk” enables us to be so much smarter than the rest of the living world then, and here Bennett sees the task as constructing, bottom up, the path from *mere* animal intelligence to our kind of brilliance by a series of explicitly described and defended steps, starting with a conveniently humble example, the famously informative waggle dance of the honeybee.

He contrasts his strategy with the more familiar top-down practice of philosophers:

One might set about answering this by supposing human talk to be other than it is in various ways, in order to see which suppositions did and which did not rob it of its right to be accounted rational. The course I shall adopt, however, is the reverse of this. I shall start with a kind of behavior which is not rational, and shall suppose it to be different in various ways in order to see which suppositions do and which do not confer on it a right to be accounted rational. (6)

This anticipates Valentino Braitenberg's "law of uphill analysis and downhill synthesis" (in his brilliant little book *Vehicles: Experiments in Synthetic Psychology* [1984]), according to which it is easier to synthesize well-understood simple elements or mechanisms into a larger entity, and predict and explain its behavior than it is to try to reverse-engineer a complex entity to see how its parts contribute to its talents.

Bennett sees three advantages to this strategy, the first of which is pure Braitenberg: "greater control over our material. It is just easier, confronted with indubitably non-rational behavior, to know where to start adding" (6). Second, there is less risk of being distracted by features of human talk that are irrelevant but are socially or emotionally important to us. (Bennett is an acute critic of the *amour propre* that can distort philosophical analysis of such topics.) Third, "we shall avoid the temptation to take the question 'what is it for a being to be rational?' in the form 'in what does *my* rationality consist?' and to try to answer it by 'introspective thought experiment'" (7)—which had been the philosophical tradition since Descartes (and including Locke, Hume, and Kant).

So although he doesn't explicitly put it this way, he sees and endorses the virtue of conducting a resolutely third-person, not

first-person, analysis of rationality. These three advantages make for what he calls “antiseptic virtue” (7). Indeed they do. The emotional and social implications of discussions of rationality are particularly hard to keep in check—witness the inflationary tendencies of the Pittsburgh school of “normativity”—and it is best, as Bennett sees, to launder them out at the beginning, so that they can be let back in only when they are called for. These are hot-button issues, and it is all too easy to be protective, when a more balanced inquisitiveness would be better. The strategy also minimizes the temptation to be taken in by the unavoidable practice of identification-by-content that is the hallmark of introspective methods. This is what I mean: if you want to talk about your own mental states, you *must* identify them by their content: “which idea? My idea of HORSE. Which sensation? My sensation of *white*.” How else? There is no way you can identify your own mental states “from the inside” as, for instance, *concept J47* or *color-sensation 294*. By taking for granted the content of your own mental states, by picking them out *by their content*, you sweep under the rug all the problems of indeterminacy or vagueness of content. Reading your own mind is too easy; reading the mind of a honeybee puts all the problems front and center.

So he starts with Karl von Frisch’s famous work (winning a Nobel Prize in 1973) on honeybees—a wise choice, since von Frisch pioneered the method of patient scientific analysis that most closely resembles Bennett’s method of philosophical analysis: cautiously taking on board the everyday language of mind, so that he can speak of what the bees *know* and *learn*, of what their dances *mean*, while scrupulously and systematically cashing out these *façons de parler* with good, hard evidence couched in terms of behavior and its observable effects. Bennett commends Frisch’s popular book *Bees: Their Vision, Chemical Senses, and Language*, first published in 1950 by Cornell University Press. Bennett makes no claims to be up on the technical literature, which he doesn’t need, since von

Frisch's bees are just a convenient starting point, soon to be abandoned in favor of science-fictional mutants that would, Bennett argues, gradually close the gap between the "language" of the bees and human talk; for example, "there could be rational bees . . . and I wish to know what we are saying when we deny that actual honey-bees are of this kind" (11).

