Special Series: Computational Psychiatry

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Clinical Psychological Science
2015, Vol. 3(3) 400–421

DOI: 10.1177/2167702614562040
cpx.sagepub.com
SAGE

Decision-Theoretic Psychiatry

Quentin J. M. Huys[1,2], Marc Guitart-Masip[3], Raymond J. Dolan[4,5], and Peter Dayan[6]

[1]Translational Neuromodeling Unit, Institute of Biomedical Engineering, University of Zürich and Swiss Federal Institute of Technology (ETH) Zürich; [2]Department of Psychiatry, Psychotherapy and Psychosomatics, Hospital of Psychiatry, University of Zürich; [3]Aging Research Center, Korolinska Institutet; [4]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London; [5]Max Planck University College London Centre for Computational Psychiatry and Ageing Research; and [6]Gatsby Computational Neuroscience Unit, University College London

Abstract

Psychiatric disorders profoundly impair many aspects of decision making. Poor choices have negative consequences in the moment and make it very hard to navigate complex social environments. Computational neuroscience provides normative, neurobiologically informed descriptions of the components of decision making that serve as a platform for a principled exploration of dysfunctions. Here, we identify and discuss three classes of failure modes arising in these formalisms. They stem from abnormalities in the framing of problems or tasks, from the mechanisms of cognition used to solve the tasks, or from the historical data available from the environment.

Keywords

Bayesian decision theory, reinforcement learning, environmental prior distributions, computational psychiatry, cognition(s)

Received 10/1/13; Revision accepted 5/26/14

Psychiatric morbidity scores very highly in the World Health Organization's tally of disease burden. In psychiatric disease, pervasive impairments in decision making conspire to move achievable current opportunities out of a person's reach and lead to violations of social norms. The cumulative consequences of missed chances and poor choices result in an environment that is ever poorer in options, making progressive decline an all too predictable outcome.

Understanding how abnormal decision making can arise, which is an obvious (though not necessary) prelude to fixing it, requires us to understand the causes of behavior as well as the fault lines where it can break. As in most such cases, there is a variety of possible causal accounts that appeal to principles cut from different cloths; in this case amplified by the fact that a multiplicity of underlying structural and functional systems underlie behavior (for a recent review, see Dolan & Dayan, 2013). Here, we provide a framework for describing some of these key facets, focusing on decision making. This analysis finds roots in the field of cognitive neuropsychiatry (Coltheart, 2007; Ellis, 1998; Halligan & David, 2001), which seeks to "explain clinical psychopathologies in terms of deficits to normal cognitive mechanisms" (Halligan & David, 2001), broadening this to include a focus on the neural substrates of those mechanisms. It also fits comfortably into nascent treatments of computational psychiatry (Huys, Moutoussis, & Williams, 2011; Maia & Frank, 2011; Montague, Dolan, Friston, & Dayan, 2012).

Our framework is built around Bayesian decision theory (BDT), which offers a coherent account of normative, instrumental, choice (Berger, 1985). BDT involves four central elements: (a) a state of the environment about which the decision making agent may only have partial information (arising from prior expectations and observations), (b) a set of actions, each of which can be executed at a state, (c) a utility function describing the net cost and

Corresponding Author:
Peter Dayan, Gatsby Computational Neuroscience Institute, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom
E-mail: dayan@gatsby.ucl.ac.uk

benefit of performing each action at each underlying true state, and (d) an inference procedure. The last of these should pick the action that maximizes the expected utility, by integrating over the distribution describing its uncertainty about the current state. The mapping from observations to action is called a policy. BDT has formally easy extensions to the case that sequences or trajectories of decisions must be made in the future—it dovetails perfectly with the theory of reinforcement learning (RL; Sutton & Barto, 1998) that underpins a vast wealth of work on the neural and psychological bases of decision making.

The framework identifies three major fault lines along which normative behavior can break and abnormalities ensue. These involve the basic building blocks determining *what problem* subjects solve, the computational processes determining *how* they solve them, and the effects of *experience*.

- Solving the wrong *problem*: People will behave differently because they inherently believe or care about different things. In substance addiction, for instance, the utility function may be so skewed toward a drug that it obliterates all the negative side effects of drug taking. But the behavior may still be consistent, that is, "optimal," with respect to that particular skewed utility function, even if this interferes with other aspects of life. Similarly, an aberrant, but very strong, prior belief can result in delusions in which the ability to take evidence contrary to the belief into account is impaired, but correctly so from the perspective of someone adhering to such a belief. Aberrant behavior can then arise from a gross mismatch between a person's internal description of the decision problem and the true state of the world.
- Solving the right problem, but poorly or wrongly: Here, the notion is that mechanisms of *inference and integration* might be suboptimal or deviant, leading to incorrect estimates of states or choices of action, even if there is nothing amiss with the internal description of the components of the problem. This could arise from frank mechanistic flaws, computational limitations, or altered computational costs. Indeed, as we will see, BDT can be sufficiently intractable for man or machine that the use of approximations, heuristics, and multiple different mechanisms with idiosyncratic domains of optimal applicability all abound. If different people employ different approximations in a single circumstance, they can behave differently.
- Solving the right problem correctly, but in an unfortunate environment: In paradigms such as learned helplessness (Maier & Watkins, 2005), healthy subjects exposed to an abnormally aversive environment generalize their experience so that their behavior in other environments becomes poor. This can be seen as a natural consequence of learning from *experience* (for instance about the prior distribution over the controllability of future environments; Huys & Dayan, 2009), which itself is a central part of the way in BDT that priors and observations lead to posteriors and hence to choices.

We discuss these in turn later. But we emphasize that they are not rigidly distinct. For instance, one might naively expect that optimal agents will ultimately, given sufficient experience, become correctly *calibrated* to their current surroundings, for instance making correct judgments about the settings of parameters associated with the environment if they use the correct model class. As we shall see, this expectation is not borne out for active agents that exert control over their own sources of information. If previous environments have led to unfortunate expectations about the current one (the third problem) or if the subjects are unable to infer correctly what they need to observe (the second problem), then they might seem to be solving the wrong task (the first problem).

Equally, the heuristics involved in addressing the intractabilities of inference might be justified by particular prior expectations about the environments in which they are to be used, and so if those priors were wrong, then the heuristics would perform poorly. This would blur the lines between the first and second fault lines. For instance, under stress, it may be appropriate to expend little cognitive effort and act quickly; but if the stress response is inappropriate for the particular environment, then this might appear like a maladaptive prior on the necessity for speed.

Note that we are at the early stages of examining these questions and issues and have only very simple illustrations. The various parts of BDT can get almost arbitrarily complicated, covering the most sophisticated inferences that we and our most powerful computers are able (and ultimately unable) to make.

As a systemic account of normal and abnormal behavior, we see BDT as playing four roles (Chomsky, 1965; Marr, 1982). First, it throws into stark light the various computational components of decision making: states, representations of those states, utilities, and policies. Second, it indicates the factors of the possibly changing decision-making environment to which an organism should be sensitive if it is to perform well. Third, it provides a yardstick against which to measure the quality or optimality of actual behavior. Finally, it provides a framework for organizing the rapidly growing evidence base in terms of an understanding of the actual algorithms and

implementations that create, maintain, infer, and otherwise satisfy the various computational demands.

The first three of these roles seem rather abstract, in that they depend on the environment that the person or organism occupies more than any particular aspect of that organism. By contrast, the fourth role is quite concrete, linking to both psychological and neural data. However, it has long been recognized (Churchland, 1986) that formally clean distinctions are not actually possible between the abstract and the concrete, or to use the terms employed by Chomsky and Marr respectively, competence and performance, or computational and algorithmic/implementational theories. For instance, the yardstick measuring the optimality of behavior should really take into account the extreme temporal and energetic demands of realizing that behavior. Those demands are exquisitely sensitive to the nature of the neural substrate. In total, this means that the experimentally addressable claims we make about behavior and its neural bases cannot be couched purely in terms of abstract claims about BDT, but must rather depend on actual facts about what people do and how they do it. Indeed, we see the most critical contribution of the abstract aspects of BDT as being to inspiration and explanation.

As a final note: A well-recognized and yet pernicious problem in psychiatry is that diseases that are commonly considered to be unitary, perhaps because of certain symptoms that are shared among most patients, nevertheless include other symptoms that are quite different. Less well celebrated is that a single symptom can arise from failures of, or abnormal settings within, different of the many systems involved (for instance, one example is developed in J. Williams & Dayan, 2005). One consequence of both of these is that single conditions or diseases will recur in multiple places in this review, with different facets arising under different fault lines.

BDT

BDT has been reviewed elsewhere, including in the context of neural RL (Dayan & Daw, 2008), and so we will be brief. The true state of the environment is described by a (random) unobserved variable $\mathbf{x} = \{\boldsymbol{x}_p, \boldsymbol{x}_s\}$. For the moment, we will think of this as comprising two major components: $\boldsymbol{x}_p$ describes the problem as a whole, determining the causal texture of the domain (Dickinson, 1980); $\boldsymbol{x}_s$ is the state of the agent *in* the domain. For a maze, for instance, $\boldsymbol{x}_p$ might include the layout of all the walls and openings, the locations of foods, the description of which foods contain which nutrients. By contrast, $\boldsymbol{x}_s$ might describe exactly where in the maze the subject presently is and its current motivational state, for example, whether hungry or thirsty or both. Most work in BDT treats $\boldsymbol{x}_p$ as given and does not make it explicit. However, we are interested in the possibility that the internal representation of the problem might be erroneous and so reify this component of the state.

At any time, the agent has a prior distribution or density $\mathcal{P}(\boldsymbol{x}) = \mathcal{P}(\boldsymbol{x}_p)\mathcal{P}(\boldsymbol{x}_s|\boldsymbol{x}_p)$ over $\boldsymbol{x}$. We write $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)$ for $\mathcal{P}(\boldsymbol{x}_s|\boldsymbol{x}_p)$, to indicate the problem dependence of the prior distribution over $\boldsymbol{x}_s$. For some parts of the state (for instance often, but not always, the motivational components), the agent might have perfect knowledge. In this case, the distribution would include delta functions. For other parts, it will be uncertain; observations $\boldsymbol{y}$ will help resolve this uncertainty via a quantity known as the likelihood $\mathcal{P}(\boldsymbol{y}|\boldsymbol{x})$. Again, for the present, we focus on observations that bear on $\boldsymbol{x}_s$ and so write the likelihood as $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}\,|\,\boldsymbol{x}_s)$ noting that different problems (i.e., different values of $\boldsymbol{x}_p$) determine different likelihood functions. The agent can apply Bayes's rule to derive the posterior distribution over the state $\boldsymbol{x}_s$ as

$$\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s|\boldsymbol{y}) = \frac{\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}\,|\,\boldsymbol{x}_s)}{\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y})}, \tag{1}$$

where $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}) = \sum_{\boldsymbol{x}_s}\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}|\boldsymbol{x}_s)$ is known as the marginal distribution over $\boldsymbol{y}$, and is the overall probability of observing $\boldsymbol{y}$ in problem $\boldsymbol{x}_p$. One can imagine receiving some further information $\boldsymbol{y}'$, with likelihood $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}'\,|\,\boldsymbol{x}_s)$. One beauty of this Bayesian formulation is that if $\boldsymbol{y}'$ is independent of $\boldsymbol{y}$ given the value of $\boldsymbol{x}_s$ (and $\boldsymbol{x}_p$), one can use the posterior $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s|\boldsymbol{y})$ from Equation 1 as the prior for absorbing this extra piece of information, such that

$$\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s|\boldsymbol{y},\boldsymbol{y}') = \frac{\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s|\boldsymbol{y})\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}'|\boldsymbol{x}_s)}{\mathcal{P}(\boldsymbol{y}')}. \tag{2}$$

We usually write $\mathcal{D}$ to represent all the data that the agent has amassed, and so the full posterior is $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s\,|\,\mathcal{D})$.

Next, the agent has available a set $\mathcal{A}$ of actions that it can perform. Along with the external actions such as making a move or pressing a lever that are the normal objects of RL, there can also be internal actions such as the choice to deploy working memory or attention, or to continue evaluating an external action (Dayan, 2012; Hay & Russell, 2011; O'Reilly & Frank, 2006).

