

in cognitive neuroscience, animal conditioning, cognitive and developmental psychology, and machine learning to outline a new theory of goal-directed decision making. Our basic proposal is that the brain, within an identifiable network of cortical and subcortical structures, implements a probabilistic generative model of reward, and that goal-directed decision making is effected through Bayesian inversion of this model. We present a set of simulations implementing the account, which address benchmark behavioral and neuroscientific findings, and give rise to a set of testable predictions. We also discuss the relationship between the proposed framework and other models of decision making, including recent models of perceptual choice, to which our theory bears a direct connection.

2.2 Introduction

Since the earliest days of both psychology and neuroscience, investigators interested in decision making and the control of behavior have recognized a fundamental distinction between *habitual* action and *goal-directed* or *purposive* action. Although this opposition has obvious roots in commonsense notions from folk psychology, its first rigorous expression emerged in a classic debate in the behaviorist era. On one side of this debate, Hull (1943), Spence (1956), and others characterized action selection as driven primarily by immediate associations from internal and environmental states to responses. On the other, Tolman (1932), McDougall (1923), and others portrayed action as arising from a process of prospective planning, involving the anticipation, evaluation, and comparison of action outcomes. Over time, this early view of habit and goal directedness as mutually exclusive accounts of behavior has given way to a more inclusive multiple-systems account, under which habitual and goal-directed control coexist as complementary mechanisms for action selection (Daw, Niv, & Dayan, 2005; Dayan, 2009; Dickinson, 1985; Bullock &

Seeded Content – **Behaviorism – Internet Encyclopedia of Philosophy**
<http://www.iep.utm.edu/behavior/>

Behaviorism was a movement in psychology and philosophy that emphasized the outward behavioral aspects of thought and dismissed the inward experiential, and sometimes the inner procedural, aspects as well; a movement harking back to the methodological proposals of John B. Watson, who coined the name. Watson's 1913 manifesto proposed abandoning Introspectionist attempts to make [consciousness](#) a subject of experimental investigation to focus instead on behavioral manifestations of intelligence. B. F. Skinner later hardened behaviorist strictures to exclude inner physiological processes along with inward experiences as items of legitimate psychological concern. Consequently, the successful "cognitive revolution" of the nineteen sixties styled itself a revolt against behaviorism even though the computational processes cognitivism hypothesized would be public and objective -- not the sort of private subjective processes Watson banned. Consequently (and ironically), would-be-scientific champions of consciousness now indict cognitivism for its "behavioristic" neglect of inward experience.

The enduring philosophical interest of behaviorism concerns this methodological challenge to the scientific bona fides of consciousness (on behalf of empiricism) and, connectedly (in accord with materialism), its challenge to the supposed metaphysical inwardness, or subjectivity, of thought. Although behaviorism as an avowed movement may have few remaining advocates, various practices and trends in psychology and philosophy may still usefully be styled "behavioristic". As long as experimental rigor in psychology is held to require "operationalization" of variables, behaviorism's methodological mark remains. Recent attempts to revive doctrines of "ontological subjectivity" (Searle 1992) in philosophy and bring "consciousness research" under the aegis of Cognitive Science (see Horgan 1994) point up the continuing relevance of behaviorism's metaphysical and methodological challenges.

Behaviorists and Behaviorisms

Behaviorism, notoriously, came in various sorts and has been, also notoriously, subject to variant sortings: "the variety of positions that constitute behaviorism" might even be said to share no common-distinctive property, but only "a loose family resemblance" (Zuriff 1985: 1) . Views commonly styled "behavioristic" share various of the following marks:

- allegiance to the "fundamental premise ... that psychology is a natural science" and, as such, is "to be empirically based and ... objective" (Zuriff 1985: 1);
- denial of the utility of introspection as a source of scientific data;
- theoretic-explanatory dismissal of inward experiences or states of consciousness introspection supposedly reveals;
- specifically antidualistic opposition to the "Cartesian theater" picture of the mind as essentially a realm of such inward experiences;
- more broadly antiessentialist opposition to physicalist or cognitivist portrayals of thought as necessarily neurophysiological or computational;
- theoretic-explanatory minimization of inner physiological or computational processes intervening between environmental stimulus and behavioral response;
- mistrust of the would-be scientific character of the concepts of "folk psychology" generally, and of the would-be causal character of its central "belief-desire" pattern of explanation in particular;
- positive characterization of the mental in terms of intelligent "adaptive" behavioral dispositions or stimulus-response patterns.

