Mind Mach (2006) 16:21–28
DOI 10.1007/s11023-006-9016-4

The attentional spotlight

Joanna J. Bryson

Published online: 27 April 2006

Abstract One of the interesting and occasionally controversial aspects of Dennett's career is his direct involvement in the scientific process. This article describes some of Dennett's participation on one particular project conducted at MIT, the building of the humanoid robot named Cog. One of the intentions of this project, not to date fully realized, was to test Dennett's multiple drafts theory of consciousness. I describe Dennett's involvement and impact on Cog from the perspective of a graduate student. I also describe the problem of coordinating distributed intelligent systems, drawing examples from robot intelligence, human intelligence, and the Cog project itself.

Keywords Artificial intelligence · AI · COG · Dennett · Memes · Robotics · Pain

Artificial Intelligence

Our culture has three sources of systematic knowledge: experimental science, mathematical or logical proofs and philosophy. Experimental science tries to approach the truth probabilistically by demonstrating evidence anyone can replicate and test themselves. Mathematical proof takes accepted starting points and applies reliable operators to build knowledge systematically. Philosophy is the least mechanistic, but perhaps the most honest, means of enquiry. Philosophy knows it relies on bright people who read remarkably fast and remember remarkably well. These people are trained in a litany of knowledge (including techniques of reasoning) which philosophers have accumulated to date. This litany contains not only positive examples but also negative ones—not only revelations, but errors. Undergraduates are trained on attractive ideas that have snared philosophers in the past and teased with the ideas that will catch most of their peers.

J. J. Bryson (✉)
University of Bath, BA2 7AY, Bath, UK
E-mail: J.J.Bryson@bath.ac.uk

Artificial intelligence allows machines to exploit all three of these approaches to knowledge acquisition. We call experimental methods *machine learning* or *reinforcement learning*, proofs are *theorem proving* or *conventional/productive planning*, and the philosophical approach is called *case-based reasoning* or occasionally *common sense reasoning*. As is the case for their academic parallel, case-based and common sense reasoning are not currently particularly popular means for trying to generate machine intelligence. But the reason for this lack of popularity is different from that normally given for philosophy's: basically, we AI researchers know that case-based reasoning is too hard.

Case-based reasoning involves memorizing huge amounts of cases, recognizing which one is applicable in the current context, knowing how and what parts of the case need to be adapted to make it match the current context well enough to use it to select an action, and more. For example, if you base an interaction for buying a train ticket on one you have recently observed, you need to know what variables have to be changed to match your own goals (e.g. if you have a different destination). As for common sense reasoning, there is a single company (Cycorp) that has been spending millions of dollars for over twenty years trying to type in enough information to make computers intelligent (Lenat, 1995). In that time, Cycorp have had to start over several times when they realized they hadn't represented the knowledge in the right way to make it sufficiently accessible. And while the knowledge they have accumulated has now become a product, that product's purchasers use it for data mining. No one thinks Cycorp have produced an intelligent entity.

Being a professional philosopher must lead to both smugness and frustration. The philosopher can watch educated colleagues from other disciplines walk straight into traps a trained philosopher could recognize from a distance, but at the same time they have trouble communicating *why* their colleagues are making a mistake. ''We philosophers have thought about that for several hundred years and it always leads to a contradiction'' just doesn't sound as convincing as experimental evidence or a proof. Yet ultimately science and mathematics are subject to exactly the same set of problems as philosophy. All three disciplines, if they start from a false premise or incorrectly apply their methodology, can wind up with (seemingly well-supported) nonsense for results. There is only one way to catch these errors: review by experts, which is why all three disciplines rely on peer review. Philosophy is the acquisition of knowledge *only* by peer review. Or, to put it another way, when a branch of philosophy stumbles on a *really good* new mechanism for acquiring knowledge, those philosophers who are researching the right areas of knowledge such that they can start applying that method get called something other than 'philosopher', like 'scientist' or 'logician'.