The agent is also assumed to have a utility (or inverse loss) function $\upsilon_{\boldsymbol{x}_p}(a,\boldsymbol{x}_s)$ which indicates the benefit of executing action a if the true state in the domain is $\boldsymbol{x}_s$. The utilities can be state-dependent—so eating food (an action) might be lucrative only if hungry (as described in $\boldsymbol{x}_s$). Utilities could include the internal costs associated with cognition (Kool, McGuire, Rosen, & Botvinick, 2010; McGuire & Botvinick, 2010). Exactly how utilities arise is

rather subtle; we touch briefly on some of the issues later.

In BDT, subjects combine the posterior distribution over states, which incorporates information about the information $\mathcal{D}$ they have experienced, with the utility, to arrive at optimal policies $\pi_{\boldsymbol{x}_p}(\mathcal{D})$ that maximize the expected utility. The dependence on the data $\mathcal{D}$ is critical to allow actions to be adaptive to all aspects of the state $\boldsymbol{x}_s$—including motivational components (e.g., allowing choices to differ depending on hunger versus thirst). Canonical equations for arriving at optimal policies are provided in the appendix. Policies might also be stochastic, just favoring somewhat more of those actions with higher expected utilities, rather than only picking the best. As we will see, various forms of learning are critical in BDT, allowing subjects to acquire good behavior.

We also describe domains with temporal dynamics, where the internal state $\boldsymbol{x}_s(t)$ changes over time t, with each action evoking its own set of stochastic transitions. For instance, foraging for food may exacerbate the need to find food, and wise foraging choices take this potential future state change into account. Here, it is necessary to determine not an individual optimal action, but rather a sequence of actions in the light of possible transitions. It turns out that intermediate variables can help to solve such problems more efficiently. The *value* $\mathcal{V}^{\pi_{\boldsymbol{x}_p}}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t))$ of the state $\boldsymbol{x}_s(t)$ (disregarding, for simplicity, the fact that this is not observed directly, but rather must be probabilistically inferred from data $\mathcal{D}(t)$) is defined as the long run reward expected to be available from that state through following policy $\pi_{\boldsymbol{x}_p}$ (which specifies what actions might be taken at the succeeding states). Typically, rewards received in the far future are valued less than rewards that arise soon—this is captured by exponential discounting, employing factor $0 \le \gamma \le 1$ at each step, although we discuss alternatives later. We are then interested in finding good policies; this leads to states having large values.

BDT requires two sorts of inference. One is to compute the posteriors such as those in Equation 2 to estimate the state $\boldsymbol{x}_s$ (and indeed $\boldsymbol{x}_p$) as well as possible. The other is to find the optimal actions. Both of these are computationally hard. The latter is particularly difficult when it is necessary to optimize over long trajectories of future actions—and becomes much harder in the face of the first problem. Neurobiologically relevant solutions have focused particularly on the case in which the state $\boldsymbol{x}$ is assumed known.

For instance, consider the case that the decision maker observes the state and knows the transition structure $\mathcal{T}$ embodying the consequences of actions, and the utilities $\upsilon_{\boldsymbol{x}_p}$. It can project its possible state forward in time using its knowledge of the transitions, and, by summing the utilities that it expects to encounter, work out which action is currently best (this is a way of implementing Equations 4 and 5 in the appendix). This is called *model-based* or goal-directed reasoning (Daw, Niv, & Dayan, 2005; Dickinson & Balleine, 2002; Doya, 1999).

There is an alternative class of so-called *model-free* or habitual approaches that trades inferential calculation for experience. They are retrospective, storing, or caching into memory the affective value $\mathcal{V}^{\pi_{\boldsymbol{x}_p}}_{\boldsymbol{x}_p}$ of past experience through the environment, and then recalling an aspect of this directly (Daw et al., 2005; Doya, 1999; Sutton & Barto, 1998). Crudely, actions can become highly valued either because they lead directly to high utilities $\upsilon \quad 0$ or because they make transitions to states predicting improved long-run utilities $\mathcal{V}^{\pi_{\boldsymbol{x}_p}}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t+1)) \quad \mathcal{V}^{\pi_{\boldsymbol{x}_p}}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t))$. The latter solves the decision problem because it allows myopia: The best action can be defined in terms of the value of the state to which it likely leads, with this value reflecting everything that is critical about the longer term, downstream, outcomes.

Model-based and model-free methods are both well-founded ways of finding optimal controls, and various implementations of each have been suggested. Although intermediate points between the two methods have been suggested, the two ends of the spectrum differ qualitatively: Model-based methods have superior statistical, but catastrophic computational characteristics; for model-free methods, it is exactly the other way around: They trade computational cost for experiential cost, rendering learning slow (Daw et al., 2005).

These characteristics have been used, at least to some extent, to distinguish the systems behaviorally (Dickinson & Balleine, 2002; Doya, 1999), because they imply that model-based control is instantly sensitive to manipulations of the subject's motivational state, or the contingencies in the environment and so leads to rapid behavioral change, whereas model-free control is slower to adapt, depending on averaging multiple experiences of the change. This is because the model-based system computes the values of actions prospectively whereas the model-free system computes them retrospectively.

Some of the various components and implementations of BDT can also be distinguished neurally (Balleine, 2005; Daw et al., 2005; Dickinson & Balleine, 2002). For instance, it has often been suggested that the ventromedial prefrontal cortex is heavily involved in evaluation associated with model-based control (Hampton, Bossaerts, & O'Doherty, 2006; Killcross & Coutureau, 2003; K. S. Smith, Virkud, Deisseroth, & Graybiel, 2012), whereas phasic dopamine neuron firing in the midbrain covaries with a prediction error δ that the model-free system uses to update predictions associated with states (D'Ardenne, McClure, Nystrom, &

Cohen, 2008; Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997) and actions (Morris, Nevet, Arkadir, Vaadia, & Bergman, 2006; Roesch, Calu, & Schoenbaum, 2007) and also causally influences learning and choice (Steinberg et al., 2013; Tai, Lee, Benavidez, Bonci, & Wilbrecht, 2012). Note, though, evidence that model-based and model-free methods can be integrated (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Gershman, Markman, & Otto, 2014; Johnson & Redish, 2007) may complicate these distinctions. One form of dopamine's influence is the apparently competing direct and indirect pathways through the striatum (controlling active or "go" responses, and selective inhibition or "nogo" in the terms of Frank, 2005), which are separately modulated by D_1 and D_2 dopamine receptors (Alexander & Crutcher, 1990; Frank, 2005; Kravitz, Tye, & Kreitzer, 2012). Consistent with the integration mentioned is evidence that these separated pathways are just as present in the dorsomedial as the dorsolateral regions of the striatum, which are implicated respectively in model-based and model-free control (Balleine, 2005).

It should come as little surprise that a function as critical as decision making should be distributed over a wide swathe of cortical and subcortical territories. However, from a clinical viewpoint, this means that we might expect problems to be intricately buried both in particular neural systems and their interactions—and thus hard to find and treat. More subtly, although we will discuss an exception later, some of the components that we have described, particularly those associated with model-based reasoning, may most parsimoniously be viewed in psychological terms, rather than in terms of their neural realizations. Thus, for instance, a simple psychological construct such as a changed prior might be represented by small manipulations to the efficiencies of an obscurely distributed welter of synapses. In those cases, the most appropriate clinical focus might also be psychological rather than neural. In turn, though, these psychological interventions could depend on explicit or implicit effects—for instance, changing a prior by directed reasoning about its inaccuracy, or just showing by example that it does not correctly capture the actual statistics of an environment.

Note that the application of BDT need not be restricted to just the choice of actions. It has also been applied to the choice of the vigor or latency of those actions (Niv, Daw, Joel, & Dayan, 2007) in the case that the subject has to maximize the rate with which it accumulates utility. In this case, it turns out that a critical role is played by the opportunity cost for any (unrewarded) passage of time; if this cost is high, then subjects should act vigorously; if it is low, then they can act slothfully. This turns out to be true both for actions that lead to maximal rewards and other, incidental, actions that are also executed. There is evidence that a signal like this is reported by relatively tonic levels of dopamine (Beierholm et al., 2013; Niv et al., 2007) with the effect of energizing action (Salamone & Correa, 2002).

Incorrect Problem

The most direct route to aberrant behavior arises from abnormalities in the definition of the problem itself, that is, from abnormalities in the prior, likelihood, or utility function, described collectively by $\boldsymbol{x}_p$. These capture the important and intuitive notions of behavioral variation arising from the fact that people may have different *a priori beliefs or explanatory schema*, or may endow the same event with different *meanings*, or may have different *aims*. These jointly define the decision-making task, and so determine its subjective—albeit not objective—optimal (and approximately optimal) solutions.

Before proceeding, we should emphasize that although they are conceptually very different, the prior $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)$ and the likelihood $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y} \mid \boldsymbol{x}_s)$ are multiplied in Equation 1, and so disambiguating their separate contributions to aberrant behavior may not always be possible. However, in all but limited circumstances, there is a key asymmetry between the effect of the two: Prior expectations can ultimately be washed away by sufficient observations. Two important exceptions to this are (a) if the prior rules out some possibilities, then they cannot be rescued by the likelihood and (b) if the prior indicates that the world changes quickly, then the weight of observations may not be able to accumulate to overwhelm the prior. We discuss abnormalities in priors and likelihoods in separate sections because of the conceptual differences.

Abnormalities in prior beliefs about state

Prior beliefs $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)$ play a special role in Bayesian formulations. They encapsulate internal beliefs that an agent might privately hold about the causal structure of observations, informed by experience accrued over various timescales. Such priors are most important in two cases. The first is because $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)$ determines the possible latent causes for any observation (Courville, Daw, Gordon, & Touretzky, 2004; Courville, Daw, & Touretzky, 2005; Fiser, Berkes, Orban, & Lengyel, 2010; Gershman & Niv, 2012). Most extremely, if some states $\boldsymbol{x}_s$ are not part of a subject's problem-dependent state space $\mathcal{X}_{\boldsymbol{x}_p}$, or are effectively prohibited by having a zero prior weight (which is one of the exceptions mentioned earlier), then these latent causes cannot possibly be inferred. If the prior deems some states possible, but overwhelmingly unlikely, then vast amounts of data $\boldsymbol{y}$ could be necessary

to infer them. Given that latent states are not ruled out, the second case in which priors exert substantial influence is in the face of a lack of information, that is, when they are strong relative to the likelihood terms $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y} \mid \boldsymbol{x}_s)$ or $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}' \mid \boldsymbol{x}_s)$ in Equations 1 and 2 (Koerding & Wolpert, 2004a; Weiss & Adelson, 1998). This corresponds to situations in which the current information $\boldsymbol{y}$ is ambiguous or outright uninformative.

The structure of problem-dependent prior distributions over state $\boldsymbol{x}_s \in \mathcal{X}_{\boldsymbol{x}_p}$ can influence decisions in various ways, including manipulating possible beliefs, generalization, representation and memory storage. Other aspects of priors are discussed in later sections.

Vetoing. Beliefs held with unusual conviction are a prominent feature of many psychiatric disorders. Such persistence may arise either because the problem-dependent state space $\mathcal{X}_{\boldsymbol{x}_p}$ allows only for states that would usually not be entertained or because the prior belief is very strong and puts vanishingly little prior belief on any alternatives. Via the multiplication in Equation 1 this may effectively nullify (if zero) or enforce (if a delta function) certain beliefs, irrespective of any observations. For instance, a prior distribution that places a high probability on threat (or rules out benign states) would force any ambiguous observation that could have either malign or benign interpretations, to lead to a posterior distribution that favors the former (O. J. Robinson, Charney, Overstreet, Vytal, & Grillon, 2012).

Equally, in aspects of depression that can be characterized by pessimistic expectations (Strunk, Lopez, & DeRubeis, 2006), subjects may have "dysfunctional beliefs" according to which negative explanations or causes are more likely a priori (Beck, Rush, Shaw, & Emery, 1979). Cognitive therapies hence focus both on reshaping patients' prior beliefs (Beck et al., 1979; J. M. G. Williams, 1992), as well as their interpretation of events (which concerns the likelihoods).

