

A ROLE FOR CONSCIOUSNESS IN ACTION SELECTION

JOANNA J. BRYSON

*Artificial Models of Natural Intelligence,
University of Bath,
Bath, BA2 7AY, United Kingdom
jjb@cs.bath.ac.uk*

This article argues that conscious attention exists not so much for selecting an immediate action as for using the current task to focus specialized learning for the action-selection mechanism(s) and predictive models on tasks and environmental contingencies likely to affect the conscious agent. It is perfectly possible to build this sort of a system into machine intelligence, but it would not be strictly necessary unless the intelligence needs to learn and is resource-bounded with respect to the rate of learning versus the rate of relevant environmental change. Support for this theory is drawn from scientific research and AI simulations. Consequences are discussed with respect to self-consciousness and ethical obligations to and for AI.

Keywords: Action selection; consciousness; learning; inhibition.

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 if they were 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, its 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 allow our present sequential AI systems to harness the power of concurrency.

In this article though, I analyze a theory that consciousness is a strategy available to agents capable of learning new behavior. Its role is to combat the combinatorics of the search for appropriate actions. I will review here evidence that there exists a class of reaction time results describing human and other animal task performances that are determined not by the cognitive complexity of the task being performed, as is generally postulated. We know this because even where the reaction times are flattened, performance stays largely the same. I postulate that delays in processing reflect an allocation of time by the learning-competent agent to online search for better solutions. 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 I am correct in this account, then the 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 behavior than the average human. Furthermore, to the extent that we are willing to call this state of awareness *consciousness*, this addresses the question of the utility of machine consciousness as well. Where machines exploit resource bottlenecks in the search for new solutions, they might also benefit from a consciousness-inspired 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 more like a conscious human than present machines, which have effectively unlimited self knowledge.

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

2. Functionalism, Evolution, Cognition and Learning

If consciousness is useful in intelligence and intelligence is useful for survival, then why are we not conscious of everything all the time? Many theories of consciousness assume that it requires some sort of metabolically 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. But others assume there may be computational utility in the limits of awareness.

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 resources it can invest in the survival of each. 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 modeling of the expected outcomes across some range of behavior alternatives — is a broadly unpopular solution ignored by plants and single-cell organisms, though both of these are capable of expressing behaviors in response to their environment. Bacteria move toward or away from substances and behave socially with other bacteria to improve their situation and prospects for preserving their genes, sometimes at the cost of self-sacrifice [Rankin *et al.*, 2010]. 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 Sec. 1 is that cognitive strategies generally, and consciousness in particular, cost time. Time for cognitive processing delays action, and 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 for the sake of perception are beaten to available conspecific females by those who do not.

Even where there are no other competing agents, the situation you are attempting to assess may change before you are yourself able to take advantage of it. For example, a strategy for crossing roads must involve reaching decisions about recognizing a safe window for crossing before that window disappears. Finally, time is bounded by lifespan and used by the duration of essential behaviors. Time is thus sufficiently precious that spending time with someone can serve as a signal of social value of a mutual bond. So strategies of decision making that cost time must produce a commensurate benefit.

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 where necessary. Norman and Shallice [1986] describe essentially an interrupt-driven theory of consciousness where that special attention is only utilized in special 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 unfamiliar, but to any surprising stimulus. This phenomenon 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]. Here in fact we get our first experimentally validated evidence for the premise described in Sec. 1, 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, it is plasticity — the ability to solve problems and take advantage of opportunities that change more rapidly than other ways of acquiring action selection rules (such as 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 one 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 arbitrarily nuanced forms of action. However, no agent computes or considers all possible actions. Organisms are not only restricted by time. Evolution limits 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 behaviorists proved while failing to validate Skinner's original naïve behaviorism, 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 determined by at least four different factors. First, as I proposed in Sec. 1 and as is suggested by reaction-time performance on some specialized tasks (cf. immediately below), individuals may allocate attention for a longer period of time when they are less certain that they know how to behave in a context. Second, as implied my account in Sec. 2, the emphasis placed on cognition by a species as a whole is a part of its adaptive suite [Thierry, 2007; Müller, 2008] and will therefore vary. 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 maximize 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 the time and/or triggers their attention — the mental presence — to be able to notice a reward schedule.

