Crude, Cheesy, Second-Rate Consciousness

Joanna J. Bryson[1,2]

Abstract. If we aren't sure what consciousness is, how can we be sure we haven't already built it? In this article I speak from the perspective of someone who routinely builds small-scale machine intelligence. I begin by discussing the difficulty in finding the functional utility for a convincing analog of consciousness when considering the capabilities of modern computational systems. I then move to considering several animal models for consciousness, or at least for behaviours humans report as conscious. I use these to propose a clean and simple definition of consciousness, and use this to suggest which existing artificial intelligent systems we might call conscious. I then contrast my theory with related literature before concluding.

1 INTRODUCTION

> "If the best the roboticists can hope for is the creation of some crude, cheesy, second-rate, artificial consciousness, they still win." — Daniel Dennett (1994), *The Practical Requirements for Making a Conscious Robot*

While leading a group building a humanoid robot in the 1990s, Rodney Brooks complained about the term *robot brain* [1]. You can have a robot hand or arm or eye or even face. But as soon as you say you have a robot brain people say "That's not a brain." The aim of this article is to make you look at some existing artificially-intelligent systems and say "You know, maybe that *is* robot consciousness."

From experience, I know this is hard to do. I was once sitting in a Cambridge, Massachusetts diner with other postdocs after Dennett had just given a seminar. The other postdocs asserted science would solve consciousness, but not in their lifetimes — not in the next hundred years. Their justification for this statement was that we knew nothing about the topic. Even if we accept this statement as fact (which I don't), they conceded that in the previous ten years there were previously many things that we'd known nothing about and had come to understand well. I believe this and more extreme beliefs about consciousness being unknowable are rooted in strong psychological desires for some aspect of human experience or action to be beyond scientific access. In general, a claim that we are "getting now closer" in science often indicates that in fact the claimant does not like the direction science is currently taking them.

While trying to understand why my colleagues were certain we were so far away from a science of consciousness, I challenged them about how a computer could prove itself conscious. Almost anyone who owns a computer can make it type or even say "I am conscious." Dennett [2] implies that our own empathy should be used to judge the achievement. But teddy bears and pet rocks do this with no intelligence at all, while sadly human history is full of people mischaracterising other people as objects.

[1] University of Bath, United Kingdom, email: j.j.bryson@bath.ac.uk
[2] The Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria

My colleagues the postdocs said that consciousness was a special sort of self-knowledge, being aware of what you are thinking. But computer programs have perfect access to all their internal states. If you set up a program correctly, you can ask it exactly what line of code — what instruction — it is executing at any time, and precisely what values are in its memory. This is in fact the job of program debugging software, such as an Interactive Development Environment (IDE). IDEs are a common type of program which are not generally considered even to be AI, let alone to be conscious [3].

If consciousness is just perfect memory and recall, then video recorders have it. If consciousness also requires access to process as well as memory, then computers have that access. Possibly some people are committed enough to these definitions that they are already convinced computers can be conscious. But in this article I will not focus on phenomenological theories of consciousness. I will look instead at a recent functionalist theory from philosophy, and relate that theory to what is known about the impact of consciousness on expressed behaviour. From this I will propose a new version of the theory that conscious experience correlates perfectly with a particular sort of search for appropriate action selection. Consciousness is a limited-capacity system for learning about potential connections between context and action. We direct it primarily to situations that are uncertain and immediate, which allows us to optimise our use of this resource in building our expertise in our current environments.

2 MULTIPLE DRAFTS AND CONCURRENCY

One well-known functionalist theory of consciousness is Dennett's multiple drafts theory, which starts from the fact that brains have many things going on in them at one time [4, 5]. In Dennett's more recent model, consciousness is a spotlight that shines on no more than one of these things at a time, at least it only shines brightly on one [6]. But why is the brain doing so many things at once? The reason is because if many processors run at the same time, more can get done quickly. In computer science, this is called *concurrency* [7].