Generalization. The structure of priors can also influence generalization. Consider an extinction experiment, where an animal might initially experience that a conditioned stimulus (CS; which is a predictor) predicts an aversive event (an unconditioned stimulus, or US), but then suddenly starts appearing without the now feared consequence. Animals do not unlearn the association between the CS and the aversive event, as can be shown by renewal, spontaneous recovery, and reinstatement experiments (Bouton, 2002). The animal thus appears to combine stability with plasticity and learn about the novel (lack of) association without overwriting and forgetting the previous one. Prior distributions can provide a structural solution to this stability-plasticity dilemma (Carpenter & Grossberg, 1988) by allowing for different (the plasticity) latent or hidden circumstances in which relationships between stimuli and outcomes only change very slowly (the stability). When some facet of the environment appears to change quickly, this prior licenses the inference that a new circumstance has become active, rather than having to overwrite, and thus lose, existing learning in the current circumstance (Courville et al., 2004; Gershman, Blei, & Niv, 2010; Gershman & Niv, 2012; Lloyd & Leslie, 2013; Redish, Jensen, Johnson, & Kurth-Nelson, 2007; Wolpert & Kawato, 1998). One can expect generalization to be promiscuous and concrete within a circumstance, but restricted and abstract between them.

The extent to which subjects adjust to change by inferring new latent causes is determined through the medium of one aspect of their prior beliefs. Variation in this aspect of the prior could capture individual variability in relapse (Clark et al., 2006; Haug et al., 2003) after extinction-based therapies. Such therapies are first-line treatment choices for many anxiety disorders and are possibly superior to pharmacology alone (e.g., Haug et al., 2003; Hofmann, Wu, & Boettcher, 2013). Contextual effects are also well known to be important in other instances of relapse, as in drug addiction (Crombag, Bossert, Koya, & Shaham, 2008). Furthermore, priors that promote overly strong generalization can also capture a rather broad set of results on learned helplessness (Lieder, Goodman, & Huys, 2013b)—an animal model for depression (Maier & Watkins, 2005). Indeed, pathological overgeneralization is a hallmark of depression (Beck et al., 1979), where low self-esteem serves as an general explanation for all perceived failures (Carver & Ganellen, 1983).

Representation. Prior distributions can also influence generalization via representation. A notion that is standard in unsupervised learning (Hinton & Sejnowski, 1999) is that the neural representation of a state such as $\boldsymbol{x}_s$ can be seen as the coordinates of that state in a coordinate system that characterizes the whole collection of possible states $\mathcal{X}_{\boldsymbol{x}_p}$, as determined exactly by the prior distribution $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)$. Representations satisfy properties such as the sparsity or mutual independence of their components; these emerge from the structure of the prior. The resulting representations can play a central role in inference. Consider a simple probabilistic reversal learning task, in which subjects have to choose one of two stimuli. One stimulus always yields rewards with a high probability, the other with a low probability. However, at times not known to the subject, the identity of the better stimulus switches. If the prior distribution mandates a representation in which the anticorrelation between the outcomes is explicit, then learning that one stimulus no longer provides reward will automatically generalize to the prediction that the other stimulus will

provide reward. Conversely, if the prior distribution mandates a representation in which the anticorrelation is not explicit, then no such generalization will be licensed.

Information storage. Prior beliefs, likely via this impact on representation, also influence the formation and preservation of memories. Chess masters are better than inexperienced players at remembering chess board configurations, but only for configurations that are likely to arise in real games (Chase & Simon, 1973). Equally, features of a story that fit with cultural expectations are more easily remembered (Bartlett, 1932). This can be understood in terms of the information content of events, or their latent causes, under the prior: The more unlikely, the more (Shannon) surprise they contain, and hence the more memory resources have to be devoted to their encoding and storage (MacKay, 2003). This clearly poses a danger to normative inference, in that the process described in Equation 2 becomes more efficient for some observations than others—potentially forming a conduit for the rapid development of maladaptively strong beliefs (Garety et al., 2005; though see Moutoussis, Bentall, El-Deredy, & Dayan, 2011). However, note also the flip side of this—events that are unlikely might seem more important targets of encoding and storage, despite those extra costs.

Abnormalities in the likelihood

The likelihood defines how experiences $\boldsymbol{y}$ relate to internal states or other unobserved (hidden) variables in the world $\boldsymbol{x}_s$. Roughly speaking, the likelihood defines not only the "experience" of low-level sensory phenomena, but, at a more abstract description level, also the "meaning" of events as defined by the effect on inference about the hidden state. As such, it is one of the key links determining the consequences of experience.

One very general fault line associated with the likelihood arises from an inability to represent the true state of affairs for observations. For instance, the probability of rare events is often overestimated (if explicit; Kahneman & Tversky, 1979) or underestimated (if experiential; Jessup, Bishara, & Busemeyer, 2008), in a way that affects risk sensitivity (Tom, Fox, Trepel, & Poldrack, 2007). Inaccuracies can also arise indirectly from noise in the representation of the events $\boldsymbol{y}$. For instance, it is often critical to estimate the rates of affectively charged outcomes. Rate estimation often depends on time estimation, which is in turn modulated by affective events (Droit-Volet & Meck, 2007) and which is notoriously uncertain (as in the substantial studies of interval timing; Gibbon, 1977). If the way that this uncertainty grows with interval length is incorrectly characterized in $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y} \mid \boldsymbol{x}_s)$, then systematic biases can ensue.

Errors in processes associated with corollary discharge have been suggested as offering a specific path to aberrant likelihoods. When subjects act on the world, they can expect particular sorts of sensory input arising from their own behavior. Corollary discharge is the general term for the neural signals that provide information about forthcoming movement; these can "cancel" out predictable input, leaving what is unpredicted to be processed (Crapse & Sommer, 2008; Stephan, Friston, & Frith, 2009). If these signals or the way that they are processed are dysfunctional—perhaps because of problems with connectivity (Stephan, Baldeweg, & Friston, 2006), with different perceptual or decision-making modules no longer being correctly calibrated with each other—then the likelihoods of the input will become erroneous because the probabilities of the actual input will not reflect the aspects of that input that are self-created. This leads to incorrect inferences about true causes in the world, and possibly delusions (Adams, Stephan, Brown, Frith, & Friston, 2013; Stephan et al., 2009). It applies not only to external actions; internally directed actions including the processes involved in model-based evaluation and episodic future thinking (Hassabis, Kumaran, Vann, & Maguire, 2007; Hassabis & Maguire, 2009; Schacter, 2012; Schacter et al., 2012), a human analogue of the well-described phenomena of hippocampal preplay in rodents (Johnson & Redish, 2007; Pfeiffer & Foster, 2013), could also be misattributed, leading to hallucinations (Bentall, 2004).

More mechanistically, schizophrenia has been related to electrophysiological responses to sensory stimuli in the visual and auditory cortex (Brenner et al., 2009) that are thought to be related to imbalanced glutamate/GABA neurotransmission as a result of a loss of GABAergic interneurons at the cortical level (Lewis, 2013; Lisman et al., 2008). This would be an example of a case in which the most parsimonious description of the problem would be neural rather than psychological. However, depending on the imbalances in different pathways, it might appear as an example of a wrong likelihood, where aberrant processing of incoming information leads to an inability to represent the true state of affairs for the observations. In this context, when the abnormal likelihood is prominent (or the prior about the "normal" perception is weak), a delusional interpretation of the reality may emerge.

Disturbances that may readily be ascribed to aberrant likelihood functions exist in many nonpsychotic diseases, too. For instance, depression risk is associated with signatures of increased neural threat signals (Roiser, Elliott, & Sahakian, 2012) even in the absence of any overt behavior or conscious perception. Equally, work using trust games has shown that patients with borderline

personality disorder are unable to perceive and correctly interpret social approaches after a bond of trust has been partially broken (King-Casas et al., 2008).

Utility function

The utility defines the subjective values of actions and states. As such, it determines what actions are subjectively optimal. An abnormal setting in the utility function implies an abnormality in the reinforcement derived from consummatory aspects of *primary* reinforcers, such as food, sex, safety, or danger. It can lead to a diverse wealth of inappropriate behavior that is nevertheless perfectly in accordance with subjects' goals. Specific particularities can have quite circumscribed effects, and indeed may contribute substantially to the normal variation in personalities and personal preferences—*de gustibus non est disputandum*.

One characteristic issue, for instance, is that the utility function might decrease or flatten for appetitive outcomes, reducing their absolute or relative attractiveness. An example of this is anhedonia, a core feature of depression (Bylsma, Morris, & Rottenberg, 2008; Hasler, Drevets, Manji, & Charney, 2004; Pizzagalli, Jahn, & O'Shea, 2005; Treadway & Zald, 2011). Clinically, it is characterized by a loss of pleasure from things people used to like, or lack of caring about those things. Similar alterations exist in the setting of schizophrenia with negative symptoms, or apathy in Parkinson's disease. Indeed, a reduced or flattened utility function might produce psychomotor retardation by affecting vigor through decreases in the opportunity cost of time, as discussed earlier (Mazzoni, Hristova, & Krakauer, 2007; Niv et al., 2007). The opposite, where certain events become more intrinsically rewarding or aversive are apparent in the euphoric states that characterize mania, or in the hyperalgesia associated with both depression or chronic pain syndromes (American Psychiatric Association, 1994; World Health Organization, 1990). In addictions, the utility function may become skewed toward drugs: In alcoholism, for instance, drink cues become potent activators of key striatal motivational areas such as the nucleus accumbens at the expense of monetary reinforcers (Wrase et al., 2007).

We include discounting under the umbrella of issues to do with the utility function, because the discount factor, γ, controls the relative utility of rewards that arrive sooner compared with those that arrive later. If someone's value of γ is near to 0, then they will have an extremely short-term outlook, favoring immediate small rewards over far larger rewards that might be a little delayed—this can be a source of apparent impulsivity. One influential notion has been that the neuromodulator serotonin controls γ, and hence the degree of patience (Miyazaki, Miyazaki, & Doya, 2012; Schweighofer et al., 2008; Tanaka et al., 2004; Tanaka et al., 2007) or willingness to wait.

However, in explicit tests, humans and other animals often show forms of hyperbolic discounting, weighing a utility $\upsilon_{\boldsymbol{x}_p}(a(\tau), \boldsymbol{x}_s(\tau))$ that will arrive $\tau - t$ time steps in the future by a function of the form $\frac{1}{k+\tau-t}$ rather than by the exponential form $\gamma^{\tau-t}$. This leads to a qualitatively different sort of impulsivity whereby choices become temporally inconsistent (Ainslie, 2001). Consider a subject at time t contemplating a choice between two options that will happen at time $2t$. The subject's preference between these (later) options at t could be different from their preference at time $2t$. This arises because hyperbolic discounting is steep in the short run and flat in the long run (compared with exponential discounting). Thus, the subject at time t might engage in expensive, apparently suboptimal, commitment behavior to prevent their future self (at time $2t$) from defecting against the choice that they currently prefer (Crockett et al., 2013).

Impulsivity is a relatively stable measure that is strongly associated with addiction. It reliably differentiates addicts from nonaddicts (Kirby, Petry, & Bickel, 1999; Petry, Bickel, & Arnett, 1998), is sensitive to acute intoxication (Tomie, Aguado, Pohorecky, & Benjamin, 1998), and characterizes addiction-relevant variation in learning (Lovic, Saunders, Yager, & Robinson, 2011) and in the function of the dopaminergic system (Buckholtz et al., 2010). However, it is also very sensitive to experimental details (Evenden, 1999; Fassbender et al., 2014) and has multiple theoretical underpinnings depending on the exact situation. We return to this in the discussion.

One might think that there should be a fact of the matter as to which utility function (and form of temporal discounting) is best, perhaps by virtue of a grounding in homeostatic considerations (Keramati & Gutkin, 2011; Savage, 2003). However, not only are claims of this character almost impossible to prove, but one should also observe that individuals and species are involved in a game-theoretic contest (J. M. Smith, 1993). The Nash equilibria in these games might well involve a mixed strategy, with different patterns of behavior such as impulsivity, which are mandated by different utility functions, stably surviving at modest population frequencies (Suomi, 2006; J. Williams & Taylor, 2006).

Incorrect Inference

The next route to deviant behavior is to think that utilities, likelihoods, priors, and states are correct, but that there are faults in inferences that are licensed from these ingredients.

A first route to incorrect inference, which has been linked to delusions (Hemsley, 1987, 1993, 2005), is if external, sensory information associated with the likelihood is

incorrectly weighted against internal, contextual, information associated with the prior. According to BDT, the optimal weighting is associated with the relative (un)certainties or (im)precisions of these quantities, so errors in calculating or using these could have this effect (Adams et al., 2013; Fletcher & Frith, 2009).

We will describe two further classes of culprits. The first lies in how the computations that lead to the assessment of the state $\boldsymbol{x}_s$ are executed. The second concerns the calculation of the policy. Here, the issue is that even if the problem, described by $\boldsymbol{x}_p$, is known, it may be intractable to infer the optimal policy by complete model-based reasoning. We consider two classes of heuristic, each of which can lead to aberrant choice.

Inference about states

Realistic cases of BDT frequently result in very substantial demands on computational and memory resources. In Equation 1, the numerator requires a multiplication for every potential state $\boldsymbol{x}_s$. As these are unobserved hidden states existing in the mind of the observer alone, there is no strict limit as to their number. Similarly, the denominator contains an integral or sum over these products: $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}) = \int d\boldsymbol{x}_s \mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s)\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}|\boldsymbol{x}_s)$. Naively evaluating Equation 1 thus not only requires many computations, but also potentially a very large memory to store the intermediate results.

The substantial computational costs of model-based reasoning make it difficult to consider more than very few hidden states. Incomplete consideration of latent causes could lead to potential internal states being overlooked, and thus facilitate inconsistencies. For instance, while occupying one motivational state, it is hard (at least for animals other than scrub-jays; Raby, Alexis, Dickinson, & Clayton, 2007) to generate expectations appropriate to another, predicted, motivational state (Loewenstein & Prelec, 1992). In addiction, for instance, sober patients may express the desire to limit consumption, and yet be unable to prevent escalating consumption once mildly intoxicated. The utility $\upsilon_{\boldsymbol{x}_p}$ of drugs in detoxified and mildly intoxicated states likely differs, and incomplete consideration of the different states might make appropriate choices difficult. Thus, choices made on the basis of the sober utility would fail to reflect the intoxicated utility (thus encouraging recidivism); choices made on the basis of the intoxicated utility would fail to reflect the actual disutility apparent in sobriety. One strategy around this is precommitment (Ainslie, 2001; Crockett et al., 2013), although this has its own attendant costs. Note also that a focus on few explanatory causes is reminiscent of perseverative thought processes such as rumination in depression or worry in anxiety.

The case that there are many observations over time $\boldsymbol{y}, \boldsymbol{y}'$, … would seem to magnify the problem. However, in Equation 2, the incorporation of a novel observation $\boldsymbol{y}'$ appears only to demand the combination of the likelihood $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{y}'|\boldsymbol{x}_s)$ with the previously computed posterior over latent state given the data up to that point $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s|\boldsymbol{y})$. It is important that this means that the actual data $\boldsymbol{y}$ can be discarded, and only its representative (sufficient) statistics determining the posterior need to be kept in memory, resulting in a drastic reduction of memory requirements. Unfortunately, this distribution can itself be very complicated, parameterizing correlations, and heuristics such as assumed density filtering or moment matching to a simpler distribution (Daw, Courville, & Dayan, 2008; Kruschke, 2006). These are not optimal and can lead to outcomes such as a disproportionate influence of early experience on inferences about the state $\boldsymbol{x}_s$.

Inference about actions

We observed earlier that given full knowledge of the transitions and utilities, optimal choices can be inferred by sums over trajectories of future actions in Equation 4. However, this is typically radically intractable. The essential difficulty is well illustrated by a game such as chess, say against an opponent whose policy is fixed. We have to choose the best sequence of actions in the light of the moves the opponent will make. Looking one move ahead means around 30 options need to be considered. The opponent has around 30 response options to each of these, and we have again the same number of options for each prior sequence available on the next move. Thus, if we were to look d steps ahead, we would have to choose between 30^d sequences to work out what to do. There are alternatives that take advantage of the recursive structure of the problem, but they still require biologically unfeasible computations on the global state space (Puterman, 2005; Sutton & Barto, 1998).

There is thus a need for computationally feasible alternatives, approximations, and heuristics. These, in possibly abnormal combinations, engender a wide variety of potential suboptimalities. We first discuss model-free solutions to the computational load faced by model-based methods, and then consider (Pavlovian) policy heuristics.

Model-free control represents one canonical approach to the computational complexities of choice. For this, experience replaces cognition: Rather than relying on a description of the consequences of actions $\mathcal{T}$, model-free solutions sample these consequences by (a) trying out an action, (b) taking note of the consequence, and (c) updating a cached value, which amounts to keeping a particular sort of running average. Habitual choices are computationally very straightforward. However, the use

of environmental samples is inefficient (Kakade, 2003), meaning that substantial experience is required in an unchanging word before the policy is appropriate. If aspects of the state or problem change then habitual behavior will be poor, and indeed inconsistent with information the agent can be shown, by other means, already to have acquired or possess.

Under normal circumstances, there is often a progressive habitization of control (Dickinson, 1985; Dickinson & Balleine, 2002), putatively because of this balance between computational and sample complexity (Daw et al., 2005), or perhaps governed by the decreasing value of the information that the computationally expensive operations of the model-based system could provide as the model-free system becomes more accurate (Keramati & Gutkin, 2011). That is, given little experience, choice is model-based, or goal-directed, but as the sample complexity bounds are satisfied, choice becomes model-free.

Alterations in the trade-off between model-free and model-based decisions is one route to psychopathology that is being actively examined. It could arise either by influencing the mechanisms determining the arbitration (Lee, Shimojo, & O'Doherty, 2014) or via influences on either of the two subsystems. In substance addiction, for instance, an increased prominence of habits could arise both via direct influences of addictive substances on phasic dopaminergic signals (Dayan, 2009; Dickinson & Balleine, 2002; Huys, Pizzagalli, Bogdan, & Dayan, 2013; Redish, 2004) and via an impairment of prefrontal goal-directed control (Chen et al., 2013; Otto, Gershman, Markman, & Daw, 2013; Redish, Jensen, & Johnson, 2008; Takahashi et al., 2011; Volkow, Fowler, Wang, Baler, & Telang, 2009). A change in this balance could also lead to reduced cognitive flexibility in a variety of psychopathologies from schizophrenia to Parkinson's disease and, eating disorders (Maia & Frank, 2011; Waltz, Frank, Robinson, & Gold, 2007). Model-based evaluation is particularly susceptible to corruption in the computations involved; adjusting appropriately to this has been postulated as leading to symptoms in schizophrenia (Moutoussis et al., 2011) and depression (Lieder, Goodman, & Huys, 2013a). Even if there is just competition between different mechanisms of choice, then hesitation, or psychomotor slowing could result, potentially leading to an overall lack of choice.

Among other factors influencing the trade-off between model-based and model-free control is the complexity of the state space—because this determines how much sampling appears to be necessary for model-free values to become correct. Thus, an incorrectly impoverished state space, stemming either from incorrect prior distributions (as discussed in the previous section) or perhaps from superficial recall from memory (J. M. G. Williams et al., 2007), would result in overgeneralization and then early dominance by what would be incorrect model-free values. Limitations in working memory have also been suggested as damaging learning, potentially in a model-based system (Collins & Frank, 2012); and this could also make for an early impetus toward habits.

Similarly, an increased tendency to generalize would effectively reduce the state space by reducing the number of differentiable states, biasing the competition between model-free and model-based responding toward the former by reducing the sample complexity. Hence, the nature of the state or action space, and the prior over these, have important consequences for the acquisition of cached values, and potentially also for the trade-off between model-based and model-free systems.

Although it is less flexible than model-based control, the sort of model-free control we have so far considered is still mutable with experience. It thus fails to capture the even more complete insensitivity to outcomes characterizing many perseverative psychopathological patterns, particularly in addiction (Vanderschuren & Everitt, 2004). It has been noted that there are more extreme forms of habit-like policies (such as direct actors, as in the actor critic rule; Barto, Sutton, & Anderson, 1983). For these, the propensity to choose an action is further divorced from any actual long-run value (which is what drives indirect actors, such as Q-learning; Watkins, 1989). The only formal requirement for propensities is that they be largest for the best available action, and so the differences can grow arbitrarily large. Direct actors of this sort can become highly resistant to change.

One potential substrate for the different sorts of model-free policy stems from the observation of a helical connection scheme between the striatum and dopamine neurons, running from ventromedial to dorsolateral regions and from the ventral tegmental area through the substantia nigra (Haber, Fudge, & McFarland, 2000; Joel & Weiner, 2000). The suggestion is that the most extreme dorsolateral region of the striatum is most actor-like, and least flexible (Belin, Jonkman, Dickinson, Robbins, & Everitt, 2009; Haruno & Kawato, 2006; Keramati & Gutkin, 2013), and indeed that the process of habitization is accompanied by a migration along this axis of the control of behavior. Any alteration of this spatialized consolidation process could lead to faster, or slower, reductions in flexibility with experience.

Another possible route to habitual behavior stems from the notion that subjects, having calculated an appropriate course of action using expensive, model-based evaluation, might just *store* the result in memory and then recall it whenever in the same state $\boldsymbol{x}_s$. This strategy, called memoization in the computer science literature, has been elaborated in various sophisticated probabilistic guises (Huys et al., in press; O'Donnell, Goodman, & Tenenbaum, 2009; Wingate, Diuk, O'Donnell, Tenenbaum,

& Gershman, 2013). Memoized actions, just like model-free habits, do not themselves change with changes in (motivational) state $\boldsymbol{x}_s$—one would need a sophisticated form of forgetting to work out exactly which actions had been invalidated and should be removed. Limitations on the capacity of goal-directed evaluations may force subjects to rely more strongly on previous solutions, which would then be incorrect in the face of changed motivational factors.

Although we have painted a rather stark division between model-based and model-free control, there is increasing work on various interactions between them. For instance, there is evidence that model-based systems might dream or preplay fictitious experience (Johnson & Redish, 2007; Pfeiffer & Foster, 2013), and that this can train model-free values (Gershman et al., 2014; Simon & Daw, 2011). This follows a venerable computational suggestion (Sutton, 1991) for improving the performance of model-free control. However, it implies a vulnerability—if the values and choices produced by the model-based system are incorrect for whatever reason, then the model-free system, which might be perfectly capable of learning a powerful set of normative behaviors given only actual experience in the world, would nevertheless be incorrectly biased by virtue of this erroneous preplay.

Worse still, because the model-free controller can undergo its slow adaptation only in the light of the experience it receives, the same problem arises if the model-based controller determines poor choices in the world, rather than just replaying fictitious poor choices. That is, even a normal, model-free controller risks ascribing a lack of reward to the environment rather than a fallacious model-based controller, and so again fail to alleviate the problem (Huys, 2007).

A converse to the model-based controller exerting its will over model-free control is the standard heuristic of replacing parts of the goal-directed decision tree with values derived from model-free experience (Campbell, Hoane, & Hsu, 2002)—that is, substituting a whole branch of the tree that would require expensive model-based evaluation with the model-free estimate of its value. The computational advantage of this is clear. If model-free learning had proceeded to its asymptote, then these values would be just what the model-based system would compute by exploring the branch. However, if the model-free values are incorrect, then this can corrupt model-based evaluation too, leading to suboptimal choice. A related possibility is that the model-based tree is initialized in memory with model-free values, with model-based processes then improving these values by progressively sampling transitions. Thus, subjects who are less able to execute these processes will automatically rely more on model-free values.

Pavlovian policy heuristics

In realistic domains, it is hard to limit the range of possible actions. This leads to a very substantial inferential burden for model-based control, and high sample complexity for model-free control. One heuristic of very widespread importance is a direct, hard-wired mapping to actions from affectively important outcomes, and, crucially, predictions of those outcomes. These so-called Pavlovian responses can be seen as an example of evolutionary programming.

Pavlovian responses (often called conditioned and unconditioned responses) can be subdivided into two broad classes. Consummatory responses are emitted in close proximity to the outcome, or even in its presence, and their nature is tightly linked to its particular features (Bolles, Holtz, Dunn, & Hill, 1980; Timberlake & Grant, 1975). Preparatory responses are elicited by CSs that predict outcomes, even at some temporal or spatial distance. Just like instrumental evaluations, Pavlovian expectations of outcomes might, in principle, involve either model-based or model-free methods (Doll, Simon, & Daw, 2012; Guitart-Masip, Huys, et al., 2012; Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009), although there is some debate as to whether Pavlovian and instrumental model-based evaluations follow the same rules and involve the same neural structures (Dayan & Berridge, 2014; M. J. F. Robinson & Berridge, 2013).

Whereas in most cases of instrumental conditioning outcomes are contingent on the choice of action, in Pavlovian conditioning, they are not. Thus, the responses are automatically elicited, can take on many forms (Timberlake, Wahl, & King, 1982), and are not adaptive to what might be required to get or avoid the outcome concerned. Instead, Pavlovian responses are directly linked to the expectations evoked by the CSs and so can organize responses associated with biologically significant outcomes such as food, water, mates and threats that it would be tremendously inefficient or even dangerous to learn from scratch. This is obviously a critical advantage (Domjan, 2005); however, it means that Pavlovian responses can thus be emitted even when instrumentally disadvantageous, as in omission schedules (Anson, Bender, & Melvin, 1969; Breland & Breland, 1961; Hershberger, 1986; Morse, Mead, & Kelleher, 1967; D. R. Williams & Williams, 1969) where the emission of, say, an approach response results in the omission of the food reward. This also happens in humans (e.g., Guitart-Masip, Huys, et al., 2012).

If they do not need to be learned, then Pavlovian behaviors must be neurobiologically hard coded. Various brain areas have been implicated. For instance, stimulation of the periaqueductal gray or the nucleus accumbens leads to species-specific, complex, topographically

organized aversive or appetitive behaviors (Bandler & Shipley, 1994; Reynolds & Berridge, 2002) modulated by cortical and neuromodulatory inputs (Faure, Reynolds, Richard, & Berridge, 2008; Faure, Richard, & Berridge, 2010; Pecina & Berridge, 2005). Furthermore, as noted, the striatum is organized along two parallel pathways, with the direct, dopamine D_1 receptor expressing, pathway promoting active "go" and the indirect pathway promoting "nogo" and expressing D_2 receptors (Alexander & Crutcher, 1990; Frank, 2005; Kravitz et al., 2012). Alterations to receptors densities, such as the D_2 downregulation seen in addiction, could directly influence the strength and probability of Pavlovian behaviors (Huys, Beck, Dayan, & Heinz, in press).

In psychopathological terms, the fact that Pavlovian responses are not contingent on the outcomes they produce implies that they are a ready source of poor choices. Subjects for whom these responses are particularly strong will thus often be found to persist in performing behaviors that can be counterproductive. This suggests one should focus on the competition between instrumental and Pavlovian responses. It is hard to know how to balance computational efficiency against inflexibility; however, it is apparent that there is substantial individual variation (Meyer et al., 2012), which at least in one task in humans correlated with medial prefrontal activation (Cavanagh, Eisenberg, Guitart-Masip, Huys, & Frank, 2013). Overly strong Pavlovian influences covary positively with addictive traits (Carter & Tiffany, 1999; Everitt & Robbins, 2005; Flagel et al., 2011; Meyer et al., 2012) and are predictive of relapse in alcohol addiction (Gruesser et al., 2004). Indeed, a stronger reliance on phasic dopaminergic signals in Pavlovian approach behavior has recently been identified as a trait risk factor for addiction, at least in rodents (Flagel et al., 2011; Meyer et al., 2012).

We might view phasic stress responses in similar terms. The idea is that stress systems evolved to deal with temporally punctate events. However, with increasing life span and quality, temporally extended stressors prevail (Korte, Koolhaas, Wingfield, & McEwen, 2005). Thus, the acute reactions to stress, which can be seen as examples of hard-wired Pavlovian responses, may no longer be appropriate. An interesting example of this comes in the meta-control (i.e., control over control) inherent in the shift from goal-directed to habitual behavior occasioned by stress (Schwabe & Wolf, 2009). Although rapid responses may be useful for acute stressors, they may specifically prevent goal-directed responses aimed at alleviating the origin of a chronic stressor. A surprising correlate of depression is an enhancement of reactive aggression (Monahan et al., 2001), with potentially substantial negative consequences in the longer term by interfering with personal relationships (Kendler, Karkowski, & Prescott, 1999). That is, evolutionarily acquired priors on behavior may worsen and prolong, rather than alleviate, particular situations.

We mentioned earlier that recent work has suggested a rich intertwining of model-based and model-free instrumental control. The interaction between Pavlovian and instrumental control has also been the topic of some interest. For instance, it has been suggested that Pavlovian influences can help complex, tree-based, decision-making tasks by taking automatic decisions whether to continue the evaluation of a subtree or to terminate it by pruning (Huys et al., 2012). The latter removes some of the computational complexity. To the extent that such heuristics are relied on implies knock-on effects on other inference and valuation mechanisms if they break (Dayan & Huys, 2008).

The fact that (Pavlovian) effects occasioned by predictions might manipulate the mechanisms by which those predictions are actually made makes for a complex inferential loop with potentially disastrous consequences. It might, for instance, exacerbate the sampling issue mentioned in the inference about states. In rumination, affectively laden stimuli or situations might lead to perseverative cognitions (possibly by biasing internal state estimation) that might in turn further strengthen the affects associated with the stimuli. Certain psychopathological states do involve a strong focus of conscious thoughts on particular objects (Gelder, Harrison, & Cowen, 2006; Sims, 2003): The hijacking of conscious explicit thoughts by affectively laden stimuli, states, or events might arise from the nefariously strong influence of Pavlovian influences on model-based calculations (Dayan & Huys, 2008).

We discussed earlier the problem of selecting between model-based and model-free instrumental controllers. Pavlovian influences add extra complexity to this choice. Indeed, there can even be game-theoretic competition, given that these different systems can make divergent estimates, based on their own idiosyncratic sources of information and processing. Accounts of drug addiction that appeal to issues such as incentive sensitization (T. E. Robinson & Berridge, 1993) are an example. The idea is that a longer-term consequence of many drugs of addiction is sensitization, that is, a functional boost in the release or effect of dopamine particularly in the ventral striatum associated with their administration, and that this aspect of dopamine might loom larger in the Pavlovian controller than in either sort of instrumental control. This disagreement between the controllers as to the utilities (or perhaps the way that these utilities mediate choices) would lead to a conflict between the assessments made about the same pharmacological outcome. As for the case of temporal discounting, a consequence of this could be apparently maladaptive commitment behavior (Dayan & Berridge, 2014; McClure, Daw, & Montague, 2003).

Finally, as depression progresses in severity it is often associated with psychomotor retardation. If patients overestimate the possibility of negative outcomes, possibly because of incorrect priors, then such sloth could arise from the Pavlovian heuristic that turns the expectation of punishment into behavioral inhibition (Carver & White, 1994; Crockett, Clark, & Robbins, 2009; Dayan, Niv, Seymour, & Daw, 2006; Gray, 1982; Guitart-Masip, Huys, et al., 2012), perhaps by inhibition of the dopaminergic system (Guitart-Masip, Chowdhury, et al., 2012; Tye et al., 2013).

Incorrect Experience

We have so far considered the case that the component of the state that describes the problem, $\boldsymbol{x}_p$, is known. However, at a longer time-scale, at least some components of $\boldsymbol{x}_p$ must also change as subjects find out more about the environment they inhabit. This has implications both for the analysis and inference about state $\boldsymbol{x}_s$, and about the decision problems that result.

In particular, unusual experiences can inspire unfortunate expectations about the future. It is here that effects at the intersection of environments and genes, and the mechanisms these build, bite most strongly. That is, characteristics of the environment are known to have pervasive behavioral and cognitive sequelae. A version of Equation 2, but applied to $\boldsymbol{x}_p$ rather than $\boldsymbol{x}_s$, illustrates one conduit for these: Past experience encapsulated in prior beliefs can shape current behavior. Many of these cases have the flavor of evolutionary or Darwinian psychiatry, with the maladaptivity resulting from mismatches between current and historical environments.

One particular issue for singular experiences is understanding their scope of application—that is, subjects have to solve the extremely difficult inference problem of working out how likely such an event is to recur, and in how confined a context the experience applies. This is a characteristic example of the case mentioned earlier in which subjects lack data, and so more deep-seated priors about these facets (sometimes called the meta-inference problem) can exert significant influence. This can be expected to lead to substantial individual differences in the effects of such experiences.

A paradigmatic example of the shaping and influence of prior beliefs is learned helplessness and variations thereof. In helplessness experiments animals are exposed to stressors that they do not have behavioral control over (Maier & Watkins, 2005; Willner, 2005). Compared with animals with control, these animals show impairments in escaping subsequent aversive stimuli and reductions in seeking rewards. The negative effect of experiencing shocks is avoided by the detection of controllability in the medial prefrontal cortex and the inhibition of the serotonergic dorsal raphé (Amat et al., 2005; Rozeske, Der-Avakian, Watkins, & Maier, 2012; Warden et al., 2012). Notably, the development of a depressive and anxious phenotype occurs in *healthy* animals. Thus, healthy, normative inference with adverse experiences can lead to prior beliefs about controllability such that a vast array of behaviors are affected adversely (Huys & Dayan, 2009). Again, the key issue comes to be generalization—in what realms do these negative experiences apply?

A related example is seen in longer-term reactions to stress. Among other sequelae, the corticosteroids released by stress lead to long-term remodeling of networks in the amygdala, prefrontal cortex, hippocampus, and hypothalamus (Arnsten, 2009; McEwen, 1998; Mitra, Jadhav, McEwen, Vyas, & Chattarji, 2005), with consequent alterations in Pavlovian responses (Grillon, Smith, Haynos, & Nieman, 2004) and likely their guidance of goal-directed mechanisms (Schwabe, Tegenthoff, Höffken, & Wolf, 2012). Stress early on in life can have very long-lasting effects, for instance accounting for a substantial proportion of the dysregulation of the stress axis seen in depression (Barr et al., 2004; Heim et al., 2000). This remodeling can be seen as baking a set of early observations about an environment into the architecture of inference and control. This could, for instance, enshrine a particular set of heuristics that would obviate subsequent expensive and thus potentially dangerous inference about aversive characteristics of the environment. However, if subsequent environments do not actually contain the threats that are implied, then flexible control will have been permanently compromised.

It is important to note that the evidence about changes in the description of the problem, $\boldsymbol{x}_p$, typically comes from observations that pertain to the states, that is, $\boldsymbol{x}_s$. Thus constraints on inference about $\boldsymbol{x}_s$, for instance because of limited working memory, can have a deleterious impact. In particular, if information about the actual inputs $\boldsymbol{y}$ pertaining to $\boldsymbol{x}_s$ has been discarded in favor of reduced statistics that are only valid for a given $\boldsymbol{x}_p$, then if $\boldsymbol{x}_p$ has changed, new experience will be necessary. This may, for instance, speak to the sloth of many pharmacological interventions in psychiatry. Antidepressants show characteristically delayed effects, often taking up to six weeks. If the effect of antidepressants is partly a change in the structure of latent causes considered as explanations (for instance, from more specific hidden causes due to a reduction in overgeneralization), then experience will be necessary for these causes to be learned about. This suggests that the accumulation of statistics over a lifetime, and the reduced chance of collecting new experience, might be one rather normative argument for a decrease of cognitive flexibility over the life span.

Discussion

We have described key elements of BDT, and illustrated a wide range of ways in which maladaptive choice can arise. We considered three broad categories: cases in which the conception of the problem or the utility are incorrect or abnormal; issues with determining the correct action given the description of the problem; and cases in which unfortunate environments, potentially coupled with priors having particular characteristics, might lead to unfortunately maladaptive expectations that do not fit the current environment. Priors of various forms are crystallized into the architecture of state spaces and inferences; and a key role is also played by heuristics that make choice possible in the face of otherwise devastating computational demands.

