

The Case for Rorts

DANIEL C. DENNETT

In the late 1960s, I created a joke dictionary of philosophers' names that circulated in xerox form, picking up new entries as it went. The first few editions were on Ditto masters — those pre-photocopy days. The 7th edition, entitled *The Philosophical Lexicon*, was the first properly copyrighted version, published for the benefit of the American Philosophical Association in 1978, and the 8th edition (brought out in 1987), still available from the APA. I continue to receive submissions of further entries, but doubt that there will ever be a 9th edition. The 8th edition lists two distinct entries for Dick Rorty:

rort, an incorrigible report, hence rorty, incorrigible.

and
 a rortiori, adj., true for even more fashionable continental reasons.

These were submitted to me years apart, inspired by two distinct epochs of Rorty's work. It may be hard to see the connecting threads between the Princeton professor whose tightly argued "Incorrigibility as the Mark of the Mental" (1970) and "Functionalism, Machines, and Incorrigibility" (1972) were aimed specifically at the smallish clan of analytic philosophers of mind, and the international man of letters described by Harold Bloom as the most interesting philosopher in the world. Can we see the stirrings of Rorty's later ideas in between the lines of his early papers in the philosophy of mind? Perhaps, but that will not be my topic.

I want to go back to Rorty's papers on incorrigibility,¹ not for historical clues about how to read the later, more widely influential Rorty, but in order to expose an excellent insight lurking in his claim that incorrigibility is the mark of the mental. It went unrecognized at the time, I think, because the reigning methodology in that brand of analytic philosophy ignored the sorts of questions that would have provoked the relevant discussion. While the incorrigibility papers were sufficiently influential — or at least notorious — to anchor an entry in the *Lexicon*, they have never been properly appreciated by philosophers of mind, myself included (of all people). I say "of all people" because Dick Rorty has always drawn explicit links between his ideas and mine, and has played a major role in drawing philosophers' attention to my work. If

be with Jefferson or FDR. When it pertains to quarks or superstrings they are, I would imagine, with figures such as Newton and Bohr. These imaginary conversations provide material for the stories we tell ourselves in order to decide whether we are progressing or regressing. (Think of Germans like Heidegger and Tillich holding imaginary conversations in 1933 about Hitler with such figures as Schiller and Fichte. Heidegger convinced himself that Hitler meant progress and thus that he was being rational in turning Nazi. Tillich convinced himself that Hitler meant regress and that he was being rational in preparing for exile.)

Putnam has, in recent years, become convinced that something like direct perceptual realism is the key to avoiding the traps into which he sees me as having fallen. So, as is evident from his Dewey Lectures, he finds McDowell's *Mind and World*, and James' *Radical Empiricism*, much more promising than I do. By contrast, those of us who see Sellars' "Empiricism and the Philosophy of Mind" as pretty much the last word philosophers need utter about perception, and as as devastating a critique of phenomenism as we shall ever have, see direct perceptual realism as an unfortunate throwback to Cartesianism. We think questions about what we really see as bad as questions about what is really real.

Putnam's recent alliance with Stanley Cavell, Cora Diamond, and James Conant – his emphasis on the Ordinary and on the need to avoid putting forward theses in philosophy – seems to me an unfortunate throwback to pre-Hegelian attempts to find something ahistorical to which philosophers may pledge allegiance. The Ordinary strikes me as just the latest disguise of the *ontos on*. I think Putnam was on the right, historicist, neo-Hegelian track in his *Reason, Truth and History*, and I wish that he had stayed on it. For that book seems to me one of the most heartening and inspiring documents of recent analytic philosophy. When testing myself for rationality, I often find myself conducting imaginary conversations with the Putnam who wrote that book.

I read *Reason, Truth and History* as an (unfortunately unsuccessful) attempt to get analytic philosophers to break away from the scientism of Reichenbach and Carnap – to get them to stop being, in Putnam's words, "too realistic about physics and too subjectivistic about ethics."³ Almost twenty years after the appearance of that book, alas, we still find philosophers wondering about the objectivity of moral values, the place of value in a world of facts, and so on – the problems that Kant took seriously and Hegel did not. As long as philosophers do so, they will use expressions like "just a story" and "just conversation," and will exhibit other symptoms of a regression to metaphysics. I regret the reappearance of those symptoms in Putnam's more recent work, but I owe a great debt to his most historicist, least metaphysical, book.

Notes

1 *Reason, Truth and History*, p. 163.

2 *Reason, Truth and History*, p. xi.

3 *Ibid.*, p. 142.

anybody was in a position to see the virtues of his position, it was I, and while I can now retrospectively see that I did indeed subliminally absorb his message and then re-invent some of it in my own terms (without sufficient acknowledgement), I certainly didn't proclaim my allegiance to, or even deign to rebut, clarify or expand upon, those claims.

If my take on this is right, it means that Dick also didn't quite appreciate the strengths of his own idea, and might even have been misled to some of his more fashionable and famous ideas by a slight misappreciation of the import of his claims about incorrigibility, but I won't pursue that surmise here. If I am right, he will have succeeded in spite of himself in making the sort of contribution to science – to our objective knowledge of the way the world, and the mind, is – that he has abjured as a philosophical aspiration. His own philosophical “conversation” turns out to be more than just conversation. He will perhaps reply that all I have shown is that today his ideas about incorrigibility have more political viability, more charismatic oomph in today's conversations than in those of the early 1970s. But I want to insist that the reason they do is that they show us something interesting about how reality may be represented.

What is the Status of Rorty's Thesis?

His central thesis is as follows:

What makes an entity mental is not whether or not it is something that explains behavior, and what makes a property mental is not whether or not it is a property of a physical entity. The only thing that can make either an entity or a property mental is that certain reports of its existence or occurrence have the special status that is accorded to, for instance, reports of thoughts and sensations – the status of incorrigibility. (1970, p. 414)

Incorrigibility is to be distinguished from infallibility. It is not that these reports could not possibly be mistaken, but just that “certain knowledge claims about them cannot be overridden” (p. 413). This immediately tilts the playing field, of course, by trading in a host of tempting but indefensible metaphysical claims for an epistemological or even sociological claim. This is just a fact, Rorty suggests, about a “linguistic convention,” about the way we treat claims, not a fact about the reality of whatever those claims are about. But at the same time his thesis is not a mere anthropological observation: certain claims cannot be overridden, he suggests, given the role they play in our shared life. (As we shall see, it is this modal claim that never got sufficient attention – from Rorty or his readers – back in the 1970s.)

