A Role for Consciousness in Action Selection

Joanna J. Bryson[1]

Abstract. This paper argues that conscious attention exists not so much for selecting an immediate action as for focusing learning of the action-selection mechanisms and predictive models on tasks and environmental contingencies likely to affect the conscious agent. It is perfectly possible to build this sort of system into machine intelligence, but it is not strictly necessary unless the intelligence needs to learn and is resource-bounded with respect to the rate of learning vs. the rate of relevant environmental change. Support of this theory is drawn from scientific research and AI simulations, and a few consequences are suggested with respect to self consciousness and ethical obligations to and for AI.

1 Introduction

Consciousness is first and foremost a culturally-evolved concept of uncertain age and origin (Dennett, 2001). As such it is not at all clear that the many things we call *consciousness* are truly aspects of a single psychological phenomenon. Even were they to be so, we would not necessarily know the phylogenetic priority between the various traits we identify with consciousness.

For the purpose of this article at least, I will focus on a completely functionalist account of consciousness and intelligence more generally. Consciousness is one evolved element of intelligence, and presumably serves a role within the cause of intelligence. I will start from the assumption that the cause of intelligence, it's essential role, is primarily to do the right thing at the right time. Intelligence survives natural selection entirely as a consequence of the advantage the actions it generates gives its host, and their outcomes in terms of the agent's (or at least, the agent's genes' Dawkins, 1982; West et al., 2007) survival and ability to reproduce.

If consciousness is adaptive in nature then it could well be useful for AI as well. This might not be true if for example consciousness is essentially a mechanism for implementing serial processing on the massively-parallel architecture which is the vertebrate brain. Since AI to date has tended to be minimally concurrent we might even in that case need some kind of "reverse consciousness" to harness the power of concurrency with our sequential systems.

In this paper though I analyse a theory that consciousness is a strategy available to agents capable of learning new behaviour to combat the combinatorics of the search for appropriate actions. I have previously argued that there exists a class of reaction time results that are determined not by the cognitive complexity of the task being performed, as is generally postulated. Rather delays in processing reflect an allocation of time by the learning-competent agent to online search for a better solutions (Bryson, 2009a,b). The amount of time allocated to this search in real-time by an individual depends on its confidence with respect to the task. The more certain an animal is, the less time it allocates to searching for a better solution or prediction concerning the situation. There are also species-specific and life-history components to the duration of the search. An assumption which we have yet to demonstrate in the laboratory is that the period of search correlates to conscious attention to the task and the feeling of awareness.

If we are correct in our accounts, this feeling-of-awareness part of consciousness can be shown to be shared with monkeys, rats and presumably many other intelligent vertebrates, though they may spend less time in this state and more in a state of "automatically" generating behaviour than the average human. Further, to the extent that we are willing to call this *consciousness*, this addresses the question of the utility of machine consciousness as well. Where machines benefit from applying resource bottlenecks to searching for new solutions, they might also benefit from a similar strategy for allocating those search resources. This would make a machine also functionally aware of a strategically-limited subset of its environment, rendering it much like a conscious human.

In this paper I seek to clarify this theory and then examine its implications. In Section 2 I describe conscious attention and cognition in an evolutionary context. In Section 3 I explain the details of and evidence for the theory. In Section 4 I describe its application to machine intelligence, and in Section 5 I briefly examine the theory's implications for self consciousness and ethical obligations.

2 Functionalism, Evolution, Cognition and Learning

If consciousness is useful to intelligence and intelligence is useful to survival, then why are we not conscious of everything all the time? Many theories of consciousness assume that it requires some sort of expensive resource which must unfortunately be limited, perhaps by metabolic cost or by the size of heads during child birth. Consciousness therefore inherits this scarcity and must be used frugally — directed with care at only the most important problems.