Concurrency is a great strategy for problems that can be taken apart into pieces. But the "hard problem" in concurrency comes when you need to combine all or even some of the answers you find back together again. This is called the problem of *coordination*. For an example, think of bees. A colony of bees can explore a large space around their hive to find flowers by having each bee fly in a random direction. They will explore even more space by using simple rules each bee can know, like "don't fly near another bee". But how much would it help the colony if only one bee finds some really good flowers? When the bees communicate by the waggle dance, a lot of bees have to stop what they are doing to be involved, and one bee has to spend a *lot* of time and energy dancing [8]. When you consider not only the cost to the bees currently engaged in the communicative task, but also the complexity of this behaviour and the time it took to evolve, you realize the dance must represent a substantial adaptive advantage to the bee colony. Some individuals sacrifice time, and the result is that on average highly-related individuals each have a better chance of finding food and bringing it home [9, 10].

How does this relate to consciousness? I suggest that self awareness only seems a significant part of consciousness because there is a significant portion of the self of which we are *not* aware. Put another way, one of the key attributes of consciousness is that it is a "bottleneck" or constraint — a limit that makes some sub-part of an otherwise uniform whole special. In the bee case, that limiting process is the communication to others when a really good source of food has been found by one bee — the recruitment of others to a single location. This same sort of communicating role has also been suggested for consciousness [11–13], but in fact I will propose a markedly different role in the following sections. First though for this section I want to return to discussing consciousness-like elements in extant AI systems.

Some approaches to artificial intelligence also have concurrent processes which normally operate more or less independently. In AI as in other disciplines such as Psychology or EvoDevo, this decomposition of the whole into some specialised subparts is called *modularity* [14, 15]. Just as in Psychology and EvoDevo, the utility of modularity in AI is that more complicated systems can be developed more simply and operate more quickly [16, 17]. The problem of coordination in AI is called *action selection* [14]. This problem emerges whenever multiple modules are contending for a single resource [18]. An example of a "resource" in this sense can be as simple as physical location. I cannot stand and give a talk at a meeting at the same time as I enjoy myself in a café, so if I want to do both I have to find some sequential ordering for my actions. Another such resource is speech — we can only say one word at a time, so words must be sequenced. And, critically for the Dennett [6] description of his attentional spotlight theory, memory. Apparently, episodic memory is a constrained resource, and only some of the things we are thinking about or perceiving will wind up in it.

3 A FUNCTIONALIST HYPOTHESIS OF CONSCIOUSNESS

Dennett [6] has arrived at the conclusion that the only common characteristic of conscious contents is "the historical property of having won a temporally local competition with sufficient decisiveness to linger long enough to enable recollection at some later time". But the question of course is, competition for what? As Dennett points out later and I will return to in later discussion, one element for humans is *public expression*. If your current thoughts made it so far as to become verbalised, they are now a part of the public awareness. In this case the "local competition" is not only internal but also external — with other speakers. The memory is not only your own but also that of any other hearers.

But most theory of mind focuses on individual consciousness. Here it may be a little harder to see why we are only conscious of one thing at a time. Perhaps the phenomenological experience of sequencing in consciousness indicates consciousness is integral to the other sequencing problem, action selection, which I mentioned before. Norman and Shallice [19] propose that consciousness is a set of extra or special resources which are brought to the problem of sequencing behaviour when the brain is either uncertain about the correct sequence (as in a new context or when working on a new task) or when such sequencing is particularly important (as in when performing a delicate operation.) This theory is very similar to my own, with the exception that I will emphasise uncertainty, not heightened control.

Norman and Shallice (and others) have always been somewhat unspecific about what the "special resources" consciousness brings to such difficult situations might be. I am going to make a specific proposal here, although I won't entirely justify it until later in this article.

My proposal is simple — I think consciousness and episodic memory are the parts of a process for adaptable action selection. This process consists of:

1. fixing an aspect of a behaviour context in the brain, and
2. allowing the brain to search for potential actions that might be best suited to this context.

This sort of action selection is exceptional — most aspects of behaviour are predicted directly by their context and do not need such a process of search. However, because human behaviour is unusually plastic, we spend quite a lot of our time doing this sort of thing, even when the next action is not particularly difficult or pressing. Perhaps due to the tools and concepts provided by language and culture, we can even use consciousness to reason about abstract concepts no immediate sensory correlates. Thus we might think about a work we are writing when driving home when the road itself does not demand full attention.

The model I have just described of interacting attention and action I derived from a model developed by researchers in human vision, Wolfe et al. [20]. The main point of their 2000 article is that when performing a new task, one doesn't learn from that performance when one can use vision rather than memory to guide the behaviour. But my hypothesis depends primarily on an incidental model they describe in that work. This model accounts for the difference in the time it takes to find some visual stimuli compared to others.

