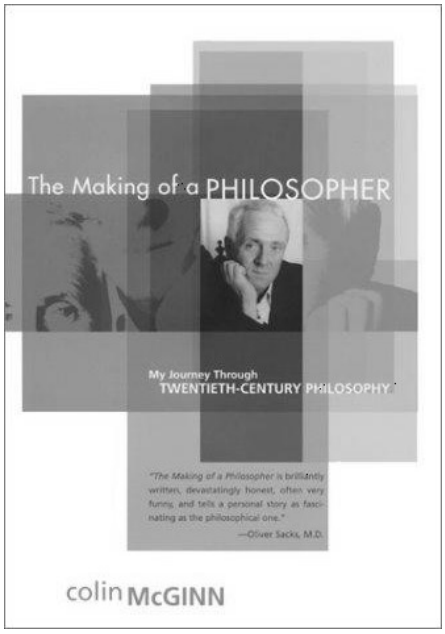
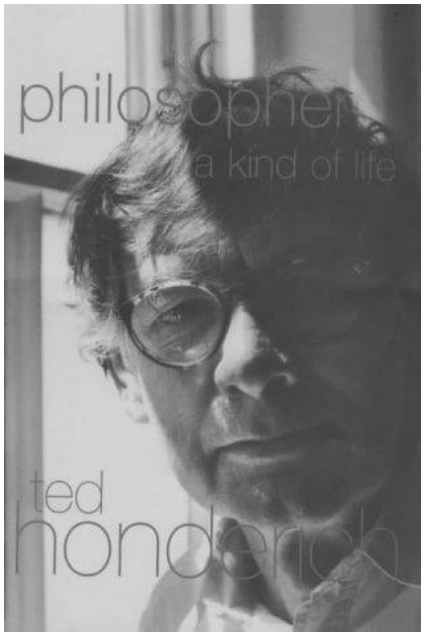
J. Andrew Ross

First-Person Consciousness

Honderich and McGinn Reviewed





Two Professors

A nice everyday conception of consciousness is as autobiography. Ongoing experience is stored and restructured as a personal narrative, and consciousness is the generic mental state that accompanies this lifelong cognitive activity. On the higher scale of human cultural achievement, the literary form of autobiography brings this form of consciousness to a natural zenith. From this perspective, it is a happy coincidence that two distinguished philosophers of consciousness have recently published philosophical autobiographies.

Correspondence: *me@andyross.net*

Journal of Consciousness Studies, **9**, No. 7, 2002, pp. ??–??

The first and more senior autobiographer is Ted Honderich,[1] who from 1988 until his retirement in 1998 was the Grote Professor of Mind and Logic at University College London. The second is Colin McGinn,[2] who since 1988 has been a professor of philosophy at Rutgers University in New Jersey. Both have written extensively on consciousness, and both weave their views on consciousness into their respective life stories. Indeed the parallels run deeper. Honderich grew up in Canada and moved to England to pursue his career in philosophy, almost all of it in London at UCL. McGinn grew up in England and moved to America to pursue his career, after spending some ten years of it at UCL alongside Honderich. And both were locked for decades in a love-hate relationship with Oxford, the parochial sun in the British philosophical firmament.

In other respects, the two protagonists seem to be opposites. Honderich is tall and gruff-voiced, a rough-hewn alpha male in the academic world, whose record of boozing, schmoozing, and womanizing stands almost scandalously proud of the philosophical pack. McGinn, by contrast, is short and modest in demeanour, and lives a quiet life as a vegetarian scholar whose main passion outside reading and writing is kayak surfing. Honderich comes from a family with German roots and has prosperous and well-connected relatives in Canada. McGinn's background is working class, with numerous coal miners in his family tree and no great financial ballast.

As for philosophy, both were close to the Oxford mainstream, in the British liberal tradition defined most prominently in recent decades by Sir Alfred Ayer, and both hold views on the philosophy of mind and consciousness that fall squarely within the Anglo-American analytical tradition of the last half-century or so. Honderich has held a variety of specific views, including a theory of *psychoneural intimacy* that seems a shade away from Davidsonian anomalous monism, and now maintains that consciousness is, in short, the experienced part of a world. McGinn has been closer to psychology, but ten years ago he came out for a position since dubbed *mysterianism* according to which we as a species are cognitively unable to get our minds around our own minds, so to speak, and must in all probability accept that own consciousness will forever remain a mystery to us.

My plan here is to review the two autobiographies briefly from the standpoint of how they motivate and illuminate their respective authors' views on consciousness, then to look at those views more closely and see how far they give us a useful view of the truth, whatever it is, about consciousness. After that, I shall stand back a little and consider some recent developments within the wider field that both authors share. The main problem here is how to accommodate first-person phenomenology in a tradition that was historically dominated by behaviourism and is still inextricably linked to third-person reductionist science. This involves discussing the work of David Chalmers. Then, I shall consider the

---

[1] **Ted Honderich**, *Philosopher: A Kind of Life*, London: Routledge, 2001, 472 pp., £30, ISBN 0415236975 (hbk).

[2] **Colin McGinn**, *The Making of a Philosopher: My Journey Through Twentieth-Century Philosophy*, New York: HarperCollins, 2002, 256 pp., $25.95, ISBN 0060197927 (hbk).

autobiographical enterprise as a source of insight into consciousness, with reference to the views of Daniel Dennett. In conclusion, I shall suggest that the unfolding history of science may be seen as the autobiography of consciousness itself.

As a declaration of interest, I am personally acquainted with both of the authors reviewed here. I met Colin and Ted in the 1970s. Colin and I are contemporaries, and we both served time as philosophers in Oxford and London. I met Ted again in August 2001 at the *Toward a Science of Consciousness* conference in Skövde, Sweden. David Chalmers I know from several recent conferences. In recent decades I have focused mainly on physics and computer science, but I share with all three a passion for the challenge of understanding consciousness.

From Boat to Grote

Ted Honderich was born in 1933 and raised in Canada. Initially, he wanted to be a writer in the manner of Hemingway or Arthur Miller, but then, as he wrote in his diary in 1957, he was 'wonderfully inspirited by A.J. Ayer's *Language, Truth and Logic*' and decided to study philosophy in England. As he reports in his autobiography: 'My sight of England, from the deck of the liner *Italia*, as we came in along the coast to Portsmouth harbour in July 1959, was wonderfully affecting.' Because Ayer was then Grote Professor of Mind and Logic at University College London, Honderich enrolled at UCL. Ayer held a regular seminar in the Grote professor's room at UCL, and there in October 1959 Honderich began his acquaintance with the subject, the role, and the room. After two years of study, he lectured for two years at the University of Sussex, then returned to teach at UCL, and stayed. In 1988 he made it to the Grote chair, and sat enthroned there until his emeritus years.

Honderich's autobiography is a large and fairly dense book that in parts rewards close reading. It is also eminently quotable, so let me save my own words for a while and quote the Grote:

> Philosophy is not any of linguistics, psychology, cognitive science or any other science. To its credit or discredit, there is hardly any Philosophy of Life in it, not much on the meaning of life, hardly any consolation. . . . [Philosophy] is the line of life owed to a certain impulse … to reduce to clarity and thereby get a systematic and comprehensive hold on the nature of one or two of the fundamental parts of reality, including human reality. . . . I suspect the truth is that our line of life . . . concentrates more on good thinking about the facts as against getting or using the facts. . . . Good thinking is getting a clear hold. That is the real impulse in philosophy (pp. 16–17).

Well, good thinking about consciousness is worth treasuring, so let's see how Honderich builds up to the theme. He declares ownership of three main 'pieces of philosophical furniture'. First, he believes in determinism:

> [E]ach of the actions in our lives and also the choosing and willing of it is an *effect*. It is the effect of a sequence of events or states or properties, each of these also being an effect. . . . Each effect is what it sounds like, something that had to happen. There was no other possibility. It wasn't just probable, to any degree (p. 7).

Second, he has a conviction about consciousness:

> The two problems here are the nature of consciousness itself and the relation of this consciousness to the brain. My conviction is that conscious events, states or properties involve what is easier to name than to analyse, a fundamental *subjectivity*. That is their essential nature. . . . A demonstrated fact of psychoneural intimacy, as I was pleased to name it, is the gift of neuroscience to philosophy. A better gift, as it seems to me, than anything from muddled physics (pp. 9–10).

Third, and less relevantly here, he believes in the principle of equality:

> [W]e should not be distracted or detained in any way from trying to make well-off in a certain sense all those who are badly-off. That is the solution to the problem of justice (p. 20).

More relevantly, Honderich's notion of the contribution of 'muddled physics' to philosophy is the history of attempts to use quantum theory to deny determinism or to explain consciousness. It is easy to sympathize with his problem here:

> [T]he interpretation of Quantum Theory, the understanding of what it comes to in terms of the world, is allowed by most of its users to be a mess. Certainly it *is* a mess, and has remained so for too long. . . . What is the mathematics or formalism of the theory *about*? (p. 8).

