Daniel Dennett

You have to give up your intuitions about consciousness

Dan Dennett was born (1942) in Boston, studied at Harvard, and took his DPhil in Oxford in 1965, where he studied with Gilbert Ryle. Since 1971 he has been at Tufts University in Massachusetts where he is Director of the Center for Cognitive Studies. In the field of consciousness studies he is best known for his rejection of the Cartesian theatre in favour of his theory of multiple drafts, and for the method of heterophenomenology, but he has a long standing interest in artificial intelligence and robots, in evolutionary theory and memetics, and in the problems of free will. He spends the summers on his farm in Maine where he thinks about consciousness while sailing, mowing the hay, or making his own cider. Among his many books are *The Intentional Stance* (1987), *Consciousness Explained* (1991), *Darwin's Dangerous Idea* (1996), and *Freedom Evolves* (2003).

Sue **Why do you think consciousness seems to be a harder problem than many other problems in science? For you, what's special about the problem of consciousness?**

Dan Human brains are just the most complicated thing that's yet evolved, and we're trying to understand them using our brains. There are people who have suggested that this was impossible. That's just nonsense, but I think the reason that we find consciousness so hard is that we have evolved a certain capacity for self-knowledge, a

certain access to ourselves which gives us subjective experience—which gives us a way of looking out at the world from where we are. And this just turns out to be very hard to understand.

How can something have that perspective? It might be just a thing, but it's a thing with a point of view, and with the capacity to reflect on that point of view and talk about it. Each one of us is trapped within a point of view. I can't ever get inside your head, and you can't ever get inside mine. The undeniable fact that we have these perspectives is not closely parallelled with anything else we know about anything else. It isn't that atoms have that sort of thing, or that molecules do, or that volcanoes or continents or trees or galaxies do; the only thing we know in the whole universe that has this feature is ourselves, and we're not even sure about each other—that's the problem of other minds.

Now, we are, in a sense, artefacts (and I mean that in the good sense of the term). We have been created by the process of evolution, both genetic and cultural. And what we're now trying to do is to reverse engineer ourselves, to understand what kind of a machine we are that this can be true of us.

Sue **Are you equating subjective experience with having a point of view?**

Dan Yes, but having a point of view is not a simple matter. There's an easy sense of having a point of view where lobsters have a point of view, and mosquitoes have a point of view. With a little stretching and pulling you might even say that a pine tree had a point of view; that is to say a pine tree responds to the world selectively—there's only some features of the environment around the pine tree that it's sensitive to and the rest of the world is indiscernible, as it were, by the pine tree.

But that's indiscernible 'as it were'. In our case there's 'real discerning'; and 'real discerning', in the eyes of many people who have thought about this, has got to be worlds away from the sort of discriminative capacities of that pine tree or that mosquito.

This creates an artefact in the bad sense of that term. To many people there's an imaginative chasm between us with our 'real discerning' and our 'real points of view', and the mere robots, or discriminating-but-not-sentient things. I think that the gap between me and a pine tree, or me and a mosquito, is huge but it's traversable by a series of steps. But I do have to say that some of the steps are quite counter-intuitive, and there's not yet in place the sort of firm 'take it or leave it' science that can force people to abandon their intuitions.

Right now it's a struggle to get people working in consciousness even to think about abandoning their intuitions. They have these powerful, seductive intuitions about how it has to be, or how it can't be, that are just wrong. Nothing new there! We've always had false intuitions about the way the world is, and counter-intuitive science has come along and changed them. But in this case, we don't yet know which intuitions to abandon and why. So the problem is very much a problem of persuasion and self-persuasion and a sort of self-manipulation of one's own imagination, which is scary to many people. So instead they try to have a theory which doesn't require them to tweak their intuitions at all, and then they end up down one cul-de-sac or another, because the theories that are not counter-intuitive are just wrong.

Sue **I imagine that you may be thinking here of the zombic hunch?**

Dan Yes. The zombic hunch is the idea that there could be a being that behaved exactly the way you or I behave, in every regard—it could cry at sad movies, be thrilled by joyous sunsets, enjoy ice cream and the whole thing, and yet not be conscious at all. It would just be a zombie.