This is also a problem familiar to AI—our best and most reliable methods become straight computer science. After all, *intelligence* can't be clearly specified or perfectly reliable, can it? Intelligence is what defines humanity, what sets us apart from the rest of the cosmos. Of course, if you are not religious about this issue, you may grant me that there's no in-principle reason that a computer can't be intelligent. But you probably don't believe that computers will become intelligent in our lifetime. By a funny coincidence, you probably don't think we'll understand consciousness in our lifetime either—you probably think that it will take about 100 years to achieve either of these.

Expect, of course, if you are reading this article, you are probably interested in Dan Dennett, and so you may well except to both understand consciousness and to see it implemented in an artifact, hopefully in the immediate future. Maybe you think we already *do* understand consciousness at least a little, and would even be willing to argue that computers or robots already have some aspects of it. If so, I agree with you.

Cog

I've been asked to write about Dennett in the context of the Cog project—the attempt by a team mostly at MIT to build a truly humanoid robot. I find this slightly embarrassing, first because the Cog project has never yet come near to what we'd hoped to be doing with it in terms of testing Dennett's ideas, and second because I was only formally on the Cog project for a year and a half. On the other hand, it was the *first* year and a half (1993–1995), and that was when a great deal of interesting work was done, at least in terms of philosophy and scoping the project. Also, the topic of Dennett and Cog is far too interesting to let go.

The head of the Cog project, Rodney Brooks, was famous for reinventing AI around the notion that if we couldn't understand insect intelligence, then we couldn't even dream about building human intelligence (Brooks, 1990). The plan was to understand and demonstrate our understanding of insects through building some, then to build and understand an iguana, a cat, and then possibly, if the cat went well, a human. Nevertheless, having successfully (but with much dispute and delay) won his tenure case modeling at least a bit of insect intelligence, Brooks decided during a year's sabbatical in 1992 that he only had one more major fight left in him, so he would leap straight to human intelligence (Freedman, 1994). To make it easier though, he decided to build a two-year-old.

Many people have told me that Brooks never really expected Cog to work. Well, in some sense, everyone knew there was a large chance of failure, but I sincerely believe that *none* of the faculty involved with the project thought that the project was doomed from the start (though some of the students did.) Certainly Brooks and Dennett *wanted* Cog to work; Brooks even went so far as to worry about details such as whether Cog might be autistic.

Multiple Drafts

In the mid to late 1980s, Brooks had invented an approach to putting AI together called *the subsumption architecture* (Brooks, 1986, 1991). Subsumption architecture was based on the theory that intelligence should be composed of a large number of simple modules which only interact with each other in very restricted ways. This approach is called *behavior-based AI* and was to some extent derived from the Minsky (1985) idea of a ''society of mind''. When I first heard Dennett was on the Cog project, I assumed that the connection was due to his interest in modular models of human intelligence, which I knew something about from having read ''Time and the Observer'' (Dennett & Kinsbourne, 1992) and a bit of *Consciousness Explained* (Dennett, 1991). We were coming at the same problem from two different directions. Dennett was concerned about how to explain the inconsistencies of conscious experience. Those of us concerned with BBAI were worried about how ordered, agent-like behavior *emerges* from a bunch of inconsistent modules. 'Emergence', meaning effectively (for AI) *arising without having been explicitly programmed in*, is very important when you are either starting from a belief system like subsumption architecture that bans you from using the obvious coordination systems, or if you are faced with a problem so complicated that you can't see how to build such a structure anyway.

The original Cog funding proposal—submitted (unsuccessfully) as an NSF grand challenge, and still available from MIT as an AI Lab Memo—specified that the multiple-drafts theory of consciousness should emerge on Cog by 1996 (Brooks & Stein, 1993, p. 16). To be fair, the dates on the timeline were a requirement of the funding agency, and were eliminated in the journal-article version of the paper. But the phrase ''multiple-drafts emergence'' is still present at the bottom of the diagram in that article.

