

This article was downloaded by: [UQ Library]

On: 05 June 2013, At: 08:44

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Inquiry: An Interdisciplinary Journal of Philosophy

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/sinq20>

Living on the Edge

Daniel C. Dennett ^a

^a Ce'nter for Cognitive Studies, Tufts University, Medford, Massachusetts, 02155-7059, USA

Published online: 29 Aug 2008.

To cite this article: Daniel C. Dennett (1993): Living on the Edge, Inquiry: An Interdisciplinary Journal of Philosophy, 36:1-2, 135-159

To link to this article: <http://dx.doi.org/10.1080/00201749308602315>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Symposium: Daniel C. Dennett's *Consciousness Explained**

Living on the Edge

Daniel C. Dennett
Tufts University

In a survey of issues in philosophy of mind some years ago, I observed that 'it is widely granted these days that dualism is not a serious view to contend with, but rather a cliff over which to push one's opponents.' (Dennett [1978, p. 252]) That was true enough, and I for one certainly didn't deplore the fact, but this rich array of essays tackling my book amply demonstrates that a cliff examined with care is better than a cliff ignored. And, as I have noted in my discussion of the blind spot and other gaps, you really can't perceive an edge unless you represent both sides of it. So one of the virtues of this gathering of essays is that it permits both friend and foe alike to take a good hard look at dualism, as represented by those who are tempted by it, those who can imagine no alternative to it, and those who, like me, still find it to be – in a word – hopeless.

The seven essays arrange themselves in such a way as to span the cliff edge handily. At one extreme, *Clark* and *Sprigge* are well over the edge, hovering, like cartoon characters held aloft by nothing but the strength of their convictions. It would be a crime to disillusion them. In the middle are *Foster*, and *Fellows* and *O'Hear*, utterly unpersuaded by my version of functionalist materialism, and willing to defend dualist (or apparently dualist) positions positively and vigorously, without begging the question against my alternative. Then there are the critics, *Lockwood*, *Seager*, and *Siewert*, whose sympathies lie (in the main) with the others, but who do not commit themselves to any solution to the problems, dualist or otherwise, but concentrate more on flaws they think they detect in my arguments.

The idea of detecting flaws in my arguments must seem risible to Clark and Sprigge, who sometimes find it sufficient rebuttal simply to paraphrase one of my claims and append an exclamation point to it. But even this is useful; it goes to confirm one of my main claims: some serious thinkers find it impossible even to *entertain* my hypotheses. Not only do their incredulous dismissals testify to the vigor of the horses I am beating

* Daniel C. Dennett, *Consciousness Explained*. Boston, MA: Little Brown & Co., 1991, xiii + 511 pp., \$27.95. ISBN 0-316-18065-3. Unprefixed page references are to this issue of the Journal.

(something about which doubts have been expressed in some quarters), but they also provide independent benchmarks for recalibrating my responses to the most persistent objections. For instance, my fictional Otto has been assailed as a stooge by some critics, in spite of the fact that his speeches are all, in fact, tightened up versions of actual objections raised against the penultimate draft. Here Otto finds friends galore. Lockwood, bless him, even describes him as my philosophical conscience!

All of the essays provide valuable clarifications and innovations – there is not a workaday or routine exercise in the lot – and I am proud to have provoked such a variety of contributions to our vision of these issues. There is a good deal of useful overlap, with several themes of mine attacked from slightly different angles, and I think the best way of exploiting this is to start with the most radical, over-the-edge (if not over-the-top) opposition, and work my way back to solid ground, occasionally skewing the order to take advantage of converging lines of attack.

Stephen R. L. Clark

Why may I not insist, against Dennett, that I do indeed know that I intend and feel, or that I know it better than I can possibly know the truth of any neurological theory he propounds? (p. 6–7)

Why not indeed? Feel free. Now what? Clark provides a marvelous bouquet of quotations, ancient and modern, to let readers see how different is the company we keep. I admire the resoluteness with which he issues his *obiter dicta*. It must be exhilarating to have such an uncomplicated and absolute faith. He recognizes that he is offering no arguments, but says that I don't either – or at best 'very few' and he does go on to concede that my technique is 'not wholly reprehensible' (p. 5), for which I am grateful.

In the main, Clark leans on Searle, and in one of the few passages in his essay that comes to grips with an argument of mine, he rather seriously misrepresents what I say in rebuttal of the Chinese Room. He says:

Dennett simply composes a set of witty conversations, which, he says, might be reproducible by computers responding solely to syntactical cues. He then encourages his readers to believe that such conversations are 'proof' that such (wholly imaginary) computers 'understand' as well as we. (p. 7)

But Searle has stipulated (because his argument requires that he do so) that a computer could indeed be programmed to produce (not 'reproduce') such a conversation responding solely to syntactical cues; Searle surely would not quarrel with any of those details, for he has allowed the defender of strong AI *carte blanche* in composing such stories, by stipulating that the resulting program might even pass the Turing Test. It is he, not I, who

introduced wholly imaginary computer programs as the test-case for his argument. I simply point out a few of the less immediately apparent implications of this (obligatory) generosity on Searle's part, and, more to the point here, do not in any way imply – let alone say (as Clark's not wholly reprehensible quotation marks suggest) – that these reflections are 'proof' that computers can understand. *What I say* is:

it is no longer *obvious*, I trust, that there is no genuine understanding of the joke going on. *Maybe* the billions of actions... produce genuine understanding in the system after all. If your response to this hypothesis is that you haven't the faintest *idea* whether there would be genuine understanding in such a complex system, that is already enough to show that Searle's thought experiment depends, illicitly, on your imagining too simple a case, an irrelevant case, and drawing the 'obvious' conclusion from it. (*Consciousness Explained*, p. 438)

I quote myself at some length here just to show how I belabored the point that I was not offering the imagined conversation as a proof of computer understanding. Clark misses this, but then he misses Searle, too. In 'Fast Thinking' (Dennett [1987, ch. 9]) I surmised that many of Searle's champions confuse Searle's actual conclusion (which I call *S*) with a look-alike conclusion (which I call *D*, and then defend, since there is something – not much, but something – to be said for it). Clark provides a confirming instance of my surmise; his parenthetical insertion of 'wholly imaginary' in the passage quoted above makes no sense coming from a defender of Searle's *S*, but would be an appropriate caveat from one who defended *D*.

The affinity between Searle and self-professed dualists like Clark has always been an unholy alliance (if I may put it that way). Searle has always insisted that his position is *not* dualistic at all, that he is a good materialist, that the presence or absence of a soul or spirit has *no* bearing on the behavioral competence of a physical body so far as he's concerned. Those who think that the presence of 'spirit' or 'soul' gives us powers to act that no computer could even mimic should thus be indifferent to Searle's strange thesis, but so desperate are they, I suppose, to find champions who will keep evil AI at bay, that they forgive Searle for conceding their chief disagreement to the opposition. This has always posed something of a diplomatic problem for Searle, who hotly denies he is a dualist whenever challenged by us hardheads, but seems to have been less eager to alert his cheering dualist supporters to the embarrassing fact that they have entirely missed his point.

The other interesting feature of Clark's essay is his criticism of Dawkins' memes on what purport to be biological grounds: disanalogies between memes and genes. Dawkins does not make a genotype/phenotype distinction for memes, nor are there identifiable loci for memes. Neither of these are clearly shortcomings of the meme concept – rather than just differences – nor does Clark give any reason for supposing these dis-

analogies couldn't be removed if there were grounds for doing so. In fact, in a commentary (forthcoming [a]) on a new paper by Dawkins (forthcoming), I note that something like the genotype/phenotype distinction might yield a readily achieved improvement in his concept. Moreover, a *functional* (not anatomical) notion of locus for memes is clearly definable – whether or not it would prove useful is an interesting question. But that is not what Clark really objects to about memes; what he really objects to is that 'It is just this sense of a divine intellect that is missing in meme-theory, and with it any respect for truth' (p. 12). But as Clark must know, this is a question-begging assertion, for some of us view this absence of the divine intellect as a *prerequisite* for any acceptable theory of meaning and truth, and respect for truth is the source of our conviction.