Now I think that many people are sure that hunch is right, and they don't know why they're sure. If you show them that the arguments for taking zombies seriously are all flawed, this doesn't stop them from clinging to the hunch. They're afraid to let go of it, for fear they're leaving something deeply important out. And so we get a bifurcation of theorists into those who take the zombic hunch seriously, and those who, like myself, have sort of overcome it. I can feel it, but I just don't have it any more.

Sue **What do you think the nature of this fear is? And, more personally, did you once have this fear yourself and have to overcome it? Or was it really quite easy for you to see that you shouldn't succumb to some slight desire to fall into the zombic hunch?**

Dan Well let me start with that first. It wasn't a momentous occasion for me when, as an undergraduate, one day it just hit me that 'Oh yeah, Alan Turing had the basic move that we could replace Kant's question of how it was possible for there to be thought, with an engineering question—let's think how can we make a thought come into existence. Oh, we could build a robot. And what would it be for a robot to have a thought?'

So, resolutely, from the third person point of view, you sneak up on consciousness from the outside, not from the inside. You keep

looking side-long at the inside, all along the way, and seeing if you can make the difference evaporate. There are powerful reasons for thinking that of course you can make the difference evaporate eventually, because it's got to; because we're part of the physical world; there's no mystery stuff; dualism is hopeless. So, since dualism is hopeless, let's see if we can figure out what the sufficient conditions are in purely material terms for there to be something that it is like something to be; something that has an inside; something that has a subjective point of view. And once I had that project clearly in my head, then it all fell into place. Now the question was just working out the details.

Sue **But you implied that sometimes the zombic hunch does tempt you...**

Dan Oh, it doesn't just tempt me. I deliberately go out of my way, every now and then, to give myself a good instance of the zombic hunch. I talk to myself, 'Come on Dan, think about it this way. Now can you feel it?' Oh, I can feel it all right. It reminds me of how you can look out on a clear night and, if you think about it right, and look at the sky and sort of tip your head just so, you can actually feel the earth in its orbit around the sun. You can see what your position is, how the earth is turning, how it's also in orbit, and it all sort of falls into place. You think 'Oh, isn't that quaint?'

This is a lovely perspective shift, but it takes knowledge and some very specific direction of attention to get into that frame of mind. Well, I think for people who have the zombic hunch and don't know how to abandon it, they have to learn to do something like that too. But they just haven't tried, and they don't want to.

Sue **Why don't they want to? What is this fear of letting go of the zombic hunch, even for people who might rationally understand the arguments for getting rid of it?**

Dan I think they're afraid that zombies would have no moral significance. Zombies would be just stuff, and you can chop stuff up, break stuff up, throw it away, burn it, whatever. It doesn't make any difference; it's just stuff. Whereas if we have immortal souls, or anything that's the moral equivalent, then we preserve our moral point of view. I think the idea of a soul is a curious fossil trace of the desire to treat ourselves as absolute.

Sue **Is it just about morality and mattering; that it matters to something or someone what we do? Or is it also about continuity—that we want to survive?**

Dan Well, I think those two are intertwined. Darwin made a great inversion of reasoning when he realized that you can have a bottom-up theory of creativity: that all the wonderful design that we see in the biosphere could be the products, direct and indirect, of a mindless, purposeless process. This simply inverts an idea that I think is as old as our species. It's what you might call a top-down theory of creativity: that it takes a big fancy thing to make a less fancy thing. Potters make pots; you never see a pot making a potter, or a horseshoe making a blacksmith. It's always big fancy, wise, wonderful things making lesser things. And so here we are; we're pretty wonderful, and so we must be made by something more wonderful still. I think it's very scary for people to give that up, and to begin to think about how our importance doesn't depend on the importance of something still more important.

You know, a good bumper sticker recipe for happiness is, find something more important than yourself to think about, but there are many such things that can replace the one big, important thing which many people think they have to have, which is God.

Sue I assume you don't believe in God. Do you think anything of the person survives physical death?