In general, where we see a variety of solutions of an apparently-adaptive trait in biology, this indicates a tradeoff between the costs and benefits of a trait, allowing the perpetuation of roughly equally-fit variation along the axis projected by this tradeoff. The best-known example of this is the tradeoff between the number of offspring an individual can have and the amount of care it can invest in each of them. Certainly the extent to which species rely on cognitive strategies for selecting appropriate actions is highly variable. Cognition — by which I mean any real-time, online modelling of the expected outcomes across some range of behaviour alternatives — is a broadly unpopular solution ignored by plants and single-cell organism, though both of these are capable of expressing behaviours in response to their environment. Bacteria move towards or away from substances and behave socially with other bacteria to improve their situation and prospects for preserving their genes, sometimes at the

[1] University of Bath, England, United Kingdom email: j.j.bryson@bath.ac.uk

cost of self-sacrifice (West et al., 2007). Plants are capable of responding not only to light and nutrients but also to pheromones of other (e.g. host) species of plants, and to direct their growth accordingly (Trewavas, 2005).

The tradeoff that follows from my proposal in the introduction is that cognitive strategies generally and consciousness cost time — time for cognitive processing delays action. Time is expensive. A delay may mean that another agent takes advantage of a situation before you. Heubel et al. (2009) demonstrate that mate competition may explain the failure of male mollies to learn to discriminate the Amazon molly *Poecilia formosa* even though 'mating' with these females gives them no fitness benefits. The time it takes to discriminate the Amazon mollies from females of the male's own species is more valuable than the cost of insemination, because those that hesitate are beaten to available conspecific females by those who do not. Even where there are no other competing agents, the situation may change before you are yourself able to take advantage of it. For example, a strategy for crossing roads must involve reaching decisions about recognising a safe window for crossing before that window disappears.

Psychometric research indicates that there is something intrinsically slow and also something noisy about biological consciousness (Norman and Shallice, 1986; Cooper et al., 1995). If this is true, then even within a highly-cognitively-resourced organism it would still be adaptive to use conscious strategies only when other mechanisms fail. Norman and Shallice (1986) describe essentially an interrupt-driven theory of consciousness where the special attention is only utilised in some circumstances, for example when a task is unfamiliar or particularly important to get right. The full version of their theory is at odds with the reports of skilled athletes, artists and musicians that their accuracy is *higher* when they are not attending to detail. However, humans and other cognitive species certainly do seem to turn our attention not only to tasks that are not familiar, but to any surprising stimulus. This phenomena underlies the popular looking-time experimental psychology paradigm for getting at what infants and other non-linguistic animals know (Spelke et al., 1992; Santos and Hauser, 2002). Again, here we see the experimentally-validated premise that organisms attend longer to things that are unfamiliar, or — in machine learning terms — that they were unable to predict.

What then is the advantage of cognitive approaches that compensates for this loss of time? Apparently, plasticity — the ability to solve problems and take advantage of opportunities that change more rapidly than other ways of acquiring action selection rules, e.g. evolution or implicit learning, can manage.

3 Timing, Awareness and Learning

In the previous sections I have argued that a fundamental cost of consciousness is time. Assuming that consciousness is engaged in some form of computation, then the source of this time penalty is combinatorics (Sipser, 2005). There are potentially-infinite combinations of contexts to consider as triggers for an uncountable set of nuanced actions. However, no agent computes all possible actions or explanations. Organisms are not only restricted by time. Evolution has limited organisms' action and perception abilities, and it further restricts their capacities to learn to associate actions and perceptions even within their species' competence. As the behaviourists proved while failing to validate Skinner's behaviourism, even simple stimulus-response conditioning does not work for all stimuli to all responses. Pigeons can learn to peck for food, but cannot learn to peck to avoid a shock. They can, however, learn to flap their wings to avoid a shock, but not for food (Hineline and Rachlin, 1969). Rats presented with 'bad' water learn different cues for its badness depending on the consequences of drinking it. If drinking leads to shocks, they condition to visual or auditory cues, but if drinking leads to poisoning they learn taste or smell cues (Garcia and Koelling, 1966). These limitations are not handicaps, but rather adaptive advantages. They should be seen as a set of prior expectations that accelerate learning in most situations that animals of a species are likely to find themselves in.