T. L. S. Sprigge

I am grateful to him for coming to the defense of 'what Dennett abusively calls "figment"' (p. 44), for he thereby renders explicit, for all to see, the covert reasoning that I have otherwise had to impute before criticizing. Here is his argument: First he draws a threeway distinction illustrated by

- (1) my image of my friend
- (2) my friend as I currently imagine him
- (3) my friend as he really is

and goes on to claim that 'the important thing here is that one should distinguish components of consciousness from the object these components function to set before me (in ways which differ significantly in perception from in thought and imagination)' (p. 44). I agree; the actual components of whatever-it-is that is the medium of conscious content must be distinguished from the intentional objects thereby constituted or represented by those vehicles (whatever they are) of content. Sprigge goes on:

Of course, if I think in a merely verbal way of a blue cow no component of consciousness will be blue in any sense at all, but if I imagine a blue cow, my image, though it is not the blue cow I am imagining, has a certain quality which I am prepared to call 'blue' myself, though pedants may argue about its proper label till the cows, of whatever colour, come home. (p. 44)

So the cow is imaginary but the image of the cow is real, and the image has a certain (real) quality: 'But the colour one can identify as a component of one's stream of consciousness is not primarily part of one's intentional world – it only becomes so in so far as it features in one's theory of the world (as it does for me, but not Dennett).' (p. 47) That is, Sprigge believes there is something real with a certain (real) quality that I do not believe in. Sprigge objects to my name for this stuff ('figment'), but I don't understand how he can also object as follows: 'All in all, it is very misleading

to represent his opponents as believing in some extra ingredient in the objective world'. (p. 47) How is it misleading? Sprigge calls himself a 'psychical realist', and has just emphasized that the difference between us is that he believes in the reality of something that I deny the existence of. It must be existence 'in the *objective* world' that is at issue, since I quite agree that figment has *subjective reality* for him – which just means he believes in it.

We agree that people can have beliefs *about* the components of their consciousness – he expresses quite a few such general beliefs of his own, as he notes. In such instances, the components themselves are the intentional objects of those (reflective) beliefs. Now the question that must divide us is this: are they *merely* intentional objects or are they also real? It depends, of course, on what you believe about the components you are thinking of. Whatever you think your own components-of-consciousness *are*, whether you are right, I claim, is *always* an open, empirical question, and the answer is never obvious – it only appears to be obvious to those who think they have privileged access to the 'inherent' nature of these components.

Consider the room full of Marilyn Monroe pictures. Here there is no doubt at all what the heterophenomenology is: a vast, regular array of identical, high-resolution photos of Marilyn Monroe. That is the *object* of your (unreflective) consciousness. Now reflect on the conscious experience itself: What components does your consciousness have, in virtue of which this is its object? Are there details *among the components* or are the components quite minimal? If the *intentional objects* of the experience at a certain moment include, say, thirty high-resolution images of Marilyn Monroe, does it follow that there are *components* of your consciousness for each of these thirty images? Notice that *you are not in a position of authority* to answer this question.

Sprigge speaks of the 'medium' in which even abstract thinking is conducted, and I agree that for there to be content, there must be vehicles of content – a medium in which that content is embodied.¹ I say the medium (in human beings) is neural processes, and hence the real qualities of these components are the qualities that neural processes have in the normal exercise of their functions. Being blue is not one of them, but encoding-blue-in-virtue-of-such-and-such-functional-properties can be. Sprigge would perhaps object that the medium of consciousness certainly doesn't *seem* to be neural processes, but if he did, I would suggest that this is because he is himself failing to make the distinction he thinks I ignore: he is mistaking the *content* for the *medium*. (This point will become clearer below, I trust, in my discussion of Seager.)

I consider Sprigge to have paid me a backhanded compliment for my arguments against qualia: he proposes to overcome them by jettisoning the standard appeal to what he calls the Humean denial of necessary con-

nections between distinct existences. He says 'once one realizes that something can both be a distinct quality of experience with its own inherent nature and also be necessarily related to certain behavioral dispositions, one is released from this unpalatable choice' (p. 52). Sprigge is right about one thing: all the qualia arguments in the literature share the assumption that there is a contradiction in any view which says that there is a necessary connection between qualia and behavioral effects on the one hand (functionalism, 'logical behaviorism') and that their identity is 'intrinsic' or 'inherent' on the other. I'm not known for my appeals to the sanctity of philosophical traditions regarding necessity and possibility, but when someone declares that a revolution in metaphysics is the way to evade my objections, I consider myself to have hit a nerve.

John Foster

I am similarly delighted to see that Foster thinks my arguments have secured one of my main conclusions: 'Dennett's reasoning is . . . impeccable: There is no way of preserving forms of non-cognitive presentation within a materialist framework.' (p. 21) Of course we draw opposite conclusions from then on: I say so much the worse for 'non-cognitive presentation' and he says so much the worse for materialism. Foster is not alone in reaching this verdict. One of the main themes of my book is that it is harder to be a good materialist than most casual materialists have thought, and it has been fascinating to me to see how many closet dualists I have driven into the open. This is progress no matter who 'wins', for it obliges materialists to take more seriously the issues that have underlain the dualist impulse all along, while also offering dualists a perspective on how they might achieve at least some of the goals they care about without acquiescing in a theoretical *cul-de-sac*.

Foster shows that I was wise not to claim to have offered a definitive refutation of dualism. I am quite sure that no such refutation is possible, but that is faint praise for dualism. We can imagine lots of non-starters of which that is true, such as the theory that says that everything is just five minutes old but arranged just *as if* the universe was some billions of years old. Foster himself is prepared to join me in dismissing epiphenomenalism as a non-serious (though irrefutable) theory.

My main objection to dualism is that it is an unnatural and unnecessary stopping point – a way of giving up, not a research program. That is quite enough for me. Foster eventually confronts this objection, and asks why I should want to avoid dualism '*at all costs?*' (p. 29) Only because that is my way of keeping the scientific enterprise going. It is a self-imposed constraint: never put 'and then a miracle happens' in your theory. Now maybe there

are miracles, but they are nothing science should *ever* posit in the course of business. *Temporarily* unexplained whatchamacallits are often a brilliant stopgap – Newtonian gravitational attraction as ‘action at a distance’ before Einstein, Mendelian genes before the discovery of the structure of DNA – but positing something which one has reason to believe *must* be inexplicable is going too far. At one point Foster discusses the prospect of treating the fundamental physical laws as ‘probabilistic’: ‘We could then think of the interventionist causal role of the non-physical mind as that of selecting between, or at least affecting the probabilities of, these physically possible options’. (p. 23) Now if, like Roger Penrose (1989), Foster supposed this ‘interventionist causal role’ was explicable in principle, he would be advocating, like Penrose, a non-conservative, expansionist materialism, but he goes on to deny this, and he notes, correctly, that this would be enough to make the view anathema to me.

This raises Foster’s main point: isn’t this way of characterizing the difference between unacceptable dualism and tolerable expansionist materialism vacuous or question-begging? Why, Foster asks, should the dualist be required to explain things more deeply than the materialist? I’d pose a more lenient demand: that the dualist offer *any* articulated, non-vacuous explanation of *anything* in the realm of psychology or mind–brain puzzles. Since I am simply proposing a constraint on what sort of theory to take seriously, it really doesn’t matter to me (except as a matter of communicative convenience) whether the term ‘dualism’ is defined in such a way as to permit varieties of dualism to meet the constraint. Indeed, Nicholas Humphrey (1992) declares that his position is, *in a certain sense*, a kind of dualism, and yet since it undertakes to meet the demands of objective science, I consider it radical, but eminently worthy of attention now – not a theory to postpone till doomsday. And if Penrose were to declare that his position, too, was really a sort of dualism, and if this understanding of the term caught on, I’d want to shift nomenclature and find some new blanket pejorative for theories that tolerate ‘and then a miracle happens’.