Dan Well, of course, many of the effects of a person's words and deeds can reverberate through human culture for some time after their death, and these can in rare cases be remarkably powerful and coherent. Abraham Lincoln is a more familiar presence *today*, better known, more recognizable, and more often thought about now, than most of the people who are actually alive today. I think that many people would *love* to have that sort of 'immortality' of effect, and would happily trade it for the more traditional prospect of a disembodied eternity in 'heaven'—an idea whose popularity is matched only by its incoherence.

But competition for admission into that pantheon guarantees that only a tiny minority will ever enter it, since the attention span of human culture is strictly limited. I wonder what the maximum value of p is, where p is the population of 'recognized immortals'. 1000? 10,000? When Elvis Presley finds his seat, does this force Dietrich Buxtehude out? That's the only sort of life after death, and it is in short supply.

Sue What do you think is your greatest contribution to consciousness studies? After all, the field has grown enormously since 1991 when you

published *Consciousness Explained*, and consciousness studies has become all the rage. Where do you see your own contribution fitting in?

Dan I think, oddly enough, perhaps my most important or influential contribution was showing people that materialism was harder than they thought—that it was more counter-intuitive than they thought. Some of the reactions to that book have fascinated me, such as the people who've come up to me and said, 'I thought I was a good materialist until I read your book, and then I began to get really queasy because I realized I have to give up a lot more of my intuitions about consciousness than I thought.'

I said 'Absolutely right! You have to embrace the counter-intuitiveness of some of these ideas. You can't just trust your common sense. There are some deeply disturbing aspects of any proper materialist theory of consciousness. So let's get on with it and expose them.'

One of my favourite sequels to the book, is that a lot of subsequent work confirms that I'm right. I think I was pretty much the first to put forward things that have now become well established phenomena, like change blindness, which I predicted. At the time, this provoked outrage or frank disbelief. People said, 'You're out of your skull there,' but I said, 'You wait, you'll see,' and sure enough, the effects are real. In fact they're much more potent than I dared claim. I sort of wish I could go back and put a little more vim into some of my statements there because, in retrospect, I was more cautious than I should have been.

Sue **Can we take change blindness as an example there? I think that if you take the findings seriously you have to wonder about every act of vision you make. All the time, in your everyday life, looking around you, you have to realize that you're conjuring something out of nothing; that you have far less information in your head than it seems. It should, and for me to some extent does, change the way you feel about your role in the world.**

Does it have that effect on you personally? Has predicting change blindness, and then realizing it was an even more powerful phenomenon than you'd thought—has that changed for you what it's like being Dan Dennett alive and looking around the world?

Dan I wish I could say 'yes', but in fact I think the answer is 'no'. I was thinking about those things even when I was an undergraduate. Here's another way of looking at what my contribution has been: consciousness looks like an insoluble mystery when you have an

inflated vision of what consciousness is, and our introspective lives tend to give us that inflated vision. We tend to think we're conscious of a lot more than we are; we tend to think that consciousness has properties that it just doesn't have. If it did have those properties, boy oh boy it would be much harder to explain than it is. So the first thing you have to do is deflate the phenomena so that you can see that they're not quite so gosh-darn wonderful—so truly mysterious—as you thought they were. Then they're sort of tamed. Then we can explain them.

Of course, there's tremendous resistance to the deflationary move. People don't like me saying that they're not conscious of as much as they thought they are, and what they are conscious of doesn't have the features that they say it has. Their reaction to this is 'Oh Dan's just denying the existence of consciousness.' No, I'm not. I'm just saying it's not what they thought it was. Now it's interesting if you look at the history of science. The term used for talking about the pre-theoretical catalogue of the properties that have to be explained was the 'phenomenology'. So Gilbert worked out the 'phenomenology of magnets', for instance. These are the phenomena, this is what has to be explained. As for the 'phenomenology of consciousness'; if you are an auto-phenomenologist, if you are an introspectionist, if you adopt the first person point of view, you're just going to get it wrong. You're going to con yourself into supposing that your consciousness has many features it just doesn't have. So the trick is to characterize the method which neutrally categorizes the phenomenology of consciousness, and then, go to work. Explain it! And when you've explained all the phenomenology, you're done. You've explained consciousness.