What goes without saying is that these incorrigible reports are “first person” reports, reports about one's own states and events, to which one is presumed to have one or another sort of “privileged access.” This term of art, once so familiar in philosophical writing about the mind, has been eclipsed for some time by other ways of attempting to characterize the crucial asymmetry: Thomas Nagel's (1974) “what it is like” formula or John Searle's (1980, 1983) championing of first person primacy, for instance. It is easy to understand Rorty's lack of sympathy for these later attempts. Far from having overlooked or underestimated the importance of the “first person point of view,” he had declared it “the mark of the mental” – but he had also provided a demystifying account of how and why it had emerged, and why it was no bulwark against creeping

"third person" materialism. Privileged access is real enough, Rorty was saying, and is indeed the premier feature of mentality, but it is no big deal, metaphysically. It is this deflationary doctrine that I want to re-examine, saving it from some Rortian excesses.

His claim may be expressed with somewhat different emphasis: what makes an entity a "first person," a thing it is like something to be, is that some of its emissions or actions are treated not just as reports, but as incorrigible reports. We vouchsafe an entity a mind by vouchsafing it a certain epistemic privilege with regard to the covert goings-on that control it. A mind is a control system whose self-reports cannot be overridden by third-person evidence.

Could there even be such a control system? One of Rorty's shrewdest observations is that our underlying materialist skepticism about this very possibility is the chief factor that propels us towards dualism and other mysterious doctrines:

Only after the emergence of the convention, the linguistic practice, which dictates that first-person contemporaneous reports of such states are the last word on their existence and features, do we have a notion of the mental as incompatible with the physical (and thus a way of making sense of such positions as parallelism and epiphenomenalism). For only this practice gives us a rationale for saying that thoughts and sensations must be *sui generis* – the rationale being that any proposed entity with which they could be identified would be such that reports about its features were capable of being overruled by further inquiry. (1970, p. 414)

It does seem at first blush as if the states and events in any material or physical control system would have to be exactly as knowable by "third persons" as by their "owner," and if this is so, then no such states and events could be thoughts and sensations. It seems to follow that in any purely material entity, first-person privilege would evaporate, at which point there would be nothing left to anchor the mental at all. Rorty does not shrink from this implication: in fact, he views his 1970 paper as explicitly arguing for a version of eliminative materialism (p. 401). He has his materialist concede that

it might turn out that there are no entities about which we are incorrigible, nearly or strictly. This discovery would be made if the use of cerebrosopes (or some similar mechanism) led to a practice of overriding reports about mental entities on the basis of knowledge of brain states. If we should, as a result of correlations between neurological and mental states, begin taking a discovery of a neurological state as better evidence about a subject's mental state than his own report, mental states would lose their incorrigible status and, thus, their status as mental. (p. 421).

He contemplates with equanimity the Churchlandish alternative:

If it came to pass that people found that they could explain behavior at least as well by reference to brain states as by reference to beliefs, desires, thoughts, and sensations, then reference to the latter might simply disappear from the language. (p. 421).

But we need to pause and disentangle some issues, for there are apparently more possibilities than Rorty discusses. First, as just noted, there's standard eliminative materialism: the triumph of neuroscience and its "cerebrosopes" would – and should – lead to the demise of mentalistic language, and we would all cease talking *as if* there were minds and mental events, a clear improvement in our conceptual scheme from the perspective of Occam's Razor. But there is another prospect: it could also happen,

for all Rorty has said, that people *mistakenly* crown neuroscience the victor, overrating the reliability of third-person theory and abandoning their linguistic practice, coming eventually to treat subjects' self-reports as unprivileged, even though they were in fact reliable enough to sustain (to justify?) the linguistic convention that mentality depends on. This would be the evaporation of the concept of mind from that culture, on Rorty's analysis, but would it also mark the death of the minds themselves? Although people's brains, their hardware, would be up to the task, their attitudes towards their own authority would shift, thereby adjusting the software running on that hardware. Could this diminish their real competence, leading to the loss of the very prowess that is the mark of the mental? As the mind-constituting practice waned, would people lose their minds? What would that be like? Could people come to view all their own first-person reports as unprivileged? What would they say – "We used to have minds, but now we just have brains"?

Would their minds cease to exist once this rush to misjudgment took place? For current Rorty, this is surely a paradigm of a misguided question, assuming, as it does, that there is a neutral standpoint from which the Truth of the ontological claim could be assessed. But many of us unre(de)constructed types may think we can take these questions about the justification and confirmation of our representations more seriously than he now allows. (In fact, he tries to soften this blow by granting scientists and other public and private investigators what he has described to me as a "vegetarian" concept of representation – not the whole ontological hog, but some sort of internal realism in which "facts" may be distinguished from "fictions" – but keep those scare-quotes handy. I think, however, that once this vegetarian concept of representation is exploited to the hilt, we will have enough of a "mirror of nature" in our hands to satisfy all but the most hysterical Realists.)

Back in 1970, the ethos of analytic philosophy let Rorty glide rather swiftly over the question of what the grounds for adopting this linguistic practice might be. In that paper he doesn't emphasize the fact that this innovation might be motivated, or defended against criticism (rightly or wrongly), but he also doesn't treat it as if it would have to be a surd memic mutation, a random happening that had huge consequences for our conceptual scheme but was itself undesigned and beyond defense or criticism. To describe the change in linguistic practices that would amount to the birth of minds, he exploits an elaboration of Wilfrid Sellars' (1963) justly celebrated just-so story about Jones, "the man who invented the concept of mind" (p. 411). Jones, Rorty reminds us, organized his shrewd observations of the comings and goings of people into a theory which postulated covert events and states in people's heads, the history of which would account for all their overt behavior. Jones then trained all the people in the fine art of making non-inferential reports about these states and events he had posited. When the training was complete, he had succeeded in transforming people from relatively inscrutable objects of theoretical analysis into reliable divulgers of their own internal workings.

According to Rorty, those who went along with Jones found that, when the behavioral evidence for what Smith was thinking about conflicted with Smith's own report of what he was thinking about, a more adequate account of the sum of Smith's behavior could be obtained by relying on Smith's report than by relying on the behavioral evidence. (p. 416).

This passage needs some emendation. Smith's report is part of the behavioral evidence, surely, and a particularly revealing part (when interpreted as a report, not as mere lip-

flapping). What Rorty means is that Smith's report, interpreted as a speech act, is recognized as providing a more adequate account than all the other behavioral evidence. He imagines that once this appreciation – it might be misappreciation – of the power of self-reports to trump other evidence was in place,

it became a regulative principle of behavioral science that first-person contemporaneous reports of these postulated inner states were never to be thrown out on the ground that the behavior or the environment of the person doing the reporting would lead one to suspect that they were having a different thought or sensation from the one reported. (p. 416)

But why should this become a regulative principle? Why turn the recognition of high reliability – what Armstrong had called “empirically privileged access” (Rorty, 1970, p. 417) – into a constitutive declaration of incorrigibility? Is this just an unmotivated overshooting of social practice, a bandwagon effect or other byproduct of enthusiasm for Jones' theory? Or might there be some deeper reason – an actual justification – for thus shifting the very criteria (to speak in 60s-talk) for the occurrence of mental phenomena?

Rorty's linguistic convention is close kin (a heretofore unacknowledged ancestor) to that by I attribute to “heterophenomenologists” (1991): deliberately permitting the subject's word to constitute the subject's “heterophenomenological world,” creating by fiat a subjective or first-person perspective whose details then become the *explicanda* for a materialist, third-person theory of consciousness. I took the existence of a widespread belief in the primacy of the first-person point of view as given, and characterized heterophenomenology as the neutral method science could – and does – use to investigate the relations between the subjective and objective. Rorty's papers suggest that the emergence of a first-person point of view is itself an effect of a similar burden-shifting move.

In his 1972 paper, Rorty hints at the point I now want to examine in more detail:

if, with respect to a very sophisticated machine, we found that certain states played roles in its behavioral economy very close to those which being frantically hungry, thinking of Vienna, etc., played in ours, then (given that the machine reported on such states and reported making no inferences to such reports) we might decide to extend the same heuristic rule to the machine's reports of those states. But if we then found that the *simplest and most fruitful* [emphasis added] explanations of the machine's behavior involved overriding these reports, we should cease to apply this rule. (1972, p. 215)

This suggests that simplicity and fruitfulness were the grounds for “extending the heuristic rule” in the first place, but why? How? Let us expand the account of this intuition pump, guiding and supporting our judgments by some facts that could only have been dimly imagined in 1972. There is today an entity in roughly pre-Jonesian position, a plausible candidate (with some optimistic projections of engineering) for elevation to first-person status: Cog, a “very sophisticated machine” indeed.

The Birth of Cog's Mind: a Just-So Story

At the AI Lab at MIT, Rodney Brooks and Lynn Andrea Stein are leading a team (of which I am a member) that is currently attempting to create a humanoid robot called

Cog. Its name has a double etymology: on the one hand, Cog is intended to instantiate the fruits of cognitive science, and on the other, it is a concrete machine situated in the real, non-virtual world, with motors, bearings, springs, wires, pulleys – and cogs. Cog is just about life-size – that is, about the size of a human adult. Cog has no legs, but lives bolted at the hips, you might say, to its stand. This paraplegia was dictated by intensely practical considerations: if Cog had legs and could walk, it would have to trail a colossally unwieldy umbilical cord, carrying power to the body and input-output to its brain, which is about the size of a telephone booth and stands to the side, along with large banks of oscilloscopes and other monitoring devices. No batteries exist that could power Cog's motors for hours on end, and radioing the wide-bandwidth traffic between body and brain – a task I took for granted in "Where Am I?" (1978) – is still well beyond the technology available.

Cog has no legs, but it has two human-length arms, with hands (three fingers and a thumb, like Mickey Mouse) on the wrists. It can bend at the waist and swing its torso, and its head moves with three degrees of freedom just about the way a human head does. It has two eyes, each equipped with both a foveal high-resolution vision area and a low-resolution wide-angle parafoveal vision area, and these eyes saccade at almost human speed. That is, the two eyes can complete approximately three fixations a second, while you and I can manage four or five. Your foveas are at the center of your retinas, surrounded by the grainier low-resolution parafoveal areas; for reasons of engineering simplicity, Cog's eyes have their foveas mounted above their wide-angle vision areas, so they won't give it visual information exactly like that provided to human vision by human eyes (in fact, of course, it will be vastly degraded), but the wager is that the information provided will be plenty to give Cog the opportunity to perform impressive feats of hand-eye coordination, identification, and search.

Since its eyes are video cameras mounted on delicate, fast-moving gimbals, it might be disastrous if Cog were inadvertently to punch itself in the eye, so part of the hard-wiring that must be provided in advance is an "innate" if rudimentary "pain" system to serve roughly the same protective functions as the reflex eye-blink and pain-avoidance systems hard-wired into human infants. Cog will not be an adult at first, in spite of its adult size. It is being designed to pass through an extended period of artificial infancy, during which it will have to learn from experience, experience it will gain in the rough-and-tumble environment of the real world. Like a human infant, however, it will need a great deal of protection at the outset, in spite of the fact that it will be equipped with many of the most crucial safety-systems of a living being. It has limit switches, heat sensors, current sensors, strain gauges and alarm signals in all the right places to prevent it from destroying its many motors and joints. The surfaces of its hands and other important parts are covered with touch-sensitive piezo-electric membrane "skin," which will trigger signals when they make contact with anything. These can be "alarm" or "pain" signals in the case of such fragile parts as its "funny bones" – electric motors protruding from its elbows – but the same sensitive membranes are used on its fingertips and elsewhere, and, as with human tactile nerves, the "meaning" of the signals sent along their attached wires depends on what the central control system "makes of them" rather than on their "intrinsic" characteristics. A gentle touch, signalling sought-for contact with an object to be grasped, will not differ, as an information packet, from a sharp pain, signalling a need for rapid countermeasures. It all depends on what the central system is designed to do with the packet, and this design is itself indefinitely revisable – something that can be adjusted either by Cog's own experience or by the tinkering of Cog's artificers.

Decisions have not yet been reached about many of the candidates for hard-wiring or innate features. Anything that can learn must be initially equipped with a great deal of unlearned design. That is no longer an issue; no *tabula rasa* could ever be impressed with knowledge from experience. But it is also not much of an issue which features ought to be innately fixed, for there is a convenient trade-off. Any feature that is not innately fixed at the outset, but rather gets itself designed into Cog's control system through learning, can then often be lifted whole (with some revision, perhaps) into Cog-II, as a new bit of innate endowment designed by Cog itself – or rather by Cog's history of interactions with its environment. So even in cases in which we have the best of reasons for thinking that human infants actually come innately equipped with pre-designed gear, we may choose to try to get Cog to learn the design in question, rather than be born with it. In some instances, this is laziness or opportunism – we don't really know what might work well, but maybe Cog can train itself up. In others, curiosity is the motive: we have already hand-designed an "innate" version, but wonder if a connectionist network could train itself up to do the task as well or better. Sometimes the answer has been yes. This insouciance about the putative nature/nurture boundary is already a familiar attitude among neural net modelers, of course. Although Cog is not specifically intended to demonstrate any particular neural net, it should come as no surprise that Cog's nervous system is a massively parallel architecture capable of simultaneously training up an indefinite number of special-purpose networks or circuits, under various regimes.

How plausible is the hope that Cog can retrace the steps of millions of years of evolution in a few months or years of laboratory exploration? Notice first that what I have just described is a variety of Lamarckian inheritance that no organic lineage has been able to avail itself of. The acquired design innovations of Cog-I can be immediately transferred to Cog-II, an evolutionary speed-up of tremendous, if incalculable, magnitude. Moreover, if one bears in mind that, unlike the natural case, there will be a team of overseers ready to make patches whenever obvious shortcomings reveal themselves, and to jog the systems out of ruts whenever they enter them, it is not so outrageous a hope, in our opinion. (But then, we are all rather outrageous people.)

One talent that we have hopes of teaching to Cog is at least a rudimentary capacity for human language. And here we run into the fabled innate language organ or Language Acquisition Device (LAD) made famous by Noam Chomsky. Is there going to be an attempt to build an innate LAD for our Cog? No. We are going to try to get Cog to build language the hard way, the way our ancestors must have done, over thousands of generations. Cog has ears (four, because it's easier to get good localization with four microphones than with carefully shaped ears like ours!) and some special-purpose signal-analyzing software is being developed to give Cog a fairly good chance of discriminating human speech sounds, and probably the capacity to distinguish different human voices. Cog will also have to have speech synthesis hardware and software, of course, but decisions have not yet been reached about the details. It is important to have Cog as well-equipped as possible for rich and natural interactions with human beings, for the team intends to take advantage of as much free labor as it can. Untrained people ought to be able to spend time – hours if they like, and we rather hope they do – trying to get Cog to learn this or that. Growing into an adult is a long, time-consuming business, and Cog – and the team that is building Cog – will need all the help it can get.

Obviously this will not work unless the team manages somehow to give Cog a

motivational structure that can be at least dimly recognized, responded to, and exploited by naive observers. In short, Cog should be as human as possible in its wants and fears, likes and dislikes. If those anthropomorphic terms strike you as unwarranted, put them in scare-quotes or drop them altogether and replace them with tedious neologisms of your own choosing: Cog, you may prefer to say, must have goal-registrations and preference-functions that map in rough isomorphism to human desires. This is so for many reasons, of course. Cog won't work at all unless it has its act together in a daunting number of different regards. It must somehow delight in learning, abhor error, strive for novelty, recognize progress. It must be vigilant in some regards, curious in others, and deeply unwilling to engage in self-destructive activity. While we are at it, we might as well try to make it crave human praise and company, and even exhibit a sense of humor.

The computer-complex that has been built to serve as the development platform for Cog's artificial nervous system consists of four backplanes, each with 16 nodes; each node is basically a Mac-II computer – a 68332 processor with a megabyte of RAM. In other words, one can think of Cog's brain as roughly equivalent to sixty-four Mac-IIs yoked in a custom parallel architecture. Each node is itself a multiprocessor, and instead of running Mac software, they all run a special version of parallel Lisp developed by Rodney Brooks, and called, simply, L. Each node has an interpreter for L in its ROM, so it can execute L files independently of every other node.² The space of possible virtual machines made available and readily explorable by this underlying architecture is huge, of course, and it covers a volume in the space of all computations that has not yet been seriously explored by artificial intelligence researchers. Moreover, the space of possibilities it represents is manifestly much more realistic as a space to build brains in than is the space heretofore explored, either by the largely serial architectures of GOFAI ("Good Old Fashioned AI," Haugeland, 1985), or by parallel architectures simulated by serial machines. Nevertheless, it is arguable that every one of the possible virtual machines executable by Cog is minute in comparison to a real human brain. In short, Cog has a tiny brain. There is a big wager being made: the parallelism made possible by this arrangement will be sufficient to provide real-time control of importantly humanoid activities occurring on a human time scale. If this proves to be too optimistic by as little as an order of magnitude, the whole project will be forlorn, for the motivating insight for the project is that by confronting and solving actual, real-time problems of self-protection, hand-eye coordination, and interaction with other animate beings, Cog's artificers will discover the sufficient conditions for higher cognitive functions in general – and maybe even for a variety of consciousness that would satisfy the skeptics.

Now we are ready to consider Rorty's thesis. At the Royal Society meeting at which I presented the first description of the Cog project, J. R. Lucas embarked on what he took to be the first step of a *reductio ad absurdum*: if a robot were really conscious, we would have to be prepared to believe it about its own internal states. This move delighted me, for not only did Lucas thereby implicitly endorse Rorty's thesis that incorrigibility was the mark of the mental; it also provided an instance in support of his canny observation that it is skepticism about incorrigibility in machines that strikes many observers as grounds for dualism. My response to Lucas was to give the invited implication a warm welcome; we would indeed be prepared to grant this incorrigibility to Cog. How so?

Cog is equipped from the outset with a well-nigh perfect suite of monitoring devices that can reveal all the details of its inner workings to the observing team. In other

words, it will be born with chronically implanted "cerebroscopes" that could hardly be improved upon. Add to this the fact that these observers are not just Johnny-come-latelies but Cog's designers and creators. One might well think then that Cog's observers would have an insurmountable lead in the competition for authority about what is going on inside Cog. The prospect of their finding it "simple and fruitful" to cede authority to Cog's own pronouncements may seem dim indeed.

But all the information visible on the banks of monitors, or gathered by the gigabyte on hard disks, will be from the outset almost as hard to interpret, even by Cog's own designers, as the information obtainable by such "third-person" methods as MRI and CT scanning in the neurosciences. As the observers refine their models, and their understanding of their models, their authority as interpreters of the data may grow, but it may also suffer eclipse. Especially since Cog will be designed from the outset to redesign itself as much as possible, there is a high probability that the designers will simply lose the standard hegemony of the artificer ("I made it, so I know what it is supposed to do, and what it is doing now!").

This is a serious epistemological problem even for traditional serial computer programs when they grow large enough. As every programmer learns, it is essential to "comment" your "source code." Comments are lines of ordinary language, not programming language, inserted into the program between special brackets that tell the computer not to attempt to "execute" them as if they were part of the program. By labeling and explaining each subassembly via helpful comments (e.g., "This part searches the lexicon for the nearest fit, and deposits it in the workspace"), programmers can remind themselves and other observers what the point or function of each such part is supposed to be. (There is no guarantee that the assembly in question actually executes its intended function, of course: nothing is more common than false advertising in the comments.) Without the handy hints about how the programmer intended the process or state to function, the very identity of the state entered when a computer executes a line of code is often for all intents and purposes inscrutable. The intrinsic or just local features of the state are almost useless guides, given the global organization on which the proper functioning of the system – whatever it is – depends.