Well, indeed, a tricky question, though hardly sufficient reason to give up on it in philosophy, especially when your life's work is a theory of determinism that seems at first glance to be simply falsified by quantum randomness. But Honderich gives up on more than physics:

> I have no love for Formal Logic, and enjoy the certainty that it has not solved or advanced any philosophical problem, and so I have not learned a lot (p. 14).

This is an embarrassing admission for a modern Anglo-American philosopher. In a tradition where some of the biggest names are Frege, Russell, Quine, and Kripke, it's hard to get by without mastering a few formulae. The consequence is that Honderich has nothing very useful to say on the deeper questions of truth and meaning.

We can check this by returning to his life story. In 1968, at 35, Honderich started his seventh year as lecturer and his fifth at UCL, working in an old terraced house that served as the philosophy department:

> My eyrie was at the back of the house. . . . If these two years and the different one that followed were not wholly unlike all my others in terms of morale, they were a nadir. . . . Another reflection on the first part of my nadir in morale is not yet perfectly manageable, and of course has to do with my actual philosophical abilities. . . . [I]t is clear that I did not have and do not have all of the things that are called philosophical strengths. Fortunately, it is also clear that that is the condition of my entire profession (pp. 157–8).

Honderich spent the academic year 1970–71 in the USA, the first semester at Yale and the second at the City University of New York. Of March 1971 he says:

> This was … the nadir of my nadir in morale, not so bearable as the rest. As recorded a month later in my diary, I could not escape a kind of frenetic thinking on my troubles, fell to weeping for a while one day, and was afraid to be alone. . . . Do I think I might have done myself in? . . . Things weren't *that* bad. They never have been. I'm ordinary enough to be saved that (p. 173).

On a personal level, the chief consolation of philosophy is the support it provides for negotiating the existential fact about life, the human predicament, that it features moments of truth when the limits of one's powers and achievements become painfully evident. Critically examined, such moments can fuel some good thinking. The sustained rhetoric of introspection in Honderich's autobiography is a pleasing result of such thinking. But its lack of system is a weakness. The book is written almost like a diary, with topics coming and going over the weeks and years, with no clear thematization, or even a summary chronology or bibliography at the end. The life passes like a dream.

That said, the life featured a central achievement: a theory of determinism. Honderich's *magnum opus* is a thick volume entitled *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes* (1988), a dry tome better not attempted by any reader who has not first enjoyed his brief, popular introduction *How Free Are You?* (1993). On the tome:

> I was confident that the book contained a resolution of the problem . . . of the human consequences of determinism. . . . [E]ach of us has two sorts of hope, including two sorts of life-hope. . . . One sort of life-hope carries the thought or is based on the idea that maybe nothing will get in the way of your desires and your nature. . . . The other sort of hope carries an additional thought, that your future is not already settled, that you have a kind of chance. . . . The way to go on, said I, was to try to give up the kind of life-hope whose contained idea has to be false if determinism is true — give it up by trying to see that the other hope you can persist in is sustaining and there are other compensations (2001, pp. 301–14).

Beyond this work and his teaching, Honderich busied himself in gentlemanly fashion over several decades with ongoing editorial duties for several book series, including a venerable RKP series called *The International Library of Philosophy and Scientific Method*, a later Routledge series called *The Arguments of the Philosophers*, and the more popular philosophy paperbacks from Penguin. His work for these series was less than zealous, and involved various elements of bad conscience whenever the signs of his relaxed approach became too evident. Yet the harvest of this lifetime of editorial experience was good: in 1991 he was commissioned to edit *The Oxford Companion to Philosophy*. After more work than he wanted, much of it sheer drudgery, the book was finally published to acclaim (Honderich, 1995). Wittgenstein biographer Ray Monk said it was 'the most authoritative single-volume reference work in philosophy yet published'. Whatever Honderich's career may have lacked in analytical depth, it made up in breadth and community service.

Another major philosophical thread in his life has been the analysis of political activism, including violence and terrorism. This resulted in a succession of more or less ephemeral books that consistently located him in the left-liberal part of the

political spectrum. Although he was no Marxist and boasted of never having read Marx, his rhetorical broadsides against conservatism, from Thatcherism generally to the specific Salisbury radicalism of his London colleague Roger Scruton (Honderich called him 'the unthinking man's thinking man'), and against centrist liberalism, for example as represented by John Rawls with his theory of justice as fairness (Honderich called it 'bumble'), must have scored points with many a Marxist. In the end, Honderich became a Labour Party activist. He was proud to work at the request of the then leader of the Labour Party, Neil Kinnock, on speeches for the 1992 general election campaign in which Kinnock, lampooned mercilessly in the tabloid press as the 'Welsh windbag', lost to the equally uncharismatic Conservative Party leader John Major. Not a great advertisement for Honderich's political savvy, perhaps, but at least a tribute to a certain kind of moral consistency.

More entertainingly, Honderich was quite a Casanova. He treats us to plenty of detail about his succession of mistresses, none of it prurient but altogether quite sufficient to establish that despite his smooth veneer he was true to form in a joke he made to his longtime Marxist colleague Jerry Cohen (now Chichele Professor of Social and Political Theory at Oxford) that women are *tarmac* — 'something rolled over or landed on in the course of life's journey'. In his own summary verdict:

> I have been a man of many women, if that uncertain description is taken to mean a man who has been for a longish time with each of many women, a succession of them. Here my life has been a bit more than middle-sized. I have been a libertine too, if one of those goes on being free from convention, and does not go in for much concealment of his freedom (pp. 27–8).

Remarkably, for a man who prides himself on good thinking and freedom from convention, his story never once reflects on his apparent compulsion to get down on the tarmac whenever he could, whatever the complications. He says he was not a sensual man and was quite conventional in sex, to the point when younger of not masturbating and finding condoms unmentionable, yet he frequently enjoyed casual sex outside the confines of his latest relationship, recommended abortions to his pregnant mistresses, and slept with several of his undergraduate students. Evidently it never occurred to him to stand back and consider the wider issues — even decades later, when trawling through his diaries to compile the catalogue of his conquests.

This should not distract us from the philosophy, but it does. No man who womanizes — and wines — so freely can be *rigorous* enough to cut it down at the coalface of knowledge. If you don't believe me, ask Colin.

From Blackpool to Broadway

Colin McGinn was born in 1950 in Hartlepool and grew up in Gillingham and Blackpool. After taking O-levels in a secondary modern school and A-levels in a grammar school, he graduated with a First in Psychology from the University of Manchester and then took a distinguished B.Phil in Philosophy at Oxford. The

latter degree was quite a challenge at first because as a philosophical neophyte he was surrounded by high-powered specialists who scorned his provincial background. But in 1974 he won the prestigious John Locke prize in philosophy — he was informed of his triumph by Professor Ayer personally and in public as he was waiting for a lecture by Saul Kripke to commence — and this put him on the road to philosophical success. In the same year he got a job as a Lecturer at University College London, alongside Honderich, who had just been promoted from Senior Lecturer to Reader.

McGinn worked for many years at UCL and wrote several books. In 1980 he spent what for him was a glorious semester at the University of California, Los Angeles, with good discussions on what were then the fashionable topics of belief and desire with David Kaplan and Keith Donnellan. In 1982 he survived a less glorious semester at the University of Southern California, where he struggled mightily with Wittgenstein and Kripke on meaning and wasted endless hours playing such video games as Pacman and Galaga in amusement arcades. These visits opened his eyes to the fact that American philosophy was flourishing independently of Oxbridge, and awakened in him the idea that maybe he should emigrate to the United States.

But it wasn't over yet in England. Just after his 1982 visit to California he well-nigh burned himself out writing a book (McGinn, 1984) that tried to correct what he saw as Kripke's misrepresentation of Wittgenstein's views on meaning (Kripke, 1982). To get away from it all in his head, he morosely wrote a novel. For years he had nursed a secondary ambition to make it as a man of letters, as he reveals in his autobiography:

> I had always had a yen to try my hand at fiction. . . . Reading the early novels of Martin Amis (before he got famous) also stimulated me; I liked his combination of literacy and vulgarity, the high and the low (p. 157).

Martin Amis graduated from Oxford with a congratulatory First in 1971, which was the year McGinn graduated, so it was natural to make comparisons. Honderich confirms this with characteristic wit: 'The envy of my small colleague Colin McGinn . . . extended even to wanting to be Martin Amis' (p. 222). Martin's autobiography (Amis, 2000), cast as a record of his relations with his father Kingsley, appeared to such fanfare in 2000 that it must have influenced Colin, whose own life story has a preface dated July 2001.