The amount of time allocated to cognition is set by at least four different factors. First, as I proposed in the Introduction and as is suggested by reaction-time performance on some specialised tasks, individuals may allocate more attention for longer when they are less certain that they know how to behave in a context. Second, as implied my account in Section 2, the emphasis placed on cognition by a species as a whole is a part of its adaptive suite (Thierry, 2007; Müller, 2008). Hauser (1999) argues that species of primates such as tamarins that chase fast prey like insects have limited learning potential because they have evolved to be disinhibited — to maximise response time at the cost of a capacity to learn. This suggestion is also supported by Bussey et al. (1998) who report that rats can only be trained to do task learning using a touch screen if an obstacle is placed in front of the screen. Being slowed down to crawl over the obstacle apparently gives them time and / or attention — the mental presence — to be able to notice a reward schedule.

A similar failure to notice reward schedules triggered my own theory of conscious attention. This time, the failure to learn is in elderly macaque monkeys. Rapp et al. (1996) show that aged rhesus macaques have two peculiarities in their task-learning performance. First, they do not exhibit a reaction-time (RT) effect traditionally attributed to computation the task requires, yet their performance is identical to younger animals that do show this RT effect. Second, the aged macaques do not learn new behaviour when their reward schedule changes, unlike the younger animals that show the RT delay.

The task concerned is transitive inference (TI). This is a standard cognitive task introduced to developmental psychology by Piaget (1954) and to experimental psychology through Bryant and Trabasso (1971). TI formally refers to the process of reasoning whereby one infers that if, for some quality, $A > B$ and $B > C$, then $A > C$. Piaget described TI as an example of concrete operational thought, but Trabasso demonstrated it in pre-concrete-operational children. It has now been demonstrated in a variety of animals as well as young children (Grosenick et al., 2007). Performance of this "pre-cognitive" version of TI has a number of associated characteristics. The one most relevant to the present discussion is the Symbolic Distance Effect (SDE) reaction-time effect. When subjects execute a transitive comparison, they operate *faster* the further away two items are in the implied sequence. For example, a correct decision on BD would be slower than one on BE, even if E is not the last item in the sequence[2] If TI were performed by simple inference, then items further apart would be expected to take *longer*, because more inferences have to be performed. That they are in fact faster helped motivate theories that transitivity learning is somehow innately sequential. Researchers have hypothesise that the subjects somehow recognise the sequential organisation of the stimuli and represent it internally in such a way that further-removed stimuli were more easily discriminated (Bryant and Trabasso, 1971; Wynne, 1998).

However, the SDE is not a reliable individual effect, only an aggre-

[2] End items are by far the easiest stimuli in TI, because unlike intervening items they are uniformly rewarded. Thus TI studies generally exclude end items from study.

gate one (McGonigle and Chalmers, 1992). This immediately throws doubt on any computational account of the SDE. Bryson and Leong (2007) demonstrate that a stimulus-action model proposed originally by Harris and McGonigle (1994) can better account for the difficulties subjects have learning the initial stimuli pairs in the first place. It is actually fantastically difficult for cognitively-limited subjects to learn that a single stimulus is good in some situations and bad in others. Getting a substantial number of individuals to pass criteria on learning the pairs requires a careful learning regime. Bryson (2009b) shows that if we assume that animals hesitate before acting on their training in proportion to their certainty about which stimulus should take precedence, then this model can replicate the SDE in aggregate and not in individuals — just as with the live subjects.