A similar failure to notice reward schedules triggered the theory of conscious attention I present here. 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 the computation the task requires. Rather than being slower on some subparts of the task than on others, they are uniformly quicker than younger monkeys on all subtasks. Yet their performance is identical to younger animals that do show the expected RT profile. Second, the aged macaques do not learn new behavior 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), which relates to reaction time. When subjects execute a transitive comparison, they operate *faster* the further apart 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.¹ 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 hypothesized that the subjects somehow recognize the sequential organization 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, SDE is not a reliable individual effect, only an aggregate 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

¹End 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.

individuals to pass criteria on learning the pairs requires an elaborate staged training regime. Bryson [2009a] shows that if we assume that animals hesitate before acting on their training in inverse proportion to their certainty about which stimulus should take precedence, then this model can express SDE in aggregate but not reliably in all individuals — just as the real SDE works with live subjects. This is the evidence for the first of the four factors determining the time allocated to cognition I named at the beginning of this section.

Why then do the elderly monkeys used by Rapp *et al.* [1996] show neither SDE nor learning when a reward schedule has changed? My hypothesis is 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 immediate action 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. A somewhat similar result is shown across task-learning time rather than across ontogeny by O'Hara *et al.* [2012], who demonstrate a loss in neophilia by keas once they are confident in task performance. This shift in allocation over an individual's ontogeny is the third of the four factors in allocating time to cognition (recall that species-general allocation was the second).

How does this relate to consciousness? Until we can replicate the no-SDE results in humans, we cannot be sure. But given both the monkey TI and the rat touchscreen 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.² 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 (implicitly, possibly even via neural degeneration) by the individual in inverse proportion to their certainty about the performance of the task.

The final, fourth factor determining time allocated to cognition is somewhat similar to the third, but we are more aware of it and find less surprising. When we are

²On a related note, Bannerman *et al.* [1995] show that hippocampal lesions (which are also associated with loss of episodic memory) do *not*, as commonly thought, block learning of water mazes. Rather, they block learning how to learn water mazes. An experienced rat can learn a new maze after lesioning, but a rat that has never learned such a maze cannot.

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; Drugowitsch *et al.*, 2012]. Again, in humans this is a conscious as well as a cognitive phenomenon.

4. How Much Machine Consciousness Does AI Need?

As I promised in Sec. 1, 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 and then attempt to demonstrate their resynthesis. If the resynthesis produces comparable results, we have a viable hypothesis of how that phenomenon decomposes. 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. Some have 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 generalizing 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 — its 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 toward 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 of the folk-psychological concept of *consciousness* than these periods of awareness and these basic capacities for learning models and correlations. I now turn briefly to discuss how some of these may follow from what I propose to be the most basic aspect of conscious attention.

First and most obviously: self-consciousness is not just consciousness, it is 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 [2008] 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 by no means believe that basic conscious attention as described in Sec. 4 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 do not 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 really *is* a person just like they are, that species without human language or culture reliably achieve self-awareness. Some individuals of social species do seem to show self-awareness, but I would not take that as indicative that every individual of that species is able to apply the rules it has learned for reasoning about others’ behavior 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 consciousness [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 coordinating societies, and it is therefore most efficient when it appropriately allocates responsibility — that is, expectations for action. Those who are aware are more likely to be responsible than those who are not (moral agency), and also are more likely to be affected by our actions toward them (moral patiency). Most of our actions such as speech and gesture have relatively little impact on someone not aware of them. Certainly in humans, only the conscious *can* be moral agents, but that does not necessarily imply that all conscious entities *must* be treated as moral agents [Bryson, 2012].

Similarly, the technical definition of suffering involves the requirement that an animal's behavior 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 new behavior while in an unfortunate situation. Thus in humans and probably other animals, this sort of conscious attention is necessary for an agent to experience suffering. But again, it is not sufficient. We know 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 anesthetics, 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, 2009b, 2010]. We can avoid uniqueness of body, and where there is uniqueness of mind we can ensure it is backed up appropriately. Furthermore, 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 artefacts 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.*, 2012]. Now as our societies change rapidly, much of this “evolution” occurs 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 [Bryson, 2009b, 2010]. 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 article I have argued that the most essential part of what we ordinarily call *consciousness* — the 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.