Sue **And is that what you call heterophenomenology?**

Dan That's heterophenomenology. Heterophenomenology is the scientific catalogue of what has to be explained.

Sue **You're very hard there on the first person point of view, but can you see no role for disciplined self-observation? I'm thinking in particular of meditation, where it's said that if you keep practising long enough, some of these things become obvious. The visual world starts to disintegrate. Our illusions of the continuity of self, the continuity of the perceived world, the simultaneity of events, all start to fall apart and you see more clearly. Do you think there's any truth in that, or do you dismiss that completely?**

Dan No, I think there is truth in it, but this is in the context of discovery not the context of justification. Every experimenter should, of course, put herself in the apparatus, and see what it's like from the inside. You should certainly treat yourself informally as a subject and see if you've overlooked something, for instance. But having done that, then you do the experiment. You use naive subjects, and you figure out some way to get what you've discovered from the first person point of view to manifest itself for neutral observers from the third person point of view. And if you can't do that, then you have to be suspicious of the insights that you thought you had.

In a sense this is obvious. Nobody in the scientific world working on consciousness would think of submitting a paper that said, 'Well I introspected under the following circumstances and these are the things that I thought.' If you think you've discovered a phenomenon, then you go out and test it using the scientific method, and that means the third person point of view.

This could be just a typical philosopher's hypersensitivity to form and rigour, if it weren't for the fact that so many people are just wrong about the results of their own introspection. People cannot prevent themselves from theorizing when they think they're observing.

Sue **One of the things that's amazed me over the years, is how systematically and deeply you are misunderstood. I'm thinking of things like heterophenomenology, the third person perspective, the zombic hunch, the Cartesian theatre. Do you understand why people find them so difficult? You write clearly; you explain things well, at least I think you do. How come you get so misunderstood all the time?**

Dan Well, I wish I knew. I wish I knew. I've got hunches about it, and here's what I think happens. I've caught myself doing this with others, so I can see how they can do it with me too. When somebody tries to tell you something which is initially very counter-intuitive for you, you put your best effort into it and then translate it into your own terms, so that you can understand it. So you're not just listening cold, you're actively translating what you're hearing into your own dialect. But of course this can horribly backfire. If somebody is trying to put forward something that really is counter-intuitive, you almost certainly get it wrong. You'll throw out the most important part and you'll turn it into one kind of nonsense or another. And if you're not alert to that, you'll think 'Look, I did my level best to understand this person, and here's what I come up with. That's just crazy, so she's just crazy.' Nobody wants to hear that maybe your level best wasn't good enough.

I also think that in a way my writing style traps me, because at least superficially its not hard to see what I'm doing. It goes down quite smoothly, not like reading Hegel or Heidegger. So people think it's easier than it is. No, it's actually very hard. I tried to make it as easy as I can, but it's still very hard, and if you make the mistake of thinking that it's actually a pretty simple idea or two, you're just wrong. But I can see why people would think that.

Sue **I'm particularly interested in one of your central arguments: the non-existence of the Cartesian theatre. You explain why there can't be a Cartesian theatre in the mind or the brain, why there isn't a show going on in the head, and why there isn't somebody watching. You call people who think they are materialists but are still trapped in imagining a Cartesian theatre, Cartesian materialists. Can you say something about what you take to be the signs of being a Cartesian materialist, and how common they are?**

Dan The sure sign of Cartesian materialism is anybody who tells the story of consciousness and doesn't go on to answer the 'and then what happens?' question. So it's as if we work so hard to get this stuff up and presented to the Queen and then what happens? We get inside the audience chamber but why does that make it consciousness? Any theory that's still got a place for the show to happen has not yet done the job.

A curious feature of this is that if you then go on and answer the 'and then what happens?' question, a lot of theorists are sure that you've left something out. Because now we're back explaining behaviours and reactions, and the effects on vocalization and memory, they want to say, 'Wait a minute, where's the consciousness?'