Perception and Memory

Subsumption architecture also specifies that an agent should have no global memory. All memory was constrained to the individual modules, and even there it could only represent very simple concepts. Essentially, each module had one piece of memory which recorded which of several previously-enumerated states that module was currently in. Whether and when the memory should change state is also previously encoded and depends almost entirely on what the module's sensors tell it about the agent's situation in the rest of the world. Which state the system is in determines exactly how the module transformed its input (based on sensing) into output (control of an action).

At the time, Dennett was also very concerned about the aspect of conscious experience that includes an apparently fully-fleshed model of reality—the apparent full awareness of the entire room around you. This complete awareness is an illusion only occasionally betrayed when, for example, you are searching for your keys and find them in the center of a table in the middle of that room, or worse, in your hand. Around the time Cog began, I heard Dennett speak publicly several times on this illusion—that we didn't really reconstruct or ''fill in'' information in the blindspots of our retinae or behind occluded objects. Rather, we only have full perception of the aspects of the scene to which we are currently attending. This idea again relates to the multiple drafts hypothesis—that consciousness involves some extra resources which are brought to bear on the bits of the problem that are currently relevant.

The attentional spotlight metaphor helps in BBAI too, in that order could emerge because the module or modules most currently relevant can inhibit other modules that might interfere with their behavior. This capacity to inhibit other modules was an early enhancement to subsumption architecture (Connell, 1990). Further, there doesn't necessarily have to be some complicated or homuncular processing underpinning the spotlight. For example we know that the fovea of the eye 'automatically' provides this sort of detailed attention to a small region of visual space while more limited visual information comes from the periphery. The visual periphery in fact specializes in things that might make you change where your fovea is fixating—like sudden motion—while having no facility for processing details like color.

What we working on Cog hoped was to find further such simple mechanisms to make organizing (BB)AI easier. But unfortunately, the brain is actually very complicated. In fact, most of the brain has at least as many connections carrying information *towards* the sensor systems as away from them. It turns out that *perception*, the interpretation of sensory information, is an extremely difficult problem that can be solved only by bringing a great deal of knowledge and experience to a set of sensory input. But knowledge and experience are things that have to be stored in memory, so obviously we needed more memory than subsumption architecture originally stipulated. Worse, we were trying to build a two-year old. Two-year olds are only just forming new perceptual categories from scratch. This means that they must be storing lots of uncategorized, relatively unspecialized knowledge while their brains seek statistical regularities in it. Ultimately, I suspect, a two-year old human is less like a behavior-based robot than a skilled adult would be.

Why Cog Can't Feel Pain

Now that I know a little more about philosophy than I did when I started my PhD, I know that Cog must have meant a great deal more to Dennett than just a way to test multiple

drafts. If you have read *Elbow Room* or *Freedom Evolves* then you probably realize that the issues I mentioned above (which AI researchers need to deal with every day) about action selection and goal arbitration are highly related to Dennett's theories of free will. I hadn't read *Elbow Room* (and couldn't have read *Freedom Evolves*) when I was on Cog. I lost interest in the question of free will as an undergraduate after I failed to hand in an essay on determinism.

It was actually only two months ago that another philosopher of AI, Dylan Evans, drew my attention to one of his favorite Dennett essays—''Why You Can't Make a Computer that Feels Pain'' (Dennett, 1978). The essay is in three sections. The first addresses the problem in the abstract, observing among other things, that some people would always intuitively dismiss the idea of a robot experiencing what they call pain, and that ''There can be no denying (though many have ignored it) that our concept of pain is inextricably bound up with (which may mean something less strong than *essentially connected with*) our ethical intuitions, our sense of suffering, obligation and evil.'' (p. 197) The middle section draws quite literally a diagram of the then-current scientific understanding of how pain is processed in humans and the mechanisms of anaesthetics and analgesics. The final section reviews then-current philosophical thought on the matter of pain and points out that many beliefs and intuitions about pain were intractably contradictory. ''Any robot instantiation of any theory of pain will be vulnerable to powerful objections that appeal to well-entrenched intuitions about the nature of pain, but reliance on such skeptical arguments would be short-sighted, for the inability of a robot model to satisfy all our intuitive demands may be due not to any irredeemable mysteriousness about the phenomenon of pain, but to the irredeemable incoherency in our ordinary concept of pain.'' (p. 228)

