

18 The Friar's Fringe of Consciousness

Daniel Dennett

Ray Jackendoff's *Consciousness and the Computational Mind* (1987) was decades ahead of its time, even for his friends. Nick Humphrey, Marcel Kinsbourne, and I formed with Ray a group of four disparate thinkers about consciousness back around 1986, and, usually meeting at Ray's house, we did our best to understand each other and help each other clarify the various difficult ideas we were trying to pin down. Ray's book was one of our first topics, and while it definitely advanced our thinking on various lines, I now have to admit that we didn't see the importance of much that was expressed therein. For instance, in my *Consciousness Explained* (1991)—which was dedicated to my colleagues Nick and Marcel, and Ray—I gave only the briefest mention of the contribution of Ray's I want to explore here: the idea that we are conscious of only an *intermediate level* of all the nested, interacting levels of representation that the brain uses to accomplish its cognitive tasks.

Ray Jackendoff (1987) argues . . . that the highest levels of analysis performed by the brain, by which he means the most abstract, are *not* accessible in experience, even though they make experience possible, by making it meaningful. His analysis thus provides a useful antidote to yet another incarnation of the Cartesian Theater as the "summit" or "the tip of the iceberg." (Dennett 1991, 278)

That antidote is still much needed by thinkers about consciousness today, and since I am probably not alone in acknowledging the point while underestimating its implications, I am going to try to saddle it with a memorable image to remind us just what adjustments to our thinking it requires. I hereby dub Ray's vision the Friar's Fringe model of consciousness—like the monk's halo of hair halfway down the crown of his head, it occupies neither the Headquarters nor the Top of the hierarchy of cognitive processes. That fringe of hair may be our chief sign that we are in the presence of a friar, but the hair isn't the source of whatever makes the friar special, and the intermediate level in Ray's

model is not where the work of semantic processing occurs. Ray argues for this in two detailed chapters in his 1987 book, drawing on phenomenological observation of our experience of music, vision, and visual imagery, and language itself, of course. He also analyzes the difficulties of other theories. His claim has since been taken up by another fine theorist, Jesse Prinz (2012). The Cartesian idea, shared by Jerry Fodor, Tom Nagel, and John Searle, that consciousness is the *source* (somehow) of all Understanding and Meaning¹ is, I believe, the greatest single cause of confusion and perplexity in the study of the mind. For some (e.g., Fodor and Nagel) it fuels the conviction that a science of the mind is ultimately beyond us, an unfathomable mystery. For others (e.g., Searle) it deflects attention from the one kind of science that could actually explain how understanding happens: a computational approach that in one way or another breaks down the whole mysterious, holistic, ineffable kaleidoscope of phenomenology into processes that do the cognitive work that needs to be done.

Ray has seen that the first step toward any viable theory of consciousness must demote consciousness from its imagined position as the ultimate Inner Control Room (where it all comes together and the understanding happens), but he doesn't quite carry through on the second step, which is embodied in the moral I draw from the demise of the Cartesian Theater:

All the work done by the imagined homunculus in the Cartesian Theater must be distributed around in space and time to various lesser agencies in the brain. (Dennett 2005, 69)

All the work. And all the play, too, for that matter: the savoring, enjoying, delighting, as well as the abhorring, being disgusted by, disdain. . . . It all has to be outsourced to lesser entities, none of which is the ego, or the person, or the Subject. Just as the phenomenon of *life* is composed, ultimately, of non-living parts (proteins, lipids, amino acids, . . .) so *consciousness* must be dismantled and shown to be the effects of non-conscious mechanisms that work sub-personally. When this step is taken, the Subject vanishes, replaced by mindless bits of machinery unconsciously executing their tasks. In *Consciousness Explained*, I described what I called the Hard Question: *and then what happens?* (255). This is the question you must ask and answer after you have delivered some item “to consciousness.” If instead you stop there, “in consciousness,” you’ve burdened the Subject with the task of reacting, of doing something with the delivery, and left that project unanalyzed. Answering the

Hard Question about the *sequelae* of any arrival in consciousness “reduces” one more bit of Cartesian magic to mere legerdemain. Can this be the right direction for a theory of consciousness to take? Resistance to this step is still ubiquitous and passionate. As so often before, Jerry Fodor finds a vivid way of expressing it:

If, in short, there is a community of computers living in my head, there had also better be somebody who is in charge; and, by God, it had better be me. (Fodor 1998, 207)

Another eloquent naysayer is Voorhees:

Daniel Dennett is the Devil. . . . There is no internal witness, no central recognizer of meaning, and no self other than an abstract ‘Center of Narrative Gravity,’ which is itself nothing but a convenient fiction. . . . For Dennett, it is not a case of the Emperor having no clothes. It is rather that the clothes have no Emperor. (Voorhees 2000, 55–56)

Exactly. If you still have an Emperor in your model, you haven’t *begun* your theory of consciousness. A necessary condition any theory of consciousness must satisfy in the end is that it portrays all the dynamic activity that makes for consciousness as occurring in an abandoned factory, with all the machinery churning away and not a soul in sight, no workers, no supervisors, no bosses, not even a janitor, and certainly no Emperor! For those who find this road to progress simply unacceptable, there is a convenient champion of the alternative option: *if you DON’T leave the Subject in your theory, you are evading the main issue!* This is what David Chalmers (1996) calls the Hard Problem, and he argues that any theory that merely explains all the functional interdependencies, all the backstage machinery, all the wires and pulleys, the smoke and mirrors, has solved the “easy” problems of consciousness, but left the Hard Problem untackled. There is no way to nudge these two alternative positions closer to each other; there are no compromises available. One side or the other is flat wrong. There are plenty of Hard Questions crying out for answers, but I have tried to show that the tempting idea that there is also a residual Hard Problem to stump us once we’ve answered all the Hard Questions is simply a mistake. I cannot prove this yet but I can encourage would-be consciousness theorists to recognize the chasm and recognize that they can’t have it both ways.²

It is one thing to declare that you are abandoning the Cartesian Theater for good, and another thing to carry through on it. Ray’s work offers a nice example of a half measure that needs to be turned into a full measure: his discussion of what he called “affects” in *Consciousness*

and the *Computational Mind* and now calls (always in scare-quotes) “feels” or “character tags.” Here is how he puts it in *User’s Guide to Thought and Meaning*:

[An earlier chapter discussed] the “character tags” that contribute the “feel” of meaningfulness and the “feel” of reality. . . . In contrast to the complexity of pronunciation and visual surfaces, these “feels” are simple binary distinctions. Is what I’m hearing meaningful or not? Is it a sentence that someone uttered, or is it “in my head”?

I’d like to look more closely at these “character tags,” which mark the overall character of the experience. I’ll contrast them with “content features” of conceptual structure and spatial structure—such as that this object belongs to the category ‘fork,’ it’s heavy and smooth, it has points, you use it to eat with, it belongs to you, it’s 17 years old, and so on. (Jackendoff 2012, 139)

The fact that he calls these items “affects” or “feels” is a bit ominous: just *whose* feels are they and how does this Subject, whoever or whatever it is, respond to them? Ray is silent on this score—that is, Ray ducks the Hard Question. But we can try to answer it for him. These “feels” are present in our phenomenology, and as such are denizens of the fringe of consciousness, byproducts of the (higher, or more central) unconscious workhouse in which conceptual and spatial structures get built and analyzed. Ray’s excellent half step forward is to dismantle the traditionally mysterious and unanalyzable “grasping” or “comprehending” by the Subject in the Cartesian Theater, outsourcing all that work to unconscious high-level processes into which “we” have no introspective access at all. Those backstage processes make all the requisite links to conceptual structures, taking care thereby of our ongoing comprehension of the words streaming through the fringe of consciousness. Those words have phonological properties we experience directly accompanied by the “feeling” that they are *meaningful* (or not). Here we have the beginnings of a nice division of labor: (almost) all the Work of Understanding has been assigned to unconscious bits of machinery, leaving only one task for the conscious Subject—appreciating the meaningfulness or noticing the meaninglessness of whatever is on stage at the moment.

Calling such a signal a “feeling” at first looks like a step backwards, back into the murky chaos of *qualia*, but the fact that the distinction is binary is encouraging, since it suggests that it does only a small job; it’s a single-throw switch, the *effects of which* are in need of delegation to some unconscious functionaries. Let’s consider some minimal reactions and then build up from there.

Alternative 1. *Discard it unopened.* If the arrival “in consciousness” engenders no further response at all, if becoming conscious doesn’t make the item even the tiniest bit “famous” or “influential,” then it never really entered consciousness at all. The Given was simply not Taken (to revert to the traditional language Wilfrid Sellars wisely urged us to abandon).