Studies that measure the time for processing are called *reaction time* (RT) studies. In vision, if you have a field of dots where some are red and one is blue, you will find the blue one very quickly, and your RT will not depend on how many red dots there are. Similarly, if there are a number of **T**s on a screen and one **L**, you will not have trouble finding the one L, and you will find it quickly no matter how many **T**s there are. *However*, if the screen has many **T**s and many **L**s, and **T**s are both red and blue, but only one **L** is blue, it will take you a relatively long time to find the one blue **L**. Further, the more distracting objects there are (red **L**s or blue **T**s), the longer it will take you to find the blue **L**.

Why is this? Vision researchers have long agreed that part of the answer is because finding an object of a particular colour or simple shape are both things problems that your eyes' concurrent systems can handle more or less by themselves. The different cells in your early visual processing can identify whether they have a blue section or a **T** shape easily, and quickly inform whatever decision system needs to know this. But apparently identifying that something is both blue *and* a **T** cannot be done this way. Wolfe and his colleagues proposed a relatively simple explanation for what happens in this case. One just randomly looks at items with one trait and checks if they also have the other trait, until one happens to look at the right one[3]. So for example, you might just look at anything blue in the field (perhaps returning multiple times to some objects) and eventually you will either see that one is also a **T** or give up. Thus the process

[3] There is an older, more complicated theory involving building a "return inhibition map" once a potential target is recognised as inadequate. Wolfe et al point out this extra mechanism is unnecessary so long as the sampling is truly random.

of recognising and visually targeting blueness or **T**ness is not very conscious, but the process of finding a conjunction, saying "is that *both* blue *and* a **T**" apparently must be.

To try to convince you of my definition of consciousness, I will now describe two more experimental psychology examples. Then I will return to the question of conscious machines. Both of my examples concern something Dennett [6] describes as "imponderable" — consciousness in non-human species.

4 ANIMAL MODELS OF CONSCIOUSNESS

4.1 'Declarative' Memory in Rats

My first scientific interest in animal consciousness came when a colleague made passing reference to declarative memory in a rat. Whether or not rats are aware, I was quite certain they didn't declare anything, which is the definition I'd learned for that term. But there is reasonably good evidence rats have explicit episodic memory. We know this from their behaviour, and from its analogies to humans in similar situations. The humans we can ask about their conscious experience.

In this case, the person who was being asked was Henry Gustav Molaison, then known as patient HM. HM had both of his hippocampuses removed to treat his severe epilepsy, and as a result lost the ability to form new episodic memories. When I was a psychology undergraduate in the 1980s, we were taught that he had lost the ability to *consolidate* short-term memories into long-term memories, but this theory proved false. At that time it was believed that when rats had their hippocampuses lesioned (destroyed) they could still consolidate their memory, but they had certain problems with navigation, so apparently hippocampuses were for navigation in rats but memory consolidation in humans. This was also wrong — the real answer is both more parsimonious and more interesting.

What HM can't do is that he can't remember an episode after that episode finishes. So you might teach him one task which he would perform successfully, but then if you distract him by going away or introducing a new task, he could not remember even having met you afterwards, let alone that you had taught him the first task. But although he had his surgery in the 1950s, HM started acquiring semantic knowledge about John F. Kennedy and rock music. Eventually, someone stopped asking HM what he remembered, and instead gave him the same sort of task the lesioned rats were successfully learning. They brought in an apparatus and said "when that light goes on, push that button". When he did so they gave him a penny. After he'd done this for some time, they distracted him by asking him to count his pennies. After this he said he didn't know what the apparatus was for. But when the light went on, he pushed the button, just as a rat would have. When they asked him why he did that, he said "I don't know."

So now that we know that rats and humans were less different than once thought, let us return the question of rat episodic memories. One of the "navigational" tasks the rats had problems with was the radial arm maze — a maze with eight arms coming out from a centre. The trick with this maze is to remember which three arms the scientists put food in, and to go to each of them and not the others because you only have a little time in the maze. Also, you can't learn to go to the three arms in a particular order, because little doors slide up and down randomly, preventing access at irregular times. The rat thus has to remember which of the three arms you've already been down *today* to make sure to go down each of them once. When the rats had no hippocampuses, they could still learn which three arms had the food day after day, just like HM could tell you about the Beatles. But on any particular day, they didn't efficiently go down those three arms once each, like a normal rat would. Rather, they acted like they couldn't remember what they'd just been doing. Just like HM. This is what my colleague had referred to as "declarative memory". The ordinary rats (the ones that still had their hippocampuses) were showing they had it by going down the three arms each once.