Reading the essay, it's clear that while pain is an interesting problem in its own right, it might be even more interesting due to its tight coupling to consciousness. Sometimes people aren't aware of pain they must be experiencing, sometimes (normally under the influence of particular drugs) they report that they *are* aware of it but don't mind it. But pain—much more so than consciousness—is something we have a vast medical literature on.

The year that the Cog project started, Dennett had working with him a visiting research fellow who was a specialist in pain, Nikola Grahek. Dennett brought Grahek to a number of meetings of the Cog group and organized for Grahek to give a seminar to us on pain's importance and mechanisms. Brooks, who tended to supervise with a very light touch, asked us if we didn't think we'd need to build something like pain into Cog, if for no other reason than for Cog's self protection. None of the students leapt at the chance though. Perhaps it was the association of pain with suffering and evil—shouldn't robots be beyond that? Or perhaps it was our parochial sense of the implausibility of a machine experiencing pain, or of the even greater implausibility of our proving it had in a dissertation. Nevertheless, if I'd seen then the relation to consciousness and grounding in medical science that I see now, I think I'd have been tempted to take the job—even knowing what I know now (and didn't then) about how hard it is to build robots.

Memetics

I assume that as a reader of a special issue on Dennett you are quite likely to have read more Dennett already than I have now, and far more than I or the other students on Cog had in 1993. I expect you are incredibly frustrated that a group of ignorants could make so little

use of the opportunity to pursue research with such a great thinker—and what's more, a great thinker who thought in the areas that most interested us.

One of my most vivid memories of Dennett from my time on Cog was the look on his face as I explained some aspect of research I was working on (I no longer remember what.) He launched into a story.

When Dennett first began working with AI graduate students, he'd been appalled by what they didn't know, and had recommended huge numbers of papers that they absolutely had to read. But then he discovered that the students would just read—and stop doing interesting things. There was always more that could be read, even more relevant things, but all the interesting things that happens with AI students happens when they are actually programming. So Dennett had learned to be very, very circumspect in recommending papers.

I've since heard him tell this story several times more—to other people. I suspect telling it is one of Dennett's strategies for inhibiting his own propensity for recommending papers.

Does Dennett's strategy make sense? Well, life is finite, knowledge is wide, and much of science is slow, arduous preparation and exploration. But if the people actually doing research don't have a thorough grounding in their discipline, how can knowledge advance?

In recent years, one of the research concerns of Dennett's I've become most interested in myself is memetics. Memetics holds that knowledge can be built by a process of Darwinian evolution. Our culture (like our organism) evolves even while the agents that replicate it, extend/diversify it and select it have anything but the full picture. On Cog, we graduate students were mostly mutators—we provided new bits of knowledge by inventing some ideas and testing others. Dennett certainly does some inventing, I don't mean to imply he isn't creative, but above all, Dennett was our source of crossover. He brought in new ideas—sometimes directly to the students, sometimes only through the professors, sometimes individually, other times as public talks, and occasionally even through his writings.

We students replicated (possibly with elaboration) what we could—what seemed to make sense to us, what seemed tractable with the technology we had, what came to mind when we were faced with a new problem—regardless of whether we remembered what had put the ideas in our minds in our first place. As I moved through my own PhD I started realizing that the role of professors in graduate research is creating a research environment not only of equipment and research obligations to funding agencies, but most importantly of the ideas and knowledge the professors find most likely to be fertile. Students will 'stumble on' ideas highly related to the ones their supervisors try to communicate before the students are ready—and they are most likely to stumble on or adopt the ones that are actually useful.