Even in traditional programs, the actual function – and hence actual identity – of a state or event may well evolve away from what is advertised in the accompanying comment, which may remain unchanged in the source code long after it has been rendered obsolete by undocumented debugging efforts. Large programs never work as intended at first – this is a regularity so unexceptional that one might almost consider it a law of nature, or the epistemological version of Original Sin. By the time they are actually made to work, the adjustments to their original design specifications are so many, and so inscrutable in combination, that nobody can say with confidence and accuracy what the "intended" function of many of the states is. And the only identity that matters in computer programs is functional identity (a point Rorty makes surprisingly well in his 1972 paper, p. 212, in the course of pursuing rather different aims).

In the case of a system like Cog, which is intended from the outset to be self-designing on a massive scale, the loss of epistemological hegemony on the part of its "third person" designers is even more assured. Connectionist training regimes, and genetic algorithms, for instance, create competences – and hence states embodying those competences – whose means are only indirectly shaped by human hands. (For that reason, programmers working in these methodologies are more like plant and animal breeders than machine-makers.)

Since, as I noted above, the meaning of signals in Cog's brain is not a function of their intrinsic properties but of their "intended" functions, and since Cog is designed to be indefinitely self-revisable in those functions, Cog's original designers have no secure hold on what the relevant boundaries are between states. What a bit of the system is "supposed to do" is the only anchor for what its meaning is, and when the designers' initial comments about those functions become obsolete, it is open for some new party to become authoritative about those boundaries. Who? Cog itself, the (unwitting) re-designer of its own states. Unlike the genius Jones of Sellars' fable, Cog need have no theory of its own operations (though in due course it might well develop such an auto-psychological interest as a hobby). Cog need only be sensitive to the pressures of training that it encounters "growing up" in a human milieu. In principle it can learn, as a child does, to generate speech acts that do divulge the salencies of its internal states, but these are salencies that are created by the very process of learning to talk about them. That, at any rate, is the theory and the hope.

And that is why I gladly defend this conditional prediction: if Cog develops to the point where it can conduct what appear to be robust and well-controlled conversations in something like a natural language, it will certainly be in a position to rival its own monitors (and the theorists who interpret them) as a source of knowledge about what it is doing and feeling, and why. And if and when it reaches this stage of development, outside observers will have the best of reasons for welcoming it into the class of subjects, or first persons, for it will be an emitter of speech acts that can be interpreted as reliable reports on various "external" topics, and constitutively reliable reports on a particular range of topics closer to home: its own internal states. Not all of them, but only the "mental" ones – the ones which, by definition, it is incorrigible about because nobody else could be in a better position than it was to say.

So it is not mere convention that guarantees (while it lasts) that there are minds in this world. There is, as Rorty claims, a convention or something like a convention in the etiology of mind, but it has a natural justification. Ceding authority to a subject-in-the-making is a way of getting it to become a subject, by putting it in a conversational milieu in which its own software, its own virtual Joycean machine (as I called it in 1991), can develop the competence to make self-reports about which it is the best authority because the states and events those self-reports are about get their function, and hence meaning, from the subject's own "take" on them.³

"If you called a horse's tail a leg, how many legs would the horse have?" Answer: "Four: calling a tail a leg doesn't make it a leg." True, and calling a machine conscious doesn't make it conscious. Many are deeply skeptical of anti-metaphysical moves such as Rorty's suggestion that a linguistic convention of incorrigibility accounts for the existence of minds, but what they tend to overlook – and what Rorty himself has overlooked, if I am right – is that the existence of such a convention can have effects over time that make it non-trivially self-fulfilling. This is really not such an unfamiliar idea – let's face it: it's Norman Vincent Peale's idea of the power of positive thinking. Or think of Dumbo, the giant-eared little elephant in the Disney cartoon. His friends the crows convince him he can fly by making up a tale about a magic feather that can give him the power of flight just as long as he clutches it in his trunk. By changing Dumbo's attitude, they give Dumbo a power that depends on attitude. Attitudes are real states of people (and elephants – at least in fables – and robots, if all goes well). Changes in conventions can bring about changes in attitudes that bring about changes in competence that are definitive of having a mind.

Could the attitudes lapse? Perhaps they could, but I have shown that Rorty

overestimated the power of "cerebroscopes" to trump first-person reports, so there is no good reason to anticipate that the triumph of neuroscience or robotics would bring about the death of the mind.

References

- Dennett, Daniel, 1978, "Where Am I?" in *Brainstorms*, Bradford Books/MIT Press.
- Dennett, Daniel, 1991, *Consciousness Explained*, New York: Little, Brown.
- Haugeland, John, 1985, *Artificial Intelligence: The Very Idea*, Cambridge MA: MIT Press.
- McGeer, Victoria, 1996, "Is Self-Knowledge an Empirical Problem? Renegotiating the Space of Philosophical Explanation," *Journal of Philosophy* XCIII, pp. 483-515.
- Nagel, T., 1974, "What is it like to be a bat?" *Philosophical Review*, 83, pp. 435-50.
- Rorty, Richard, 1965, "Mind-body Identity, Privacy, and Categories," *Review of Metaphysics*, 19, p. 24-54.
- Rorty, Richard, 1970, "Incorrigibility as the Mark of the Mental," *Journal of Philosophy*, 67, pp. 399-424.
- Rorty, Richard, 1972, "Functionalism, Machines, and Incorrigibility," *Journal of Philosophy*, 69, pp. 203-20.
- Searle, J., 1980, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, 3, pp. 417-58.
- Searle, J., 1983, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge Univ. Press.
- Schmid, Wilfrid, 1963, "Empiricism and the Philosophy of Mind," in *Science, Perception and Reality*, New York: Humanities Press, pp. 127-96.

Notes

- 1 See also Rorty, 1965, which I will not discuss, although it is an important paper in the history of the philosophy of mind.
- 2 For more details on the computational architecture of Cog, see my "The Practical Requirements for Building a Conscious Robot," in the *Philosophical Transactions of the Royal Society*, (1994), 349, pp. 133-46, from which this brief description is excerpted, or for more up-to-date information, consult the World Wide Web site for the Cog Shop at MIT.edu/projects.
- 3 These points grew out of discussion with Victoria McGeer, of a paper she presented at the Society for Philosophy and Psychology meeting in Vancouver in 1993; the successor to that paper, McGeer, 1996, carries these points further.