References

- Alonso, D., Fuentes, L. J. and Hommel, B. [2006] “Unconscious symmetrical inferences: A role of consciousness in event integration,” *Conscious. Cogn.* **15**(2), 386–396.
- Bannerman, D. M., Good, M. A., Butcher, S. P., Ramsay, M. and Morris, R. G. M. [1995] “Distinct components of spatial learning revealed by prior training and NDMA receptor blockade,” *Nature* **378**, 182–186.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. and Cohen, J. D. [2006] “The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks,” *Psychol. Rev.* **113**(4), 700.
- Bryant, P. E. and Trabasso, T. [1971] “Transitive inferences and memory in young children,” *Nature* **232**, 456–458.
- Bryson, J. J. [2000] “A proposal for the Humanoid Agent-builders League (HAL),” in *AISB’00 Symp. Artificial Intelligence, Ethics and (Quasi-)Human Rights*, ed. Barnden, J. pp. 1–6.
- Bryson, J. J. [2004] “Consciousness is easy but learning is hard,” *Philos. Mag.* (28), 70–72.
- Bryson, J. J. [2009a] “Age-related inhibition and learning effects: Evidence from transitive performance,” in *The 31st Ann. Meeting of the Cognitive Science Society (CogSci 2009)* (Lawrence Erlbaum Associates, Amsterdam), pp. 3040–3045.
- Bryson, J. J. [2009b] “Building persons is a choice,” *Erwägen Wissen Ethik* **20**(2), 195–197, commentary on Anne Foerst, *Robots and Theology*.
- Bryson, J. J. [2010] “Robots should be slaves,” in *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. Wilks, Y. (John Benjamins, Amsterdam), pp. 63–74.

- Bryson, J. J. [2012] "Patience is not a virtue: Suggestions for co-constructing an ethical framework including intelligent artefacts", in *The Machine Question: AI, Ethics and Moral Responsibility*, D. Gunkel, S. Torrance and J. J. Bryson (eds.), AISB, Birmingham, UK.
- Bryson, J. J. and Leong, J. C. S. [2007] "Primate errors in transitive 'inference': A two-tier learning model," *Anim. Cogn.* **10**(1), 1–15.
- Bussey, T. J., Warburton, E. C., Aggleton, J. P. and Muir, J. L. [1998] "Fornix lesions can facilitate acquisition of the transverse patterning task: A challenge for 'configural' theories of hippocampal function," *J. Neurosci.* **18**(4), 1622.
- Cooper, R., Shallice, T. and Farrington, J. [1995] "Symbolic and continuous processes in the automatic selection of actions," in *Hybrid Problems, Hybrid Solutions, Frontiers in Artificial Intelligence and Applications*, ed. Hallam, J. (IOS Press, Amsterdam), pp. 27–37.
- Dawkins, R. [1982] *The Extended Phenotype: The Gene as the Unit of Selection* (W. H. Freeman & Company, Oxford).
- Dennett, D. C. [1978] "Why you can't make a computer that feels pain," *Synthese* **38**(3), 415–456. Automation-Theoretical Foundations of Psychology and Biology, Part I (Jul., 1978), pp. 415–456.
- Dennett, D. C. [2001] "Are we explaining consciousness yet?" *Cognition* **79**, 221–237.
- Dennett, D. C. [2008] "Can we really close the Cartesian theater? Is there a homunculus in our brain?" in *Proc. Vienna Conf. Consciousness*, ed. Dittami, J. (University of Vienna Press, Vienna), available online.
- de Waal, F. B. M. [1996] *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* (Harvard University Press, Cambridge, MA).
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N. and Pouget, A. [2012] "The cost of accumulating evidence in perceptual decision making," *J. Neurosci.* **32**(11), 3612–3628.
- Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D. and Walker, M. P. [2007] "Human relational memory requires time and sleep," *Proc. Natl. Acad. Sci.* **104**(18), 7723.
- Garcia, J. and Koelling, R. A. [1966] "The relation of cue to consequence in avoidance learning," *Psychon. Sci.* **4**, 123–124.
- Grosenick, L., Clement, T. and Fernald, R. [2007] "Fish can infer social rank by observation alone," *Nature* **445**, 429–432.
- Harris, M. R. and McGonigle, B. O. [1994] "A model of transitive choice," *Q. J. Exp. Psychol.* **47**(3), 319–348.
- Haskell, M., Wemelsfelder, F., Mendl, M. T., Calvert, S. and Lawrence, A. B. [1996] "The effect of substrate-enriched and substrate-impooverished housing environments on the diversity of behaviour in pigs," *Behavior* **133**, 741–761.
- Hauser, M. D. [1999] "Perseveration, inhibition and the prefrontal cortex: A new look," *Curr. Opin. Neurobiol.* **9**, 214–222.
- Heubel, K. U., Rankin, D. J. and Kokko, H. [2009] "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.
- Hineline, P. N. and Rachlin, H. [1969] "Escape and avoidance of shock by pigeons pecking a key," *J. Exp. Anal. Behav.* **12**, 533–538.
- Martin, N. and Alsop, B. [2004] "Transitive inference and awareness in humans," *Behav. Proc.* **67**, 157–165.
- McGonigle, B. O. and Chalmers, M. [1992] "Monkeys are rational!" *Q. J. Exp. Psychol.* **45**(3), 189–228.
- Müller, G. B. [2008] "Evo-devo as a discipline," in *Evolving Pathways: Key Themes in Evolutionary Developmental Biology*, ed. Minelli, A. (Cambridge University Press, New York), pp. 5–30.