We made many simplifications to link the various threads. Particularly egregious was the separation of two discrete components of state: $\boldsymbol{x}_p$, which describes the structure of the problem, the prior, likelihood and utility; and $\boldsymbol{x}_s$, which describes the current circumstance within this problem that the agent occupies. This is too simple because, given imperfect knowledge, these bleed into each other. Furthermore, in the section on learning $\boldsymbol{x}_p$, it was also apparent that we need priors over the nature and evolution of this (which can lead to problems in conditions such as posttraumatic stress disorder)—which, in our Bayesian formulation, can be considered as hyperpriors, described by splitting $\boldsymbol{x}_p$ into more stable (the hyperpriors) and less stable (the problem) components. A direction for the future would be to specify a richer hierarchy of components to the full state $\boldsymbol{x}$ and consider how information flows through observation and learning; and how choice is normatively determined.

Although we organized our discussion around the three categories mentioned, there are actually many potential interactions between them (only some of which we described); there are also cases in which there are different possibilities for where problems arise, which we cannot yet resolve. For instance, as we noted, because it is the product of the prior and the likelihood that determines the posterior, it is not always straightforward to disentangle them. Thus, in autism, there is a well-documented fractionation of experience, with a particular focus on details. Alternative accounts suggest that this arises from a weak "top-down" prior, or an overly strong "bottom-up" likelihood (Brock, 2012; Happe & Frith, 2006; Mottron, Dawson, Soulieres, Hubert, & Burack, 2006; Pellicano & Burr, 2012). Thus, aberrant beliefs might arise from opposite alterations of either; and one can envisage compensatory modifications to priors that exactly make up for any malfunction of the likelihood. Abnormal prior beliefs may be primary or compensatory for abnormal likelihoods. Experimentally, prior and likelihood can be disentangled by varying the information $\boldsymbol{y}$, for instance in learning tasks where the effect of the prior should vanish with increasing amounts of evidence (Stankevicius, Huys, Kalra, & Series, 2014).

Equally, we noted that delusional interpretations can arise as a result of overly weak influence of the likelihood (perhaps arising from a specific form of glutamate/GABA imbalance; Lewis, 2013; Lisman et al., 2008). From an empirical Bayesian viewpoint, priors can be seen as arising from accumulated likelihoods. Thus one can understand the observation that whereas newly debuting psychotic patients often have unsystematic delusions (i.e., the patient does not know where the "signals" come from) and can accept that they are an abnormal experience, as the disorder goes untreated and turns chronic, patients often develop an explanation that puts together all the elements of the delusion (i.e., the patients "knows" who is sending the signals and why). This might arise as an incorrect prior is learned through iterative substitutions of more and more elaborated delusional posteriors. Such a learning mechanism may be related to the notion that a key factor determining the prognosis of schizophrenia is the duration of untreated illness (Dell'Osso, Glick, Baldwin, & Altamura, 2013) and the fact that higher doses of antipsychotics are required to control symptoms in chronic patients (Kahn et al., 2008; Lieberman et al., 2005).

As another example, because of its pervasive effects on behavior, it may be hard to assign any measured changes strictly to alterations in the utility function itself—downstream mechanisms may well themselves be the subject of pathological changes (though see Koerding & Wolpert, 2004b). Take observations of steep discounting, that is, temporal impulsivity. This could arise directly as a consequence of a particular utility function. However, discount factors can arise normatively as being determined by the reachability and stability of environments (Kurth-Nelson, Bickel, & Redish, 2012); thus impulsivity could arise from early observations of instability (Kidd, Palmeri, & Aslin, 2013; J. Williams & Dayan, 2005). Impulsivity could also arise as a result of an inferential inability to build a deep decision tree of future states and actions, leaving only proximal outcomes as reliably expected. Finally, impulsivity could also arise from overly strong Pavlovian approach tendencies to a proximal (small) reward (Carter & Tiffany, 1999; Dayan et al., 2006; Flagel et al., 2011). These different sources of impulsivity can be distinguished behaviorally—one of the benefits of the BDT analysis is that it makes crystal clear the requirement to do so. The answer matters because it would be important to separate out the causes for steep utility functions in addiction (Kirby et al., 1999; Petry et al., 1998).

Similarly, an anhedonic inability to enjoy previously pleasant events may be because the primary reward with which they are associated is no longer reinforcing, but it may also be because they are no longer effectively associated with the primary reward. One recent report attempted to assess this directly and suggested that

anhedonia in depression was associated more with the primary reinforcement than with learning (Huys et al., 2013) although there certainly also is evidence for alterations in downstream processes, including learning (Hasler et al., 2004; Huys et al., 2013; Treadway & Zald, 2011). Similarly, patients with addiction show enhanced ventral striatal responses to drug-associated cues—which might just be related to exposure or learning, but also a reduced response to other rewards (Wrase et al., 2007).

Furthermore, as we have briefly illustrated in a few cases, the various mechanisms can feed off each other in a deleterious manner. Take the example of the effect of an unfortunate prior. It is often the case that such priors are overwhelmed by experience so that learners can become appropriately calibrated to their environments. However, if a subject interprets an extreme traumatic event as something that is in danger of recurring, they might avoid interacting with the world in any way that would allow them to discover that it actually will not. This issue can be seen as afflicting the trade-off between exploration and exploitation, and is analogous to the slow extinction of active avoidance (Moutoussis, Bentall, Williams, & Dayan, 2008). Agents that can influence or determine their own experience, will often normatively fail to become well calibrated, and so behave in a maladaptive manner.

Some forms of depression may also involve miscalibration, if incorrectly negative priors about the environment (Beck et al., 1979; Strunk et al., 2006) prevent the very exploration that would show that those priors are not true. As already discussed earlier, the delayed effect of SSRIs may be related to the need to relearn these wrong priors. It is interesting that SSRIs produce positive biases in the processing of emotional information already at early stages of treatment (Harmer & Cowen, 2013). The increased efficiency of an SNRI may be related to the known ability of noradrenaline to increase attention (Chamberlain & Robbins, 2013), which in turn may facilitate the relearning process.

In sum, as is apparent from the many unresolved issues in this review, it is the very early days for using the tools of BDT as a route to dividing up the various sources of problems. The most important task is to define behavioral paradigms that discriminate between the various failures (Maia & Frank, 2011; Montague et al., 2012), whence it will become possible to think about any possible means of ameliorating the conditions that are revealed.

Appendix

A Formalization of BDT

We use the notation and descriptions from our initial description of BDT. If optimal, the agent's policy should be to choose the action that maximizes its expected utility, averaged over its uncertainty about what the state actually is:

$$\begin{aligned} a^*_{\boldsymbol{x}_p}(\mathcal{D}) &= \mathrm{argmax}_{a\in\mathcal{A}}\left\{\mathcal{E}[\upsilon_{\boldsymbol{x}_p}(a,\boldsymbol{x})]_{\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s|\mathcal{D})}\right\} \\ &= \mathrm{argmax}_{a\in\mathcal{A}}\left\{\textstyle\sum_{\boldsymbol{x}_s}\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s\mid\mathcal{D})\upsilon_{\boldsymbol{x}_p}(a,\boldsymbol{x}_s)\right\} \end{aligned} \tag{3}$$

We mainly described BDT in the case that the agent uses all its information from the past $\mathcal{D}$ to make a single decision $a^*_{\boldsymbol{x}_p}(\mathcal{D})$. In many circumstances, the agent occupies a part of the environment that has dynamics, in that the state can change over time t according to a stochastic action-dependent transition $\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t+1)\mid\boldsymbol{x}_s(t);a_t) = \mathrm{T}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t+1),\boldsymbol{x}_s(t);a(t))$. There is a natural extension of the posterior Equation 2 to this case, which typically leads to a decay over time in the influence of past information. There is also a natural extension to the definition from Equation 3 of the quantity that needs to be optimized

$$\boldsymbol{a}^*_{\boldsymbol{x}_p}(\mathcal{D}(\mathrm{t})) = \mathrm{argmax}_{\boldsymbol{a}\in\mathcal{A}^\infty}\left\{\mathcal{E}\left[\sum\nolimits_{\tau=t}^{\infty}\gamma^{\tau-t}\upsilon(a(\tau),\boldsymbol{x}_s(\tau))\right]_{\mathcal{P}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t)|\mathcal{D}(t))}\right\} \tag{4}$$

where $0 \le \gamma < 1$ is a discount factor that controls the relative impact of proximal and distal rewards in the expectation, $\boldsymbol{a}$ is a whole trajectory of actions (living in the space $\mathcal{A}^\infty$ of such trajectories), and the distribution of $\boldsymbol{x}_\tau$ evolves with the actions chosen, the data received and the transitions $\mathcal{T}$.

Standard treatments of Equation 4 are based on dynamic programming (Bellman, 1957). For instance, if the states $\boldsymbol{x}_s$ are perfectly known or identifiable from the observations $\boldsymbol{y}$, and the policy $\pi_{\boldsymbol{x}_p}(a;\boldsymbol{x}_s) = \mathcal{P}_{\boldsymbol{x}_p}(a\mid\boldsymbol{x}_s)$ is state dependent, with no explicit dependence on time, then the expectation on the right-hand side of Equation 4 over future utilities $\upsilon_{\boldsymbol{x}_p}$ can then be written recursively:

$$\begin{aligned} \mathcal{V}^{\pi_{\boldsymbol{x}_p}}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t)) = \sum_a \pi_{\boldsymbol{x}_p}(a;\boldsymbol{x}_s)\big[&\upsilon_{\boldsymbol{x}_p}(a(t),\boldsymbol{x}_s(t)) + \\ \sum_{\boldsymbol{x}_s(t+1)} &\mathcal{T}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t+1),\boldsymbol{x}_s(t);a(t))\gamma\mathcal{V}^{\pi_{\boldsymbol{x}_p}}_{\boldsymbol{x}_p}(\boldsymbol{x}_s(t+1))\big] \end{aligned} \tag{5}$$

There is also a generalization of this to the case that the states $\boldsymbol{x}$ are not perfectly known (Kaelbling, Littman, & Cassandra, 1998).

Author Contributions

All the authors participated in formulating and writing the article.

Acknowledgments

We are grateful to Karl Friston, Read Montague, and Klaas Enno Stephan for discussions and debate and to Tiago Maia and the reviewers for thoughtful comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by the German Research Foundation (Q.J.M.H.: Deutsche Forschungsgemeinschaft, DFG, FOR 1617: Grant RA1047/2-1), the Gatsby Charitable Foundation (P.D.) and the Wellcome Trust (R.J.D.: Senior Investigator Award 098362/Z/12/Z; the Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust 091593/Z/10/Z).

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, *4*, 47.

Ainslie, G. (2001). *Breakdown of will*. Cambridge, England: Cambridge University Press.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*, 266–271.

Amat, J., Baratta, M. V., Paul, E., Bland, S. T., Watkins, L. R., & Maier, S. F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience*, *8*, 365–371.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anson, J. E., Bender, L., & Melvin, K. B. (1969). Sources of reinforcement in the establishment of self-punitive behavior. *Journal of Comparative and Physiological Psychology*, *67*, 376–380.

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*, 410–422.

Balleine, B. W. (2005). Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits. *Physiology and Behavior*, *86*, 717–730.

Bandler, R., & Shipley, M. T. (1994). Columnar organization in the midbrain periaqueductal gray: Modules for emotional expression? *Trends in Neurosciences*, *17*, 379–389.

Barr, C. S., Newman, T. K., Schwandt, M., Shannon, C., Dvoskin, R. L., Lindell, S. G., . . . Higley, J. D. (2004). Sexual dichotomy of an interaction between early adversity and the serotonin transporter gene promoter variant in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 12358–12363.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics*, *13*, 834–846.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression* (1st ed.). New York, NY: Guilford.

Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Düzel, E., Dolan, R., & Dayan, P. (2013). Dopamine modulates reward-related vigor. *Neuropsychopharmacology*, *38*, 1495–1503.

Belin, D., Jonkman, S., Dickinson, A., Robbins, T. W., & Everitt, B. J. (2009). Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. *Behavioural Brain Research*, *199*, 89–102.

Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Bentall, R. P. (2004). Abandoning the concept of schizophrenia. In J. Read, L. R. Mosher, & R. P. Bentall (Eds.), *Models of madness* (pp. 195–208). New York, NY: Routledge.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Berlin, Germany: Springer Verlag.

Bolles, R. C., Holtz, R., Dunn, T., & Hill, W. (1980). Comparisons of stimulus learning and response learning in a punishment situation. *Learning and Motivation*, *11*, 78–96.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*, 976–986.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681–684.

Brenner, C. A., Krishnan, G. P., Vohs, J. L., Ahn, W.-Y., Hetrick, W. P., Morzorati, S. L., & O'Donnell, B. F. (2009). Steady state responses: Electrophysiological assessment of sensory function in schizophrenia. *Schizophrenia Bulletin*, *35*, 1065–1077.

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Trends in Cognitive Sciences*, *16*, 573–574; author reply 574–575.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., . . . Zald, D. H. (2010). Dopaminergic network differences in human impulsivity. *Science*, *329*, 532.

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, *28*, 676–691.

Campbell, M., Hoane, A. J., & Hsu, F.-H. (2002). Deep blue. *Artificial Intelligence*, *134*, 57–83.

Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, *21*, 77–88.

Carter, B. L., & Tiffany, S. T. (1999). Meta-analysis of cue-reactivity in addiction research. *Addiction*, *94*, 327–340.

Carver, C. S., & Ganellen, R. J. (1983). Depression and components of self-punitiveness: High standards, self-criticism, and overgeneralization. *Journal of Abnormal Psychology*, *92*, 330–337.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending

reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319–333.

Cavanagh, J., Eisenberg, I., Guitart-Masip, M., Huys, Q. J. M., & Frank, M. J. (2013). Frontal theta overrides Pavlovian learning biases. *Journal of Neuroscience*, *33*, 8541–8548.

Chamberlain, S. R., & Robbins, T. W. (2013). Noradrenergic modulation of cognition: Therapeutic implications. *Journal of Psychopharmacology*, *27*, 694–718.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Chen, B. T., Yau, H.-J., Hatch, C., Kusumoto-Yoshida, I., Cho, S. L., Hopf, F. W., & Bonci, A. (2013). Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. *Nature*, *496*, 359–362.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*. Cambridge, MA: MIT Press.

Clark, D. M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., Grey, N., . . . Wild, J. (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *74*, 568–578.

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*, 1024–1035.

Coltheart, M. (2007). Cognitive neuropsychiatry and delusional belief. *Quarterly Journal of Experimental Psychology*, *60*, 1041–1062.

Courville, A. C., Daw, N., Gordon, G. J., & Touretzky, D. S. (2004). Model uncertainty in classical conditioning. In S. Thrun, L. Saul, & B. Schoelkopf (Eds.), *Advances in neural information processing systems* 16 (pp. 977–984). Cambridge, MA: MIT Press.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2005). Similarity and discrimination in classical conditioning: A latent variable account. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* 17 (pp. 313–320). Cambridge, MA: MIT Press.

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews: Neuroscience*, *9*, 587–600.

Crockett, M. J., Braams, B. R., Clark, L., Tobler, P. N., Robbins, T. W., & Kalenscher, T. (2013). Restricting temptations: Neural mechanisms of precommitment. *Neuron*, *79*, 391–401.

Crockett, M. J., Clark, L., & Robbins, T. W. (2009). Reconciling the role of serotonin in behavioral inhibition and aversion: Acute tryptophan depletion abolishes punishment-induced inhibition in humans. *Journal of Neuroscience*, *29*, 11993–11999.

Crombag, H. S., Bossert, J. M., Koya, E., & Shaham, Y. (2008). Context-induced relapse to drug seeking: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3233–3243.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, *319*, 1264–1267.

Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind* (pp. 431–452). Oxford, England: Oxford University Press.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*, 1204–1215.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1704–1711.

Dayan, P. (2009). Dopamine, reinforcement learning, and addiction. *Pharmacopsychiatry*, *42*(Suppl. 1), S56–S65.

Dayan, P. (2012). How to set the switches on this thing. *Current Opinion in Neurobiology*, *22*, 1068–1074.

Dayan, P., & Berridge, K. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision and revelation. *Cognitive, Affective and Behavioral Neuroscience*, *14*, 473–492.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective and Behavioral Neuroscience*, *8*, 429–453.

Dayan, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. *PLoS Computational Biology*, *4*, e4.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Networks*, *19*, 1153–1160.

Dell'Osso, B., Glick, I. D., Baldwin, D. S., & Altamura, A. C. (2013). Can long-term outcomes be improved by shortening the duration of untreated illness in psychiatric disorders? A conceptual framework. *Psychopathology*, *46*, 14–21.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge, England: Cambridge University Press.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*, 67–78.

Dickinson, A., & Balleine, B. W. (2002). The role of learning in motivation. In C. R. Gallistel (Ed.), *Learning, motivation & emotion: Steven's handbook of experimental psychology* (pp. 497–533). New York, NY: John Wiley.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the human brain. *Neuron*, *80*, 312–325.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*, 1075–1081.

Domjan, M. (2005). Pavlovian conditioning: A functional perspective. *Annual Review of Psychology*, *56*, 179–206.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, *12*, 961–974.

Droit-Volet, S., & Meck, W. H. (2007). How emotions colour our perception of time. *Trends in Cognitive Sciences*, *11*, 504–513.

Ellis, H. D. (1998). Cognitive neuropsychiatry and delusional misidentification syndromes: An exemplary vindication of the new discipline. *Cognitive Neuropsychiatry*, *3*, 81–89.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*, 348–361.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*, 1481–1489.

Fassbender, C., Houde, S., Silver-Balbus, S., Ballard, K., Kim, B. K., Rutledge, K. J., . . . McClure, S. M. (2014). The decimal effect: Behavioral and neural bases for a novel influence on intertemporal choice in healthy individuals and in ADHD. *Journal of Cognitive Neuroscience*, *26*, 2455–2468.

Faure, A., Reynolds, S. M., Richard, J. M., & Berridge, K. C. (2008). Mesolimbic dopamine in desire and dread: Enabling motivation to be generated by localized glutamate disruptions in nucleus accumbens. *Journal of Neuroscience*, *28*, 7184–7192.

Faure, A., Richard, J. M., & Berridge, K. C. (2010). Desire and dread from the nucleus accumbens: Cortical glutamate and subcortical GABA differentially generate motivation and hedonic impact in the rat. *PLoS ONE*, *5*, e11223.

Fiser, J., Berkes, P., Orban, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, *14*, 119–130.

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., . . . Akil, H. (2011). A selective role for dopamine in stimulus-reward learning. *Nature*, *469*, 53–57.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*, 48–58.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72.

Garety, P. A., Freeman, D., Jolley, S., Dunn, G., Bebbington, P. E., Fowler, D. G., . . . Dudley, R. (2005). Reasoning, emotions, and delusional conviction in psychosis. *Journal of Abnormal Psychology*, *114*, 373–384.

Gelder, M., Harrison, P., & Cowen, P. (2006). *Shorter Oxford textbook of psychiatry*. Oxford, England: Oxford University Press.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*, 197–209.

Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, *143*, 182–194.

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning & Behavior*, *40*, 255–268.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*, 279–325.

Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Oxford, England: Clarendon.

Grillon, C., Smith, K., Haynos, A., & Nieman, L. K. (2004). Deficits in hippocampus-mediated Pavlovian conditioning in endogenous hypercortisolism. *Biological Psychiatry*, *56*, 837–843.

Gruesser, S. M., Wrase, J., Klein, S., Hermann, D., Smolka, M. N., Ruf, M., . . . Heinz, A. (2004). Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. *Psychopharmacology*, *175*, 296–302.

Guitart-Masip, M., Chowdhury, R., Sharot, T., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Action controls dopaminergic enhancement of reward representations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 7511–7516.

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, *62*, 154–166.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience*, *20*, 2369–2382.

Halligan, P. W., & David, A. S. (2001). Cognitive neuropsychiatry: Towards a scientific psychopathology. *Nature Reviews: Neuroscience*, *2*, 209–215.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience*, *26*, 8360–8367.

Happe, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*, 5–25.

Harmer, C. J., & Cowen, P. J. (2013). "It's the way that you look at it"—A cognitive neuropsychological account of SSRI action in depression. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *368*, 20120407.

Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks*, *19*, 1242–1254.

Hasler, G., Drevets, W. C., Manji, H. K., & Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology*, *29*, 1765–1781.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1726–1731.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *364*, 1263–1271.

Haug, T. T., Blomhoff, S., Hellstrom, K., Holme, I., Humble, M., Madsbu, H. P., & Wold, J. E. (2003). Exposure therapy and sertraline in social phobia: 1-year follow-up of a randomised controlled trial. *British Journal of Psychiatry*, *182*, 312–318.

Hay, N., & Russell, S. J. (2011). *Metareasoning for Monte Carlo tree search* (Tech. Rep. UCB/EECS-2011-119). Berkeley: University of California, Berkeley, EECS Department.

Heim, C., Newport, D. J., Heit, S., Graham, Y. P., Wilcox, M., Bonsall, R., . . . Nemeroff, C. B. (2000). Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *Journal of the American Medical Association*, *284*, 592–597.

Hemsley, D. (1987). An experimental psychological model for schizophrenia. In H. Hafner, W. F. Gattaz, & W. Janzarik

(Eds.), *Search for the causes of schizophrenia* (pp. 179–188). New York, NY: Springer.

Hemsley, D. R. (1993). A simple (or simplistic?) cognitive model for schizophrenia. *Behaviour Research and Therapy*, *31*, 633–645.

Hemsley, D. R. (2005). The development of a cognitive model of schizophrenia: Placing it in context. *Neuroscience and Biobehavioral Reviews*, *29*, 977–988.

Hershberger, W. A. (1986). An approach through the looking-glass. *Animal Learning and Behavior*, *14*, 443–451.

Hinton, G. E., & Sejnowski, T. J. (1999). *Unsupervised learning: Foundations of neural computation*. Cambridge, MA: MIT Press.

Hofmann, S. G., Wu, J. Q., & Boettcher, H. (2013). D-Cycloserine as an augmentation strategy for cognitive behavioral therapy of anxiety disorders. *Biology of Mood and Anxiety Disorders*, *3*, 11.

Huys, Q. J. M. (2007). *Reinforcers and control. Towards a computational aetiology of depression*. London, England: UCL, University of London, Gatsby Computational Neuroscience Unit. Retrieved from http://www.gatsby.ucl.ac.uk/~qhuys/pub.html

Huys, Q. J. M., Beck, A., Dayan, P., & Heinz, A. (in press). Neurobiological structure and computational understanding of addictive behaviour. *Phenomenological Neuropsychiatry: Bridging the Clinic and Clinical Neuroscience*.

Huys, Q. J. M., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*, 314–328.

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology*, *8*, e1002410.

Huys, Q. J. M, Lally, N., Paul Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (in press). The interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences, USA*.

Huys, Q. J. M., Moutoussis, M., & Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Networks*, *24*, 544–551.

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood and Anxiety Disorders*, *3*, 12.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science*, *19*, 1015–1022.

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: An analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, *96*, 451–474.

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, *27*, 12176–12189.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*, 99–134.

Kahn, R. S., Fleischhacker, W. W., Boter, H., Davidson, M., Vergouwe, Y., Keet, I. P. M., . . . Grobbee, D. E. (2008). Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: An open randomised clinical trial. *Lancet*, *371*, 1085–1097.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292.

Kakade, S. M. (2003). *On the sample complexity of reinforcement learning*. London, England: University of London.

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry*, *156*, 837–841.

Keramati, M., & Gutkin, B. S. (2011). A reinforcement learning theory for homeostatic regulation. In J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. C. N. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* 24 (pp. 82–90). Cambridge, MA: MIT Press.

Keramati, M., & Gutkin, B. (2013). Imbalanced decision hierarchy in addicts emerging from drug-hijacked dopamine spiraling circuit. *PLoS ONE*, *8*, e61489.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, *126*, 109–114.

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex*, *13*, 400–408.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, *321*, 806–810.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*, 78–87.

Koerding, K., & Wolpert, D. M. (2004a). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.

Koerding, K. P., & Wolpert, D. M. (2004b). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 9839–9842.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*, 665–682.

Korte, S. M., Koolhaas, J. M., Wingfield, J. C., & McEwen, B. S. (2005). The Darwinian concept of stress: Benefits of allostasis and costs of allostatic load and the trade-offs in health and disease. *Neuroscience and Biobehavioral Reviews*, *29*, 3–38.

Kravitz, A. V., Tye, L. D., & Kreitzer, A. C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature Neuroscience*, *15*, 816–818.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review*, *113*, 677–699.

Kurth-Nelson, Z., Bickel, W., & Redish, A. D. (2012). A theoretical account of cognitive effects in delay discounting. *European Journal of Neuroscience*, *35*, 1052–1064.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, *81*, 687–699.

Lewis, D. A. (2013). Inhibitory neurons in human cortical circuits: Substrate for cognitive dysfunction in schizophrenia. *Current Opinion in Neurobiology, 26C*, 22–26.

Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., . . . Hsiao, J. K. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, *353*, 1209–1223.

Lieder, F., Goodman, N., & Huys, Q. J. M. (2013a, February). *Controllability and resource-rational planning*. Paper presented at COSYNE, Salt Lake City, UT.

Lieder, F., Goodman, N., & Huys, Q. J. M. (2013b, July). *Learned helplessness and generalization*. Paper presented at the Cognitive Science Conference, Berlin, Germany.

Lisman, J. E., Coyle, J. T., Green, R. W., Javitt, D. C., Benes, F. M., Heckers, S., & Grace, A. A. (2008). Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends in Neurosciences*, *31*, 234–242.

Lloyd, K., & Leslie, D. S. (2013). Context-dependent decision-making: A simple Bayesian model. *Journal of the Royal Society Interface*, *10*. Retrieved from http://rsif.royalsocietypublishing.org/content/10/82/20130069

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, *107*, 573–597.

Lovic, V., Saunders, B. T., Yager, L. M., & Robinson, T. E. (2011). Rats prone to attribute incentive salience to reward cues are also prone to impulsive action. *Behavioural Brain Research*, *223*, 255–261.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge, England: Cambridge University Press.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*, 154–162.

Maier, S. F., & Watkins, L. R. (2005). Stressor controllability and learned helplessness: The roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. *Neuroscience and Biobehavioral Reviews*, *29*, 829–841.

Marr, D. (1982). *Vision*. New York, NY: Freeman.

Mazzoni, P., Hristova, A., & Krakauer, J. W. (2007). Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *Journal of Neuroscience*, *27*, 7105–7116.

McClure, S. M., Daw, N. D., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, *26*, 423–428.

McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, *840*, 33–44.

McGuire, J. T., & Botvinick, M. M. (2010). Prefrontal cortex, cognitive control, and the registration of decision costs. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 7922–7926.

Meyer, P. J., Lovic, V., Saunders, B. T., Yager, L. M., Flagel, S. B., Morrow, J. D., & Robinson, T. E. (2012). Quantifying individual variation in the propensity to attribute incentive salience to reward cues. *PLoS ONE*, *7*, e38987.

Mitra, R., Jadhav, S., McEwen, B. S., Vyas, A., & Chattarji, S. (2005). Stress duration modulates the spatiotemporal patterns of spine formation in the basolateral amygdala. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 9371–9376.

Miyazaki, K., Miyazaki, K. W., & Doya, K. (2012). The role of serotonin in the regulation of patience and impulsivity. *Molecular Neurobiology*, *45*, 213–224.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., . . . Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. Oxford, England: Oxford University Press.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, *16*, 72–80.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience*, *9*, 1057–1063.

Morse, W., Mead, R. N., & Kelleher, R. (1967). Modulation of elicited behavior by a fixed-interval schedule of electric shock presentation. *Science*, *157*, 215–217.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*, 27–43.

Moutoussis, M., Bentall, R. P., El-Deredy, W., & Dayan, P. (2011). Bayesian modelling of jumping-to-conclusions bias in delusional patients. *Cognitive Neuropsychiatry*, *16*, 422–447.

Moutoussis, M., Bentall, R. P., Williams, J., & Dayan, P. (2008). A temporal difference account of avoidance learning. *Network*, *19*, 137–160.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*, 507–520.

O'Donnell, T. J., Goodman, N. D., & Tenenbaum, J. B. (2009). *Fragment grammars: Exploring computation and reuse in language* (Tech. rep.). Cambridge, MA: MIT, Computer Science and Artificial Intelligence Laboratory.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*, 283–328.

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, *24*, 751–761.

Pecina, S., & Berridge, K. C. (2005). Hedonic hot spot in nucleus accumbens shell: Where do mu-opioids cause increased hedonic impact of sweetness? *Journal of Neuroscience*, *25*, 11777–11786.

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*, 504–510.

Petry, N. M., Bickel, W. K., & Arnett, M. (1998). Shortened time horizons and insensitivity to future consequences in heroin addicts. *Addiction*, *93*, 729–738.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*, 74–79.

Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry*, *57*, 319–327.

Puterman, M. L. (2005). *Markov decision processes: Discrete stochastic dynamic programming*. Wiley Series in Probability and Statistics. New York, NY: John Wiley.

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, *445*, 919–921.

Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, *306*, 1944–1947.

Redish, A. D., Jensen, S., & Johnson, A. (2008). A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral and Brain Sciences*, *31*, 415–437; discussion 437–487.

Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review*, *114*, 784–805.

Reynolds, S. M., & Berridge, K. C. (2002). Positive and negative motivation in nucleus accumbens shell: Bivalent rostrocaudal gradients for GABA-elicited eating, taste "liking"/"disliking" reactions, place preference/avoidance, and fear. *Journal of Neuroscience*, *22*, 7308–7320.

Robinson, M. J. F., & Berridge, K. C. (2013). Instant transformation of learned repulsion into motivational "wanting." *Current Biology*, *23*, 282–289.

Robinson, O. J., Charney, D. R., Overstreet, C., Vytal, K., & Grillon, C. (2012). The adaptive threat bias in anxiety: Amygdala-dorsomedial prefrontal cortex coupling and aversive amplification. *NeuroImage*, *60*, 523–529.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews*, *18*, 247–291.

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, *10*, 1615–1624.

Roiser, J. P., Elliott, R., & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology*, *37*, 117–136.

Rozeske, R. R., Der-Avakian, A., Watkins, L. R., & Maier, S. F. (2012). Activation of the medial prefrontal cortex by escapable stress is necessary for protection against subsequent inescapable stress-induced potentiation of morphine conditioned place preference. *European Journal of Neuroscience*, *35*, 160–165.

Salamone, J. D., & Correa, M. (2002). Motivational views of reinforcement: Implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behavioural Brain Research*, *137*, 3–25.

Savage, T. (2003). The grounding of motivation in artificial animals: Indices of motivational behavior. *Cognitive Systems Research*, *4*, 23–55.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*, 603–613.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*, 677–694.

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., & Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews: Neuroscience*, *10*, 885–892.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2012). Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal-directed action in the human brain. *Journal of Neuroscience*, *32*, 10146–10155.

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, *29*, 7191–7198.

Schweighofer, N., Bertin, M., Shishida, K., Okamoto, Y., Tanaka, S. C., Yamawaki, S., & Doya, K. (2008). Low-serotonin levels increase delayed reward discounting in humans. *Journal of Neuroscience*, *28*, 4528–4532.

Simon, D. A., & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, *31*, 5526–5539.

Sims, A. C. P. (2003). *Symptoms of the mind: And introduction to descriptive psychopathology* (3rd ed.). Philadelphia, PA: Saunders.

Smith, J. M. (1993). *Evolution and the theory of games*. New York, NY: Springer.

Smith, K. S., Virkud, A., Deisseroth, K., & Graybiel, A. M. (2012). Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 18932–18937.

Stankevicius, A., Huys, Q. J. M., Kalra, A., & Series, P. (2014). Optimism as a prior belief about the probability of future reward. *PLoS Computational Biology*.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*, 966–973.

Stephan, K. E., Baldeweg, T., & Friston, K. J. (2006). Synaptic plasticity and dysconnection in schizophrenia. *Biological Psychiatry*, *59*, 929–939.

Stephan, K. E., Friston, K. J., & Frith, C. D. (2009). Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin*, *35*, 509–527.

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy*, *44*, 861–882.

Suomi, S. J. (2006). Risk, resilience, and gene–environment interactions in rhesus monkeys. *Annals of the New York Academy of Sciences*, *1094*, 52–62.

Sutton, R. (1991). Dyna, an integrated architecture for learning, planning and reacting. *Sigart Bulletin*, *2*, 160–163.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tai, L.-H., Lee, A. M., Benavidez, N., Bonci, A., & Wilbrecht, L. (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nature Neuroscience*, *15*, 1281–1289.

Takahashi, Y. K., Roesch, M. R., Wilson, R. C., Toreson, K., O'Donnell, P., Niv, Y., & Schoenbaum, G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience*, *14*, 1590–1597.

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, *7*, 887–893.

Tanaka, S. C., Schweighofer, N., Asahi, S., Shishida, K., Okamoto, Y., Yamawaki, S., & Doya, K. (2007). Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. *PLoS ONE*, *2*, e1333.

Timberlake, W., & Grant, D. L. (1975). Auto-shaping in rats to the presentation of another rat predicting food. *Science*, *190*, 690–692.

Timberlake, W., Wahl, G., & King, D. (1982). Stimulus and response contingencies in the misbehavior of rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*, 62–85.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Tomie, A., Aguado, A. S., Pohorecky, L. A., & Benjamin, D. (1998). Ethanol induces impulsive-like responding in a delay-of-reward operant choice procedure: Impulsivity predicts autoshaping. *Psychopharmacology*, *139*, 376–382.

Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: Lessons from translational neuroscience. *Neuroscience and Biobehavioral Reviews*, *35*, 537–555.

Tye, K. M., Mirzabekov, J. J., Warden, M. R., Ferenczi, E. A., Tsai, H.-C., Finkelstein, J., . . . Deisseroth, K. (2013). Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature*, *493*, 537–541.

Vanderschuren, L. J. M. J., & Everitt, B. J. (2004). Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science*, *305*, 1017–1019.

Volkow, N. D., Fowler, J. S., Wang, G. J., Baler, R., & Telang, F. (2009). Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology*, *56*(Suppl. 1), 3–8.

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological Psychiatry*, *62*, 756–764.

Warden, M. R., Selimbeyoglu, A., Mirzabekov, J. J., Lo, M., Thompson, K. R., Kim, S.-Y., . . . Deisseroth, K. (2012). A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. *Nature*, *492*, 428–432.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. Cambridge, England: King's College, Cambridge University.

Weiss, Y., & Adelson, E. H. (1998). *Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision*. Cambridge, MA: MIT, Artificial Intelligence Laboratory.

Williams, D. R., & Williams, H. (1969). Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior*, *12*, 511–520.

Williams, J., & Dayan, P. (2005). Dopamine, learning, and impulsivity: A biological account of attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology*, *15*, 160–179; discussion 157–169.

Williams, J., & Taylor, E. (2006). The evolution of hyperactivity, impulsivity and cognitive diversity. *Journal of the Royal Society Interface*, *3*, 399–413.

Williams, J. M. G. (1992). *The psychological treatment of depression*. New York, NY: Routledge.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*, 122–148.

Willner, P. (2005). Chronic mild stress (CMS) revisited: Consistency and behavioural-neurobiological concordance in the effects of CMS. *Neuropsychobiology*, *52*, 90–110.

Wingate, D., Diuk, C., O'Donnell, T., Tenenbaum, J., & Gershman, S. (2013). *Compositional policy priors*. Cambridge, MA: MIT Press.

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*, 1317–1329.

World Health Organization. (1990). *International classification of diseases*. Geneva, Switzerland: Author.

Wrase, J., Schlagenhauf, F., Kienast, T., Wüstenberg, T., Bermpohl, F., Kahnt, T., . . . Heinz, A. (2007). Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *NeuroImage*, *35*, 787–794.