Why then do the elderly monkeys used by Rapp et al. (1996) show neither SDE nor learning when a reward schedule has changed? I speculate that as monkeys advance in age, the probability increases that they have learned well the tasks available in their environment, and so the probability they will benefit from inhibiting acting decreases. Their very survival to an advanced age effectively increases their certainty in their actions — their age correlates to their probability of being correct. Here this regularity is detected and addressed physiologically rather than cognitively, with a reduction of neurological capacity for inhibition. It comes at a cost of reducing their capacity for learning if the environment does change in unexpected ways.

How does this relate to consciousness? Until we can replicate the no-SDE results in humans, we can not be sure. But given both the monkey and the rat results it seems intuitive that the lack of SDE correlates with the lack of conscious attention. Few would argue that consciousness plays an intrinsic role in some forms of learning. Yet implicit learning can evidently take place and people can act in response to things they learn without having an explicit model of what they are doing. Some researchers report detectable differences in the quality or reapplicability of what is learned implicitly (Martin and Alsop, 2004; Alonso et al., 2006), but to at least a superficial level the differences are often indistinguishable in the context of the task learned itself (Siemann and Delius, 1993).

What I am claiming here is that there exists a class of learning tasks that are only likely to be achieved when conducted with conscious attention. This class includes at a minimum the capacity to detect better strategies even during the performance of familiar tasks. This learning takes time, and this time is allocated by the individual in proportion to their certainty about the performance of the task. This is the third factor in the allocation of time for cognition mentioned at the beginning of this section.

The final, fourth factor is similar, but one we are more aware of and find less surprising. When we are aware there is a need for a rapid decision, we can make one. When we do so, we are also more likely to make errors (Shadlen and Newsome, 1998; Bogacz et al., 2006). Again, in humans this is a conscious as well as a cognitive phenomena, but not one I will touch on further in this article.

4 How Much Machine Consciousness Does AI Need?

As I promised in the Introduction, this paper is not about every aspect of consciousness. One of the advantages of AI and simulations more generally is that we can decompose evolved entities into their constituent parts, then attempt to demonstrate their resynthesis. If the resynthesis produces comparable results, we have a viable hypothesis. If our model is the simplest one that accurately describes the natural phenomenon it models, then it should be taken seriously.

The previous sections argue that conscious awareness — presence in the moment such as is linked to the formation of episodic memory — is correlated with the ability to learn not only episodes but also new reward schemes for task learning. Dennett has called consciousness a spotlight; my theory shifts the metaphor slightly to that of a searchlight. Action selection would in many cases go forward in the same way without the searchlight, except that it would in fact be *faster* in the darkness. The process of search requires not only special cognitive capacities but also time.

From a computational or machine learning perspective the advantages of this kind of system is easy to justify. Suppose we have a system which learns, but it cannot learn fast enough to build a complete model of its environment. This might be either because its environment keeps changing, or its life is short and its environment is complex, or because its rate of action depends on the complexity of its model so it needs to keep its model simple by constantly generalising it and forgetting something of the past. At any rate, the system needs to choose a subset of its environment to concentrate its learning ability — it's learning *attention* — on. What would be a good set of criteria? Two obvious ones would be:

1. It should focus attention on the actions it is currently taking. This makes sense because any action it takes now it is likely to need to take again in the future — the things that it is acting upon are quite likely to be of some significance to it.
2. It should focus attention longer on things that it attends to but cannot predict.

If we combine these rules with the predispositions we find in nature to focus attention at least briefly on unexpected, loud or novel sounds or visual motion, then we might get quite an effective model of animals like grazing deer or cows. If we added in a drive to actively explore the manipulation of novel situations and affordances, we could simulate more creative species like predators or primates.