Indeed McGinn's prose style owes a lot to Martin Amis, and many of the most amusing words in his vocabulary are straight from the Amis *oevre*. McGinn's novel was called *Bad Patches* and written as the first-person story of an unfortunate antihero who suffers gruesome mishaps, works with a pair of stooges called Fock and Fack, and makes out with a female dentist. McGinn commissioned an agent to try to publish it, but no-one was interested. I haven't read the book, of course, but in my mind's eye I can already see Martin's style prints all over it. Martin and I were friends for a while as undergraduates, and I too was fascinated by his early novels. Indeed many years later I too wrote a novel, and my agent also failed to find a publisher for it. Colin may agree that such an enterprise is

born of the sort of dark night of the soul for which the best consolation is good philosophy.

In the summer of 1982, McGinn applied without much hope for the prestigious post of Wilde Reader in Mental Philosophy at the University of Oxford, formerly held until his untimely death by the legendary and charismatic Gareth Evans. Miraculously, it seemed, McGinn was offered the job, apparently because the psychologists saw in him a kindred spirit, more sympathetic than the philosophically stronger candidate Chris Peacocke, who as an All Souls genius may have been rather too Olympian for them. Once McGinn was established as the new mental philosopher, he began working hard on the interface of philosophy and psychology. There he had his greatest insight in philosophy:

> [T]wo sets of thoughts were mingling in my mind at this time: the potential unknowability of reality, and the deeply puzzling nature of the mind–brain relation. . . . One night, as I lay in bed turning these things over in my mind . . . the two sets of ideas locked together. It was one of those flashes of insight that I had read about in other people's memoirs. Maybe the reason we are having so much trouble solving the mind–body problem is that reality contains an ingredient that we cannot know. . . . [I]f we could remedy this ignorance the solution to the problem would be immediate and uncontroversial. . . . We are suffering from what I called 'cognitive closure' with respect to the mind–body problem (p. 182).

The result was his paper 'Can We Solve the Mind–Body Problem?' (reprinted in McGinn, 1991) that first presented the *mysterian* position for which he was later popularized in *Scientific American*. This bleak and rather disappointing insight is McGinn's defining achievement as a philosopher. Most of his other philosophical work, so far as I can see, is of technical or educational interest but not historic, and his early writings made little impact. Said Ted, unkindly: 'McGinn . . . distinguished himself not only as the Wilde Reader in Oxford but also the Wilde Writer' (p. 365). It is a relief to add that McGinn's life story takes no shots at Honderich.

In 1988 McGinn spent a semester at City University of New York. He realized it was time to move out of an increasingly stifling Oxford scene and leave the Thatcherized remnants of British philosophy behind. Soon after his return to England, he was offered a post at Rutgers University in New Jersey and took it. The rest was plain sailing. Living in New York, enjoying street life on Broadway and kayak surfing off Long Island to keep the looming ghost of critical self-consciousness at bay, he found the peace of mind to write his life story.

From Dualistic Identity to Existence

Let us now look briefly at Honderich's views on consciousness. Here I face a methodological problem. Nowhere in his autobiography does Honderich state what those views are in a way that escapes the morass of second thoughts, jargon, and posturing *isms* that envelops too much in the philosophy of mind. Everything is qualified with doubts and nuanced to the competing views of other thinkers. Life is too short to go through all the works he mentions in passing, especially

since he seems to think most of them are wrong, so I shall rest content here with a few brief notes.

Let me illustrate the problem thus. In the year 1971–72, Honderich edited a volume entitled *Essays on Freedom of Action* (Honderich, 1973) whose theme was whether determinism was compatible or incompatible with freedom, on which he says:

> My assembled contributors … had given a majority vote for the answer of Compatibilism, with [Donald] Davidson, [Daniel] Dennett, and [Anthony] Kenny to the fore. . . . David Wiggins, with the aid of a symbol or two of formal logic and 29 substantial footnotes, followed by further material attached to an asterisk, had proved to his satisfaction that much was to be said for the gloom and bravery on the other side. . . . My situation was still one of being inclined to join David Wiggins in the gloom of Incompatibilism, if not at all in the bravery about determinism (2001, pp. 182–3).

See what I mean? But let's plough on. In the year 1977–78, Honderich and Myles Burnyeat edited a volume entitled *Philosophy As It Is* (Honderich, 1979) that included the celebrated paper 'Mental Events' by Donald Davidson, on which Honderich says:

> [W]hat did Donald Davidson mean by saying mind and brain were *identical*? What did he mean if he also said that he was *not* reducing mind to brain, not embracing Eliminative Materialism, not joining those Australians for whom conceiving the *Art of the Fugue* was nothing but a complex physical event — no sweetener for the pill? I had managed to write my brief preface without finding out (p. 227).

When Honderich tries to tackle these questions directly, this is the result:

> The main idea in Identity Theories of mind and brain . . . was that a conscious or mental event, an event with the property of subjectivity, was identical with a brain event. But what did that come to? . . . The first answer was that the conscious event had only the property or properties of subjectivity. . . . This was the madness of mentalizing the brain. . . . Things were just as bad if you started at the other side. . . . This was the absurdity of Eliminative Materialism. . . . There was another answer to the question. . . . You could say the conscious event had *both* the property of subjectivity and also neural properties. Indeed, that seemed to be what was in the minds of such sensible persons as Professor Davidson. . . . Clearly a dualism of two kinds of properties remained when the dualism of two events had been discarded. This deserved the name of being a Dualistic Identity Theory. . . . This was the prideful start of the paper 'Psychophysical Lawlike Connections and Their Problem'. My friend Alastair Hannay honourably published it in his journal *Inquiry* (pp. 244–6).

This seems not to be going anywhere much beyond Davidson's ideas, but let's pursue the thread further anyway. Here's the next gobbit:

> Another anomaly on my mind was Anomalous Monism, the best-known and most intriguing version of the idea that mind and brain do not merely go together but are one thing. It was owned by the aforementioned Don Davidson, the Pied Piper of Berkeley, California. . . . This Dualistic Identity Theory has more to it. . . . The first proposition is the humdrum one that there are causal relations between mental and physical things. . . . The second is that wherever there are causal relations between

> things, the things are connected as a matter of natural or scientific law, nomically connected. . . . The third proposition is that there are no lawlike connections between mental and physical things. Mental things are not a matter of law but are *anomalous* (p. 261).

We can skip the further quotes for this story. Honderich argued that this boiled down to epiphenomenalism and published his claim in *Analysis*. In the next issue, Peter Smith of Sheffield University said that Honderich was confused. In the next, Honderich replied that Smith was confused. In the next, Smith replied that Honderich was extremely confused. In the next, Honderich said something unintelligible (to me at least) about mauve slippers. I spare you the references in *Analysis*. Who cares? Academic catfights are sometimes as daft as they seem.

Honderich picked many such fights. One such arose from the big book by Karl Popper and neurophysiologist John Eccles called *The Self and Its Brain* (Popper & Eccles, 1977):

> The book . . . announced that the Self or Self-Conscious Mind was not tied to the brain, but was its proprietor, somehow free-floating and magnificent. . . . It proved possible for me . . . to concentrate . . . on the basic reason given for the doctrine. This was a piece of Californian nonsense owed mostly to one Benjamin Libet of that state's university. Wrapped up in much experimental evidence in nine scientific papers, it was about a conscious sensation occurring on its own before the brain caught up with it. My refutation, wrapped up in much conceptual clarification, had been accepted as an article by *The Journal of Theoretical Biology*, thereby establishing to its readers that philosophy was not merely the handmaiden of science (pp. 271–2).

In fact, the story is a little more complicated. Libet and his colleagues said:

> [A] dissociation between the timings of the corresponding 'mental' and 'physical' events would seem to raise serious though not insurmountable difficulties for the . . . theory of psychoneural identity (Libet *et al*., 1979, p. 222).

On the basis of Libet's early results, Popper and Eccles said:

> This antedating procedure does not seem to be explicable by any neurophysiological process. Presumably it is a strategy that has been learnt by the self-conscious mind . . . to play tricks with time (1977, p. 364).

It took some years get this clear. To summarize, Libet's most intriguing result was that when a stimulus was applied to the skin of certain patients, it took about half a second before they were consciously aware of that stimulus, yet the patients themselves had the subjective impression that there was no delay at all in their becoming aware of the stimulus. The patients apparently referred the perception of the skin-touching backwards in time by about half a second (Penrose, 1989). Daniel Dennett gives a good account of the story that concludes:

> Where does this leave Libet's experiments with cortical stimulation? As an interesting but inconclusive attempt to establish something about *how the brain represents temporal order* (1991, p. 162).

The full story is long, tangled, and irrelevant here. As I see it, all this illustrates the dangers of waxing too rhetorical about consciousness before the basic science is firmly in place.

To return to Honderich, he continued to engage with relish in academic fights. In 1992–93, he published 'The Union Theory and Anti-Individualism' in *Mental Causation*, a collection of papers edited by Heil and Mele. This paper was the first of several in which his 'long-held truths' about the nature of consciousness and its relation to the brain were defended against what he hoped were passing fashions. The paper featured clashes with Harvard professor Hilary Putnam and California professor Tyler Burge, but it is hard to make sense of the details from his rather florid autobiographical account, as it breaks off abruptly with a paragraph about ongoing editorial work for the *Oxford Companion to Philosophy*. Apparently a righteously angry female contributor mailed him a large brown envelope addressed in large letters to *Gross Yob Honderich*.