Among these features, not even Zuriff's "fundamental premise" is shared by all (and only) behaviorists. Notably, Gilbert Ryle, Ludwig Wittgenstein, and followers in the "ordinary language" tradition of analytic philosophy, while, for the most part, regarding behavioral scientific hopes as vain, hold views that are, in other respects, strongly behavioristic. Not surprisingly, these thinkers often downplay the "behaviorist" label themselves to distinguish themselves from their scientific behaviorist cousins. Nevertheless, in philosophical discussions, they are commonly counted "behaviorists": both emphasize the external behavioral aspects, deemphasize inward experiential and inner procedural aspects, and offer broadly behavioral-dispositional construals of thought.

with an aversive event such as toxin-induced illness (*conditioned aversion*; Adams, 1982; Adams & Dickinson, 1981; Colwill & Rescorla, 1985a, 1988) or by inducing a change in motivational state (Balleine, 1992; Balleine & Dickinson, 1994; Dickinson & Dawson, 1989). Under appropriate circumstances, this intervention results in a rapid shift in behavior either away from or toward the actions associated with the relevant outcome. Such a shift is interpreted as reflecting goal-directed behavior because it implies an integration of action-outcome knowledge with representations of outcome reward value.

Another key experimental manipulation involves breaking the causal contingency between a specific action and outcome. Here, typically, the animal first learns to associate delivery of a certain food with a particular action but later begins to receive the food independently of the action. The upshot of this “contingency degradation” is that the animal less frequently produces the action in question (Colwill & Rescorla, 1986; Dickinson & Mulatero, 1989; Williams, 1989). Such behavior provides evidence that actions are being selected based on (appropriately updated) internal representations of action-outcome contingencies, thus meeting the criteria for goal directedness.

The same definition for goal directedness extends to decisions involving sequences of action (Daw et al., 2011; Ostlund, Winterbauer, & Balleine, 2009; D. A. Simon & Daw, 2011b). An illustrative example, introduced by Niv, Joel, and Dayan (2006), involves a rat navigating through a two-step T maze, as shown in Figure 2.1 (lower right). The animal in this scenario must make a sequence of two left-right decisions, arriving by these at a terminus containing an item with a particular incentive value. A goal-directed decision at S_1 would require retrieval of a sequence of action-outcome associations – linking a left turn at S_1 with arrival at S_2 and a left turn at S_2 with cheese – as well as access to stored information about the incentive value of the available outcomes. Building on this simple example, Niv et al. (2006)

rewards. The rats had access to two levers. When a rat pressed the right lever and then the left, a bit of sucrose was delivered. When the levers were pressed in the opposite order, the rat received polycose. The sequences left-left and right-right, meanwhile, yielded no reward. Following training, one of the food rewards was devalued through satiety. When presented with the two levers in this setting, rats tended to execute the sequence yielding the nondevalued food more frequently than the opposite sequence. Ostlund et al. (2009) also showed analogous changes in sequence production following contingency degradation.

Two further standard operationalizations of goal-directed decision making derive from the classic research championed by Tolman. In the *latent learning* paradigm (Blodgett, 1929), rats run a compound T maze as shown in Figure 2.1 (upper right), until they reach the box labeled “exit.” After several sessions, a food reward is placed in the exit box. After the animals discover this change, there is an immediate reduction in the frequency of entrances into blind alleys. Animals suddenly take a much more direct path to the exit box than they had previously. In *detour* behavior, as described by Tolman and Honzik (1930), rats run a maze configured as in Figure 2.1 (left). When the most direct route (Path 1) is blocked by a barrier at location A, the animals tend to opt for the shortest of the remaining paths (Path 2). However, when the block is placed at location B, animals take the third path. In each of these cases, a change in action-outcome contingencies triggers immediate adjustments in behavior, providing a hallmark of goal-directed decision making.