Of course a pressing concern from an AI perspective is — where in the action-selection process should the inhibition happen? The answer might seem to be obviously somewhere towards the beginning, since if a new perspective or alternative *is* discovered in the time allocated, selection can be improved. However note that in real animals and children, "looking" knowledge is not perfectly correlated with acting knowledge (Santos and Hauser, 2002), and indeed some kinds of learning experiences do not seem to affect action selection until after a night's sleep (Ellenbogen et al., 2007). If neuroscience research like Shadlen's is representative of more complex tasks, then it really may be simply a general and ubiquitous slowing of the action selection process, and the advantages of insight may just be happenstance where they occur in time. It seems to me likely that a candidate action is chosen quickly and then its execution is inhibited while the perceptual cues that elicited that response and the expectations driven by the intended action are allowed to play themselves out in the agent's working memory to see if alternative strategies become more attractive or alternative explanations seem more likely. If a better resolution does emerge the agent might be described as experiencing insight as it flushes its old plan and selects a new one.

5 Implications: Self Knowledge, Language and Ethics

Obviously there are many other aspects to the public concept of *consciousness* than these periods of awareness and basic capacities for learning models and correlations. I would now briefly like to talk about how some of these may follow from what I propose to be the most basic aspect of conscious attention.

The most obvious claim is that self consciousness isn't just consciousness, it's consciousness of the self, something that obviously requires a capacity for consciousness *and* a concept of self. In our culture, acquisition of the self concept is of course facilitated by language and shaped by culture. I stand in complete agreement with the recent work of Dennett (2009) and more generally with the Extended Mind Hypothesis (Wheeler, 2010) that consciousness and cognition more broadly are significantly enhanced, extended by and dependent on material and social culture. But I do not think that this essential aspect of consciousness attention requires language or culture. Further, I doubt that consciousness is necessary for AI to exploit language and culture where those are able to be learned by brute force rather than in a systematic, task-driven way. I would argue that Google Search is absolutely an AI application that exploits human culture, but I don't see a reason to refer to Google Search as conscious.

To return to self consciousness, I doubt also given the difficulty that children and even adults have in learning that every person *is* a person just like they are, that species without human language or culture do reliably achieve self awareness. Some individuals of social species do seem to show self consciousness, but I wouldn't take that as indicative that every individual is able to apply the rules it has learned to reason about others' behaviour to reasoning about its own. Google on the other hand has many searchable representations of itself and treats itself exactly like any other company or web presence. Thus self-awareness is neither necessary nor sufficient for conscoiusness(Bryson, 2004).

One impediment to relatively simple explanations of attention and self concept such as those above is that our culture has an enormous amount of moral and ethical associations linked with consciousness. It is easy to imagine why there would be a confounding of consciousness with ethical obligation. Ethics is an evolved mechanism for sustaining societies, and it is most efficient when it appropriately allocates responsibility. Those who are aware are more likely to be responsible than those who are not, and also are more likely to be affected by our actions towards them. Most of our actions such as speech and gesture have relatively little impact on someone not aware of them. Only the conscious can be moral agents, but that does not necessarily imply that *all* conscious entities must be treated as moral agents.

Similarly, the technical definition of suffering involves the requirement that an animal's behaviour changes for the worse even after the end of the disphoric situation (Haskell et al., 1996). Clearly by the definitions given above this could only happen if the agent was learning (or attempting to learn) new behaviour while in the unfortunate situation. Thus this sort of conscious attention is necessary for an agent to experience suffering. But again, it is not sufficient. Even humans when in particular neurological states will not suffer even if they experience severe pain (Dennett, 1978). It is hard to comprehend some of the effects of anaesthetics, but easier to imagine building a machine able to learn to perform tasks but not to suffer.