Alternative 2. *Log it in “short-term memory.”* This suffices to elevate the item to the status of reportability, whether or not the person reports it (saying something like “Hey, weird, I just had this feeling that “agnostic” was meaningful!”) This is a start, but just what *is* short-term memory, and what does it do? (The Hard Question again: and then what happens?) The answer, I propose, is that putting an item in short term memory permits it to reverberate for a while in the Global Neuronal Workspace (Baars 1989; Dehaene et al. 1998; Dehaene and Naccache 2001) where it can contribute to a host of other ongoing projects of conceptual structure refinement, action guidance, and so forth. It is influential enough to be reportable, noticeable, memorable—at least for a short period.

Alternative 3. *Draw “conclusions” from it.* Among the contributions it can make while echoing back and forth in short term memory is to influence what happens next in some of these projects. To take the case in point, a “feeling” of *meaningfulness* will typically *not* disrupt ongoing projects the way its opposite, a “feeling” of *meaninglessness* does. The gist of its normal influence is *All is well. Carry on!*, in contrast to *Abort! Caution!*, the typical (but not universal) gist of its opposite. The latter may also initiate a new project, the formation and deliverance of a public speech act along the lines of “Hang on there, it sounded like you just said ‘turnip voting highway.’ What did you mean?” *The role of consciousness in this instance is to serve as the expeditor or interface between a struggling central conceptual structure analyzer and some outside source, another person.*

This is the role that accounts for the most striking feature of the Friar’s Fringe model of consciousness: the *intermediate* level of the contents to which “we” have access. When I say “we,” I mean the first-person *and* the second-person. Our facility of conscious access has been designed (by a combination of genetic evolution, cultural evolution and individual learning histories) to be a user-friendly interface between persons. When Ned Block speaks of “access consciousness” and we ask ourselves “access for whom?,” the best answer is: access for other people. Your consciousness is other folks’ avenue to what’s going on in your head, and it has some of the features it has because everything has to be couched in terms

that can be communicated to other people readily. (Cf. Chris Frith's recent discussions of similar ideas.)

Just as the desktop screen on your laptop has been designed to convey to the user only the readily digestible, intuitively "natural" aspects of what is going on in your laptop, the requirements for entrance into the Friar's Fringe (which isn't a neuroanatomical place, of course, but a functional category) are that an item have content that is readily communicable to others.

But what about the fabled ineffability of some contents in consciousness? Isn't this variety of *incommunicability* a hallmark of the "qualia" of experience? This is the inevitable byproduct of the user-friendliness condition: our capacity to report on any topic bottoms out at a lowest level, and whenever that level is reached in an attempt to convey "what it is like," a null result occurs: "I can't describe it; it's an ineffable something." Ineffable, but somehow identifiable. This is a feature that is particularly striking in cases of the tip-of-the-tongue phenomenon, which is a kind of temporary ineffability: we can't find the word (yet) but we can say a lot about what it *isn't* and a little about the linguistic neighborhood (it's two syllables with the stress on the first) in which it will be found. Temporary ineffability is the ubiquitous phenomenon that provides the best support for this treatment of ineffability, as simply the current limit of analysis. Ear training, courses in wine tasting, and the like can move the boundaries, deepening individuals' access to their inner goings on. The Fringe's boundaries are neither sharp nor permanent, in most regards. There are many "flavors" of ineffability, and we can tell them apart but not say how. (Since we can't say how, it is deeply misleading to say they have "flavors," even in scare-quotes, since that implies we know—it's by "taste"—precisely what we don't know: how we do it.)

Alternative 4. *Monitor.* In a different circumstance the role of consciousness might be entirely internal or first-personal, provoking the redirection of conceptual analysis machinery down new avenues in search of meaningfulness. The traditional idea of consciousness as a *monitor* of one's ongoing activities is not in itself mistaken; it is only when the monitor is allowed to work away intelligently, unreduced and undistributed, that it constitutes a bad homunculus, a postponer of theory. When we talk to ourselves, either aloud or in silent soliloquy, "we" have experientially direct access to the words' identities, their sounds and emphases, as Ray points out, and to their meaningfulness or meaninglessness, but not to the unconscious machinery that does all the heavy lifting, both

producing the speech acts and analyzing them, nor to the factors that are controlling that machinery. Monitoring our own thought, we can *hope* for an insightful breakthrough, but not command one.