- Norman, D. A. and Shallice, T. [1986] "Attention to action: Willed and automatic control of behavior," in *Consciousness and Self Regulation: Advances in Research and Theory*, eds. Davidson, R., Schwartz, G. and Shapiro, D. (Plenum, New York), pp. 1–18.
- O'Hara, M., Gajdon, G. K. and Huber, L. [2012] "Kea logics: How these birds solve difficult problems and outsmart researchers," in *Logic and Sensibility*, ed. Watanabe, S. (Keio University Press, Japan), pp. 23–38.
- Piaget, J. [1954] *The Construction of Reality in the Child* (Basic Books, New York).
- Rankin, D. J., Rocha, E. P. C. and Brown, S. P. [2010] "What traits are carried on mobile genetic elements, and why?" *Heredity* **106**(1), 1–10.
- Rapp, P. R., Kansky, M. T. and Eichenbaum, H. [1996] "Learning and memory for hierarchical relationships in the monkey: Effects of aging," *Behav. Neurosci.* **110**(5), 887–897.
- Santos, L. R. and Hauser, M. D. [2002] "A non-human primate's understanding of solidity: Dissociations between seeing and acting," *Dev. Sci.* **5**, F1–F7.
- Shadlen, M. and Newsome, W. [1998] "The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding," *J. Neurosci.* **18**(10), 3870–3896.
- Siemann, M. and Delius, J. D. [1993] "Implicit deductive reasoning in humans," *Naturwissenschaften* **80**, 364–366.
- Sipser, M. [2005] *Introduction to the Theory of Computation*, 2nd edition (PWS, Thompson, Boston, MA).
- Spelke, E. S., Breinlinger, K., Macomber, J. and Jacobson, K. [1992] "Origins of knowledge," *Psychol. Rev.* **99**, 605–632.
- Thierry, B. [2007] "Unity in diversity: Lessons from macaque societies," *Evol. Anthropol.* **16**, 224–238.
- Trewavas, A. [2005] "Green plants as intelligent organisms," *Trends Plant Sci.* **10**(9), 413–419.
- West, S. A., Griffin, A. S. and Gardner, A. [2007] "Evolutionary explanations for cooperation," *Curr. Biol.* **17**, R661–R672.
- Wheeler, M. [2010] "Minds, things, and materiality," in *The Cognitive Life of Things: Recasting the Boundaries of the Mind*, eds. Malafouris, L. and Renfrew, C. (McDonald Institute for Archaeological Research, Cambridge).
- Whitehouse, H., Kahn, K., Hochberg, M. E. and Bryson, J. J. [2012] "The role for simulations in theory construction for the social sciences: Case studies concerning divergent modes of religiosity," *Religion, Brain Behav.* **2**(3), 182–224.
- Wynne, C. D. L. [1998] "A minimal model of transitive inference," in *Models of Action*, eds. Wynne, C. D. L. and Staddon, J. E. R. (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 269–307.