In later years, Honderich mellowed somewhat. In November 1994, he presented a paper at a philosophical meeting in Copenhagen in which he discussed John Searle's book *The Rediscovery of the Mind* (Searle, 1992):

> My Copenhagen paper looked at [Searle's] way of seeking to state the truth about consciousness . . . by . . . relying on what he called humble and obvious truths about the mind. One was that a conscious event has a special mode of existence. It exists only as *somebody's* conscious event. It depends for its existence on a 'first person', an 'I'. But what did that mean if it was not a dive into the deep and murky philosophical water? . . . What about the natural idea noticed in his famous paper 'What Is It Like to Be a Bat?' by Thomas Nagel . . . the idea that when something is conscious, there is a way it is like to be that thing. . . . Isn't it inevitable that what we understand by what it is like to be something is *what it is like to be something conscious* or indeed *what it is like to be conscious*? But this is a disaster. . . . The sad conclusion of my paper, which thereafter went into the *American Philosophical Quarterly*, was that humble and obvious truths were no great help in trying to understand consciousness (pp. 355–6).

In spring 1996, preparing for the 1996–97 lectures of the Royal Institute of Philosophy, Honderich boiled down his thoughts on consciousness to a short list:

(1) Conscious events are physical events.
(2) Conscious events are in our heads.
(3) Conscious events are not merely cells.

However, this was not entirely satisfactory: 'It sounded like something awful heard of before in the history of philosophy, in connection with souls, egos and selves — *conscious stuff*, maybe a relative of ectoplasm' (p. 364). In the first RIP lecture, he finally said that *my consciousness consists in the existence of a world*.

> Despite the initial suspicion, my saying my perceptual consciousness consists in a world is indeed not philosophical disaster. . . . We need to be guided by the idea of *consciousness as existence* (p. 371).

Honderich pursued this theme in a series of papers. In May 2002, he mailed me an electronic preprint of the latest installment. The basic idea is that what it is to

be conscious of certain things is for those and related things in a certain way to exist. Whether we can do much more with this idea or not, I like the idea that consciousness consists in a world. I had a similar idea some years ago, independently, and built a lot of formal logic around it. But I spare you the details. Let's just say all these ideas represent work in progress.

The Mysterious Flame

For my money, McGinn's best book — and indeed the best elementary introduction by anyone to the philosophy of consciousness — is *The Mysterious Flame* (1999). Because it presents his views on consciousness more fully than does his rather thin autobiography, I shall quote from it exclusively in this section.

> The central topic of this book is the explanation of consciousness. Suppose I had asked you to imagine waking from a coma without having a brain in your head. You would have been rightly perplexed. Having a brain is what makes it possible to have a mental life. The brain is the 'seat of consciousness'. . . . The machinery of the brain allows the mind to work as it does and to have the character it does. . . . I argue that the bond between the mind and the brain is a deep mystery. Moreover, it is an ultimate mystery, a mystery that human intelligence will never unravel (pp. 4–5).

Thanks in part to the Amis apprenticeship, McGinn's style is clear and direct — a refreshing change from Honderich's ponderous constructions (in a *New Society* review of his early book on punishment, Giles Playfair said, 'Honderich is, to say the least, an ungifted writer of English prose'). McGinn also has the knack of drilling down precisely onto perplexing issues:

> Isn't there some kind of violation of the uniformity of nature in the fact that brains produce consciousness? Brains seem very similar to other parts of animal bodies, being basically a big collection of cells organized according to biochemical principles. Yet there is a yawning chasm between the natures of these entities, because brains produce consciousness and those other meaty organs do not, not even a little bit (p. 9).

> Some people like to harp on the complexity of the brain, as if this gave a clue to its mental productivity. But sheer complexity is irrelevant: merely adding more neurons with more synaptic connections doesn't explain our problem a bit. The problem is how *any* collection of cells, no matter how large and intricately related, could generate consciousness (p. 11).

> What do electricity and cells have to do with conscious subjectivity? How could a conscious self exist *inside* such a soggy clump? It begins to seem that we are all djinns, each magically ensconced in our own personal brain lamps, waiting to be rubbed into life (p. 17).

When it comes to presenting the main historical doctrines on mind and brain, or rather to presenting the straw men that he wishes to cut down with *samurai* swordstrokes of pure reason, McGinn's exposition is clear as a bell:

> Materialism says there is nothing more to the mind than the brain as currently conceived. The mind is made of meat. It *is* meat, neither more nor less. … According to materialism, we are under an illusion about the nature of the mind (p. 18).

> Dualism . . . is best interpreted as the belief that there is no logical relation between brain and mind. There is no possibility of reducing the mind to the brain, because they are separate realms. There are indeed empirical and contingent relations between the two — correlations between mental and physical processes have been discovered — but there is no necessary link between consciousness and the brain (pp. 23–4).
>
> There are two major problems with dualism, the 'zombie problem' and the 'ghost problem'. The zombie problem is that dualism allows us to subtract the mind from the brain while leaving the brain completely intact. . . . The ghost problem is the converse of the zombie problem. If the mind is separate from the body, then not only can the brain exist without the mind but the mind can exist without the brain (pp. 25–7).

But the argument soon homes in on the despairing theme that we may never understand consciousness. For what it's worth, I find the arguments he presents to be bloodless and unconvincing. The general drift is toward the *a priori* position that we do not have reasonable grounds to expect in advance that we will find a satisfactory explanation, despite all the progress we see elsewhere in science and the massive advances we have made recently in the detailed understanding of brain physiology and cognitive function. Here is an example of such an *a priori* argument:

> We certainly cannot infer that *since* we understand the physical world so well it is only a matter of time until we understand consciousness, because consciousness is so different from what has so far yielded to our understanding (p. 36).

The basis for his pessimism is in large part the Chomskian view that the mind is highly modular, with specific innate capabilities. This view has been skillfully popularized in an evolutionary context by Steven Pinker (1997). As McGinn puts it:

> The prevailing view in cognitive psychology today is that the human mind consists of separate faculties, each dedicated to certain cognitive tasks: linguistic, social, practical, theoretical, abstract, spatial, and emotional. The mind is thus as highly structured as the body. . . . Every mental faculty has limits to its achievements and acuity, and necessarily so. . . . We can, it is true, do more with our minds than apes can, but that does not mean that we somehow magically escape the constraints of biology (pp. 40–2).

This model of the mind as a Swiss army pocketknife, a multifunctional kludge, provides a tempting reason to deny the explicability of consciousness, but McGinn offers — apparently unwittingly — a large piece of ammunition that I think we can use to argue *for* the ultimate explicability of how the djinn is ensconced in the meat. He presents the ammo as a general consideration on the logic of thinking:

> Perhaps the most basic aspect of thought is the operation of *combination*. This is the way in which we think of complex entities as resulting from the arrangement of simpler parts. There are three aspects to this basic idea: the atoms we start with, the laws we use to combine them, and the resulting complexes. We find these three basic elements in everything from physics to language to mathematics. . . . The big question

> is this: Is the mode of derivation of the mind from the brain comprehensible according to this kind of combinatorial model? . . . The answer is clearly 'No.' (pp. 56–8).

This stacking of parts into wholes suggests a Lego brick model of mind that uses a single basic strategy for every representational task. If you have enough Lego bricks, or enough neurons in your brain, you can model just about anything. To use another metaphor, just as the information revolution has swept all before it by digitizing any and every content area into the binary logic of bit streams, so the biological breakthrough of endowing a species with a big, labile neuronet that can do nifty combinatorics has swept that species into dominance all over the natural landscape. And the reason I think this is ammunition *for* the explicability of the mind–brain connection is that there is no end of historical cases where the apparently simple *qualitative* nature of some puzzling natural phenomenon yielded — often unexpectedly — to a somewhat more complex conception of it in terms of the quantitative behaviour of otherwise familiar entities. For me, the explanations of heat and the phases of matter in terms of molecular motion and of life and heredity in terms of organic chemistry both offer close enough parallels to suggest that the Lego brick mind, despite its humble evolutionary ancestry, can perform prodigious feats of explanation. It is certainly not clear that the proper answer to McGinn's rhetorical question is no.

McGinn offers another piece of ammunition to his critics when he stresses the humble nature of consciousness:

> Consciousness is not the evolutionary pinnacle, not the most impressive piece of organism design to date. Consciousness, I believe, is biologically primitive and simple, comparatively speaking (p. 62).

This suggests that the whole mind–brain mystery is a mountain from a molehill. The puzzle looks big to us because we are close to it. Indeed, we are each as close as we can be to our own personal copy of the puzzle. *Minds are us*, we have to admit, and then we can't stand back far enough to be objective. Yet McGinn wants to say we face a more substantial cognitive limit:

> To grasp what I am saying about cognitive limitations it is vital to distinguish sharply between the brain as an objective entity in the world and our conception of the brain. The key point is that the objective nature of the brain is not exhausted by our conception of it (p. 66).