There's a bi-modal distribution between people who think that any theory of consciousness that leaves out the first person is a hopeless theory, and those who think that any theory of consciousness that *doesn't* leave out the first person is a hopeless theory. You've got to leave the first person out of your final theory. You won't have a theory of consciousness if you still have the first person in there, because that was what it was your job to explain. All the paraphernalia that doesn't make any sense unless you've still got a first person in there, has to be turned into something else. You've got to figure out some way to break it up and distribute its powers and opportunities into the system in some other way. So the Cartesian materialist is the one who describes large parts of the machine, but it's still inhabited.

Sue I think I see it in things like, 'and then it's displayed', and 'then it enters consciousness'. Would you count those as signs of being a Cartesian materialist?

Dan Those are certainly danger signs, unless the person goes on and cashes that out very carefully.

Or there might be a theory that says, 'And then our brains tell us *blah blah blah*'. So who's this 'us'?

Sue I like your idea that you deliberately throw yourself into the zombic hunch. Perhaps I should sometimes throw myself into the Cartesian theatre more willingly than I do.

I do sometimes get into it. I can get quite upset thinking about—say—the brownness of this desk here; the 'how it is to me'. I have the very powerful conviction that I am in here experiencing this ineffable, unique, private sensation of the brownness. Can you help me? I know quite a bit about your theory, but when I get really badly into that feeling, how can I get out of it?

Dan The way I recommend is to ask yourself, 'What am I pointing to? What am I ostending when I say *this*?' What I think you'll find is that you can start elaborating a sort of catalogue of the facts that matter to you at this moment. Maybe it's the particular deliciousness of this taste in my mouth; so what is that deliciousness? Well I'd like some more, and I can recall it at a later date, and so on. We're going to take care of all that. We're going to include your disposition to want some more, your capacity to recollect, and even the likelihood that you will find yourself pleasurably recollecting this experience of it. There's a huge manifold of reactive dispositions that you're pointing to when you're saying 'This very yumminess right now,' and what you have to do is recognize that however indissoluble, however unanalysable, however intrinsically present that all seems to you, what has to be explained is that it *seems* to you, not that it is so.

Sue But you have said there that it *seems* so to me. Presumably you would say that we not only have to explain why it seems so to me but why it seems there is a me to whom it seems.

Dan Yes, exactly. Those are the two halves you've got to explain.

And people—wonderfully conveniently for them, and inconveniently for the truth—forget that it seems that way to the zombie too.

Sue Do you think you have free will?

Dan Yes.

Sue And what do you mean?

Dan I mean that in all the things that matter to me I can make a decision based on my consideration of what matters the most and why. I wouldn't have free will if I were obsessive, or were an addict, or were seriously deranged so that I couldn't keep track of reasons, or if I had a memory disorder that meant that I couldn't keep track of a project from minute to minute. Then my free will would be pointless.

The model that we want to have for free will is of an agent that is autonomous, not in some metaphysical sense, but in the sense of being able to act on the reasons that matter to the agent, and who's got the information that is needed to act in a timely fashion. In order to appreciate this you have to realize what brains are for. Brains are for generating expectations about the future. The simplest imaginable thing a brain is for, is for ducking an incoming brick. You see the brick coming. You see it's heading for you. You expend a little energy to duck so it doesn't hit you. There's a lot of things to avoid in life; and there's a lot of things to try to accomplish, but let's take the verb 'avoid'. It's key in one particular regard, that the word inevitable comes from it. Inevitable means unavoidable, and it doesn't really make sense in a context where there isn't avoiding. Where there's avoiding, there are things that are inevitable and there are things that are, if you like, 'evitable'. What makes it possible to avoid things is having some foreknowledge of what's going to happen. So, that's what our brains are for.

If you've got that equipment and it gets used, you can have reasons for acting that are good reasons, that are your reasons. You didn't make them out of nothing, you made them out of all the information and all the values that you've ever considered and reflected upon and decided upon, and for better or for worse, you've come up with a particular set of values, and now you're ready to act.

Just take a simple case of a chess player who makes a move. Why did he make that move now? I might say, 'Well the clock was running, I had to decide sooner or later. OK, enough thinking, it's time to move. This is my move. It may not be the best move. I may live to regret it. I may discover a better move in a few seconds. I may get checkmated. But I wasn't deceived about the position of the board. I wasn't deceived about the rules. I wasn't deceived about the point of the game. That was the best I could come up with. So that's my free will, that's my move.'

Sue In all of that description of free will, you kept saying 'I, I, my, my'. I want to know how you relate this to your ideas that there's no audience

in the Cartesian theatre, that the self is a benign user illusion and so on. Who is it who's having the free will?

Dan The agent.

Sue By 'the agent' do you mean the whole body?

Dan Sure.

Sue Then isn't it important that you distinguish that view from what many people feel about free will, which is that *they*, the audience in the Cartesian theatre, the little special conscious me inside here, is the one that has the free will?

Dan One of the most curious ironies to me, in my earlier writing on this, is that the most important sentence in my book *Elbow Room*, I put in parentheses, and so nobody paid much attention to it. I said—and it was meant ironically—that if you make yourself really small you can externalize virtually everything.

The imaginative pressure to think of yourself as very small is easy enough to find. When I raise my arm, well what is it? There must be some part of my brain that is sort of sending out the signal and then my arm is obeying *me*, and then when I think about the reasons why, it's very natural to suppose that my *reason store* is over there somewhere, and I asked my reason store to *send* me some good reasons. So the imagery keeps shrinking back to a singularity; a point, a sort of Cartesian point at the intersection of two lines and that's where I am. That's the deadly error, to retreat into the punctate self. You've got to make yourself big; really big.

One aspect of this has been very nicely expressed in recent years by Andy Clark, in *Being There*, that we offload a lot of our minds into the world, we then do our thinking using those peripheral devices as part of our equipment. We don't have to do it all in our heads, we can do it out there with slide rules, or calculators, or laptops, or with a little help from our friends.

In fact I think most of us who manage to live moral lives, lives that we're not ashamed of, in fact rely a great deal more on the support of our friends than we readily acknowledge.

Sue You told me that you had many of your ideas as an undergraduate, and that what you've been doing since has really been fleshing those out and explaining them, but I'd like to know this: has anything happened in your life as a philosopher of consciousness that's really changed you, or changed the way you feel about yourself?

Dan I haven't had any conversion experiences that I can think of! But I certainly think my interactions with people outside philosophy have had a huge effect. Probably the first five years of my career, back in the upper Neolithic and the late '60s, I was still hanging out most of the time with philosophers, and a rather small, but cherished and fascinated percentage of the time, I spent talking to people of other disciplines. Gradually I came to realize that I learned more, and found it philosophically more interesting, talking to people about artificial intelligence, biology, neuroscience, and psychology, than I did talking with my fellow philosophers. So over the years I went where the fun was for me, and it fed on itself. I got invited to more and more non-philosophical occasions and conferences, and I read more and more articles and books, to the point now where I read the philosophy as a duty. It staggers me to realize how much less fun it is to read most philosophy than it is to read good biology or good psychology or good artificial intelligence. And so that's made a huge difference to me.

Of course, this means that a lot of people in the field of philosophy say, 'Well you know Dennett's just no longer a philosopher, he may have been a philosopher once but he isn't now.' Well, I don't want to argue about that, but if so, then maybe all philosophers should cease being philosophers and try to do the sort of thing that I'm doing, because I think I'm getting philosophical results, and making philosophical progress. I think it's far better than the sort of vacuum philosophy that we all used to engage in back in the old days.

Sue **And what *is* philosophy?**

Dan Philosophy is what you do when you don't yet know what the right questions are to ask.

CONVERSATIONS *ON* CONSCIOUSNESS

Susan Blackmore

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2005

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Data to follow
ISBN 0 19 280622-X
EAN 978-0-19-280622–2
1 3 5 7 9 10 8 6 4 2
Typeset by RefineCatch Limited, Bungay, Suffolk
Printed in Great Britain by
Clays Ltd., St Ives plc