This sheds light on Foster’s claim that my ‘prior rejection of dualism’ is ‘the basis’ for my denial of the inner theater.

What enables Dennett to represent his functionalist approach as correct is that, given the falsity of Cartesian dualism, there is no possibility of finding a ‘central conceptualizer and meaner’ to be the subject of the irreducible cognitive states and activities which our initial (anti-functionalist) intuitions envisage; and without such a subject, there is no serious rival to an account of cognition along functionalist lines. (p. 21)

Yes, one could say that I diagnose dualism as a sort of false crutch for the imagination; it gives people the illusion that they can understand how there

could be a 'central conceptualizer and meaner'. A case in point: Foster asks

If the mind is a non-physical substance, what is its intrinsic nature? But it seems to me that the dualist has no special problems here. To begin with, I do not myself see why the dualist needs to admit that there is anything more to the nature of the mind than what introspection can reveal. (p. 29)

Recall the challenge I put to Sprigge about the room full of Marilyns. I claimed that we are not authoritative about the 'components' of our conscious experience. Is Foster disagreeing? What does introspection reveal to him about the contents of his mind in this instance? Does it include lots of high-resolution Marilyns? I wonder if Foster would endorse the speech I put in Otto's mouth:

You argue very persuasively that there aren't hundreds of high-resolution Marilyns in the brain, and then conclude that there aren't any anywhere! But I argue that since *what I see* is hundreds of high-resolution Marilyns, then *since*, as you argue, they aren't anywhere in my brain, they must be somewhere else – in my nonphysical mind! (*Consciousness Explained*, p. 359)

Either way, the dualist 'has special problems'; endorsing Otto's reliance on his introspection exposes the vacuity of dualistic 'investigation', and denying it leaves dualism with no avenues of exploration.

Foster's account of my view is largely accurate and sympathetic, in spite of his deep opposition to it, but there is one point where he slightly misrepresents it. He is right that I argue against any such thing as 'distinctively presentational, non-cognitive awareness' (p. 20), but it is not quite right to say that my view is 'that consciousness, whether sensory or introspective, is purely cognitive – a matter of acquiring beliefs or making judgments' (*ibid.*), since I grant that contentful states of all sorts occur, and have effects that are in their varying degrees and ways constitutive of consciousness. The disruptive effect of pain, for instance, is not 'cognitive' in itself; it is surely one of the most weighty factors in the family of dispositional factors that make something pain. In my discussion of Lockwood, I will review and expand upon the elements of my view that make Foster *almost* right.

Fellows and O'Hear

It is always reassuring to encounter accurate summaries of one's views by one's critics, and Fellows and O'Hear come up with some expressions that actually improve on the source text. For instance their version of the Cartesian Theatre is *non pareil*:

In the Cartesian Theatre, what seems to me is. If I seem to be a unitary self, then I am. If I seem to be seeing a bent stick or a pink elephant, then I am seeing such

things – not in the external world, it is true, but in my private theatre none the less. (p. 77)

Their account of how I think heterophenomenology opens up an alternative to this vision is right on target: ‘. . . the fact that we are suffering from illusions with respect to our mental life does not show that there are illusory objects (qualia, selves) of which we are aware. For Dennett, it shows only that we are inclined to have and affirm false beliefs of various sorts.’ (p. 78)

They go on to attribute to me the ‘thesis that there is no more to experience than thinking, or that seeing is believing’ (p. 79). This echoes Foster’s reading (as just noted) and, again, I will defer my response of this crucial point to the discussion of Seager, and concentrate first on their alternative and its difficulties.

Dennett will say to us that our thought that we are more than zombies is an illusion we suffer from owing to a defective set of metaphors we use to think about the mind.

Exactly right.

We hope to show, by contrast, that Dennett’s replacement metaphors leave out something which is essential to us as human beings. (p. 75)

What might that be? ‘. . . a seeming whose presence to our consciousness makes all the difference between human life and zombiehood or, quite simply, between animate and inanimate existence’ (ibid.). Real seeming, in short. They cite Wittgenstein’s claim that sensation is ‘not a something but not a nothing either,’ and go on to suggest that ‘the not-nothing which conscious experience is, is not the same thing as the judgments we make about it’ (p. 79). They think, then, that I am not Wittgensteinian enough, whereas I think that this is one instance in which I am more Wittgensteinian than St Ludwig himself, who chickened out in this oft-quoted passage. Or perhaps it is just his followers who have wanted to read more into his uncharacteristically awkward phrase than he meant. I suppose one can read a sort of vitalist subtext into some of Wittgenstein’s comments, but that is the dark side of his inimitable chiaroscuro. In any event, Fellows and O’Hear support their suggestion not with further reflections from Wittgenstein, but – in a somewhat jarring juxtaposition – by a novel interpretation of Searle’s Chinese Room.

They see exactly the point of my rejection of Searle’s Chinese room, and they accurately recount my argument against it, but say it misses the point. Searle’s argument has to do ‘with the ascent from formal or syntactic operations to semantical ones’ (p. 81). Their discussion betrays a variety of naïve misconceptions about computer science in general and AI in particular, but let’s concentrate on the conclusion, not the path: programmes, they say, are *intrinsically* syntactical and only *extrinsically*

semantical or meaning-bearing; 'symbol shuffling by itself does not give any access to the meanings of the symbols' (see p. 82). Access to what or to whom? I would say (deliberately adopting Searle's slanted terminology) that symbol shuffling is what *makes there be* something that could have access to, want access to, need access to, the meanings of symbols. They say '[c]omputer programmes need minds to read them, hence minds cannot be computer programmes merely' (p. 82). This just begs the question, of course, but it hints at the bottom line of their objection:

Dennett cannot, then, on pain of circularity, say that it is just another text or set of texts which gives texts meaning. At least some texts will have to have the meaning-conferring properties of selves and agents. (p. 87)

Aha! They would break the threatened circle (or regress) with a Central Meaner. Or, being good Wittgensteinians, if not a *Central Meaner*, then an *animate* meaner over and above (somehow) the merely apparent meaner to be found in a zombie or robot. There is some very dubious animism or vitalism hinted at (e.g. in the concession about what a Frankenstein could, in principle, do with 'living tissue' [p. 84]). I will forbear trotting out the usual objections to this vitalistic theme, since they are presumably as familiar and unpersuasive to Fellows and O'Hear as they are familiar and conclusive to me. Instead, I will simply highlight my alternative.

Philosophers often maneuver themselves into a position from which they can see only two alternatives: infinite regress versus some sort of 'intrinsic' foundation – a Prime Mover of one sort or another. There is always another alternative, which naturalistic philosophers should look on with favor: a *finite* regress that peters out without marked foundations or thresholds or essences. Here is an easily avoided paradox: every mammal has a mammal for a mother – which implies an infinite genealogy of mammals (which cannot be the case). The solution is not to search for an essence of mammalhood that would permit us in principle to identify the Prime Mammal, but rather to tolerate a finite regress that connects mammals to their non-mammalian ancestors by a sequence that can only be partitioned arbitrarily. The reality of today's mammals is secure without foundations. (For more on this theme in my work, see Dennett [forthcoming (b)].) In this instance, the solution is to show via a *finite* regress (or progress, if one works 'bottom up') how it can be the case that 'intrinsically syntactic' mechanisms ultimately compose systems that deserve to be called semantic engines, capable of 'seeing [each other's] outpourings in semantical terms'. (p. 83).