Good point, but not a problem. The soggy clump of cells is a red herring. The functioning brain has a feature that does not show up well in blithe philosophical talk of meat or neurons, or even in many clinical or laboratory investigations of brains, namely a surrounding electromagnetic field. This field offers a much more plausible candidate to serve as the substrate of the mind or the seat of the soul than any macromolecular, microtubular, or neural substrate. The objective nature of the brain is that it generates a complex and dynamic electromagnetic field that interacts with the neurons in subtle ways that are not yet well understood. On this issue, the best reference I know is a recent *JCS* paper (McFadden, 2002). An electromagnetic field that interacts with neurons through coherence and phase locking (see the work of Wolf Singer and others cited by McFadden)

has nontrivial quantum properties (like laser beams) and should therefore be seen as a photonic field. Given the undeniable fact that we have a long way to go before we exhaust such fertile conceptions, it is premature to abandon all hope and go canoeing with the mysterians.

Interestingly, the field theory of consciousness that may emerge from the work of McFadden and others suggests a view of the mind that surprisingly echoes the radical dualism of Popper and Eccles. McGinn characterizes this general kind of dualism in theistic terms:

> The picture is that God created your soul and adjoined it to your body for the duration of your mortal life. . . . The brain is merely the organ or instrument of consciousness, not its cause or origin. According to this kind of theistic dualism, the brain is a mediation device used by the soul to influence the movements of the body. The soul itself is a thing apart (pp. 83–4).

The theism is scientifically unhelpful, of course:

> The hypothesis of God simply pushes the question back, either because he himself has complex design or because he is himself a conscious being. . . . The second problem is that . . . sentience is by no means unique to human beings. . . . The third problem is that the mind of an organism is manifestly causally dependent upon its brain, no matter how hard it is to penetrate the nature of this dependence (p. 87).

A materialist dualism obtained by identifying the mind with a photonic field obviates the first problem and turns the second and third into benefits. It also puts us right on the issue of how minds can act on brains:

> What is crucial to hyperdualism is the denial that brains cause consciousness to exist. Rather, consciousness exists in its own right in its own dimension of reality. The brain is not generative; it is merely transducive. As it were, the brain listens to consciousness instead of uttering consciousness. . . . This raises two big questions: How could disembodied consciousness cause anything? and How could the physical sequence of events in the material universe be disrupted by what is going on in the parallel mental universe? (pp. 91–2).

A more practical question is whether we can clarify the phenomenology of photonic fields sufficiently to avoid other problems, such as panpsychism:

> According to panpsychism, the reason we are stumped by the question of how the material brain produces consciousness is that we ignore the fact that consciousness is pervasive in nature. Matter is throbbing with consciousness in all of its manifestations; the brain simply steps the mental volume up high enough for us to notice its presence (p. 95).

If photonic fields are supposed to explain consciousness, this is by no means a trivial problem, since such fields are just about everywhere in nature. The explanation must be a long and tortuous story, and then we are almost back where we started:

> Granted that atoms do not have full-blown mental states, might they not have mental states in a degraded or attenuated sense? . . . No, the idea must be that rocks have what are sometimes called *proto*mental states, states that can *yield* conscious states while not themselves *being* conscious states. . . . The problem with this theory [is that

> it] merely says that matter has *some* properties or other, to be labeled 'protomental,' that account for the emergence of consciousness from brains. But of course *that* is true! It is just a way of saying that consciousness cannot arise by magic; it must have some basis in matter (pp. 98–9).

The photonic theory is still an embryo, too delicate for the cut and thrust of philosophical posturing, but it does bring relativistic and quantum insights about time and indeterminacy right into the heart of the story, where they can help us explain the elementary facts of consciousness. Such a theory may seem like overkill, but that's not the problem that worries McGinn, who stoutly maintains there will be no kill at all:

> If a theory provided a fully adequate explanation of the mind–brain link, it would not really matter how crazy it appeared to us to be. The problem is that no matter *how* crazy we allow ourselves to be, we can never account for the elementary facts of consciousness (p. 104).

This is a *non sequitur* of numbing grossness (to echo Peter Strawson's words about something Kant once said — words that McGinn quotes approvingly in his autobiography as brilliant logical swordplay). But before we toss McGinn's work out with the trash, let's review some of the elementary facts he has in mind.

> [T]here is this strange incongruity in the relation between mind and world: the world outside us is essentially spatial and we represent it that way in our every experience, yet our experience is itself essentially nonspatial. . . . The nonspatiality of consciousness is connected with another feature of it, namely its imperceptibility. Consciousness enables us to perceive the world, but it is not itself a perceptible thing. . . . This is surely part of the reason for the famed infallibility of introspection: you can't be wrong about your conscious states because there is no sense in the idea of these states moving out of range of the introspective faculty (pp. 111–14).

OK so far — nonspatiality, imperceptibility, and infallibility.

> Consciousness appears to be *transparent* to the subject of consciousness. The question is: Is this impression correct? . . . Is every property that is *intrinsic* to my consciousness revealed to my faculty of self-knowledge? (p. 140).

The word *intrinsic* is tricky. McGinn says that the Freudian unconscious, for example, is extrinsic to consciousness, but the *computational* unconscious is intrinsic. This is important, since the computational substructure of conscious thoughts is in a fuzzy zone that may or may not surface in consciousness, as we can see from the routine performance of learned tasks or from the phenomenon of blindsight. But perhaps the most puzzling fruit of consciousness is the sense of self:

> I may not be certain that there is an external world, or even that I have a body, but I am certain that I exist. . . . *Cogito, ergo sum*. . . . If we cannot understand states of consciousness, then it is hardly likely that we will be able to understand the nature of the *subject* of those states. That subject is simply defined to *be* what has those mysterious conscious states. … The deeper question here is how a bunch of cells can become a self *anyway*: What converts biological tissue into that self whose existence so impressed Descartes? The fact is that there are no scientific criteria for the

> appearance of selves; all we have are shaky intuitions about when to declare the onset of selfhood (pp. 156–62).

This is a more complex question that surely involves a lot of cultural baggage. No theory of brainwaves is going to help us decide when abortion is permissible, for example, or where to set the age of responsibility for children. Yet biologists happily talk about the immunological self and psychologists about logical stages in the development of self, so we don't need to be too mysterian about this.

Another basic issue is how far a computational model of mind can take us. This is ground well covered by John Searle and all his debaters. McGinn sides with Searle:

> Mental processes are *not* identifiable with symbol-manipulating algorithms. There are two big problems with the theory. The first is that minds *do* respond to meaning and not just to syntax. . . . My mental processes involve the manipulation of meanings, not merely strings of syntax. I am a *semantic* manipulator, as well as a syntactic one. . . . The second point is that running a program does not guarantee sentience; in fact, it is neither necessary nor sufficient for sentience. It is not necessary because sentience in general does not involve symbolic manipulations (pp. 182–3).

The point seems clear, but there are subtle issues here that proponents of functionalism and machine intelligence can use to fight back. Those who talk too confidently about the computational intractability of semantics are guilty of the sort of reification of meaning that offended Wittgenstein in his later years. Meaning is a treacherous quagmire: even street signs mean what they say whether I read them or not, and even I don't always mean what I say. But we digress. For McGinn, the outcome is clear:

> What follows from all this is not that a robot could not be conscious. What follows is that a robot could not be conscious *in virtue* of being a computer — that is, in virtue of running computer programs (p. 185).

Be that as it may, the more interesting question is whether something analogous holds for *any* purported mechanism of consciousness. For example, it may soon be debatable whether a robot could be conscious *in virtue* of interacting functionally with a dynamically configured photonic field around its core processor. Only time will tell.

To return to McGinn's argument, this much is true:

> The mechanism of consciousness is a mystery. But then how are we to *say* whether an inorganic brain could be conscious? If we knew what made *our* brains conscious, then we could ask whether that property could exist in an inorganic system. But we are in the dark on the question simply because we don't know what makes our brains conscious (p. 197).

This much, however, is not:

> Speaking loftily, it is just a matter of bad cognitive *luck* that we cannot solve the mind–body problem; our minds happen not to have been engineered that way (p. 214).

McGinn's bad luck certainly, but if someone had said something similar about, say, the matter–energy problem a hundred years ago, Einstein would have proved them wrong just three years later. Indeed McGinn seems tantalizingly close to such a revelation:

> My whole point has been that mind and brain form an indissoluble unity *at the level of objective reality*. . . . Objectively we are naturally constituted from smoothly meshing materials, as seamless as anything else in nature. We only *seem* comical because we cannot grasp what this unified reality consists of (p. 230).

But no, his grip is gone. Someone else must do the job.

Toward a Science of Consciousness

Forgive me, but I need to review some history before we can go on to see how to build over all the work of Honderich and McGinn — and many other philosophers — on consciousness. The key figure here will be David Chalmers, who is clearly the leading philosopher of consciousness to have emerged in the last ten years.