2.3.1 Toward a computational account

Our interest in the present work is in understanding the computations and mechanisms that underlie goal-directed decision making, as it manifests in behaviors like the ones just described. Given the recent success of temporal-difference models in

research on habit formation, one approach might be to draw from the same well, surveying the wide range of algorithms that have developed in artificial intelligence, machine learning, and operations research for solving multistep decision problems based on preestablished contingency and incentive knowledge (see Bertsekas & Tsitsiklis, 1996; Puterman, 2005; Russell & Norvig, 2002; Sutton & Barto, 1998). We do believe that it is important to consider such procedures for their potential biological relevance,¹ and later we will circle back in order to do so. However, the theory we present draws its inspiration from a rather different source, looking to previous research in neuroscience, psychology, and computer science that has invoked the notion of a *probabilistic generative model*. In order to set the scene for what follows, we will briefly unpack this construct and highlight previous work in which it has been applied.

Generative models in psychology and neuroscience. Over recent years, a broad formal perspective has taken root within both cognitive and neural research, in which probabilistic inference plays a central organizing role. A recurring motif, across numerous applications of this perspective, is that of inverse inference within a generative model. The basic idea emerged first in research on visual perception. Early on, Helmholtz (1860/1962) characterized vision as a process of unconscious inference, whose function is to diagnose the environmental conditions responsible for generating the retinal image. In recent years, this perspective has found expression in the idea that the visual system embodies a generative model of retinal images, that is, an internal model of how the ambient scene (objects, textures, lighting, and so forth) gives rise to patterns of retinal stimulation. More specifically, this generative model encodes a conditional probability distribution, $p(\text{image}|\text{scene})$. The inference of which Helmholtz spoke is made by inverting this

¹As detailed in the General Discussion, the idea that we pursue also has precedents in machine learning, although it does not yet figure among the standard approaches to solving sequential decision problems.

generative model using Bayes' rule, in order to compute the posterior probability $p(\text{scene}|\text{image})$ (Dayan, Hinton, Neal, & Zemel, 1995; Kersten, Mamassian, & Yuille, 2004; Knill & Richards, 1996; Yuille & Kersten, 2006).

The influence of this generative perspective has gradually spread from perception research to other fields. In particular, it has played an important role in recent work on motor control. Here, the generative (or forward) model maps from motor commands to their postural and environmental results, and this model is inverted in order to establish a mapping from desired effects to motor commands (Carpenter & Williams, 1995; Jordan & Rumelhart, 1992; Kilner, Friston, & Frith, 2007; Körding & Wolpert, 2006; Rao, Shon, & Meltzoff, 2007; Wolpert, Doya, & Kawato, 2003; Wolpert, Ghahramani, & Jordan, 1995). Beyond motor control and perception, theories centering on probabilistic inference over generative models have figured in numerous other realms, including language (Chater & Manning, 2006; Xu & Tenenbaum, 2007), memory (Hemmer & Steyvers, 2009), conceptual knowledge (Chater & Oaksford, 2008; Griffiths, Steyvers, & Tenenbaum, 2007), perceptual categorization (Yu, Dayan, & Cohen, 2009), and—significantly—causal learning and the learning of action-outcome contingencies (Blaisdell, Sawa, Leising, & Waldmann, 2006; Glymour, 2001; Gopnik et al., 2004; Gopnik & Schulz, 2007; Green, Benson, Kersten, & Schrater, 2010; Sloman, 2005; Tenenbaum, Griffiths, & Niyogi, 2007).

One exciting aspect of the generative approach in psychology is that its terms can be transposed, in very much the same mathematical form, into accounts of the underlying neural computations. The notion of inverse inference within a generative model has played a central role in numerous recent theories of brain function, both in visual neuroscience (Ballard, Hinton, & Sejnowski, 1983; Barlow, 1969; T. S. Lee & Mumford, 2003; Rao & Ballard, 1999) and elsewhere (Dayan et al., 1995; Friston, 2005; Knill & Pouget, 2004; Mumford, 1992, 1994).

Goal-directed decision making as inverse inference. Our central proposal in the present work is that goal-directed decision making, like so many other forms of human and animal information processing, can be fruitfully understood in terms of probabilistic inference. In particular, we propose that goal-directed decisions arise out of an internal generative model, which captures how situations, plans, actions, and outcomes interact to generate reward. Decision making, as we characterize it, involves inverse inference within this generative model: The decision process takes the occurrence of reward as a premise and leverages the generative model to determine which course of action best explains the observation of reward.