The last trump played by Fellows and O'Hear is an unflinching defense of the reality of the *self*: '[E]ach of us can say that unless. . . there was a genuine sense of an "I" as the centre of my experience, there would be no way to fix or locate a particular set of thoughts or texts as the place from

which I operated.' (p. 88) What puzzles me about this argument is that it seems quite obvious to me that everything they say about indexicality and confusion about location applies to zombies – we just have to add scare-quotes to avoid begging the question. Zombies are presumably as subject to the disruptive disorders of Multiple Personality Disorder and scatter-brainedness as we are, after all, and when they manage to rise above these afflictions, it is because they have a well-designed 'sense of an "I"'. A zombie can 'wonder' where he is, 'discover' that the hand in the yellow glove is not his own hand after all, and 'recognize' himself in a mirror. I don't see that Fellows and O'Hear have offered any reasons for dismissing the solution to the problem of indexicality I proposed: you are what you control. It's as simple as that. Or as I put it: 'How come I can tell you all about what [goes] on in my head? Because that is what I am . . . a knower and reporter of such things in such terms is what is me' (p. 410, emphasis removed).

Finally, after an excellent summary of my position, they say that I fail to explain thereby 'how it is that Otto could be said to be mistaken unless there was an Otto who was, in the sense Dennett wants to rule out, the subject of the deception' (p. 89). Let's consider this objection in more detail. Suppose Otto were a mere zombie. Then he would be mistaken in his pseudo-beliefs (the pseudo-beliefs he 'expresses' in the speeches I have given him). If an unconscious zombie can have a 'belief', he can have a mistaken 'belief'. Fellows and O'Hear, like Searle and many others, have already conceded that there would be a certain explanatory robustness and systematicity to scare-quoted talk about the 'beliefs' and other pseudo-mental states of zombies. It falls to them therefore to show that there is a *further* problem with the attribution of particular sorts of beliefs – e.g. indexical beliefs or higher-order beliefs or mistaken beliefs – and I see no grounds being given.

Michael Lockwood

The most strikingly pre-emptive criticism I have encountered to my theory is simple: I have (quite obviously) not explained consciousness *at all*; I have *left out* consciousness; I have 'side-stepped' the central puzzle.² I have missed the whole point. Now this would be a rather strange sort of neglect, even by the standards of neuropsychology – to set out to write a book explaining consciousness, and to write a book that actually accomplishes *something* but nevertheless entirely overlooked the project it putatively set out to cover. If I told you someone had written a book entitled *The Arab-Israeli Conflict Solved*, and that it somewhat surprisingly neglected to mention the fact that the Arabs and Israelis dispute the right to certain

tracts of land in the Middle East, you would probably conclude that the author must be insane (or post-modern, if that means anything different).

But so alien is my explanation of consciousness to many readers that they do not even recognize it as a *flawed* explanation, or a *refuted* explanation of consciousness – they don't see that I have even *tried* to tackle what they consider the nub. Sometimes the confidence with which this view is expressed amuses me. It reminds me of an encounter I just learned about from a colleague. A team of medical educators recently returned to Boston from Mexico, where they had been engaged in a project to teach the principles of birth control to uneducated women in remote areas. Armed with all the latest audio-visual equipment, they had held large groups of these women spellbound with their videotapes of microscopic sperm wriggling their way towards an ovum, computer-animated diagrams of conception, and so forth.

After one presentation, one of the spectators was asked her opinion. 'It's really very interesting how you people make babies', she replied, 'but here we don't do it that way. You see, our men have this milky fluid that comes out of their penises. . .' It is almost embarrassingly obvious to Sprigge and Clarke, for instance, that I *could not* be talking about consciousness – the consciousness they know so well – and so they are tempted to conclude that I must be a very different sort of animal!

Which is harder to credit: that I would write a book that didn't even try (in spite of its title), or that they (and so many of them!) would fail to see an attempt as an attempt? I could point to those who *do* see my book as offering not only an explanation but a good one, but the other side will clearly view them as taken in by my slippery rhetoric, lulled to sleep by tricks and examples. Somebody is missing something big; who is missing what?

Faced with such a curious question, I find Lockwood's essay a godsend, for he sees exactly what my attempt at explanation is, and sees that it *is* such an attempt, and such a radical one that he can scarcely believe that I mean it. But he gives me the benefit of the doubt, thank goodness. His first line of attack concerns the consciousness of animals and infants. This is a frequently voiced objection to my theory of consciousness as a culture-borne virtual machine: isn't my theory refuted by the obvious fact that animals bereft of culture (and, of course, newborn human infants) are conscious? Lockwood, appealing as so many do to Nagel's 'what it is like to be' formula, says:

Consciousness in this sense is presumably to be found in all mammals, and probably in all birds, reptiles and amphibians as well. (p. 66)

It is the 'presumably' and 'probably' that I want us to attend to. Lockwood gives us no hint as to how he would set out to replace these terms with

something more definite. I'm not asking for certainty. Birds aren't just *probably* warm-blooded, and amphibians aren't just *presumably* air-breathing. Nagel confessed at the outset not to know – or to have any recipe for discovering – where to draw the line as we descend the scale of complexity (or is it the cuddliness scale?). This embarrassment is standardly waved aside by those who find it just obvious that there is something it is like to be a bat or a dog, equally obvious that there is *not* something it is like to be a brick, and unhelpful *at this time* to dispute whether it is like anything to be a fish or a spider (to choose a few standard candidates for the median).

Fellows and O'Hear put the same point somewhat more circumspectly: animals and human infants seem to be conscious perfectly well without the mediation of any culturally acquired 'software'. (p. 80)

I agree; they seem to be. But are they? And what does it mean to say they are or they aren't? It has passed for good philosophical form to invoke mutual agreement here that we know what we're talking about even if we can't explain it yet. I want to challenge that standard methodological assumption. I claim that this question has no *clear* pre-theoretical meaning, and that since this is so, it is ideally suited to play the deadly role of the 'shared' intuition that conceals the solution from us. *Maybe* there really is a huge difference between us and all other species in this regard; *maybe* we should consider the idea that there could be unconscious pains (and that animal pain, though real, and – yes – morally important, was unconscious pain); *maybe* there is a certain amount of generous-minded delusion (which I once called the Beatrix Potter syndrome) in our bland mutual assurance that, as Lockwood puts it, 'Pace Descartes, consciousness, thus construed, isn't remotely, on this planet, the monopoly of human beings' (p. 66).

How, though, could we ever explore these 'maybes'? We could do so in a constructive, anchored way by first devising a theory that concentrated exclusively on *human* consciousness – the one variety about which we will brook no 'maybes' or 'probablys' – and then *look and see* which features of that account apply to which animals, and why. There will still be plenty of time to throw out our theory if and when we find it fails to carve nature at the joints, and we might just learn something interesting.

Forget culture, forget language. The mystery begins with the lowliest organism which, when you stick a pin in it, say, doesn't merely react, but actually *feels* something. (p. 67)

Indeed, that is where the *mystery* begins if you insist on starting *there*, with the assumption that you know what you mean by the contrast between merely reacting and actually feeling. In an insightful essay on bats (and whether it is like anything to be a bat), Kathleen Akins (forthcoming) shows that Nagel inadvisedly assumes that a bat *must* have a point of view.

There are many different stories that can be told from the vantage point of the various subsystems that go to making up a bat's nervous system, and they are all quite different. It is tempting, on learning these details, to ask ourselves 'and where in the brain does the bat *itself* reside?', but this is an even more dubious question in the case of the bat than it is in our own case! There are many parallel stories that could be told about what goes on in you and me. What gives one of those stories about *us* pride of place at any one time is just that it is the story you or I will tell if asked (to put a complicated matter crudely). When we consider a creature that isn't a teller – has no language – what happens to the supposition that one of *its* stories is privileged? The hypothesis that there is one such story that would tell us (if we could understand it) what it is actually like to be that creature dangles with no evident foundation or source of motivation – except the dubious tradition appealed to by Lockwood, and Fellows and O'Hear.

Bats, like us, have plenty of relatively peripheral neural machinery devoted to 'low level processing' of the sorts that are routinely supposed to be entirely unconscious in us. And bats have no machinery analogous to our machinery for issuing public protocols regarding their current subjective circumstances, of course. Do they then have some other 'high level' or 'central' system that plays a privileged role? Perhaps they do and perhaps they don't. Perhaps there is no role for such a level to play, no room for any system to perform the dimly imagined task of elevating merely unconscious neural processes to consciousness. Lockwood says 'probably' all birds are conscious, but *maybe* some of them – or even all of them – are rather like sleepwalkers, or non-zimbo zombies! The hypothesis is not new. Descartes notoriously held a version of it, but it is Julian Jaynes (1976) who deserves credit for resurrecting it as a serious candidate for further consideration. It may be wrong, but it is not inconceivable – except to those who cling to their traditions as if they were life-rafts.

And what of the 'one great blooming, buzzing confusion' of infant consciousness? (James [1890, p. 462]) Well, vivid as James's oft-quoted (and misquoted) phrase is – a rival on the philosophy hit parade for Nagel's formula – it manifestly presumes more than any cautious investigator would claim to be able to support. That the inchoate human brain is unorganized to some degree is not in doubt; that this incipient jumble of competing circuits is experienced *as* anything at all by the infant is the merest presumption. It may be, and then again, it may not. The standard working assumption appealed to by Lockwood and Fellows and O'Hear doesn't let us consider these as open hypotheses, in spite of the considerable scientific grounds for doing so. At least some animals and infants seem to be conscious in just the way we adults are, but when we adopt an investigative strategy that first develops an articulated theory of adult human con-

sciousness and then attempt to apply it to other candidates (as I do in the last chapter), it turns out that appearances are misleading at best.

In particular, the very idea of there being a dividing line between those creatures 'it is like something to be' and those that are mere 'automata' begins to look like an artifact of our traditional presumptions. Since in the case of adult human consciousness there is no principled way of distinguishing when or if the mythic light bulb of consciousness is turned on (and shone on this or that item), debating whether it is 'probable' that all mammals have *it* begins to look like wondering whether or not any birds are *wise* or reptiles have *gumption*. Of course if you simply will not contemplate the hypothesis that consciousness might turn out *not* to be a property that sunders the universe in twain, you will be sure that I must have overlooked consciousness altogether, since I entertain and even defend this hypothesis.

Lockwood recognizes that my defense of the scarcely credible hypothesis involves denying 'the reality of the appearance itself' (p. 68). Like Siewert, whose views I will turn to next, Lockwood appreciates the pivotal role of the example of blindsight in my campaign against 'real seeming'. The standard presumption is that blindsight subjects make judgments (well, guesses) in the absence of any qualia, and I use this presumption to build the case that ordinary experience is not all that different.

Dennett's position, in effect, is that it is only in *degree* that normal sight differs from blindsight. Normal sight carries with it far greater confidence in the corresponding judgments, and is of vastly greater discriminative power; but there is, in the end, no *qualitative* difference between that and blindsight. (ibid.)

Exactly. Lockwood aptly presents my Marilyn's case as a supporting argument, and grants: 'Here, then, is a concrete instance of an illusion of phenomenology' (p. 69). It is only the extension of my claim from this example that he cannot accept, because he cannot see how I could explain 'the activity of turning the "spotlight of attention" on to the deliverances of our senses' (ibid.). He says:

So what are we supposed to be doing? Simply generating new judgments and checking the old ones against them? Surely not. Judgments, Otto would insist, are too anaemic, too high-level, too *intellectual* to do duty for the *substance* of sensation and perception. (p. 69)

Lockwood's Otto has just echoed the objections of Foster, Sprigge, and Fellows and O'Hear to my suggestion that 'seeing is believing'. A nice thing about Otto (even when it is somebody else putting the words in his mouth!) is that he *actually* suffers from the failures of imagination that I only *suspect* other philosophers of succumbing to. In this instance, Otto has usefully betrayed the source of his error: he is thinking of judgments on the mistaken model of a short, simple sentence you might say to yourself

(with conviction) in, oh, less than a minute's-worth of silent soliloquy. Such a judgment is pretty thin gruel, compared to the zing of real seemings. (As Lockwood adds, 'I hear Otto asking: "Is an orgasm merely a judgment, or bundle of judgments?"')

What are judgments, then, if they are not to be modeled on sentences expressed to oneself? Haven't I myself on occasion called them *propositional episodes*? Yes, and I beg to remind my fellow philosophers that propositions, officially, are not the same as sentences in any medium, and as abstractions they come in all 'sizes'. *There is no upper bound* on the 'amount of content' in a *single* proposition, so a single, swift, rich 'propositional episode' *might* (for all philosophical theory tells us) have so much content, in its brainish, non-sentential way, that an army of Prousts might fail to express it exhaustively in a library of volumes.

Is it 'remotely credible' that 'seeing is believing'? Lockwood's Otto is incredulous because he has fallen for some covert (or 'sophisticated') version of the following: Seeing is like pictures, and believing is like sentences, so since a picture is worth a thousand words, seeing could not be believing! If you think that the contrast between 'merely verbal' and 'imagistic' (Sprigge) secures a distinction between (informational) *content* and *quality*, you are dismissing a major theoretical option without trying properly to imagine it. (For more on this theme, see Dennett forthcoming [c].)

Charles Siewert

Siewert, in his scrupulous, ingenious essay, shares with the other authors the reluctance to abandon qualia (or, in his terms, 'visual quality') and, like Lockwood, he sees my discussion of blindsight as a weak link in my chain of arguments. I have claimed that there is no gradual story that can be coherently told that takes us from actual blindsight to zombiehood. Siewert accepts the challenge:

But now if we can imagine a sort of minor loss of consciousness with conscious-like responsiveness intact in the case of unprompted blindsight, then why not suppose this sort of loss gradually augmented, so that the variety and extent of consciousness diminishes finally to nothing, while the behavior remains that of a conscious human being? If there is no conceptual obstacle to this, we arrive in piecemeal fashion at the notion of a totally unconscious morpho-behavioral homologue to ourselves – the dreaded zombie . . . (p. 100)

This is just the sort of examination of an intuition pump that I recommend. Can the crucial knobs be turned or not? He sees that the burden of proof here is delicately poised – judgments of conceivability or inconceivability are too easily come by to 'count' without something like a

supporting demonstration, and that will have to involve a careful survey of possible sources of illusion or confusion.

Let us first catalogue the differences that have to be traversed as we move from actual blindsight to the target of zombiehood (for the moment, just partial zombiehood – visual zombiehood). Actual blindsight subjects need to be prompted. They claim they see nothing (in their blind fields), and moreover, they don't spontaneously volunteer any judgments, or modulate their nonverbal actions on the basis of visual information arising from the blind region. Actual blindsight subjects exhibit sensitivity to a very limited or crude repertoire of contents: they can guess well about the shape of simple stimuli, and very recently (since *Consciousness Explained* was published) evidence of color-discrimination has been secured by Stoerig and Cowey (1992), but there is still no evidence that blindsight subjects have powers beyond what can be manifest in binary forced choice guessing of particular widely-dispersed properties. (No one has yet shown *delicate* color discriminations in blindsight – or even the capacity to tell a red cup on a green saucer from a green cup on a red saucer, for instance.)

My contention is that what people have in mind when they talk of 'visual' consciousness, 'actually seeing' and the like, is nothing over and above some collection or other of these missing talents (discriminatory, regulatory, evaluative, etc.); I don't know where to 'draw the line' – leave that to those who disagree with me – but I should think that any believer in visual properties is going to become embarrassed at some point in the traverse from actual blindsight to partial zombiehood. Let us look at Siewert's path. He asks you to imagine having a blindsight scotoma, but noticing that 'you are on occasion struck by the thought, as by a powerful hunch or presentiment, that there was something just present (say, an *X*) in the area corresponding to your deficit' (p. 98). This thought, though conscious, would not be an instance of a conscious *visual* experience, however accurate and reliable, he claims, and he supposes that I would say it is a 'terrible error' (ibid.) to claim to be able to imagine this prospect. But I agree that it is readily imaginable. I agree that if I found myself having such hunches, and grew to rely on them, I would still be unlikely to consider them instances of *visual* consciousness – but only because they are (as imagined by Siewert) so poor in content: *an X* or even *an X suddenly moving left to right, and currently just about there* or even *a pink X*. The paradigmatic presentiment is a content-sparse propositional episode, while vision is paradigmatically rich.

Siewert finds this line of mine unpersuasive. He even sees as 'evasive' (p. 102) what I consider to be an essential move. Let me review the bidding: we're talking about an intuition-pump transition, excellently oriented and posed by Siewert, and I have drawn attention to a feature that is, I claim, doing the dirty work: the tacit assumption that the 'amount of content' or

whether a discriminative talent is '*high-grade*' (Siewert's term) makes no difference. *This* is the knob we must turn this way and that to see what happens. Let's try, slowly.

Can I imagine having a presentiment, lacking all visual *quality*, but with a full serving of visual *content*?

(A) It suddenly occurred to me that there was a wad of crumpled paper lying on the floor, shaped remarkably (if viewed from this angle) like a sleeping kitten, except that (it suddenly occurred to me) the sun was glinting off the edges *just so*, and this led me to have the further hunch that if I squinted, the wad of paper seemed to be exactly the same color as my bedspread over there... but of course there was nothing *visual* about my experience – I'm blind!

What effect does speech (A) have on your intuitions? As the content rises, as the visual competence becomes higher and higher grade, do you find yourself less willing to take the subject's word for what it is like? Perhaps you find yourself tempted to declare that nobody could have presentiments that rich in content without their being somehow based on, or at least accompanied by, visual qualities. That gives away the game, however, since it implies that there couldn't be a visual zombie after all; anyone who could pass all those 'behavioral' vision tests would *have* to have visual qualities 'on the inside'. Anybody who said (A) to me would arouse my suspicion that they were suffering from some sort of hysterical linguistic amnesia.³ So I have trouble imagining *myself* asserting (A) – except as a joke. But I can imagine it. That is, I can imagine finding myself in the curious position of wanting to say that my current hunch had all that content (and more – much more than I could express in ordinary conversational time) while at the same time wanting to insist that nevertheless my experience was strangely missing something – something I might want to call visual *quality*. But when I imagine myself in this circumstance, I find myself hoping that I would also have the alertness to question my own desire so to speak. 'Gosh! Maybe I'm suffering from some strange sort of hysterical semantic amnesia!' After all, some color-blind people are oblivious of their affliction, and I suppose there could be an opposite condition, a sort of visual hypochondria, or what we might call 'acute vision nostalgia'. ('Oh yes, I can still make visual *judgments*, color judgments and the like, but, you know, things just don't look the way they used to look! In fact, things don't look to me like anything at all! I've lost all visual *seeming* – I'm blind, in fact!') What I have a *very* hard time imagining is what could induce me to think I could choose ('from the inside') between the hypothesis that I really *had* lost all visual *quality* (but none of the content), and the hypothesis that I had succumbed to the delusion that other people, no more gifted at visual discernment than I, enjoyed an extra sort of visual *quality* that I sadly lacked.

If I found myself in the imagined predicament, I might well panic. In a weak moment I might even convert, and give up my own theory. But that is just to say that in fact I not only can imagine that my theory is wrong; I can even imagine myself coming to believe that it is wrong. Big deal. I can also imagine myself having the presence of mind in these bizarre straits to take seriously the hypothesis that my own theory favors: I'm deluding myself about the absence of visual quality. That might even seem obvious to me.

So it is not, as Siewert realizes, a *simple* question of what he or I can and can't imagine. I have argued against the familiar idea among philosophers that blindsight offers a clean example of visual *function* without visual *quality* – a secure first step towards taking the concept of zombies seriously. I don't consider myself to have given a conclusive a priori argument against this idea, but just to have offered a plausible alternative account that explains (I claim) the same primary phenomena – and the secondary phenomena: the tendency of philosophers to overlook my account.

Siewert sees that there is an ominous stability (ominous by his lights) to my position, and he diagnoses its dependence on an epistemological position of mine he calls 'third-person absolutism'. As one who thinks absolutism of any sort is (almost!) always wrong, I heartily dislike the bloodcurdling connotations of this epithet, but I think he's got my epistemological position clear. Whatever the position is called, it is not a rare one. It is, in fact, the more or less standard or default epistemology assumed by scientists and other 'naturalists' when dealing with other phenomena, and for good reason. As he notes, he has yet to work out the details of a defense of an alternative that doesn't slide into solipsism or something equally bad.

While waiting for him to compose a justification for his novel epistemology, with its 'distinctive warrant', I propose a pre-emptive strike – at any rate a glancing blow. According to Siewert's neutral epistemology, certain things are conceivable that are not (or not clearly) conceivable according to standard 'third-person absolutism'. And so . . .? Would this show that these things are actually possible, or would it show that this novel epistemology is too lenient? Should science take these newly conceived possibilities seriously? Why? The 'neutrality' of his proposed vantage-point midway between traditional first-person authority and traditional third-person objectivity is fragile. At least the old infallibility doctrine had a certain self-supporting *chutzpah* in its favor.

William Seager

Seager's essay is constructed around the attempt to present me with a dilemma: 'either his verificationism must be so strong as to yield highly

implausible claims about conscious experience, or it will be too weak to impugn the reality of actual phenomenology.' (p. 123) Like others (e.g. some of the commentaries in *Behavioral and Brain Sciences* on Dennett and Kinsbourne [1992] and Baars and McGovern [forthcoming]), Seager wants to diagnose the central move of my theory as a more or less standard verificationist power play – too strong to command assent. In the months that have intervened since my book went to press, I've composed several corrective passages that put my apparent arch-verificationism in better light. (Those darn Multiple Drafts – there's just no keeping up with them!) They are tailor-made, it turns out, to meet Seager's objections, so, with apologies for self-quotation, I will repeat them below, since I think it is valuable to get all this point and counterpoint together in a single place.

As Seager shows very clearly, in his careful discussion of the color phi case, the difference between his H1 and H2 is that while experience is 'generated' (p. 125) in both cases, this happens either before (Orwellian) or after (Stalinesque) the binding of the mid-trajectory color shift. And he notes that my account 'does not require consciousness at all' (p. 126) – that is, does not require the 'generation of experience' as a separate component in the manner of H1 and H2. He finds this 'disturbing' (ibid.), since it seems to him to imply that all is dark inside. But of course! We know that to be true. There is nothing 'luminous' (as Purvis put it) going on in the brain.

The conviction that this extra, well-lit process must occur is, of course, a persistent symptom of Cartesian Materialism, and Seager's view is illuminated (if I may put it that way) by Lockwood's ingenious dramatization. Lockwood imagines a version of my Multiple Drafts Model that retains the Cartesian Theater, as a stage (or at least a 'pool of light') in which an 'avant-garde' play is performed, complete with inconsistent flashbacks, revisions, and tampered 'instant replay' video. This does indeed preserve, as Lockwood claims (pp. 60–61), all the curious temporal features I used in order to disparage the Cartesian Theater, without abandoning the presentation process. I welcome this elaboration, for it lays bare the fundamental problem: Lockwood's troupe is not avant-garde enough! Why should they bother with the boring and bourgeois ritual of actually *presenting* the play (with actors in costume, etc.) when all they really have to do is send their real-time script-revisions directly to the critics, libraries, and subscribers? The content is all there, in time to have its apposite effects (their play will seem to have been performed), and you save a fortune on lighting!). There has to be some extra role for presentation in this special medium, and Lockwood offers nothing but 'common sense' in favor of the need for such a process or such a medium.⁴

Seager tries to fill this gap by developing what he considers an embarrassment, if not a *reductio*. He formulates:

(H3) There is conscious experience

(H4) There is no conscious experience, but (false) memories of conscious experience are being formed continuously . . .

and notes that it follows that I cannot distinguish these. He is right; H3 and H4 are just a different way of stating the apparently rival hypotheses that you are conscious or that you are a zimbo who only mistakenly thinks he's conscious, and my conclusion is indeed that the *apparent* difference between these hypotheses is an artifact of bad concepts. I should have made this more explicit in the book, but it took a critic, Bruce Mangan (forthcoming), to distill the essence of the point. Consciousness, he proposes, is 'a distinct information-bearing medium':

He points out that there are many physically different media of information in our bodies: the ear drum, the saline solution in the cochlea, the basilar membrane, each with its own specific properties, some of which bear on its capacity as an information medium (e.g., the color of the basilar membrane is probably irrelevant, but its elasticity is crucial). 'If consciousness is simply one more information-bearing medium among others, we can add it to an already rather long list of media without serious qualms.'

But now consider: all the *other* media he mentions are *fungible* (replaceable in principle without loss of information-bearing capacity, so long as the relevant physical properties are preserved). As long as we're looking at human 'peripherals' such as the lens of the eye, or the retina, or the auditory peripherals on Mangan's list, it is clear that one could well get by with an artificial replacement. So far, this is just shared common sense; I have never encountered a theorist who supposed an artificial lens or even a whole artificial eye was impossible; getting the artificial eye to yield vision *just* like the vision it replaces might be beyond technological feasibility, but only because of the intricacy or subtlety of the information-bearing properties of the biological medium.

And here is Mangan's hypothesis: when it comes to prosthetic replacements of media, all media are fungible in principle *except one*: the privileged central medium of consciousness itself, the medium that "counts" because representation in *that* medium is conscious experience. What a fine expression of Cartesian materialism! I wish I had thought of it myself. Now neurons are, undoubtedly, the basic building blocks of the medium of consciousness, and the question is: are they, too, fungible? The question of whether there could be a conscious silicon-brained robot is really the same question as whether, if your neurons were replaced by an informationally-equivalent medium, *you* would still be conscious. Now we can see why Mangan, Searle, and others are so exercised by the zombie question: they think of consciousness as a 'distinct medium', not a distinct system of content that could be realized in many different media. . . . The alternative hypothesis, which looks pretty good, I think, once these implications are brought out, is that, first appearances to the contrary, consciousness *itself* is a content-system, *not* a medium. And that, of course, is why the distinction between a zombie and a really conscious person lapses, since a zombie has (by definition) exactly the same content-systems as the conscious person. (Dennett [forthcoming (d)])

So I do not shrink from the apparently embarrassing implication Seager adduces. He goes on, in any case, to offer further arguments against my supposed verificationism. The first concerns time scale. He can see no reason why the difference in time scale between the absent-minded driving

case and the cutaneous rabbit case should lead me to describe them in different terms. I say the driving case is best described as 'rolling consciousness with swift memory loss', and Seager quite properly asks why we shouldn't conceive of the cutaneous rabbit in just the same (Orwellian) way. There is indeed only a difference in degree – in elapsed time – but for that very reason, of course, also in collateral effects; and my claim is that these collateral effects are just the differences in degree that eventually yield us the only difference that can be made out. Consider the following parallel:

. . . certain sorts of questions about the British Empire have no answers, simply because the British Empire was nothing *over and above* the various institutions, bureaucracies and individuals that composed it. The question 'Exactly when did the British Empire become informed of the truce in the War of 1812?' cannot be answered. The most that can be said is 'Sometime between December 24, 1814 and mid-January, 1815.' The signing of the truce was one official, intentional act of the Empire, but the later participation by the British forces in the Battle of New Orleans was another, and it was an act performed under the assumption that no truce had been signed. Even if we can give precise times for the various moments at which various officials of the Empire became informed, no one of these moments can be singled out – except arbitrarily – as the time the Empire itself was informed. Similarly, since You are nothing over and above the various subagencies and processes in your nervous system that compose you, the following sort of question is always a trap: 'exactly when did I (as opposed to various parts of my brain) become informed (aware, conscious) of some event?' Conscious experience, in our view, is a succession of states *constituted by* various processes occurring in the brain, and not something over and above these processes that is *caused by* them. (Dennett and Kinsbourne [1992b, pp. 235–6])

There is nothing *other than* the various possible, normal or abnormal, collateral effects of various content-determinations that could count towards (or against) any particular verdict regarding the relative timing of consciousness, so when those effects are reduced to near zero, there is nothing left to motivate a verdict.

This point is even better illustrated in response to Seager's discussion of dreams. He astutely observes that much of the theoretical apparatus of my 1976 paper, 'Are Dreams Experiences?' foreshadow the analyses in *Consciousness Explained* – right down to my apparently outrageous suggestion, way back then, that one might dream a dream backwards but remember it back-to-front, a bit of elbow room for the brain that is not just possible in principle, but (I now claim) necessary in practice. What about forgotten dreams? Is it the case that no test could reveal whether we had them or not? No, there are lots of imaginable tests that could determine whether or not, while you slept, a particular narrative was activated, composed, re-activated, rehearsed, etc. – while remaining entirely inaccessible to waking recollection in the morning. What *would* be beyond

testing is the apparent distinction between this all going on entirely unconsciously and going on in the consciousness of dreams.

So I *can* appeal to the findings of sleep researchers (of the future – as Seager says, the REM findings are nowhere near enough) to remove ‘forgotten dreams from the realm of the unverifiable’ (p. 130) – but the price, which I for one will gladly pay, is that by ‘dreams’ we have to equivocate (apparently) between conscious dreams and unconscious (zomboid) dreams. I am not sure why he thinks I hold there would be no way of investigating these hypotheses. I have already stipulated that all the various contents in all the narrative threads can be (in principle) identified, and their vehicles traced, timed, and located, so there will be no bar at all to the discovery of what Seager calls episodes of narrative spinning. There will be less reason than ever for calling them conscious, of course, and this was the germ of truth in Norman Malcolm’s notorious claims.

This still seems verificationist, of course, but the appearance is misleading, and I now have a new way of clarifying my position, thanks to Lockwood. In a debate with me at Amherst College some months ago, Michael came up with a wonderful phrase (which appears in slightly revised form in his essay in this journal): consciousness, he said (with an air of reminding his audience of the obvious) is ‘the “leading edge” of . . . memory’. ‘Edge? Edge?’, I replied, ‘What makes you think there is an edge?’ and my response to him on that occasion has since grown into a separate paper, ‘Is Perception the “Leading Edge” of Memory?’ (forthcoming [e]). It also provoked me to compose yet another little story, which I have used to stave off this misconception in another reply to critics (Dennett [forthcoming (d)]):

You go to the racetrack and watch three horses, Able, Baker and Charlie, gallop around the track. At pole 97 Able leads by a neck; at pole 98 Baker, at pole 99 Charlie, but then Able takes the lead again, and then Baker and Charlie run ahead neck and neck for awhile, and then, eventually all the horses slow down to a walk and are led off to the stable.

You recount all this to a friend, who asks ‘Who won the race?’ and you say, ‘Well, since there was no finish line, there’s no telling. It wasn’t a real race, you see, with a finish line. First one horse led and then another, and eventually they all stopped running’. The event you witnessed was not a real race, but it was a real event – not some mere illusion or figment of your imagination. Just what kind of an event to call it is perhaps not clear, but whatever it was, it was as real as real can be.

Notice that verificationism has nothing to do with this case. You have simply pointed out to your friend that since there was no finish line, there is no fact of the matter about who ‘won the race’ because there was no race. Your friend has simply attempted to apply an inappropriate concept to the phenomenon in question. That’s just a straightforward logical point, and I don’t see how anyone could deny it. You certainly don’t have to be a verificationist to agree with it. I am making a parallel claim: the events in the brain that contribute to the composition of conscious experiences all have locations and times associated with them, and these can be

measured as accurately as technology permits, but *if* there is no finish line in the brain that marks a divide between preconscious preparation and the real thing – *if* there is no finish line relative to which pre-experienced editorial revision can be distinguished from post-experienced editorial revision – the question of whether a particular revision is Orwellian or Stalinesque has no meaning.

There would be a finish line if there were, in the brain, a *transduction* of information into a new medium, but I have argued that there is no such transduction. The functions or competences that together compose what we think of as definitive of consciousness eventually come to apply to some of the various contents that float by in our brains; it is access to these functions, and nothing else, that puts contents into our streams of consciousness (in contrast to our streams of unconsciousness). There is a stream of consciousness, but there is no bridge over the stream!

I have tried, in these responses, to repay in kind the respect these critics have paid to my book. Better than ever I appreciate how hard it is to make oneself take seriously views one finds outrageous, how easy it is to be tempted by cheap caricature. These critics have set a good example, time and again coming up with keenly observed and constructively expressed versions of doctrines of which they are deeply skeptical. As one who is all too often deeply disappointed and embarrassed by the way my fellow philosophers snipe at each other, I would like to express my deep satisfaction with the way this encounter has come out.

When I have pounced with glee on telling turns of phrase in my opponent's essays, I hope I have managed to be as fair as they have been with me. My belief is that it is in relatively casual and unguarded choices of expression that we philosophers tend to betray what is really moving us, so opportunistic 'pouncing' is an ineliminable part of philosophical method. What is required to keep it from deteriorating into cheap debating tricks and sea-lawyering is, on the side of the pouncer, a proper attention to the principle of charity, and on the side of the pouncee, a willingness to listen, to *entertain* the other side's points before composing rebuttals – or (wonder of wonders) concessions. It is a pleasure and an honor to count these philosophers as not just the loyal opposition, but as fellow investigators on what must be, in the end, a common project.

NOTES

- 1 For more on the 'medium' of consciousness see below, in the discussion of Seager.
- 2 In his review in *Science*, Dale Purves (1992) claims I 'sidestep' the question of consciousness in my book because what consciousness *is* is a 'luminous and immediate sense of the *present*, about which we are quite certain'. He never attempts to unpack that metaphor of luminosity, and in the end he allows as how 'metaphor is not enough' – I quite agree.
- 3 Or perhaps that variety of temporal lobe epilepsy for which pronounced 'philosophical interest' is known to be a defining symptom. See, e.g., Waxman and Geschwind (1975).
- 4 He suggests that I am wrong about 'filling in' and cites Ramachandran's recent research as bearing on this. I welcome the attention drawn to Ramachandran's work, for in fact it ends

up supporting my view, not undermining it. The issues are much too involved to do justice to here, but Churchland and Ramachandran (forthcoming) present the attack in great detail, and I reply in kind in three papers (Dennett [1992], forthcoming [a] and [c]).

REFERENCES

- Akins, K., forthcoming. 'What is it Like to be Boring and Myopic?', in B. Dahlbom (ed.), *Dennett and his Critics: Demystifying Mind*, Oxford: Blackwell.
- Baars, B., and McGovern, K., forthcoming. 'Does Philosophy Help or Hinder Scientific Work on Consciousness?', in *Consciousness and Cognition*.
- Churchland, P. S. and Ramachandran, V. S., forthcoming. 'Filling In: Why Dennett is Wrong', in B. Dahlbom, ed., *Dennett and his Critics: Demystifying Mind*, Oxford: Blackwell.
- Dawkins, R., forthcoming. 'Viruses of the Mind', in B. Dahlbom (ed.), *Dennett and his Critics: Demystifying Mind*, Oxford: Blackwell.
- Dennett, D. C., 1976. 'Are Dreams Experiences?', *Philosophical Review* 85 (April), pp. 151-71.
- Dennett, D. C., 1978a. 'Current Issues in the Philosophy of Mind', *American Philosophical Quarterly* 28, pp. 249-61.
- Dennett, D. C., 1987. *The Intentional Stance*, Cambridge, MA: MIT Press.
- Dennett, D. C., 1992. 'Filling in vs. Finding out: a Ubiquitous Confusion in Cognitive Science,' H. Pick, P. Van den Broek, D. Knill (eds), *Cognition, Conceptual, and Methodological Issues*, Washington, DC: American Psychological Association.
- Dennett, D. C., forthcoming (a). 'Back From the Drawing Board', in B. Dahlbom (ed.), *Dennett and his Critics: Demystifying Mind*, Oxford: Blackwell.
- Dennett, D. C., forthcoming (b). 'Self Portrait', for S. Guttenplan (ed.), *Companion to the Philosophy of Mind*, Oxford: Blackwell.
- Dennett, D. C., forthcoming (c). 'Seeing is Believing - or Is It?', in K. Akins (ed.), *Perception, Vancouver Studies in Cognitive Science 5*. Oxford: Oxford University Press.
- Dennett, D. C., forthcoming (d). 'Caveat Emptor' (reply to my critics) in *Consciousness and Cognition*.
- Dennett, D. C., forthcoming (e). 'Is Perception the "Leading Edge" of Memory?', in A. Spadafora (ed.), *Memory and Oblivion*, Locarno Conference, Locarno, Switzerland, October, 1992.
- Dennett, D. C., and Kinsbourne, M., 1992. 'Time and the Observer: The Where and When of Consciousness in the Brain', *Behavioral and Brain Sciences*, 15, pp. 183-200.
- Dennett, D. C., and Kinsbourne, M., 1992b. 'Escape from the Cartesian Theatre' (reply to commentators), *Behavioral and Brain Sciences* 15, pp. 234-47.
- Humphrey, N., 1992. *A History of the Mind*, London: Chatto & Windus; New York: Simon & Schuster.
- James, W., 1890. *The Principles of Psychology*, Cambridge, MA: Harvard University Press (1983 ed.).
- Jaynes, J., 1976. *The Origins of Consciousness in the Breakdown of the Bicameral Mind*, Boston: Houghton Mifflin.
- Mangan, B., forthcoming. 'Dennett, Consciousness, and the Sorrows of Functionalism', in *Consciousness and Cognition*.
- Penrose, R., 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford: Oxford University Press.
- Purvis, Dale, 1992. 'Consciousness Redux', *Science* 257, pp. 1291-2.
- Stoerig, P., and Cowey, A., 1992. 'Wavelength Discrimination in Blindsight', *Brain* 115, pp. 425-44.
- Waxman, S. G., and Geschwind, N., 1975. 'The Interictal Behavior Syndrome of Temporal Lobe Epilepsy', *Archives of General Psychiatry* 32, pp. 1580-6.

Received 23 November 1992

Daniel C. Dennett, Center for Cognitive Studies, Tufts University, Medford, Massachusetts 02155-7059, USA