RESPONSE TO DANIEL DENNETT

I agree entirely with Daniel Dennett's criticism of my suggestion, in "Incorrigibility as the Mark of the Mental," that the incorrigibility of certain first-person reports should be thought of as the result of adopting a "convention." "Convention" is an unQuinean, unDavidsonian, notion. I should have been more wary of it. It would have been more consistent and more prudent to have spoken of a habit of reliance on such reports - a habit adopted for obvious and good reasons - than of a convention.

When I referred, in a passage Dennett quotes, to a "regulative principle of behavioral science" I should have spoken of a persistent, but revisable, habit of behavioral

scientists. Speaking in this way would have made it clear that I wanted to talk about convenience rather than convention, and about the best way to cope with behavior rather than about ontological commitment. On my view, of course, any and every pattern of linguistic practice is an attempt to cope with the behavior (either linguistic or non-linguistic) of things. That is why I have no use for the analytic-synthetic, the fact-vs.-convention, or the "matter of fact-vs.-no matter of fact" distinctions.

Dennett does seem to have a use for such distinctions. For he raises the question "Why turn the recognition of high reliability [of introspective reports] into a constitutive declaration of incorrigibility? Is this just an unmotivated overshooting of social practice . . .? Or might there be some deeper reason – an actual justification – for thus shifting the very criteria (to speak in 60s-talk) for the occurrence of mental phenomena?"

In posing these questions Dennett seems to want to let go of 60s-talk, which was still pervaded by the distinctions I have just deplored, with one hand while holding on to it with the other. He shares this ambivalence with many other contemporary philosophers: they would not be caught dead invoking the analytic-synthetic distinction, but nonetheless want to preserve a distinction between assertions made for the sake of convenience and assertions that have "an actual justification."

As I see it, one can describe any true assertion as a convenient tool for coping with reality, or as a good move in a language-game, or even as a reasonably accurate representation of reality, just so long as one does make invidious distinctions between kinds of assertions (so that true political or literary or moral judgments, for example, are tools and moves but not representations, for example, whereas true physical theories are all three). Describing true assertions as representations of reality, or as corresponding to reality, is harmless if the metaphors of representing and corresponding are not pressed.

Not pressing them is the pragmatic cash-value of using what Dennett calls a "vegetarian" notion of representation. I would prefer, however, to describe this not as a notion of representation – a rather complex and novel one, which may require a theory of "internal realism" to explicate – but simply as a dead metaphor. There is no harm in saying of good tools and good moves that they are also good representations, but nothing interesting is conveyed by this choice of idiom, and its employment should not tempt us to construct theories about how representation works. For it is no more useful to ask "what bits of a physical theory represent what bits of non-linguistic reality?" than to ask "who passed the law of gravity?"¹

Another way of putting the difference between Dennett's views and mine is that mine allow no room for the notion of "more than just conversation" which he invokes when discussing the "more fashionable and famous ideas" which, in my later years, I have been "misled" into adopting. I see no way to make a principled distinction between conversation about politics and literature, on the one hand, and scientific inquiry on the other except in sociological terms. (The natural scientists, for example, can predict better than the other conversationalists, are more likely to agree among themselves, and so on.) But I take it that Dennett wants to say that there are (what he calls) "deeper reasons – actual justifications" why these sociological facts obtain. That makes him what I call a "scienticist." Scienticists think that they are paying a high compliment when they say of someone, as Dennett says of me, that he "succeeded in spite of himself in making . . . [a] contribution to science – to our objective knowledge of the way the world, and the mind, is." I regard this compliment as like a decoration bestowed by a king who, with any luck, will soon be forced into exile by a citizenry

exasperated with monarchic pretensions. The medal is gaudy, and its award was a nice gesture, but it does not mean much.

Part of my ambition, to paraphrase Freud, is to help it come to pass that where epistemology and metaphysics were, sociology and history shall be. So, as Dennett correctly says, I want to "trade in a lot of tempting but indefensible metaphysical claims for an epistemological or even sociological claim." "Sociological" is a much better term than "epistemological" because the fact that, in Dennett's words, "certain claims [for instance, first-person reports of thoughts and sensations] cannot be overridden . . . given the role they play in our shared life" is, given the Myth of Jones, a socio-historical fact. Epistemology has always had pretensions to ahistoricity.

Dennett says that he wants to save my neo-Sellarsian "deflationary doctrine" about first-person reports of the mental from "some Rortian excesses." But I have trouble seeing just what these excesses are supposed to be. This is because I have trouble seeing what it is to "take these questions about the justification and confirmation of our representations more seriously than he [Rorty] now allows." Presumably my lack of seriousness, and at least one of my excesses, consists in not facing up to the question "Would minds cease to exist if the sociological facts changed?"

Dennett correctly says "for current Rorty, this is surely a paradigm of a misguided question." He suggests that I would find it misguided because it assumes "that there is a natural standpoint from which the Truth of the ontological claim could be assessed." I would prefer to say that it is misguided because it tries to drive a wedge between being an assertion that has an unquestioned and useful role in our language-game and being an assertion which is ontologically correct. It abandons the vegetarian, philosophically banal, ontological attitude which Arthur Fine has called "natural" in favor of invidious comparisons between various sorts of entities. It tries to make the invidious distinction I described above as "scientistic": the distinction between true assertions which are good moves and good tools but not necessarily representations, and assertions which are all three.

I take it that the example of Cog's possible future incorrigibility is supposed to give me reason to take questions like "Would they still have minds? When did they start having minds?" more seriously. But I am not sure exactly how the argument from Cog goes. I am happy to agree that "there is a high probability that the designers [of Cog] will simply lose the standard hegemony of the artificer ('I made it, so I know what it is supposed to do, and what it is doing now!')." As I said at the outset, I quite agree that "it is not mere convention that guarantees that there are minds in the world" but rather success of the sort of training program which Sellars' Jones, and Cog's interlocutors, conduct. But I am not sure about the claim that "the saliences of its [Cog's] internal states . . . are saliences that are created by the very process of learning to talk about them."

The question "does talking about X's create the X's, or were they there already?" is one I have discussed elsewhere.² The line I take is that although in some cases the question is easily and commonsensically answered (mountains were there already, bank accounts weren't), in many other cases the question is pointless. It is pointless because the choice of answer makes no difference. I cannot see that it matters whether the pre-Sellarsians had minds and Jones simply trained them to report on them, or whether the training created the minds. Here, as in the case of Cog, I should have to be told more about why the question is being raised.

So, though I can agree (barring some quibbles with the term "conventions") with Dennett's penultimate claim that "changes in conventions can bring about changes in

attitudes that bring about changes in competence that are definitive of having a mind," I would not be perturbed to be told that all that is brought about by the changed attitudes is a change in competence that is definitive of having knowledge that one has a mind. When it comes to his final sentence, I am inclined to say that the relevant triumph of neuroscience might be described either as "the death of the mind" or as "the obsolescence of mentalistic reports," and that it would not make much difference which description is chosen. If one does not care about whether or not a good tool or good move is also a good representation of "the way the world, and the mind, is," then one will not care about the choice between these alternative descriptions. My attitude is: give us the tools, make the moves, and then say whatever you please about their representational abilities. For what you say will be, in the pejorative sense, "merely philosophical."³

I want now to turn away from Cog and mentality to some issues about scientism. This will permit me to take up a line of thought found in the final pages of Akeel Bilgrami's paper, that I leave undiscussed in my response to him. Bilgrami says that in the contemporary academy there are some bad people, whom he calls "bullshitters," and who do not value truth. In a paper called "Faith in Truth," which contains considerable discussion of my views, Dennett has said much the same. Like Bilgrami, he is appalled by the "postmodernist" types who seem not to know the difference between seriousness and frivolity. The fear and loathing of "postmodernism" which is explicit in "Faith in Truth" can be found between the lines of the paragraph in "The Case for Rorts" in which Dennett refers to my "more fashionable and famous ideas."

In "Faith in the Truth," Dennett makes clearer what he takes these ideas to be. There he criticizes what he calls my "attempt to show that philosophers' debates about Truth and Reality really do erase the gulf [between being serious and being frivolous], really do license a slide into some form of relativism." In the end, Dennett continues, the Rortian view is that "it is all just conversations, and [that] there are only political or historical or aesthetic grounds for taking one role or another in an ongoing conversation."

In this article, echoed in a later paper called "Postmodernism and Truth," Dennett joins the chorus of people who see "postmodern relativism" as a subversive and dangerous movement, and who see me as aiding and abetting this movement. Obviously, I prefer Dennett's avuncular warnings to the scornful ridicule of my more virulent critics. But, being avuncular in my turn, I would caution both Dennett and Bilgrami against aiding and abetting the Blimpishness which characterizes many polemics against "postmodernism" by analytic philosophers. I see both philosophers as exhibiting the sort of cultural chauvinism which I call "scientism." The sort of chauvinism I have in mind is illustrated by the many viewings-with-alarm we are getting nowadays about the insidious influence of "fashionable Continental ideas." ("By gad, sir! The enemy is at the gate! Time for all decent chaps to rally round! We must defend Truth and Science against those frivolous, deconstructing, relativists!")

I have no wish to cast doubt on the distinction between the frivolous and the serious. That is a serious and important distinction. It is well exemplified in the contrast between the silliest, least literate, members of academic departments of literature and honest, hard-working, intellectually curious, laboratory scientists – just as the distinction between self-righteous priggery and tolerant conversability is well exemplified by the contrast between the sulkiest, least literate, members of analytic

philosophy departments and honest, hard-working, intellectually curious, literary critics.

Neither of these distinctions, however, has any connection with the difference of philosophical opinion between those who do and those who do not believe that truth consists in accurate representation of the intrinsic nature of reality. This latter difference can also be described as that between people who think that justification to all comers is the only goal of inquiry and those who think that there is an additional goal, namely getting things right. People who hold the latter view typically hold the view which I call "scientific": they believe that this goal is often achieved by natural science but not by those who debate political or literary matters.⁴ This difference in philosophical outlook divides people who can see Dennett's point when he contrasts "just conversation" with something better, from people like me, who cannot. So we are the ones who agree with Brandom that "Conversation is the highest good for discursive creatures". So we are puzzled by the term "just."

Akeel Bilgrami sees a connection between philosophical error and bullshit that I cannot spot. He thinks it is "a matter of some importance in our culture, especially our academic culture, that we see the nature and the great importance of truth as a value in another sense than the moral value of truth-telling." I suspect that Dennett (and probably James Conant as well) would agree with Bilgrami on this point. They might also agree with him when he goes on to say that "the bullshitter" is prepared to speak and write in the requisite jargon, without any goal of getting things right. But Bilgrami, as far as I can see, tells us nothing more about how to tell bullshitters from non-bullshitters. All we can do is watch for indications of whether their actions are directed to this goal.

But what indications are those? What behavioral evidence is relevant? I doubt that there is more hope of accumulating relevant behavioral evidence here than there is when attempting to answer the question "Is he saved?" or "Does he love the Lord his God with all his heart and soul and mind?". The question "Do you value truth?" seems to me as about as pointless as these latter questions.

Nevertheless, I quite agree with Bilgrami that there is a difference between the sort of people he calls "bullshitters" and others. This difference, however, has nothing to do with a person's goals. In particular, it has nothing to do with whether she thinks of herself as trying to make contributions to "our objective knowledge." Rather, the people whom Bilgrami describes as bullshitters are distinguished by being unconversable, incurious, and self-absorbed.

Unconversable people are the ones you cannot talk profitably with on matters of common interest, no matter how hard you try: you finally are forced to conclude that persistent failure to get on the same wavelength is their fault rather than yours. More specifically, you tell the serious inquirers from the frivolous "bullshitters" by finding out who makes a serious effort to hitch his jargon, his interests, and his goals, up with yours — who is willing to go to considerable effort to build conversational bridges. Someone who seems to be making as sincere and determined an effort to do these things as you yourself are will count as "serious." Somebody who doesn't may reasonably be called "frivolous," though perhaps "self-centered" or "intolerant" are more appropriate terms. This test will work on Cog as well as on protoplasmic language-users, and will work no matter what reply either sort of interlocutor makes to catechismic questions like "Do you value truth?" "Do you have faith in truth?" and "Do you aspire to objective knowledge of how the world is?".

Asking such questions of someone one suspects of being frivolous is like a departing

representative of the British Colonial Office testing whether a certain native can be trusted to help run his country by inquiring as to whether he is a good Anglican. Such an official (someone like David Low's cartoon figure, Colonel Blimp) sees an obvious connection between reading from the Book of Common Prayer, dressing for dinner, not shooting foxes, abiding by various other British customs, and being a decent, trustworthy chap. Dennett and Bilgrami see an obvious connection between having the right "realist intuitions" and therefore making the right scientific noises – reading from the analytic philosopher's version of the Book of Common Prayer, so to speak – and being a non-bullshitter. Dennett's conviction that I am aiding and abetting bullshitting can, I think, only be explained by this sort of chauvinism.

Cultural chauvinism consists of the view that one who does not conform to certain traditional practices (sartorial, sexual, gustatory, conversational, or sacramental, for example) is likely to lack such desirable features as seriousness, decency and trustworthiness. I think that Dennett and Bilgrami are guilty of cultural chauvinism when they assume that people who differ from them on philosophical questions, or in the sort of compliments they offer laboratory scientists, lack some sort of moral probity.

The test of the relevant sort of moral probity is whether or not one does one's honest best to break out of one's own parochial little language-game (Foucauldian culture criticism, possible-world semantics, Scientology, superstring theory, British middle-class morality, Anglican worship, Brandomian social-practice semantics, whatever). In testing for such probity, the right question to ask is whether the person in question does his best to find a way to talk about matters of common concern with people who are not accustomed to playing his own preferred game.

These are the sorts of reasons why I think that questions like Bilgrami's "Does she value truth?" and Dennett's "Does she have faith in truth?" are irrelevant to the distinction between seriousness and frivolity. I cordially agree with Bilgrami and Dennett that much current conversation among academics (particularly those accustomed to using the term "postmodern" in full seriousness) is jargon-ridden, profitless, and an unfortunate diversion of libidinal energy from more worthwhile projects. But I do not think that colleagues who go in for these profitless activities are at fault because they fail to grasp the need to get things right. They are trying to get things right too, but the things in question are artifacts which nobody else can see much use for. Their unconversability, and their social uselessness, are results of their failure to see any need to convince a larger circle of the utility of their new toys.

Rather than saying that these people do not value Truth, I would say that they do not have enough intellectual curiosity. They do not try hard enough to find out what is going on elsewhere in the intellectual world. They do not attempt a Gadamerian fusion of horizons. Similar rebukes apply to members of the many little cults (or, if you prefer, "schools") which have grown up within analytic philosophy. Like the worst of the "cultural studies" Foucauldians, lots of analytic philosophers think that if they can make sufficiently clever moves within their own cult's language-game they need not worry about what anybody else in academia, or the larger world, is saying or doing.

To sum up, I see an important sociological distinction between incurious cultists and more conversable sorts of people – a distinction that is important for our practical decisions about whom to talk with about what. But I do not think that this sociological difference reflects the difference between valuing and failing to value something called "Truth" or something else called "Reason." To believe that it does seems to me as chauvinistic as the view that moral probity depends upon belief in the existence of a divinity to whom we owe obedience. One important discovery of recent centuries is

that atheists can be just as decent chaps as theists. It is time to follow this up with the realization that literary critics can be just as rational as experimental physicists, even those literary critics who remain in what Dennett calls (in "Faith in Truth") "flatfooted ignorance of the proven methods of scientific truth-seeking and their power."⁵

The religious chauvinism we loathe when it appears in national politics should not be mimicked by a scientific chauvinism in academic politics. Carnap's and Popper's concern with "the demarcation problem" is, unfortunately, still alive and well within analytic philosophy. But that philosophical tradition will never become mature enough to make a contribution to the conversation of the intellectuals until it gets over this youthful obsession. Until it does, it is likely to retain its jejune self-image as "more scientific," and therefore more morally virtuous, than non-analytic philosophy. It will continue to combine juvenile arrogance with Blimpish self-satisfaction.

There is, to be sure, a sense in which analytic philosophy is indeed more scientific than other kinds of philosophy. Most analytic philosophers are puzzle-solvers, in the sense in which Kuhn said that natural scientists were puzzle-solvers.⁶ They find contradictions between our intuitions, and ways of resolving those contradictions, just as natural scientists find contradictions between theories and observations, and then think of ways of resolving those contradictions. Non-analytic philosophers, on the other hand, typically do not solve puzzles.

They do other things. Some of them try to change our intuitions (by, for example, getting us to think of the idea that true beliefs are accurate representations of reality as an optional metaphor rather than as an important insight). Others tell stories about the history of thought (of the grand, *geistesgeschichtlich*, Hegel-Heidegger, sort). Still others (Derrida, for example) offer remarkable new readings of old philosophical texts. There is plenty of room in the intellectual world for all these activities, and it is hard (unless one thinks that no intuition should ever be erased, nor any text recontextualized) to see why they should be thought of as in competition with analytic philosophy. But as long as analytic philosophers cling to the chauvinist idea that they, together with their colleagues in the natural sciences, have a special relation to "Truth" (valuing it more, for example, or having more faith in it) than their more "literary" colleagues lack, they will be tempted by the unconvertibility, and the arrogant frivolity, that they deery in others.

Notes

¹ Does this mean that there is, as Dennett says, a vegetarian use of "mirror of nature" which could satisfy "all but the most hysterical Realists"? Sure. One can vegetarianize any dead metaphor simply by refusing to press it – refusing to analyze its meaning, determine its transcendental conditions of possibility, or otherwise philosophize about it. But, as I see it, the hysteria is not about realism but about scientism – about the need to make the natural sciences look good by setting them apart from the rest of high culture with the help of heavyweight philosophical, rather than lightweight sociological, distinctions.

Where Dennett sees a scale with hysterical realism at one end and what he thinks of my own hysterical, albeit fashionable, anti-realism at the other, and himself holding to the virtuous non-hysterical mean, I see a fairly sharp break between the people who want to make natural science wondrous and different from the rest of high culture, and those who, like myself, view it as a portion of the conversation which serves certain purposes (e.g.,

prediction) and does not serve others (e.g., figuring out what to do with our lives). I return to this topic below, in my concluding paragraphs.

- 2 See "John Searle on Realism and Relativism" in my *Truth and Progress*.
- 3 I see the question about whether Cog's trainers are "getting it to *become* a subject by putting it in a conversational milieu" as like the question "Do we make our children, or our slaves, subjects by putting them in a conversational milieu?" Maybe we do, or maybe we just gain access to their pre-existing subjectivity. How could it matter? I cannot see how the question would come up unless one thought that the question of whether fetuses, or illiterate slaves, have rights is to be answered by figuring out whether they contain an ineffable whatsis called "subjectivity" or "personhood." Those who do think so hope that metaphysics will guide us when we make moral and political decisions. This hope strikes me as pathetic.
- 4 This difference is discussed at length in my "Is Truth a Goal of Inquiry: Donald Davidson vs. Crispin Wright," included in my *Truth and Progress*.
- 5 Dennett says that this ignorance deprives these people of "the leverage provided by scientist's faith in the truth." Lack of this leverage explains the fact that whereas "again and again in science, yesterday's heresies have become today's new orthodoxies," "no religion exhibits that pattern in its history." Historians of religion will have bones to pick here.
- 6 This is not true, however, of the greatest and most imaginative analytic philosophers – those who, like Sellars, Kripke and Davidson, start by moving the pieces around and wind up knocking over the chessboard.