Logical positivism had its roots in the late nineteenth century in the work of the Austrian physicist Ernst Mach, whose work was important for Einstein's special theory of relativity (1905). From those roots, the Vienna Circle arose after the First World War and created a tradition that venerated Wittgenstein's *Tractatus* (1922) and gave rise to Ayer's *Language, Truth and Logic* (1936). Wittgenstein had been inspired by Frege and worked with Russell, but the *Tractatus* is a unique work with a deeper ambition. He later repudiated that ambition and spent the following decades working out a more pragmatic view of language and thought. Ayer's work flowed directly from the approach that Russell defined and remained consistent as Oxford orthodoxy evolved toward Wittgenstein's later views.

The positivist movement had its effect on the sciences, not only in physics but also in psychology, where J.B. Watson founded behaviourism and B.F. Skinner and others continued it into the 1960s. In philosophy, too, the influence was still strong in the 1960s. In particular, Willard Van Orman Quine, who was not only Skinner's friend but also an accomplished mathematician, pushed on with Russell's work in mathematical logic and created a new philosophical puzzle, the *indeterminacy of translation*, that meshed well with a behaviourist outlook. Quine and Davidson set the philosophical tone in Oxford in the 1970s, when I was there. The positivist tradition left its mark on cognitive science in the idea that the brain can reasonably be modelled by any computational black box that when fed with appropriate input produces the right output. That mark is evident in Douglas Hofstadter's brilliant and idiosyncratic work (1979) on Gödel's incompleteness theorems and the computational approach to the mind in Artificial Intelligence. After early research in mathematics at Oxford, Chalmers worked with Hofstadter's team in Indiana.

The phenomenological tradition began in Germany in the nineteenth century from the work of Edmund Husserl, who lost a debate with Frege on the

foundations of arithmetic. We can see Husserl's work as an attempt to recreate in a rigorous and scientific manner what Hegel had sketched in his verbose but visionary *Phänomenologie des Geistes* (1807). Husserl's most famous student was Martin Heidegger, whose notoriously obscure book *Sein und Zeit* (1927) founded the existentialist movement most famously associated with Jean-Paul Sartre. Many observers imagine that despite its initial ambitions, the phenomenology of Husserl and his followers is no longer relevant to psychology. Yet any modern science of consciousness has a historical link there. As the perceptive critic Thomas Metzinger says:

> [T]he idea of a 'science of consciousness' is anything but a new idea, especially from the viewpoint of the philosopher. For example, the whole phenomenological movement (and its demise) can be understood in these terms (Metzinger, 1995, p. 4).

In philosophy, too, the phenomenological tradition diverged so far from the Anglo-American tradition that the two can now be seen as quite separate cultural movements, between which understanding is at best limited. To illustrate this claim, I can hardly do better than quote Honderich, reporting on a 1998 BBC radio show in which he participated:

> I pleased myself and some others by first saying of Continental Philosophy that I was like many British philosophers in not allowing my ignorance of it entirely to obstruct my judgement. It was a different kind of thing from ours, and aspired more to the condition of literature or intellectual show-business. It was only disgraceful by our standards (p. 3).

As an aside, one may wonder whether the works of Honderich and others, if not most recent Anglo-American philosophical writings on consciousness, will seem disgraceful to future scientists of consciousness.

Now I can explain the relevance of all this history. Both positivism and phenomenology influenced the development in the 1920s of quantum mechanics by Bohr, Einstein, Dirac, Heisenberg, Schrödinger, and others. The message of positivism was that in science only facts count, and anything that cannot be verified or falsified can be cast out as metaphysics. The facts about black body radiation and electron orbitals in atoms were not consistent with the usual interpretation of classical physics. So out with the classical metaphysics and start again! The message of phenomenology was that the way to start again was to create a systematic account of the experimentally observable phenomena, no more. That account would create its own theoretical frame, and with it a new metaphysical conception of reality.

By and large, this has happened. The new metaphysics is not yet as stable as we might like, as Honderich and McGinn would both insist, but we are making progress. The theoretical work of John Bell in the 1960s and its experimental investigation by Aspect and others in the 1980s have greatly clarified the situation, and enabled us to clean up the interpretational mess left by the pioneers. The new *consistent histories* approach to quantum phenomena championed by Omnès (popularized in Omnès, 1999, and Lindley, 1996) improves on all its

predecessors, and in particular on the Everett *many-worlds* picture favoured by David Chalmers.

So, back to Chalmers. To set the stage, let me again proceed with a series of quotations, this time from (Chalmers, 1996). First, his stated aim:

> In developing my account of consciousness, I have tried to obey a number of constraints. The first and most important is to *take consciousness seriously*. . . . The second . . . is to *take science seriously*. . . . The third constraint is that I take consciousness to be a natural phenomenon, falling under the sway of natural laws. If so, then there should be *some* correct scientific theory of consciousness, whether or not we can arrive at such a theory (pp. xii–xiii).

He starts from Nagel's idea of *what it is like* and introduces qualia:

> We can say that a being is conscious if there is *something it is like* to be that being . . . we can say that a mental state is conscious if it has a *qualitative feel* — an associated quality of experience. These qualitative feels are also known as phenomenal qualities, or *qualia* for short (p. 4).

Dennett's vociferous objections to qualia (for example, in Dennett, 1991) notwithstanding, Chalmers makes extensive use of qualia. I think the proper course is to remain agnostic about them for a while and see where they take us. If we don't like the destination, we can always come back and throw them out.

Next, Chalmers distinguishes two quite distinct concepts of mind, the *phenomenal* and the *psychological*. On the phenomenal concept, mind is characterized by the way it *feels*; on the psychological concept, mind is characterized by what it *does*. Aspiring phenomenologists face a linguistic problem that psychologists do not share, namely that our *language* for phenomenal qualities is derivative on our nonphenomenal language. The result is that our progress in the physical and cognitive sciences has not shed significant light on the question of how and why cognition and consciousness are related. For Chalmers, a useful concept here is *supervenience*, which formalizes the intuitive idea that one set of facts can fully determine another set of facts:

> B-properties *supervene* on A-properties if no two possible situations are identical with respect to their A-properties while differing in their B-properties (p. 33).

> The position we are left with is that almost all facts supervene logically on the physical facts (including physical laws), with possible exceptions for conscious experience, indexicality, and negative existential facts. To put the matter differently, we can say that the facts about the world are exhausted by (1) particular physical facts, (2) facts about conscious experience, (3) laws of nature, (4) a second-order 'That's all' fact, and perhaps (5) an indexical fact about my location (p. 87).

Facts (4) and (5) here are key, I believe, in reconstructing a logical concept of consciousness in the framework of modern physics, where the relativity to the observer of both time and determinacy is reflected in the formalism. But let us not digress. Back to supervenience:

> The failure of consciousness to logically supervene on the physical tells us that no reductive explanation of consciousness can succeed. Given any account of the

> physical processes purported to underlie consciousness, there will always be a further question: Why are these properties accompanied by conscious experience? (p. 106).

Chalmers presents the following argument against physicalism (p. 123):

(1) In our world, there are conscious experiences.
(2) There is a logically possible world physically identical to ours, in which the positive facts about consciousness in our world do not hold.
(3) Therefore, facts about consciousness are further facts about our world, over and above the physical facts.
(4) So materialism is false.

On this basis, Chalmers argues that to bring consciousness within the scope of a theory of everything in fundamental physics, along the lines envisaged by Stephen Hawking and Steven Weinberg, 'we need to introduce *new* fundamental properties and laws' (p. 126). He calls his view *naturalistic dualism*. He claims that that if one takes consciousness seriously, then property dualism is the only reasonable option.

This much would presumably be endorsed warmly by Honderich. However, Chalmers disagrees with McGinn:

> *Mysterianism*. Those unsympathetic to reductive accounts of consciousness often hold that consciousness may remain an eternal mystery. . . . Such a view has been . . . developed by McGinn. . . . Such a view can be tempting, but it is premature. To say that there is no reductive explanation of consciousness is not to say that there is no explanation at all (p. 379).

As a preliminary to his effort toward a nonreductive explanation, Chalmers introduces a new idea, or rather an old one in a new context:

> The primary nexus of the relationship between consciousness and cognition lies in *phenomenal judgments*. . . . Phenomenal judgments are often reflected in *claims* about consciousness: verbal expressions of those judgments (p. 173).

Using it, Chalmers disputes Dennett's claim to have explained consciousness (in Dennett, 1991) by explaining phenomenal judgments, in effect reductively. He argues that Dennett exploits the knife-edge between the phenomenal and psychological realms, and points out that what we need to explain are not the judgments but experiences themselves. We shall return to this distinction between the words and the referents of those words — between syntax and semantics — in the context of the view of consciousness as computation.