For details and full referencing of the above experiments, see Carlson [21]. But the main point here for my argument, is that rats seem to have a special episodic memory, like humans. Also like humans, rats lose that memory if they lose their hippocampuses.

4.2 Absent-Mindedness in Macaques

From the above I hope we can accept that animals as much like us as rats have at least part of what we normally think of as consciousness, and that they use it for remembering things and choosing their actions. Of course, rat awareness is probably quite different from primate awareness. In a controversial set of experiments, Rolls [22] found evidence that while rats occupy their hippocampuses primarily with information about their present location, primates have more representations of the location they are *looking* at. Thus perhaps a rat is *only* self conscious, while a monkey can think about things at other locations.

I will now move on to the third experimental psychology study, on the effect of aging. One of the standard tasks studied in animal cognition is called *transitive inference*. You may remember this from math — if $A > B$ and $B > C$, then $A > C$. Science has shown surprisingly that many animals (even rats and pigeons) find the $A > C$ inference easily — *if* they can learn the two premises. However, it is very, very hard to learn two different premises involving B, one in which it is good and one in which it is bad. Thus animals (and young children) require a great deal of training to memorise the original, adjacent pairs.

The experiment I am about to describe once again depends on reaction time. There are a number of characteristic effects that happen when animals (including humans) learn a sequence of pairs such as: $A > B; B > C; C > D; D > E; E > F$. One characteristic is that the further apart two stimuli are from each other in that chain, the *faster* the animal is at making their choice. This is called the Symbolic Distance Effect (SDE). So due to the SDE, the reaction time for answering $B?E$ is on average shorter than that for answering $B?D$.

As described earlier, reaction times are normally associated with cognition. Historically, researchers have been trying to discover what computation animals might be performing that does transitive inference yet goes faster as a chain gets longer [23, 24]. But the theory of consciousness I presented above provides a different explanation. My theory predicts that the more uncertain animals are about their next action, the longer they hesitate. This allows their brain to search for a better, more certain solution, using a process like I described above for vision.

I came to this theory for two reasons. One is that I have spent some time researching mistakes children and monkeys make in performing transitive inference, and wound up supporting a model of the underlying process that explains everything *except* the SDE. Therefore I [25] — as well as some other people [26] — think the SDE is not dependent on the transitive reasoning. The second reason is even simpler — the SDE can go away and the animals still perform transitive inference correctly. Rapp et al. [27] have shown that elderly rhesus macaques perform transitive inference more quickly than their juniors and just as accurately. However, they have no SDE. All their transitive decisions are at the same reaction time, which is faster than *any* of the younger monkey's decisions.

If old monkeys can perform transitive inference without an SDE, do then what is it for? Do the older monkeys pay any penalty? Yes: they don't notice if the rewards change on one of their pairs. Because of an error in their experimental design, Rapp and his colleagues started rewarding all their monkeys on the pair $B?D$ at chance, so most of the monkeys (the younger ones) stopped performing $B > D$ and rather went to chance on choosing B or D. But the old monkeys, who hadn't been hesitating, also didn't notice the change in reward and kept choosing B.

This is just one experiment and there's clearly a lot more work to be done. But I put forward as a hypothesis that the older lab monkeys are more likely to go into "auto-pilot" mode on a simple lab task. This could be adaptive for them, since if they'd lived that long in the wild they'd probably already know how to perform most tasks. Further, they might be losing scramble competitions (the way rhesus macaques forage) to younger, more agile monkeys in their troop [28]. Thus learning is probably less important than speed for elderly monkeys. Of course, we can't be sure that they are performing their transitive inference decisions without conscious awareness, because we can't ask them directly about their memory. But hopefully we will find a way to extend this research into human subjects.

5 DO WE HAVE CONSCIOUS MACHINES YET?

Now I return to the question of whether we have already achieved machine consciousness. Maybe not the full rich human pageantry of narrative with qualia, meta-reasoning and everything, but perhaps what Dennett has called "crude, cheesy, second-rate artificial consciousness" [29, p. 137]. What I have proposed above (taken all together) is that calling something "conscious" requires several things:

1. There must be multiple, concurrent candidate processes for conscious attention.
2. There must be some special process applied to a selected one of these processes.
3. This special process must achieve some function, probably concerning sequencing actions. And,
4. as a side effect, the object of this attention will normally be recorded in episodic memory, at least for a while.

Do any machines meet these criteria? I think probably yes. As pathetic as they are compared to humans or our science fiction, I think many of the humanoid robot systems which engage in dialog with human users and attempt to select objects from table tops can probably be thought of as meeting all these criteria in a crude, cheesy sort of way. Such robots are at MIT, Georgia Tech and the University of Birmingham, to name just a few [30–32].

If you think on a larger, Chinese-room sort of scale for a cognitive system, we might also see AI playing a part in other kinds of consciousness. For example, the Internet employs massive concurrency to create a world-wide database of useful information. If someone wants to act on a piece of that information, they employ a search engine to limit their view of all that data to say ten URLs with context on a single web-page. Under the definition of consciousness above, a page enters the consciousness of the system as a whole at the same time it enters the consciousness of the human being who is doing the final selection of the page to be viewed.

Notice that the browser or search-engine on their own would *not* be conscious, because both require the human to do the actual sequencing. However, the human, the browser *and* the chosen search-engine provider (e.g. Google) all retain explicit memory of the selected Internet item and some summary details about its selection, at least for some time. The browser will use this memory to suggest that page to the person again; the search company will use this memory to make it more likely this page is shown to other people who search, and the human will use the information for whatever they originally intended (or possibly something else). Thus in a way a single action selection mechanism is used concurrently by three different cognitive systems. And I think the two forms of consciousness that have AI elements are not too unlike what Dennett [6] refers to as "the publication competence". They are making public conscious information, and this he describes as the final arbiter of what, for a human, is conscious.

6 WHAT THIS THEORY IS NOT

Note that this theory is entirely agnostic about qualia, self representation and so forth. The phenomena described by Lenggenhager et al. [33] for example could well correlate to the sorts of information frequently used by the conscious search process as part of its action selection.

This work is not identical to the currently-popular Global Workspace Theory (GWT) [11, 13]. As I said earlier, while my theory does relate to some coordinated effort between brain systems, the same could be said of any mental process. But I do not believe that *any* process in the brain is global, for simple reasons of combinatorics [34]. I have recently come to believe that processes like those described by Shanahan [13] could well determine the highest-level task- or goal-selection algorithms in autonomous systems, systems that in animals largely correlate to chemical / hormonal regulation systems, [35, 36]. This is an important part of action selection and also one that may be combinatorially accessible. But it is not the same as detailed, dextrous action selection. Much AI experimentation with spreading-activation systems of action selection has shown that these systems do not scale to any sort of complex action selection such as is displayed by mammals [37, 38].

This is not to say I dislike all or even most of the content of the current GWT as described by [39]. My theory covers a far smaller range of the conscious phenomena, but also an aspect which Baars does not concentrate on. The main purpose for consciousness to Baars is to integrate a large variety of information sources. The main purpose of consciousness for me is to allocate an appropriate amount of time to learning about and searching for the next action. These theories may be perfectly compatible. Baars' mechanisms could well be seen as the *how* of consciousness, and the *why is it like that?* Here my theory has focussed on primarily on the *when* and the *what is it for?*

7 CONCLUSION

The goal of this article has been to convince you that there may already be a robot consciousness, at least to the same extent that there are already robot hands and robot legs. Part of the reason we have trouble understanding consciousness is because the term has origins in folk-psychology and as such covers a large range of phenomena, some of which are probably not particularly related [40]. What I have done here is concentrate on two criteria for consciousness Dennett [6] identifies:

1. that it is something that happens to one candidate process among many, and
2. that it creates a lasting impression in something like episodic memory.

From this I have proposed that consciousness is part of a particular process of action selection — one that is triggered by uncertainty and allows for the exploration and association of new actions in a particular context. This is in contrast to the majority of action selection, which is more-or-less reducible to stimulus-response, possibly also with some automated arbitration [41]. From this I have been able to argue that we can find evidence of consciousness not only in animals but also in *existing* AI systems.

None of my arguments are meant to belittle consciousness in any way, although obviously as a functionalist I am happy if they help demystify it. I am not claiming consciousness is emergent, epiphenomenal or being otherwise antirealist. Rather, consciousness is a central process to the part of intelligent behaviour I am most happy to call "cognitive".

Explaining how something works is by no means the same as explaining it away. Similarly, by disassociating consciousness from mystic ideas of soul I do not deny the central role of a concept of self in current human morality, nor the critical importance of moral behaviour to any social species. Even the crude, cheesy, second-rate artificial consciousnesses I have described are not I think belittled by that description — anything but. I think clarifying our concepts on cognition can help us appreciate the progress we have already made in AI as well as improve our approaches. Hopefully as we develop more informed perspectives on intelligence, we will begin building more useful — and more conscious — cognitive systems.

Acknowledgements

This article was originally commissioned for and presented to The Second Vienna Conference on Consciousness (September 2008) as a discussion of Daniel Dennett's contribution to that meeting. Thanks to John Dittami for the invitation; to Wayne Christiansen and Alejandro Rosas Lopez for comments on an earlier draft; and to Dittami and Dennett for their comments at the meeting. This research is supported by a fellowship from The Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria. Thanks to the University of Bath for providing me the sabbatical leave to pursue it.

REFERENCES

[1] Rodney A. Brooks and Lynn Andrea Stein. Building brains for bodies. *Autonomous Robots*, 1(1):7–25, 1994.

[2] Daniel C. Dennett. *The Intentional Stance*. The MIT Press, Massachusetts, 1987.

[3] Joanna J. Bryson. Consciousness is easy but learning is hard. *The Philosophers' Magazine*, (28):70–72, Autumn 2004.

[4] Daniel C. Dennett. *Consciousness Explained*. Little Brown & Co., Boston, MA, 1991.

[5] Daniel C. Dennett and Marcel Kinsbourne. Time and the observer: The where and when of consciousness in the brain. *Brain and Behavioral Sciences*, 15:183–247, 1992.

[6] Daniel C. Dennett. Can we really close the cartesian theater? Is there a homunculus in our brain? In John Dittami, editor, *Proceedings of the Vienna Conferences on Consciousness*. University of Vienna Press, 2009. in press, available from the Web.

[7] Michael Sipser. *Introduction to the Theory of Computation*. PWS, Thompson, Boston, MA, second edition, 2005.

[8] Karl von Frisch. *The Dance Language and Orientation of Bees*. Harvard University Press, Cambridge, MA, 1967.

[9] Ellouise Leadbeater and Lars Chittka. Social learning in insects — from miniature brains to consensus building. *Current Biology*, 17(16):703–713, 2007.

[10] Ivana Čače and Joanna J. Bryson. Agent based modelling of communication costs: Why information can be free. In C. Lyon, C. L Nehaniv, and A. Cangelosi, editors, *Emergence and Evolution of Linguistic Communication*, pages 305–322. Springer, London, 2007.

[11] Bernard J. Baars. *In the theater of consciousness: The workspace of the mind*. Oxford University Press, USA, 1997.

[12] Peter Carruthers. The cognitive functions of language. *Brain and Behavioral Sciences*, 25(6):657–674, December 2003.

[13] Murray P. Shanahan. Global access, embodiment, and the conscious subject. *Journal of Consciousness Studies*, 12(12):46–66, 2005.

[14] Joanna J. Bryson. Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12(2):165–190, 2000.

[15] Joanna J. Bryson. Modular representations of cognitive phenomena in AI, psychology and neuroscience. In Darryl N. Davis, editor, *Visions of Mind: Architectures for Cognition and Affect*, pages 66–89. Idea Group, 2005.

[16] Richard Samuels. The complexity of cognition: Tractability arguments for massive modularity. In P. Carruthers, S. Laurence, and S. Stich, editors, *The Innate Mind: Structure and Contents*, pages 107–121. Oxford University Press, 2005.

[17] Marc W. Kirschner, John C. Gerhart, and John Norton. *The Plausibility of Life*. Yale University Press, New Haven, CT, 2006.

[18] Bruce Mitchell Blumberg. *Old Tricks, New Dogs: Ethology and Interactive Creatures*. PhD thesis, MIT, September 1996. Media Laboratory, Learning and Common Sense Section.

[19] Donald. A. Norman and Tim Shallice. Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, and D. Shapiro, editors, *Consciousness and Self Regulation: Advances in Research and Theory*, volume 4, pages 1–18. Plenum, New York, 1986.

[20] Jeremy M. Wolfe, Nicole Klempen, and Kari Dahlen. Postattentive vision. *The Journal of Experimental Psychology: Human Perception and Performance*, 26(2):293–716, 2000.

[21] Niel R. Carlson. *Physiology of Behavior*. Allyn and Bacon, Boston, seventh edition, 2000.

[22] Edmund T. Rolls. Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9:467–480, 1999.

[23] Peter E. Bryant and Thomas Trabasso. Transitive inferences and memory in young children. *Nature*, 232:456–458, August 13 1971.

[24] Thomas R. Shultz and Abbie Vogel. A connectionist model of the development of transitivity. In *The 26^{th} Annual Meeting of the Cognitive Science Society (CogSci 2004)*, pages 1243–1248, Chicago, August 2004. Lawrence Erlbaum Associates.

[25] Joanna J. Bryson and Jonathan C. S. Leong. Primate errors in transitive 'inference': A two-tier learning model. *Animal Cognition*, 10(1):1–15, January 2007.

[26] Brendan O. McGonigle and Margaret Chalmers. Monkeys are rational! *The Quarterly Journal of Experimental Psychology*, 45B(3):189–228, 1992.

[27] Peter R. Rapp, Mary T. Kansky, and Howard Eichenbaum. Learning and memory for hierarchical relationships in the monkey: Effects of aging. *Behavioral Neuroscience*, 110(5):887–897, October 1996.

[28] Lynne A. Isbell. Contest and scramble competition: patterns of female aggression and ranging behavior among primates. *Behavioral Ecology*, 2(2):143–155, 1991.

[29] Daniel C. Dennett. The practical requirements for making a conscious robot. *Philosophical Transactions: Physical Sciences and Engineering*, 349(1689):133–146, October 15 1994.

[30] Deb K. Roy and Alexander P. Pentland. Learning words from sights and sounds: A computational model. *Cognitive Science*,

26(1):113–146, 2002.
[31] Cynthia. Breazeal, Matt Berlin, Andrew Brooks, Jesse Gray, and Andrea L. Thomaz. Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, 54(5):385–393, 2006.
[32] Nick Hawes, Aaron Sloman, Jeremy Wyatt, Michael Zillich, Henrik Jacobsson, Geert-Jan Kruijff, Michael Brenner, Gregor Berginc, and Danijel Skočaj. Towards an integrated robot with multiple cognitive functions. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI-07)*, pages 1548–1553, 2007.
[33] Bigna Lenggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke. Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099, 24 August 2007.
[34] Joanna J. Bryson. Language isn't quite *that* special. *Brain and Behavioral Sciences*, 25(6):679–680, December 2002. commentary on Carruthers,"The Cognitive Functions of Language", same volume.
[35] Joanna J. Bryson and Emmanuel A. R. Tanguy. Simplifying the design of human-like behaviour: Emotions as durative dynamic state for action selection. In J. Vallverdú and D. Casacuberta, editors, *The Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*. IGI Global, Hershey, PA, 2009. in press.
[36] Philipp Rohlfshagen and Joanna J. Bryson. Improved animal-like maintenance of homeostatic goals via flexible latching. In *Proceedings of the AAAI Fall Symposium on Biologically Inspired Cognitive Architectures*, 2008.
[37] T. Tyrrell. An evaluation of Maes's bottom-up mechanism for behavior selection. *Adaptive Behavior*, 2(4):307–348, 1994.
[38] Eddy J. Davelaar. Sequential retrieval and inhibition of parallel (re)activated representations: A neurocomputational comparison of competitive queuing and resampling models. *Adaptive Behavior*, 15(1):51—71, March 2007.
[39] Bernard J. Baars. Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. In S. Laureys, editor, *The Boundaries of Consciousness: Neurobiology and Neuropathology*, volume 150, chapter 4, pages 45–53. Elsevier, 2005.
[40] Daniel C. Dennett. Are we explaining consciousness yet? *Cognition*, 79:221–237, 2001.
[41] Tony J. Prescott. Forced moves or good tricks in design space? Landmarks in the evolution of neural mechanisms for action selection. *Adaptive Behavior*, 15(1):9–31, March 2007.