Philosophers get largely ignored by Bennett, aside from brief passing acknowledgments of Kant, Wittgenstein, Descartes, Locke, Hume, and Popper, but it is clear that he is deliberately avoiding tussles with a host of contemporary philosophers, indirectly criticizing their positions, and occasionally exhibiting some lessons learned, from Ryle, for instance:

I take "This behavior manifests rationality" and cognate sentences to express a non-relational claim about the behavior itself, and not a relational claim about what causes the behavior, or underlies it, or is projected from the secret soul into the public world by it. I do not deny that there are private mental states, and philosophical problems about them. I merely restrict my attention to the criteria which underlie our everyday belief that human beings indulge in reasoning processes while honey-bees and earthworms do not, or—what may be the same belief—that humans are rational while honey-bees and earthworms are not. These criteria are plainly behavioural in nature, and involve no Cartesian speculations. (10)

Having constructed his base camp, he proceeds to look closely at the bees. Their waggle dances clearly have the function of transmitting information about the location of food from scout bees to others, and von Frisch has worked out a well-confirmed sketch of the system implemented. But do the bees understand the system? Might the bee dances be more like shivering when cold or

frightened (which can be informative to any observer without being an act of communication)? The difference appears to be that the waggle dance is in some sense conventional, not a purely physiological symptom of inner state. There are problems, however.

The trouble is that the idea of a conventional association essentially involves the idea of intentions and reasons: to describe as “conventional” the correlations between food discoveries and subsequent dances is to say that the bees have certain reasons. (14)

Bennett goes on:

Now, I shall argue that bees do not in fact have reasons for their behaviour, and thus that their dances are not literally symbolic, and thus that the dances do not constitute language. (15)

Here I think the main strength of Bennett’s method is also its weakness. As a piece of *conceptual analysis* it is hard to fault the claim that *conventionality* presupposes *reasons had by reasoners*, but this blinds him to the prospect of an intermediate position which is not readily articulated without a helping of initially counterintuitive innovation—a “move” that I have only recently perfected (to my satisfaction): drawing a quite sharp distinction between *doing things for reasons* and *having reasons for doing things*. There are reasons aplenty in nature (I call them free-floating rationales): trees do things for reasons, fungi do things for reasons, bacteria do things for reasons, and we human beings do things—sneeze, shiver, cough—for reasons, but they do not (usually) *have* reasons for doing these things. Only reasoners *have* reasons for doing (some of) the things they do. Bees aren’t reasoners, as Bennett demonstrates quite forcefully, but that does not stop them from doing things (unwittingly)

for reasons. Reasons, in these cases, are not just causes; they are *raisons d'être*, uncovered and endorsed by natural selection (Dennett, forthcoming a, b). Shivering when cold has both a complex physiological cause and a reason: it helps maintain body temperature. The complex physiological causes controlling the bee dance are one thing; the rationale uncovered and promoted by natural selection for the bee dance is communication.

Consider a similar phenomenon, involving communication between members of different species. You have probably seen video of gazelles being chased across the plain by a predator and noticed that some of the gazelles are leaping high into the air during their attempts to escape their pursuer. This is called stotting (or sometimes pronking). Why do gazelles stot? It is clearly beneficial, because gazelles that stot seldom get caught and eaten. This is a regularity that can baffle. No account of the actions of all the proteins in all the cells of all the gazelles and predators chasing them could reveal why this regularity exists. For that we need the branch of evolutionary theory known as costly signaling theory (Zahavi 1987; Fitzgibbon and Fanshawe 1988). The strongest and fastest of the gazelles stot in order to advertise their fitness to the pursuer, signaling in effect "don't bother chasing me; I'm too hard to catch; concentrate on one of my cousins who isn't able to stot—a much easier meal!" and the pursuer takes this to be an honest, hard-to-fake signal and ignores the stotter. This is the free-floating rationale and *need not be appreciated by either gazelle or lion*. That is, the gazelle may be entirely oblivious to why it is a good idea to stot if you can, and the lion may not understand why it finds stotting gazelles relatively unattractive prey, but if the signaling wasn't honest, costly signaling, it couldn't persist in the evolutionary arms race between predator and prey. (If evolution tried using a "cheap" signal, like tail flicking, which every gazelle, no matter how frail or lame, could send, it wouldn't pay for lions to pay attention to

it, so they wouldn't.) These explanations in terms of free-floating rationales are not reducible to explanations at lower levels, such as the molecular level, but it is also important to recognize that even though the explanation of why and how stotting works is from the intentional stance (in terms of what the lion is rational to conclude from the stotting of the gazelle), the individual lion or gazelle need not understand the meaning of stotting for it to work.

By not quite nailing the distinction between there being reasons and an agent's having reasons, Bennett misses the intermediate cases that naturalism requires, in order to account for the gradual emergence of function by the march of evolution and by the development of the individual. For note that Bennett's hard line on conventionality would rob young children of the practice of communicating. Children perform thousands of speech acts (or proto-speech acts if you insist) before they show clear evidence of *having* reasons for what they say.

Still, Bennett's all-or-nothing approach has its virtues, since it permits him to imagine a series of distinct steps in the direction of turning the bees' behavior into properly rational behavior, and the first step comes within a whisker of making the missing distinction: "we can say of honey bees that their dancing behaviour is covered by rules, but not that honey-bees have rules according to which they dance" (15). "Covered" by rules. He goes on to articulate the difference: "Apian dances are regular; human talk is rule-guided." He explicitly avoids the mistake of thinking that any behavior that has a complex explanation in terms of physical causes cannot also have "mental predicates" applied to it and sees the first step towards bees *having rules* (if not yet reasons): "For a creature to be correctly said to have a rule, it is necessary that it should be able to break the rule" (17).

How could this behavior be observed and confirmed? Long before Brandom and Haugeland and the Pittsburgh school of

normativity drew attention to the importance of the role of *criticism*, Bennett sketched it out crisply:

A necessary condition for this is that there should be a recognizable kind of performance which a bee goes through if and only if it has just observed a dance, or a post-dance foraging flight, which it knows to be in breach of the rules. (18)

As his use of the verb “knows” demonstrates, Bennett is now moving firmly away from the puritanical strictures of behavioristic language and acknowledging that he is going for a “structural” account that is both objective and naturalistic, on the one hand, and mentalistic on the other. He calls the behavior denial behavior—acknowledging that it is not fully fledged denial of the sort a human being can engage in, but it is on that path, a ‘move from descriptions which are rules to descriptions which refer to rules.’”

But there are more steps to come. In turn he introduces and justifies

1. past-tense and future-tense dance types
2. allowing the bees to talk of danger as well as food.
3. minimal compositionality (all the future talk shares a feature and all the food talk shares a feature, etc.), so that if you know the rules you can mix and match to compose dances that mean *there was food at location x* and *there will be danger at location y* and so forth.

These steps have the cumulative effect of getting the bees in touch with the reasons, making reasons themselves (via their “mental” representation) objects in the bees’ cognition in the same way that food locations, fellow bees, obstacles, and the like are objects of their cognition.

Denial actions have some interesting properties. Occasions in which a dance provokes denials *must be* controversial, the denials themselves subject to rebuttal or criticism! Why? Because otherwise a particularly influential bee could single-handedly destroy the whole communicative edifice by just issuing blanket denials, a weapon of mass destruction: “the entire community of bees could lose its whole stock of reasons for any claim as the result of a busy afternoon’s work by a single irresponsible bee” (68). The only way the criticism of a denial could have any punch would be if we endow the bees with something like the distinction between soundness and validity: “what we need is a kind of denial which is performed when and only when some bee ‘knows’ that a bad reason has been given for some claim” (58). And for this to be possible, that bee must have some appreciation of the fact that the valid form is *universally* valid; not just this one time but always. Now that would be one rational bee! The Sellarsian give-and-take in the space of reasons is given a derivational foundation in Bennett’s thought-experimental exercise.

Eventually, having constructed this imaginary edifice of bee prowess, he reflects on what he has done:

Returning to the main thread of the argument, we must see where rationality fits into all this. All our *prima facie* cases of rationality or intelligence were based on the observation that some creature’s behaviour was in certain dependable ways successful or appropriate or apt, relative to its presumed wants or needs. . . . With the introduction of universal and dated statements, there are canons of appropriateness whereby we can ask whether an apian act is appropriate not to that which is particular and present to the bee but rather to that which is particular and past or to that which is not particular at all but universal. That is what generalising and talking about the past have in common. (85)

Here, in 1964, Bennett has laid out most of the insights that I struggled to expose to light in my various articles on the intentional stance beginning with "Intentional Systems" (Dennett 1971). He expanded on his ideas in *Linguistic Behaviour* (1976). My 1983 target article in *Behavioral and Brain Sciences*, "Intentional Systems in Cognitive Ethology: the 'Panglossian Paradigm' Defended," provoked a lengthy and sympathetic commentary by Bennett, the first time either of us attempted to assay this convergence. His commentary, "Cognitive Ethology: Theory or Poetry?" (1983: 356ff.), chided me for abjuring the rigorous road to an explicitly articulated theory of the "conceptual structures" that must form the foundations of cognitive ethology in favor of crowd-pleasing (or scientist-pleasing) metaphors and analogies:

He [Dennett] encourages them to go on believing that the conceptual foundations of cognitive ethology are rather easy to lay—a few broad strokes of the brush, or slaps of the trowel, and there you are. Really, it is much harder and more laborious than that. I shall sketch the sort of thing that is needed, and point out some things in Dennett's paper that suffer from the lack of any proper foundations. (356)

The foundations Bennett then described actually differed from my own view only in emphasis and in a continued blindness to the utility for science of explanations that allude to free-floating rationales:

But what if every event can be explained mechanistically, that is, in terms of its subject's intrinsic properties, with no mention of any property of the form A/G? [Actor/Goal] I answer that it is all right to bring x under a teleological generalization if the latter captures a class of events that is not covered by

any one generalization of a mechanistic sort. Where there is a contest between one teleological and one mechanistic generalization (or even, perhaps, two or three of the latter), mechanism wins because it is more basic, uses concepts of wider applicability, and so on (see Taylor 1964, 29). But if a teleological generalization does work for us—giving us classifications, comparisons, contrasts, patterns of prediction that mechanism does not *easily* [my italics] provide—then that justifies us in employing it. This, I submit, is the *Grundgesetz* of the whole theory of teleological explanation and thus of the intentional stance. (356)

Looking at the example of the vervet monkey's leopard-alarm call that I had discussed, he claims that if it turns out that there is a fairly narrow range of physically different stimulus patterns that provoke it,

In that case, the generalization "Whenever it is in (what it registers as being) a leopard-threatening situation it does a leopard-avoiding thing" should be relinquished: The intentional stance has no honest work to do here, because all its work is equally done by something that is preferable to it because lower level. (Whether the S-R pattern is hard-wired or a result of learning is quite irrelevant, so far as I can see.) (356)

Using the same reasoning, Bennett would be obliged to banish the intentional stance for explaining the (existence and stability of) stotting, or so I interpret his claim here. My response to Bennett in the same issue of *Behavioral and Brain Sciences* is worth repeating almost in full, if only to provide the context for a few further admissions and corrections on my part. Not mentioning *Rationality*, I began by acknowledging that Bennett's 1976 book

is indeed full of insights that should be of interest and value to ethologists; in fact it discusses, in greater detail, virtually every topic of the target article. (Embarrassing note: Bennett and I, working entirely [*sic*] independently, arrived at a slew of similar conclusions at about the same time; it took our students and colleagues to put us in touch with each other's work a year or so ago. Now if there turns out to be someone named Cennett!)

Bennett grants that my "conclusions" are acceptable to him. Moreover, he is not claiming (so far as I can see) that his theory permits explanations, predictions, or verdicts that are inaccessible to me, given my way of doing business. Indeed, the accounts he provides in his commentary (e.g. of when and why to talk of the goal of leopard avoidance, what settles the issue of whether a high-order attribution to Tom [the vervet] is correct) are very much what I would have said, and to some extent have said on other occasions. The difference is that he claims to derive his conclusions the hard (and proper) way—from a rigorous, precise, articulated theory of conceptual structures—while I obtain the same results by what seems in contrast to be a slapdash, informal sort of thinking that I explicitly deny to be a theory in the strict sense of the term. Bertrand Russell (1919, 71) once excoriated a rival account by noting it had all the advantages of theft over honest toil; Bennett, I am grateful to say, finds a variation on this theme: I stand accused of poetry.