Attention

But Dennett was also more than an extra professor—not to belittle that role. But there's no question that students actively sought his attention and interest. Students not only filter their professors' ideas but also use their professors as filters.

I was at a going-away party that Dennett attended—one of a very few professors who turned up—and one very bright engineering student from another laboratory succeeded several times in engaging Dennett's attention and approbation in competition with a nearby (untenured) Harvard professor. At some point late in the evening the student asked his partner (a psychologist who was seated next to me) whether the ''old guy'' was famous.

I told his partner Dennett's name, and she turned to her partner and said simply ''Don't worry; he's famous.''

This may sound like a story of mindless vanity, but it isn't—everyone involved was quite smart and were employing excellent strategies. The engineer went on to interview successfully for a prestigious academic position a short time later—who knows how much good having a more accurate assessment of some of his ideas may have done him, and now in turn his students? He successfully identified someone important to impress, and his query helped him check the validity of his own self-evaluations.

People love to deride the famous for any flaw, but the fact is that fame is difficult to come by and a good indicator of deserved success. Further, once one becomes famous and therefore a target to be impressed, one is exposed to a great deal of information, which only makes one's opinion that much more valuable.

Conclusion

If we can think of science in the large—of *philosophy* in its original sense, encompassing all of philosophy, science and logic—if we can think of that as a distributed modular system (where the modules are the scientists) then we have a problem much like the one that I described before for BBAI. How do we know when one module has come upon a good solution, given that tracking all the other modules would take more resources than any module has, and that ideally each module should devote as many of its resources to doing its own work as possible? The answer is that each module, each philosopher, must track *a few* other modules, and then we must let evolution slowly select the best ideas from each of these clusters.

But what Dennett has done for Cog is a bit more than that. Dennett *is* an attentional spotlight. He brings the usefully analyzed and aggregated assessments of a huge number of researchers to bear on any problem he attends to. Further, if his attention is great enough to warrant written or spoken mention, then other resources—relevant papers sent by their authors, bright students, even funding—also have an increased probability of coming to that attention's focus.

As I said in this article's introduction, the truth is that every discipline of knowledge acquisition *is* still a form of philosophy. As a culture, we gather a breadth of information, slowly accumulating facts and fallacies—attempting to ban the mistakes and build on the successes. What Dennett brought (and presumably still brings) the students on Cog is a myriad of information, suggestions, paper references, and questions. He also brings his enthusiasm, faith, and personal connections. He brings his attention.

References

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, RA-2*, 14–23.

Brooks, R. A. (1990). Elephants don't play chess. In: P. Maes (Ed.), *Designing autonomous agents: theory and practice from biology to engineering and back*. (pp 3–15). Cambridge, MA: MIT Press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.

Brooks, R. A., & Stein, L. A. (1993). *Building brains for bodies*. Memo 1439, Cambridge, MA: Massachusetts Institute of Technology Artificial Intelligence Lab.

Connell, J. H. (1990). *Minimalist mobile robotics: a colony-style architecture for a mobile robot*. Cambridge, MA: Academic Press. also MIT TR-1151.

Dennett, D. C. (1978). Why you can't make a computer that feels pain. In *Brainstorms*, pages 190–229. Bradford Books, Montgomery, Vermont. page numbers are from the 1986 Harvester Press Edition, Brighton, Sussex.

Dennett, D. C. (1991). *Consciousness explained*. London, UK: Allan Lane, The Penguin Press.

Dennett , D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Brain and Behavioral Sciences, 15*, 183–247.

Freedman, D. H. (1994). Bringing up RoboBaby. *Wired*, *2*(12), pages 74,76,78,80–81.

Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM, 38*(*11*), 33–38.

Minsky, M. (1985). *The society of mind*. New York, NY, Simon and Schuster Inc.