These are, of course, the apt and familiar responses we make to “feelings” of meaninglessness or its opposite, but notice that once we have catalogued a few of them (the highlights from an apparently inexhaustible list of possibilities), we can leave the *feeling* out of it, and just have the binary switch or flag as the triggerer of this family of responses. The feeling is, as Ray says, ineffable—it has no content beyond just the bare sense of meaninglessness or meaningfulness—and we have, arguably, captured that content in our catalogue of appropriate responses. The feeling is not doing any work. One might put it this way (tempting fate): a *zombie*, lacking all feelings or qualia, who is equipped with a binary switch with the input-output conditions we have just described doesn't lack anything important; it can monitor its own cognition for signs of meaninglessness, and react appropriately when they are uncovered just as we conscious folk do; it can tell others about the “phenomenology” of its own experiences of meaningfulness and meaninglessness, and that account will gybe perfectly with our accounts, since there is nothing more to these “feelings” than this.

These binary character tags are the easiest cases. Ray did well to put the term “feelings” in scare-quotes, since they are best considered as only feelings *pro tem*, on their way to the junkyard once we answer the Hard Question about what happens next when “we” have them. Once we get used to the move, we can start tackling all the more complicated, multi-dimensional aspects of our experience and deconstructing them in similar fashion.³

Notes

1. Ray's innovation in his *User's Guide to Thought and Meaning* of using a rather sacred font for philosophical terms that are meant to be particularly deep and portentous, is irresistible.

2. I can offer intuition pumps to render my claim at least entertainable by those who find it frankly incomprehensible at first. See especially “The Tuned Deck,” in Dennett (2003), (from which some material in the previous paragraphs is drawn) and Dennett (2005, 2013).

3. My favorite example of this kind of further deconstruction (effing the ineffable, we might call it) is David Huron's analysis of the “qualia” of musical scale tones, in *Sweet Anticipation* (2006). What does the “stability” of *do*, the tonic, amount to, compared to the “instability” of *ti*, the leading tone, and which families

of metaphors, adjectives, and adverbs, tend to go with which families of tones? With patient and experimentally tested analysis, Huron demonstrates the *composition* of the heretofore ineffable qualia of *re* and *mi* and *sol* and *fa*, showing that however “atomic” and unanalyzable they seem to be at first, their perception and appreciation is a task that can be outsourced to unconscious neural responses (Huron 2006, 145).

References

- Baars, Bernard J. 1989. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Dehaene, Stanislas, and Lionel Naccache. 2001. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79 (1–2): 1–37.
- Dehaene, Stanislas, Michel Kerszberg, and Jean-Pierre Changeux. 1998. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America* 95 (24): 14529–14534.
- Dennett, Daniel, 1991. *Consciousness Explained*. Boston: Little Brown.
- Dennett, Daniel. 2003. Explaining the “magic” of consciousness. *Journal of Cultural and Evolutionary Psychology* 1 (1): 7–19.
- Dennett, Daniel. 2005. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.
- Dennett, Daniel. 2013. *Intuition Pumps and Other Tools for Thinking*. New York: Norton.
- Fodor, Jerry. 1998. The trouble with psychological Darwinism. Review of Steven Pinker’s *How the Mind Works* and Henry Plotkin’s *Evolution in Mind*. *London Review of Books*, January 22, 1998, 11–13. Reprinted in *In Critical Condition*, edited by Jerry Fodor, 203–214. Cambridge, MA: MIT Press, 2000.
- Huron, David. 2006. *Sweet Anticipation*. Cambridge, MA: MIT Press.
- Jackendoff, Ray. 1987. *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
- Jackendoff, Ray. 2012. *A User’s Guide to Thought and Meaning*. Oxford: Oxford University Press.
- Prinz, Jesse. 2012. *The Conscious Brain: How Attention Engenders Experience*. Oxford: Oxford University Press.
- Voorhees, Burton. 2000. Dennett and the deep blue sea. *Journal of Consciousness Studies* 7 (3): 53–69.

Structures in the Mind

**Essays on Language, Music, and Cognition in Honor of
Ray Jackendoff**

Ida Toivonen, Piroska Csúri, and Emile van der Zee, editors

The MIT Press
Cambridge, Massachusetts
London, England

© 2015 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu.

This book was set in Times by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Structures in the mind : essays on language, music, and cognition in honor of Ray Jackendoff / edited by Ida Toivonen, Piroska Csúri, and Emile van der Zee.

pages cm

Includes bibliographical references and index.

ISBN 978-0-262-02942-1 (hardcover : alk. paper) 1. Psycholinguistics.

2. Cognitive science. 3. Neurolinguistics. 4.

Cognition. I. Jackendoff, Ray, 1945- honoree. II. Toivonen, Ida. III.

Csúri, Piroska. IV. Zee, Emile van der.

P37.S846 2015

401'.9-dc23

2015009287

10 9 8 7 6 5 4 3 2 1