In fact, my own opinion is that we are obliged when we make intelligent machines to make ones we are not obliged to (Bryson, 2000, 2009c, 2010). We can avoid uniqueness of body, and where there is uniqueness of mind we can ensure it is backed up appropriately. Further, any machine we build we will have built, and even if it acquires new goals we will have determined the means by which it acquires them. In this, machines and artifacts more generally are fundamentally different from the agents that evolved naturally along with us, including other people. In my opinion we should always view ourselves as essentially responsible for machines. The human condition is the process of children aging and becoming responsible first for themselves, then for their parents, but I see no reason to replicate this process with AI. Originally, our ethical systems co-evolved with our societies (de Waal, 1996; Whitehouse et al., 2011). Now as our societies change rapidly, much of this 'evolution' is through deliberated legislation. I believe the most stable solution for human society is to value humanity over robots and maintain our responsibility for the machines we make. Otherwise there will be a moral hazard for people to commit violence and vandalism through their machines. Whether the machines are capable of learning while they are acting has little impact on the consequences for human society if we allow each other to displace our responsibility onto our creations.

6 Conclusion

In this paper I have argued that the most essential part of what we ordinarily call *consciousness* — that part that generates awareness of the moment and episodic memory — is a learning system associated with but not necessary for action selection in mammals. It provides a capacity for learning subtle contingencies in action selection — for noticing (for example) that a reward schedule has changed within an apparently-familiar task. I have suggested that the reason we are not conscious of everything at all times is simple combinatorial complexity — the fact that learning takes time and time is valuable.

I have suggested that machines will need this sort of attention only to the extent that they need to learn new skills or models and that they are limited in their ability to learn. In this case they would also need a heuristic for focusing their available capacity. Again only in this case, the heuristic that has evolved for us is likely to be useful for them as well — to allocate attention on the actions you actually perform, and for a time in proportion to your uncertainty about your next action. Consciousness allows you to predict changes in your immediate environment, including those expected to result from your action.

Finally, I have argued that this sort of attention is necessary but not sufficient for a variety of other phenomena we associate with consciousness — particularly ethical phenomena. It is however neither necessary nor sufficient for self concept in AI, but almost certainly precedes it in human and animal cognition.

ACKNOWLEDGEMENTS

I would like to thank the referees for their comments which helped improve this paper.

REFERENCES

Diego Alonso, Luis J. Fuentes, and Bernhard Hommel. Unconscious symmetrical inferences: A role of consciousness in event integration. *Consciousness and Cognition*, 15(2):386–396, June 2006.

Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathon D. Cohen. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700, 2006.

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

Joanna J. Bryson. A proposal for the Humanoid Agent-builders League (HAL). In John Barnden, editor, *AISB'00 Symposium on Artificial Intelligence, Ethics and (Quasi-)Human Rights*, pages 1–6, 2000.

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

Joanna J. Bryson. Crude, cheesy, second-rate consciousness. In Mark Bishop, editor, *The Second AISB Symposium Computing and Philosophy*, pages 10–15, Edinburgh, April 2009a. SSAISB.

Joanna J. Bryson. Age-related inhibition and learning effects: Evidence from transitive performance. In *The* 31^{st} *Annual Meeting of the Cognitive Science Society (CogSci 2009)*, pages 3040–3045, Amsterdam, 2009b. Lawrence Erlbaum Associates.

Joanna J. Bryson. Building persons is a choice. *Erwägen Wissen Ethik*, 20(2):195–197, November 2009c. commentary on Anne Foerst, *Robots and Theology*.

Joanna J. Bryson. Robots should be slaves. In Yorick Wilks, editor, *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, pages 63–74. John Benjamins, Amsterdam, March 2010.

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.

Timothy J. Bussey, E. Clea Warburton, John P. Aggleton, and Janice L. Muir. Fornix lesions can facilitate acquisition of the transverse patterning task: A challenge for "configural" theories of hippocampal function. *Journal of Neuroscience*, 18(4):1622, 1998.

Richard Cooper, Tim Shallice, and Jonathon Farringdon. Symbolic and continuous processes in the automatic selection of actions. In John Hallam, editor, *Hybrid Problems, Hybrid Solutions, Frontiers in Artificial Intelligence and Applications*, pages 27–37. IOS Press, Amsterdam, 1995.

Richard Dawkins. *The Extended Phenotype: The Gene As the Unit of Selection*. W.H. Freeman & Company, 1982.

Frans B. M. de Waal. *Good Natured: The origins of right and wrong in humans and other animals*. Harvard University Press, Cambridge, MA, 1996.

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

Daniel C. Dennett. Are we explaining consciousness yet? *Cognition*, 79:221–237, 2001.

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.

Jeffrey M. Ellenbogen, Peter T. Hu, Jessica D. Payne, Debra Titone, and Matthew P. Walker. Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences*, 104 (18):7723, 2007.

J. Garcia and R. A. Koelling. The relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4:123–124, 1966.

L. Grosenick, T.S. Clement, and R.D. Fernald. Fish can infer social rank by observation alone. *Nature*, 445:429–432, 2007.

Mitch R. Harris and Brendan O. McGonigle. A model of transitive choice. *The Quarterly Journal of Experimental Psychology*, 47B (3):319–348, 1994.

M. Haskell, F. Wemelsfelder, M. T. Mendl, S. Calvert, and A. B. Lawrence. The effect of substrate-enriched and substrate-impoverished housing environments on the diversity of behaviour in pigs. *Behavior*, 133:741–761, 1996.

Marc D. Hauser. Perseveration, inhibition and the prefrontal cortex: A new look. *Current Opinion in Neurobiology*, 9:214–222, 1999.

Katja U. Heubel, Daniel J. Rankin, and Hannah Kokko. How to go extinct by mating too much: Population consequences of male mate choice and efficiency in a sexual-asexual species complex. *Oikos*, 118(4):513–520, 2009. ISSN 0030-1299.

P. N. Hineline and H. Rachlin. Escape and avoidance of shock by pigeons pecking a key. *Journal of Experimental Analysis of Behavior*, 12:533–538, 1969.

Natasha Martin and Brent Alsop. Transitive inference and awareness in humans. *Behavioural Processes*, 67:157–165, 2004.

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

Gerd B. Müller. Evo-devo as a discipline. *Evolving Pathways: Key Themes in Evolutionary Developmental Biology*, pages 5–30, 2008.

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.

Jean Piaget. *The Construction of Reality in the Child*. Basic Books, New York, 1954.

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.

Laurie R. Santos and Marc D. Hauser. A non-human primate's understanding of solidity: Dissociations between seeing and acting. *Developmental Science*, 5:F1–F7, 2002.

M.N. Shadlen and W.T. Newsome. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998.

M. Siemann and J. D. Delius. Implicit deductive reasoning in humans. *Naturwissenshaften*, 80:364–366, 1993.

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

Elizabeth S. Spelke, K. Breinlinger, J. Macomber, and K. Jacobson. Origins of knowledge. *Psychological Review*, 99:605–632, 1992.

Bernard Thierry. Unity in diversity: Lessons from macaque societies. *Evolutionary Anthropology*, 16:224–238, 2007.

Anthony Trewavas. Green plants as intelligent organisms. *Trends in plant science*, 10(9):413–419, 2005. ISSN 1360-1385.

Stuart A. West, Ashleigh S. Griffin, and Andy Gardner. Evolutionary explanations for cooperation. *Current Biology*, 17:R661–R672,, August 21 2007.

M. Wheeler. Minds, things, and materiality. In Lambros Malafouris and Colin Renfrew, editors, *The cognitive life of things: Recasting the boundaries of the mind*, Cambridge, May 2010. McDonald Institute for Archaeological Research.

Harvey Whitehouse, Ken Kahn, Michael E. Hochberg, and Joanna J. Bryson. The role for simulations in theory construction for the social sciences: Case studies concerning divergent modes of religiosity. *under revision*, 2011.

Clive D. L. Wynne. A minimal model of transitive inference. In C. D. L. Wynne and J. E. R. Staddon, editors, *Models of Action*, pages 269–307. Lawrence Erlbaum Associates, Mahwah, NJ, 1998.