I plead *nolo contendere*, for it seems to me that, aside from differences in expository style and organization, Bennett and I are not just arriving at the same conclusions (for the most part); we are doing the same thing. If Bennett has a theory, it is not—had better not be, for the reasons just reviewed—a theory directly about internal processes. The sort of behavioral evidence he relies on to anchor his claims simply won't carry theory that far. So his theory is, like my instrumentalism, a

theory of “conceptual structures,” as he says. The methodological difference I see is strictly in the format of presentation, with Bennett’s theory being, like many other philosophical theories, “a system of definitions propounded and defended” (Shwayder 1965). I think the idea that there is a proper theory to be developed here is a philosophical fantasy. Getting clear about something does not always mean producing a clear theory of it—unless we mean something quite strange by “theory.” (I stand in awe of the systematic knowledge about automobiles good mechanics and automotive engineers have, but I don’t think they have or need a theory of automobiles—certainly not a theory that yields formal definitions of the main concepts of their trade.) (382)

I should not have said my work was developed “entirely” independently of Bennett’s. I definitely (mis)read *Rationality* when it came out or soon thereafter, and just look how much of my thinking about the intentional stance is prefigured there. But I am not alone in having been scooped by Bennett. Sellars’s work on the space of reasons was roughly contemporaneous with Bennett’s, but so far as I know Bennett’s ideas have not been featured as such in more recent work by the Pittsburgh school. To the sensitized twenty-first-century eye, foreshadowings of late Quine and early Millikan can also be detected in *Rationality*, but I doubt if Quine ever paid much if any attention to the little red book from Routledge and Kegan Paul. Millikan, however, did, without making all that much of it. There is a longish endnote in her *Language, Thought, and Other Biological Categories* (1984, 338) which says in part, “If performing inferences is tantamount to having reasons and having reasons marks off rational creatures from others, then I am agreeing with Bennett about *why* bees are not rational.” And in her *White Queen Psychology and Other Essays for Alice* (1993, 79),

she notes approvingly: "Bees, as Bennett (1964) has observed, are not rational."

BIBLIOGRAPHY

- Bennett, Jonathan (1964). *Rationality*. London: Routledge and Kegan Paul.
- (1976). *Linguistic Behaviour*. Cambridge: Cambridge University Press.
- (1983). Cognitive Ethology: Theory or Poetry? Commentary on Dennett 1983, 356ff. *Behavioral and Brain Sciences* 6: 356–358.
- Braitenberg, Valentino (1984). *Vehicles: Essays in Synthetic Psychology*. Cambridge, MA: MIT Press.
- Dennett, Daniel (1971). Intentional Systems. *Journal of Philosophy* 78 (February 25): 87–106.
- (1983). Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended (with commentaries). *Behavioral and Brain Sciences* 6: 343–390.
- FitzGibbon, C. D. and Fanshawe, J. H. (1988). Stotting in Thomson's Gazelles: An Honest Signal of Condition. *Behavioral Ecology and Sociobiology* 23(2) (August): 69–74.
- Frisch, Karl von (1950). *Bees: Their Vision, Chemical Senses, and Language*. Orig. pub., Ithaca, NY: Cornell University Press.
- Malcolm, Norman (1959). *Dreaming*. London: Routledge and Kegan Paul.
- Millikan, Ruth (1984). *Language, Thought, and Other Biological Categories*. MIT Press.
- (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Russell, Bertrand (1919). *Introduction to Mathematical Philosophy*. London: Routledge. Repr. 1993.
- Shwayder, David (1965). *The Stratification of Behaviour: A System of Definitions Propounded and Defended*. London: Routledge and Kegan Paul.
- Taylor, Charles (1964). *The Explanation of Behaviour*. London: Routledge and Kegan Paul.
- Winch, Peter. *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge and Kegan Paul.
- Zahavi, A. (1987). The Theory of Signal Selection and Some of Its Implications. In Delfino, V. P. (ed.), Bari, April 9–14, 1985. Bari: Adriatici Editrici, 305–327.