Having established that we cannot expect to create a reductive theory of consciousness, Chalmers sets about looking for a nonreductive theory:

> The most promising way to get started in developing a theory of consciousness is to focus on the remarkable *coherence* between conscious experience and cognitive structure. The phenomenology and the psychology of the mind do not float free of each other; they are systematically related (p. 218).

For Chalmers, the central correlation between physical processing and experience is the coherence between consciousness and *awareness*. What gives rise directly to experience is not oscillations or temporally extended activity or high-quality representations, but the process of direct availability for global control. This relates to the global workspace theories of Baars and others.

Chalmers considers at length the suggestion that consciousness arises in virtue of the functional organization of the brain. This leads to an extended but rather inconclusive discussion of the protean science of information:

> This treatment of information brings out a crucial link between the physical and the phenomenal: whenever we find an information space realized phenomenally, we find the same information space realized physically. And when an experience realizes an information state, the same information state is realized in the experience's physical substrate (p. 284).

> A conscious experience is a realization of an information state; a phenomenal judgment is explained by another realization of the same information state. And in a sense, postulating a phenomenal aspect of information is all we need to do to make sure those judgments are truly correct: there really *is* a qualitative aspect to this information, showing up directly in phenomenology and not just in a system of judgments (p. 292).

> It is sometimes suggested from within physics that information is fundamental to the physics of the universe. . . . This 'it from bit' view is put forward by [John Archibald] Wheeler. . . . To each fundamental feature of the world there corresponds an information space, and wherever physics takes those features to be instantiated, an information state from the relevant space is instantiated (pp. 302–3).

This is deep stuff indeed, but necessary. A science of consciousness must engage fundamental physics, and I believe that information will be central in that engagement. Wheeler only glimpsed the new realm, but David Deutsch has recently done great work to help establish a revolutionary new science of quantum information (see Deutsch, 1997, and Nielsen, 2000). However, Chalmers may have travelled too far with the prequantum computationalists:

> I . . . argue that the ambitions of artificial intelligence are reasonable . . . there is a nonempty class of computations such that the implementation of any computation in that class is sufficient for a mind, and in particular, is sufficient for the existence of conscious experience (p. 314).

This defence of the strong AI claim is natural enough for a former colleague of Hofstadter but it puts Chalmers into direct conflict with John Searle, not to mention both Honderich and McGinn. At first blush, it looks like a mere confusion of computation with consciousness, but the issues are too subtle for summary dismissal here.

At last, Chalmers brings us back to quantum physics, where we started with positivism and phenomenology:

> I . . . argue that we can reconceive the problems of quantum theory as problems about the relationship between the physical structure of the world and our

> experience of the world, and that consequently an appropriate theory of consciousness can lend support to an unorthodox interpretation of quantum mechanics (p. 334).

The unorthodox interpretation Chalmers has in mind is the Everett interpretation (see also Lockwood, 1989; Deutsch, 1997), on which he says:

> Everett's view is sometimes called a *many-worlds* interpretation . . . but the view I am discussing is more accurately a *one-big-world* interpretation. . . . On this view, if there is any splitting, it is only in the minds of observers. As superpositions come to affect a subject's brain state, a number of separate minds result, corresponding to the components of the superposition. Each of these perceives . . . a *miniworld*, as opposed to the *maxiworld* of the superposition. . . . Everett calls his view a *relative-state* interpretation: the state of a miniworld . . . only counts as the state of the world *relative* to the specification of an observer (p. 347).

But this is an old view. The modern view of complex macroscopic systems and their state statistics does not support the idea that the coherent miniworlds in a superposition would be big enough to be perceived by a human mind as worlds (so at last the Schrödinger's cat nightmare is banished — see Lindley, 1996). At best, we may be able to construct a sense in which the popping of such miniworlds out of superposition would correspond to the appearance in time of individual qualia in an atomized phenomenal manifold. This way we could make qualia more respectable and retrospectively vindicate Chalmers' faith in them. Wittgenstein liked the idea that a cloud of philosophy could condense into a drop of grammar, and indeed his own early philosophy of truth conditions condensed into the binary grammar of bit strings. Soon, perhaps, the philosophy of worlds multiplying and collapsing will condense into the physics of popping qubits in a quantum foam of experience. But to pursue this here would led us too far afield.

Autobiography as Philosophy

We have flown through a cloudscape of cosmic dimensions, and it is time to land again in the familiar world of human lives. How do autobiographies help us to understand consciousness?

Daniel Dennett claims to explain consciousness (Dennett, 1991), and the argument is temptingly close to what we want. For Dennett, to be conscious is to run a virtual machine in the brain that spins an ongoing autobiography from the accumulating increments of experience. So let's look closer. Dennett starts with this problem:

> Events in consciousness . . . are *experienced* by an *experiencer*, and their being thus experienced is what makes them . . . *conscious* events. . . . And the trouble with brains, it seems, is that when you look in them, you discover that *there's nobody home* (p. 29).

To avoid the sort of problems the behaviourists had with phenomenology, Dennett introduces *heterophenomenology*. If *autophenomenology* is my theoretical account of my own subjective experience, heterophenomenology is my

account of someone else's reported inner experience. This is analogous to fiction, where the reader lets the text define a fictional world, and whatever the author says, so long as it makes sense, is satisfied in that world. Thus the scientific heterophenomenologist allows that the experimental subject's reports define a *heterophenomenological world* that satisfies whatever the subject says, so long as the subject remains coherent and consistent.

Dennett proposes a *multiple drafts* model of the mind. According to this model, all varieties of perception — indeed, all varieties of thought or mental activity — are accomplished in the brain by parallel, multitrack processes of interpretation and elaboration of sensory inputs. These drafts are autobiographical histories. And like schoolbook history, they can falsify the facts. Dennett distinguishes two kinds of falsification: *Orwellian* and *Stalinesque*. Orwellian revisions of history use artful confabulation to create false memories of experiences that never occurred. Stalinesque revisions stage false experiences that obscure and replace the original facts. It seems that our cognitive processes use either or both of these ploys to clean up our memories and redraft our the stories of our selves.

All this may look fine and dandy, but it seems to leave consciousness untouched. Here Dennett offers a good biological analogy:

> There is a nice parallel between . . . the origins of sex and the origins of consciousness. There is almost nothing *sexy* (in human terms) about the sex life of flowers, oysters, and other simple forms of life, but we can recognize . . . the foundations and principles of our much more exciting world of sex. Similarly, there is nothing particularly *selfy* (if I may coin a term) about the primitive precursors of conscious human selves, but . . . we must begin at the beginning (pp. 172–3).

Essentially, Dennett tells an evolutionary story here. But rather than base it on genes, he uses Dawkins' more fanciful notion of memes, a notion now well established in modern folklore due in part to the efforts of Sue Blackmore (Blackmore, 1999). Dennett concludes:

> Human consciousness is *itself* a huge complex of memes (or more exactly, meme-effects in brains) that can best be understood as the operation of a '*von Neumannesque*' virtual machine *implemented* in the *parallel architecture* of a brain that was not designed for any such activities. The powers of this *virtual machine* vastly enhance the underlying powers of the organic *hardware* on which it runs . . . (p. 210).

A machine with a von Neumann architecture is centralized, linear, and serial, indeed a fairly literal implementation of the ideally minimalist architecture of a Turing machine. A virtual machine is a software construction that can run on quite different machines. For example, a Java virtual machine can be delivered through a browser to run on your desktop hardware, whatever model you have. In this sort of way, Dennett proposes, culture delivers a serial self through language to our cerebral wetware, where it sits and grows and enslaves the little cognitive demons that make up our biologically evolved modular mind. Here Dennett coins another richly evocative term:

> In our brains, there is a cobbled-together collection of specialist brain circuits, which, thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well designed virtual machine, the *Joycean machine* (p. 228).

However, the Joycean stream of consciousness so familiar to lovers of *Ulysses* is a dangerous metaphor for a philosopher who has set his face against Cartesian dualism:

> There is no single, definitive 'stream of consciousness,' because there is no central Headquarters, no Cartesian Theater where 'it all comes together' for the perusal of a Central Meaner. Instead of such a single stream (however wide), there are multiple channels in which specialist circuits try, in parallel pandemoniums, to do their various things, creating Multiple Drafts as they go. Most of these fragmentary drafts of 'narrative' play short-lived roles in the modulation of current activity but some get promoted to further functional roles, in swift succession, by the activity of a virtual machine in the brain. The seriality of this machine (its 'von Neumannesque' character) is not a 'hard-wired' design feature, but rather the upshot of a succession of coalitions of these specialists (pp. 253–4).

At last we come to Dennett's big claim about consciousness:

> Anyone or anything that has such a virtual machine as its control system is conscious in the fullest sense, and is conscious *because* it has such a virtual machine (p. 281).

This sets Dennett against both Honderich and McGinn, but presumably puts him in some kind of agreement with Chalmers, despite Dennett's failure to explain *experience* (as opposed to judgments about experience) and his emphatic repudiation of qualia. Chalmers says that the implementation of certain computations is sufficient for a mind and for the existence of conscious experience. The computations performed by Joycean virtual machines would seem to make them good candidates for such mind machines. So despite the complete absence both of new physics and of a role for qualia in the autobiographical view, we seem to have made progress. If new physics can explain qualia, perhaps we can allow them back. Otherwise, why not let them go?

Science and the Self

Consciousness and the self are related. The self is what *has* consciousness. To the extent that we *are* conscious, given that our consciousness can be as limited as both Freudian orthodoxy and heterophenomenology suggest, we have a more or less successfully running inner narrative of a self that serves as the more or less sharp focus of all our internal and external experiences. Dennett is stolidly biological about the self:

> But the strangest and most wonderful constructions in the whole animal world are the amazing, intricate constructions made by the primate, *Homo sapiens*. Each normal individual of this species makes a *self*. Out of its brain it spins a web of words

> and deeds, and, like the other creatures, it doesn't have to know what it's doing; it just does it (p. 416).

Of course, we philosophers *do* have to know what we're doing, otherwise we're out of a job. And despite what both Chalmers and Dennett say, I believe that what we're doing is a lot more than programming. A self is an autobiographical wrapper for a pre-existing entity, an entity that accumulates physical experience and only later, after infection with the memes of culture, puts those experiences into words. This entity may or may not be the photonic field that I guess it could be, but whatever it is, physics will have a lot to say about it. We need some fundamental new science here. On that I agree with McGinn.

To return to McGinn's mysterianism for a moment, many years ago he made the following tentative (and tangled) approach toward suggesting how we might look for an explanatory theory of consciousness:

> There has to be more to consciousness than there seems to be or else it could not depend upon the physical world in the way we know it does. . . . It may help to bring this idea into focus if I contrast it with two proposals made by Thomas Nagel. . . . The first proposal . . . is that subjective experience might be describable in objective (though nonphysical) terms, and that such an 'objective phenomenology' might put us in a better position to understand the physical basis of experience. … Nagel's second proposal . . . is that the real nature of conscious states might just consist in states of the brain. . . . The kind of hidden structure I envisage would lie at neither of the levels suggested by Nagel: it would be situated somewhere between them. Neither phenomenological nor physical, this mediating level would not (by definition) be fashioned on the model of either side of the divide, and hence would not find itself unable to reach out to the other side. Its characterization would call for radical conceptual innovation (which I have argued is probably beyond us). (McGinn, 1991, pp. 103–4.)

Not surprisingly, Dennett pounces on this suggestion triumphantly:

> The 'software' or 'virtual machine' level of description . . . is exactly the sort of mediating level McGinn describes: not explicitly physiological or mechanical and yet capable of providing the necessary bridges to the brain machinery on the one hand, while on the other hand not being explicitly phenomenological and yet capable of providing the necessary bridges to the world of content, the world of (hetero-) phenomenology. We've done it! We *have* imagined how a brain could produce conscious experience (Dennett, 1991, p. 434).

I can imagine McGinn replying that heterophenomenology is not autophenomenology: the uniquely vivid quality of *my* experience remains unexplained.

This reply is devastating to any theory of consciousness that fits within science as we now know it, because that science is built from a third-person perspective. Before we can crack the problem of first-person experience, we need a logico-physical frame that can accommodate the asymmetry between the first-person and third-person perspectives. Insistence upon the importance of this problem is essentially Chalmers' great contribution to the debate.

A first move toward a solution is to assert that first-person consciousness is instantiated *uniquely* in a world. Only first-person consciousness requires an

autophenomenological analysis. Third-person consciousness is tractable in psychology or in heterophenomenology. This drastically simplifies the task of building a theory of worlds. A world is an entity with the logico-physical property of being epistemically centred on a unique self. Each self constructs its own evolving world. To recall Chalmers' take on Everett, we can call such a world a *miniworld*. Our miniworlds are largely congruent, but they differ in where they locate their central point. The corresponding *maxiworld* is all of reality as we know it. The central self for the maxiworld is not a personal self but a *cosmic* self. Our miniworlds are modelled after the maxiworld, rather like tabletop globes are modelled after Planet Earth.

In this view, the full concept of consciousness itself is reflected in the vast universe of science. The universe feels like pure consciousness. Consciousness and the cosmos are complementary. Pure consciousness is *timeness* (to exapt a word coined by Llinás, 2001, p. 120) and the cosmos is a spatial manifold (with four extended and perhaps seven compact dimensions), yet spacetime is a single reality. Moreover, as our science develops, our concept of consciousness expands to reflect it. So, to complete the circle of my story, the history of science is the autobiography of consciousness. No wonder McGinn despaired of understanding consciousness!

Let the final thought here go to Honderich. In the coda to his autobiography he says:

> A human life, any human life that has lasted a while, has a fullness that can seem greater than that of any other single subject-matter. … Each life or entire consciousness and carry-on, in a sense that may one day be made explicit, is *a world*, a world going on through time and one that includes other people and more (p. 389).

Consciousness reflects the existence of a world. The logic and physics of worlds is the logic and physics of consciousness. A human life is a microcosm, a drop in the cosmic ocean.

References

Amis, Martin (2000), *Experience* (London: Jonathan Cape).

Ayer, Alfred J. (1936), *Language, Truth and Logic* (London: Gollancz, 2nd edn, 1946).

Blackmore, Susan (1999), *The Meme Machine* (Oxford: Oxford University Press).

Chalmers, David J. (1996), *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press); see also Chalmers' homepage — www.u.arizona.edu/~chalmers — for updates of the book's ideas and numerous links to other sites of philosophical interest.

Dennett, Daniel C. (1991), *Consciousness Explained* (London: Allen Lane).

Deutsch, David (1997), *The Fabric of Reality* (London: Allen Lane).

Heil, John and Mele, Alfred (ed. 1993), *Mental Causation* (Oxford: Oxford University Press).

Hofstadter, Douglas R. (1979), *Gödel, Escher, Bach: An Eternal Golden Braid* (New York: Basic Books).

Honderich, Ted (ed. 1973), *Essays on Freedom of Action* (London: Routledge & Kegan Paul).

Honderich, Ted (1988), *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes* (Oxford: Oxford University Press).

Honderich, Ted (1993), *How Free Are You? The Determinism Problem* (Oxford: Oxford University Press)

Honderich, Ted (ed. 1995), *The Oxford Companion to Philosophy* (Oxford: Oxford University Press).

Honderich,Ted (2001), *Philosopher: A Kind of Life* (London: Routledge).

Honderich, Ted and Burnyeat, Myles (ed. 1979), *Philosophy As It Is* (London: Allen Lane).
Kripke, Saul A. (1982), *Rules and Private Language* (Oxford: Basil Blackwell).
Libet, B., Wright, E.W., Feinstein, B. and Pearl, D.K. (1979), 'Subjective referral of the timing for a conscious sensory experience', *Brain*, **102**, pp. 193–224.
Lindley, David (1996), *Where Does the Wierdness Go? Why Quantum Mechanics Is Strange, But Not as Strange as You Think* (New York, Basic Books).
Llinás, Rodolfo R. (2001), *I of the Vortex: From Neurons to Self* (Cambridge, MA: MIT Press).
Lockwood, Michael (1989), *Mind, Brain and the Quantum: The Compound 'I'* (Oxford: Basil Blackwell).
McFadden, Johnjoe (2002), 'Synchronous firing and its influence on the brain's electromagnetic field: Evidence for an electromagnetic field theory of consciousness', *Journal of Consciousness Studies*, **9** (4), pp. 23–50.
McGinn, Colin (1984), *Wittgenstein on Meaning: An Interpretation and Evaluation* (Oxford: Basil Blackwell).
McGinn, Colin (1991), *The Problem of Consciousness: Essays Towards a Resolution* (Oxford: Basil Blackwell).
McGinn, Colin (1999), *The Mysterious Flame: Conscious Minds in a Material World* (New York: Basic Books).
McGinn, Colin (2002), *The Making of a Philosopher: My Journey Through Twentieth-Century Philosophy* (New York: HarperCollins).
Metzinger, Thomas (ed. 1995), *Conscious Experience* (Paderborn: Schöningh/Exeter: Imprint Academic).
Nielsen, Michael A. and Chuang, Isaac L. (2000), *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press).
Omnès, Roland (1999), *Understanding Quantum Mechanics* (Princeton: Princeton University Press).
Penrose, Roger (1989), *The Emperor's New Mind – Concerning Computers, Minds, and the Laws of Physics* (Oxford: Oxford University Press).
Pinker, Steven (1997), *How the Mind Works* (London: Allen Lane).
Popper, Karl R. and Eccles, John C. (1977), *The Self and Its Brain* (New York: Springer International).
Searle, John (1992), *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).
Wittgenstein, Ludwig (1922), *Tractatus Logico-Philosophicus* (London: Routledge and Kegan Paul, trans. Pears, D.F., McGuinness, B.F., 1961).