Although this specific idea is new to psychology and neuroscience, it has a number of direct and indirect precedents in machine learning, as we later detail (Attias, 2003; Botvinick & An, 2009; G. F. Cooper, 1988; Dayan & Hinton, 1997; Hoffman, Freitas, Doucet, & Peters, 2009; Shachter & Peot, 1992; Toussaint & Storkey, 2006; Verma & Rao, 2006b). In what follows, we draw many of our raw materials from such work, but also reshape them to yield an account that makes maximal contact with existing psychological and neuroscientific theory.

Overview. The ensuing presentation is divided into three main sections, corresponding to the three levels of theoretical analysis famously proposed by Marr (1982; see also Jones & Love, 2011). We begin in the next section by considering the computational problem underlying goal-directed control. The succeeding section moves on to consider the algorithm or procedure involved in solving that computational problem. Finally, in a third section, we consider the level of neural implementation. Following these three core sections of the paper, we discuss the relationship between the present ideas and earlier work, and consider directions for further development.

2.4 Reframing the computational problem

In building a formal theory, we take as our point of departure an insight recently expressed by Daw et al. (2005; see also Dayan & Niv, 2008), which is that goal-directed decision making can be viewed as a version of *model-based reinforcement learning*. The “model” referred to in this term comes in two parts: a *state-transition function*, which maps from situation–action pairs to outcomes, and a *reward function*, which attaches a reward value to each world state. Model-based reinforcement learning refers to the project of discovering an optimal (reward-maximizing) *policy*, or mapping from states to actions, given this two-part model (Sutton & Barto, 1998).

To state this more formally: Model-based reinforcement learning begins with a set of givens, which include a set of states, S ; a set of actions, A ; a state-transition function $T(s \in S, a \in A, s' \in S)$, which specifies the probability of arriving in state s' after having performed action a in state s ; and a reward function $R(s)$, which assigns a scalar reward value to each state. The computational problem is then to choose a policy $\pi(s, a, t) = p(a|s, t)$ that maximizes expected cumulative reward over steps of action t up to some planning horizon T :

$$\operatorname{argmax}_{\pi} E \left[\sum_{t=1}^T p_t(s|\pi) R(s) \right]. \quad (2.1)$$

Our objective is to reframe this problem in terms of probabilistic inference. As a first step in that direction, the problem’s ingredients, as well as their interrelations, can be represented in the form of a probabilistic graphical model (see Bishop, 2006; Koller & Friedman, 2009; Pearl, 1988). Figure 2.2A begins construction of this model with an initial set of three nodes. The node S represents a variable indicating the decision maker’s current situation or state.² This node is shaded to indicate

²Representing state as a multinomial variable is obviously a massive simplification. However, the graphical model formalism can accommodate richer representations of state, including factored

- robots and systems* (pp. 2382–2387). Beijing, China.
- Vigorito, C. M., & Barto, A. G. (2010). Intrinsically motivated hierarchical skill learning in structured environments. *IEEE Transactions on Autonomous Mental Development*, 2(2), 132–143.
- Voicu, H., & Schmajuk, N. (2002). Latent learning, shortcuts and detours: a computational model. *Behavioural Processes*, 59(2), 67–86.
- Vul, E., & Pashler, H. (2008). Measuring the crowd within probabilistic representations within individuals. *Psychological Science*, 19(7), 645–647.
- Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, 48(3), 332–367.
- Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3), 326–339.
- Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decisionmaking. *Annual Review of Neuroscience*, 30, 31–56.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411, 953–956.
- Wallis, J. D., & Miller, E. K. (2003). From rule to response: neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology*, 90(3), 1790–1806.
- Walton, M. E., Kennerley, S. W., Bannerman, D. M., Phillips, P. E. M., & Rushworth, M. F. S. (2006). Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural Networks*, 19(8), 1302–1314.
- Watkins, C. J. C. H. (1989). *Learning from delayed rewards* (Unpublished doctoral dissertation). Cambridge University, Cambridge, England.
- Weiss, Y., & Pearl, J. (2010). Belief propagation. *Communications of the ACM*, 53(10), 94.
- White, I. M., & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal