Perception, Action and Utility: The Tangled Skein

Samuel J. Gershman[1] and Nathaniel D. Daw[2]

[1] Neuroscience Institute and Department of Psychology, Princeton University

[2] Center for Neural Science and Department of Psychology, New York University

June 5, 2011

1 Introduction

Normative theories of learning and decision-making are motivated by a computational-level analysis of the task facing an animal: *what* should the animal do to maximize future reward? However, much of the recent excitement in this field originates in *how* the animal arrives at its decisions and reward predictions—-algorithmic questions about which the computational-level analysis is silent. Answers to these algorithmic questions are essentially claims about *dynamics* (learning/adjustment across trials or the transformation of decision variables within trials). These depend crucially on the computational quantities and how they are manipulated algorithmically, the focus of this chapter.

At the computational level, statistical decision theory — and its relatives utility theory and reinforcement learning — seems to offer a clear framework for the integration of perception and action. Actions are chosen to optimize expected future utility. These expectations are with respect to latent states of the task, and the role of perception is to analyze stimuli so as to form probabilistic beliefs about these states. This framework supports a sequentially staged view of the problem — perception guiding evaluation — which has more or less implicitly licensed two influential streams of work in neuroscience, studying either stage in isolation.

The question of this chapter is what becomes of this seemingly clean computational separation at the algorithmic and implementational levels. The computations mandated by decision theory are, in general, intractable, and many different approximations replace them in practice, with underappreciated consequences for the interdependency of perception and action. Work in approximate inference, for instance, considers realizable (e.g., variational or Monte Carlo) approximations to the expected value computation, and involves trading off the relative cost of approximations in the action and sensory stages, or even entirely collapsing the two in various ways. These approximations induce a variety of dynamic patterns (in learning, inference and decision stages) that are not anticipated by a purely computational-level analysis, and we present evidence that many behavioral and neural phenomena are naturally explained in terms of the dynamics of approximate algorithms. In the following we first review the classical view, before attempting to unravel the tangled skein of perception, action, and utility that arises in practice.

2 The classical view

We begin by formalizing mathematically the problem of decision under uncertainty.[1] Consider a utility function $U(s, a)$ (or more pessimistically, its negation, the loss function) that depends on the agent's action a and a hidden state of the task s, which may be continuous and high-dimensional. For an agent seeking to maximize expected utility (or minimize loss), the optimal action is given by

$$a^* = \underset{a}{\operatorname{argmax}} \, \mathbb{E}_\pi[U(s, a)]$$

$$= \underset{a}{\operatorname{argmax}} \int_s \pi(s|x)U(s, a)ds, \tag{1}$$

where $\pi(s|x)$ is the agent's posterior belief about the hidden state conditional on observing sense data x, given by Bayes' rule:

$$\pi(s|x) = \frac{p(x|s)p(s)}{\int_s p(x|s)p(s)ds}. \tag{2}$$

In other words, the agent posits a generative model of the hidden and observed task variables in the form of a joint distribution $p(s, x)$, and then uses Bayes' rule to "invert" the generative model and infer the hidden state. The agent then considers what utility it can expect to achieve for an action, which is an average of utilities for taking the action in each state, weighted by the state's posterior probability. The generic role commonly imputed to the perceptual system is the computation of this posterior belief (Lee and Mumford, 2003; Knill and Pouget, 2004; Friston, 2005).

The assumptions of statistical decision theory are, in various forms, pervasive throughout psychology, neuroscience, economics and ecology (not to mention statistics and engineering, where they were first developed; see, for example, Berger, 1985). They are the basis of signal detection theory and drift-diffusion models in psychophysics (Green and Swets, 1966; Bogacz et al., 2006); of optimal control theory in sensorimotor control (Kording and Wolpert, 2006; Trommershäuser et al., 2008); of Bellman's equation in reinforcement learning (Sutton and Barto, 1998; Dayan and Daw, 2008); of subjective expected utility theory in economics (Von Neumann and Morgenstern, 1947; Savage, 1954); and of foraging theory in behavioral ecology (McNamara and Houston, 1980; Stephens and Krebs, 1986). More recently, neuroscientists have begun to probe the brain for signatures of these assumptions, in particular the neural computations of utilities and posterior probabilities (Glimcher, 2003).

We focus on two aspects of decision theory that have important implications for its implementation in the brain:

1. Decision theory implies a strong form of separation between probabilities and utilities. In particular, the posterior must be computed *before* (and hence independently of) the expected utility. This assumption is sometimes known as *probabilistic sophistication* (Machina and Schmeidler, 1992; Bossaerts et al., 2008). It means that I can state how much enjoyment I

---

[1]Our treatment is more precisely characterized as Bayesian decision theory, which is the most widely used special case of statistical decision theory. For discussion of alternatives to Bayesian decision theory, see Berger (1985). Some of these alternatives are considered in the context of vision by Maloney and Zhang (2010), and in the context of two-alternative forced choice tasks by Zacksenhouse et al. (2010).

would derive from having a picnic in sunny weather, independently of my belief that it will be sunny tomorrow.

2. The integral over s in Equations 1 and 2 is generally intractable. Although most investigations of decision theory in psychology and economics make Gaussian or multinomial assumptions about $\pi(s|x)$, these assumptions are not generally applicable to real-world decision-making tasks, where distributions may not take any convenient parametric form. Even if they do, the integrals may still resist closed-form solutions. This means that if the brain is to perform the necessary calculations, it must employ some form of approximation.

While statistical decision theory has been criticized on many other grounds (see, for example, Kahneman and Tversky, 1979; Camerer, 1998), we focus on these aspects because they highlight the algorithmic and implementational commitments of the theory. Statistical decision theory, to be plausibly implemented in the brain, requires segregated representations of probability and utility, and a mechanism for performing approximate inference. We discuss each of these requirements in turn. Then, in Section 5, we suggest how these requirements may in fact be intimately connected.

3   Segregation of probability and utility in the brain?

In this section, we discuss evidence in favor of the classical view that perception (state inference) and utility are computed and represented independently. We then describe several challenges to this view based on behavioral and neural data.

3.1   State inference

The basic sequential staging of state inference to action, via utility, seems in accord with the gross organization of at least the posterior half of cortex. Visual information arrives in occipital cortex and progresses anteriorly through the dorsal visual stream on the way toward neurons with direct control of effectors in primary motor cortex, just in front of the central sulcus. This progression is thought to instantiate a hierarchically staged inference process: for example, in the influential architecture proposed by Riesenhuber and Poggio (1999), object representations are built up through a hierarchy of receptive fields, with orientation- and position-tuned simple cells (in V1) at the bottom of the hierarchy view-tuned cells (in inferotemporal cortex) at the top of the hierarchy. Contemporary probabilistic models (e.g., Lee and Mumford, 2003; Friston, 2005) recapitulate this framework in the language of statistical inference: Visual cortex is inverting an internal "image model" to recover the latent objects that generated the sensory inputs. These models have in common an essentially inferential characterization of visual cortex.

Theories of this sort, if they consider utility at all, view it as entering at a fairly late stage of the processing stream; and experiments on vision throughout the hierarchy tend to consider perceptual variables in isolation of affective ones. All this is licensed, or indeed mandated, by the segregation of probability and utility in statistical decision theory (Bossaerts et al., 2008).

If the role of visual perception is ultimately to estimate a distribution over states, does this estimate have a discrete neural correlate? Perhaps the most detailed and compelling answer comes from

studies of motion perception. Here the state being inferred is the direction of dot motion given noisy sensory information. Newsome et al. (1989) recorded from direction-selective neurons in visual area MT while monkeys performed what is now known as the "random dots" task: Monkeys view a display of randomly and coherently moving dots intermixed, and must report the direction of the coherent motion to earn a juice reward. Neurons in MT appear to report instantaneous motion energy; aggregating their spikes over time allows estimating the state: i.e., the motion direction. Indeed, varying the coherence of motion energy, the experimenters showed a remarkable correspondence between physiological discrimination curves from aggregated spikes and psychophysical discrimination curves, from monkeys' behavioral reports. A follow-up study by Salzman et al. (1990) established the causal nature of this relationship by showing that microstimulation of these neurons systematically biased the perceived direction of motion.

The brain appears to perform an analogous integration of instantaneous motion energy into an aggregate state estimate one synapse downstream from MT, in the lateral intraparietal area (LIP). There, Shadlen and Newsome (2001) observed direction-selective neurons that appeared to integrate motion evidence over time in a manner that predicted the timing of behavioral responses. These and other data were interpreted in terms of classical signal detection concepts (Gold and Shadlen, 2002), with LIP neurons reporting a log-likelihood ratio for motion direction (the "weight of evidence") based on sensory evidence provided by inputs from MT.[2] It is worth noting here that while these signal detection are computational-level quantities, it is principally their *algorithmic* dynamics (e.g., log-likelihood accumulation) that are the objects of study.

So far so good. On this view, the visual system (areas V1 and MT) furnishes LIP with sensory evidence, which in turn computes a quantity (the weight of evidence) directly related to the posterior distribution over motion direction. Note, however, that the utility function in the dots task is essentially trivial (e.g., one for a correct report of motion direction, and zero for an incorrect one). As mentioned, statistical decision theory licenses the search for state estimation while ignoring utility; on the other hand, because of the simple form of the utility function, the overall expected utility is just proportional to the likelihood in the random dots task, and so these functions cannot be distinguished.

3.2   Utility

What about utility? A mostly separate body of work studies the mapping from states and actions to their utilities. Although there are a number of neural players involved in valuation, the most detailed work concerns the putative involvement of the midbrain dopamine system and its targets in learning the utilities that follow different states and actions. In this work, in a sort of mirror image of the simplistic utility function used in perceptual investigations such as the random dots task, the state is typically taken as unitary and wholly observed, posing no difficulties of ambiguous or uncertain perception.

Even given perceptual certainty, research on valuation generally assumes — both theoretically and in its experimental designs — that there is additional, irreducible stochasticity in the consequences subsequent to a state: for instance, a particular (wholly identifiable) picture might be followed by

---

[2]The log-likelihood ratio is equivalent to the log-posterior odds ratio under a uniform prior.

juice on half of trials, chosen randomly. In Equation 1 this stochasticity might be characterized by folding it into the state (e.g., defining two states for the picture, differentially rewarded albeit perceptually indistinguishable). However, in the interest of separating perceptual uncertainty from stochasticity in state and outcome dynamics, outcome stochasticity can equivalently be characterized by introducing a nested level of averaging in Equation 1, over the outcomes according to their probability given the state.[3] Much work concerns incremental learning rules for taking a running average over experienced outcomes' utilities, so as to learn an estimate of their expectation; this is the purview of reinforcement learning (see Niv, 2009, for a recent review). In particular, prominent accounts of the responses of dopamine neurons suggest that they carry a "reward prediction error" signal for updating such a running average (Schultz et al., 1997). The targets of these neurons, notably in the striatum and prefrontal cortex, are believed to be involved in valuation and action selection (Montague et al., 2006).

Insofar as this strategy already suggests an alternative to the staged representation of probabilities and utilities — since it directly learns utilities in expectation over outcome stochasticity, and so forgoes representing their decomposition into separate outcome probabilities and utilities — these theories suggest an alternative to the simple, two-stage mapping of decision theory onto the brain. However, again distinguishing uncertainty about the perceptual state from stochasticity in its consequences, these mechanisms do seem to fit the bill as the complement to the perceptual state inference systems discussed above: they map a given (assumed known) perceptual state and action to its utility (the latter in expectation over outcome stochasticity).

Indeed, numerous authors (Daw et al., 2006; Dayan and Daw, 2008; Braun et al., 2010; Gershman and Niv, 2010; Rao, 2010) have argued that the full problem of decision making under perceptual uncertainty—when both the perceptual state and the utilities are unknown and non-trivial—can essentially be treated by the composition of these two mechanisms: a perceptual state inference mechanism along the lines of Gold and Shadlen (2002), and a state utility learning mechanism of the sort described by Schultz et al. (1997). These ideas recapitulate the two-stage probability-utility architecture envisioned by statistical decision theory.

3.3   Challenges

The full story, however, is not so simple. First, abundant evidence indicates that reward modulation occurs at all levels of the visual hierarchy, including V1 (Shuler and Bear, 2006; Serences, 2008) and even before that in the lateral geniculate nucleus (Komura et al., 2001; O'Connor et al., 2002). For example, Shuler and Bear (2006) trained rats to associate monocular stimulation with liquid reward, and found that V1 neurons altered their firing patterns to predict the timing of reward. Second, the visual system appears to sacrifice probabilistic fidelity increased sensitivity to behaviorally relevant (i.e., highly rewarding) stimuli. Machens et al. (2005) measured the sound ensembles that are preferentially encoded by grasshopper auditory receptor neurons, and found that the distribution of optimal stimulus ensembles diverged from the distribution of natural sounds. Specifically, the ensembles were concentrated in a region of stimulus space occupied by mating signals. This finding

---

[3]More generally, states may occur in sequence and the aggregate value will depend on a series of nested expectations over each successive state, as in the Bellman equation defining utility in a Markov decision process (Sutton and Barto, 1998).

also has important implications for approximate inference schemes, which we revisit in Section 5.2.

More specific to the particular ideas and areas discussed here, if visual responses in the lateral geniculate nucleus are already reward-modulated, the idea of far-downstream LIP as a pure representation of posterior state probability is dubious. Indeed, other work varying rewarding outcomes for actions (Platt and Glimcher, 1999; Sugrue et al., 2004) shows that neurons in LIP are indeed modulated by the probability and amount of reward expected for an action — probably better thought of as related to expected utility rather than state probability per se. (Recall that the dots task essentially confounds these two quantities.) Moreover, area LIP is a poor candidate for a purely perceptual representation of perceived state, since it is a transitional area involved not just in visual perception but also action, specifically saccade control. (Monkeys in these experiments use saccades to signal their motion judgments.) But recall that area LIP is only one synapse downstream from the instantaneous motion energy representation in MT. If it already represents expected utility there seems to be no candidate for an intermediate stage of pure probability representation.

A different source of contrary evidence comes from behavioral economics. The classic Ellsberg paradox (Ellsberg, 1961) revealed prefences in human choice behavior that are *not* probabilistically sophisticated. The example given by Ellsberg involves drawing a ball from an urn containing 30 red balls and 60 black or yellow balls in an unknown proportion. Subjects are asked to choose between pairs of gambles (A vs. B or C vs. D) drawn from the following set:

	red	yellow	black
A	$100	$0	$0
B	$0	$100	$0
C	$100	$0	$100
D	$0	$100	$100

Experimentally, subjects prefer A over B and D over C. The intuitive reasoning is that in gambles A and D, the probability of winning $100 is known (unambiguous), whereas in B and C it is unknown (ambiguous). Hence, this pattern of preferences is sometimes known as *ambiguity aversion*, and has been repeatedly demonstrated in a variety of paradigms (see Loewenstein et al., 2008, for a recent review). Crucially, there is *no* subjective probability distribution that can produce this pattern of preferences. This is widely regarded as violating the assumption of probability-utility segregation in statistical decision theory.

In summary, despite the fact that researchers have worked separately on perception and utility, data from both physiology and behavior speak against a clean separation of systems. Addressing these discrepancies theoretically is a major task. The data appear to demand algorithms with richer dynamical interactions between perceptual and motivational systems. In Section 5, we discuss several possible approaches.

4   Approximate inference

In this section we discuss a key algorithmic challenge for decision theory: calculating expectations under the posterior. When the hidden state s is high-dimensional and continuous, analytical

tractability is elusive and some form of approximation must be used. This is a common problem in statistics and machine learning, where the approximations tend to fall into one of two classes:

- *Monte Carlo approximations* (Robert and Casella, 2004), where the posterior is approximated by a set of samples $s^{1:M}$. The law of large numbers guarantees that as M gets larger, the approximation converges to the true posterior. Generally, one cannot draw samples directly from the posterior, so two variations of the method are typically used. In *importance sampling*, samples are drawn from a proposal distribution $g(s)$ and then weighted according to:

$$w^m \propto \frac{p(x|s^m)p(s^m)}{g(s^m)}.$$ (3)

  The Monte Carlo approximation of expected utility is then approximated by:

$$\mathbb{E}_\pi[U(s,a)] \approx \sum_{m=1}^{M} w^m U(s^m, a).$$ (4)

  For certain generative models, an online version of importance sampling, known as *particle filtering*, is also possible, where the weights are updated after each observation. The second variation is known as *Markov chain Monte Carlo* and involves drawing the samples from a Markov chain whose stationary distribution is the posterior. Thus, sampling from the chain for long enough will eventually produce (unweighted) samples from the posterior.

- *Variational approximations* (Jordan et al., 1999), where the posterior is approximated by a more tractable distribution q chosen from a family of distributions $\mathcal{Q}$ so as to minimize the Kullback-Leibler (KL) divergence between q and π:

$$q^* = \underset{q \in \mathcal{Q}}{\operatorname{argmin}} \operatorname{KL}(q||\pi),$$ (5)

  where the KL-divergence is defined as

$$\operatorname{KL}(q||\pi) = \int_s q(s) \log \frac{q(s)}{\pi(s|x)} ds.$$ (6)

  The KL-divergence is minimized (to zero) when $q = \pi$, but generally π is not in $\mathcal{Q}$. The expected utility can then be approximated by $\mathbb{E}_q[U(s,a)]$, which by design should be tractable.

All of these variations have been recently explored in different contexts. Sanborn et al. (2010) suggested that human categorization phenomena are well-described by a version of particle filtering, and Daw and Courville (2008) have made a similar argument in the context of animal conditioning. Markov chain Monte Carlo approximations have been suggested as explanations for the dynamics of perceptual multistability (Gershman et al., 2009). Variational methods, while underappreciated, have begun to be considered as possible explanation for various associative learning phenomena (Daw et al., 2008; Sanborn and Silva, 2009).

None of these algorithmic models has yet been explored systematically in the context of decision theory. However, several experimental findings are suggestive. Most Bayesian models assume that

the posterior changes gradually as more information is acquired, and consequently the expected utility should change gradually as well. In a choice task, for example, the animal is tasked with estimating the expected utility associated with each choice, where the hidden state represents a scalar association parameter governing the relationship between choice and reward (Kruschke, 2008). Assuming humans and animals are "soft" maximizers, gradual changes in expected utility imply gradual changes in choice behavior. However, careful analysis reveals that, at the individual subject level, choice behavior appears to change abruptly over the course of learning (Gallistel et al., 2004). One explanation, proposed by Daw and Courville (2008), is that subjects use a Monte Carlo approximation like particle filtering to approximate the posterior, and abrupt changes arise from the stochastic nature of the sampling process when only a small number of samples are used.

Another suggestive experimental finding is known as *highlighting* (Medin and Bettger, 1991; Kruschke, 1996). In the balanced version of the experimental design, subjects are presented with 3 cues (A, B and C) and two outcomes (R and S). In the first phase, subjects observe $AB \to R$ 3 times as often as $AC \to S$. In the second phase, the same contingencies are preserved, but the proportions are reversed: now $AC \to S$ occurs 3 times as often as $AB \to R$. When tested with A alone, subjects predict R (a primacy effect), whereas when tested with the novel compound BC, subjects predict S (a recency effect). As shown by Kruschke (2006), this combination of primacy and recency effects is very difficult to explain by normative Bayesian models. Instead, Kruschke showed that an approximate learning algorithm, which he named "locally Bayesian learning," could explain these effects by propagating sufficient statistics (rather than full posterior distributions) between layers in a neural network. The basic insight was that the local propagation induces sequential dependencies even when the underlying generative model lacks such dependencies. Later work by Sanborn and Silva (2009) showed that locally Bayesian learning could be interpreted as a variational message passing algorithm, and suggested other elaborations in the same family of algorithms (see also Daw et al., 2008).

5   New vistas

We have so far treated approximate inference as an inherently probabilistic computation, consistent with the decision theoretic framework. However, what is mainly interesting about these inference schemes is that they suggest ways of weakening or abandoning the separation between probabilities and utilities. We discuss two possible versions of this revision. The first answer attempts to reframe decision theory purely in terms of inference—that is, to dispense with a privileged concept of utility and deal entirely with probabilities. In this setting, utilities influence probabilistic approximations in the same manner as other variables. The second answer attempts to stay within the decision theoretic framework, but ups the ante by treating the approximations themselves as decisions, with costs determined by the computational effort required by the approximations.

5.1   Decision-making as probabilistic inference

A rich vein of recent work in machine learning has explored the idea that decision problems can be reframed as inference problems (Dayan and Hinton, 1997; Toussaint and Storkey, 2006; Hoffman

et al., 2009; Vlassis and Toussaint, 2009; Theodorou et al., 2010). Although these approaches differ in their precise mathematical formulation, the common idea is that by transforming the utility function appropriately, one can treat it as a probability density function parameterized by the action and hidden state. Consequently, maximizing the "probability" of utility with respect to action, while marginalizing the hidden state, is formally equivalent to maximizing the expected utility. That is, sensory inference is optimization — of the posterior probability of the causes, given the data — and the optimization underlying utility maximization can be framed in parallel terms.

Although this is more or less an algebraic maneuver, it has profound implications for the organization of decision-making circuitry in the brain. In machine learning, the importance of this insight has been that the stable of approximations and tricks that have been developed to tame difficult problems of probabilistic inference can also be applied to action optimization. The neuroscientific version of this insight is that what appear to be dedicated motivational and valuation circuits may instead be regarded as parallel applications of the same underlying computational mechanisms over effectively different likelihood functions.

One version of this idea has been explored by Botvinick and An (2009). Building on earlier work in computer science (Cooper, 1988; Tatman and Shachter, 1990), Botvinick and An argued that the dorsolateral prefrontal cortex (DLPFC) could be thought of as computing action policies by iteratively updating probabilistic beliefs over action nodes in a kind of directed graphical model known as an "influence diagram." They proved that this algorithm converges to the optimal policy, and then went on to show that it could reproduce several behavioral signatures of goal-directed behavior commonly associated with the DLPFC (see also Daw et al., 2005).

The most detailed articulation of common mechanisms for inference and decision, however, has come from the work of Karl Friston and his colleagues (see Friston, 2010, for a recent review). Friston has argued that many aspects of neural computation can be subsumed under a single "free-energy principle." To understand this principle, let us return briefly to the variational approximation described in Section 4. A basic identity from probability theory defines the relationship between the KL-divergence and the marginal likelihood $p(x) = \int_s p(x|s)p(s)ds$:

$$\mathrm{KL}(q||\pi) = \log p(x) + \mathcal{F}(x, s), \tag{7}$$

where $\mathcal{F}(x, s)$ is known as the *free-energy* in statistical physics. In the context of perceptual inference in neuroscience, the marginal likelihood can be understood as a measure of how well the brain's internal model explains its sensory inputs, after integrating out the causes of the inputs. Since the KL-divergence is always non-negative, the negative free-energy is a lower bound on the log marginal likelihood. Thus, optimizing an internal model by minimizing free-energy (with respect to the model parameters) is equivalent to maximizing a lower bound on the log marginal likelihood.

Friston and colleagues (Friston et al., 2009, 2010) formulate the decision optimization problem in these terms. There are at least two separable claims here. The technical thrust of the work is similar to the ideas discussed above: if one specifies a desired equilibrium state distribution (here playing the role of the utility function), then this can be optimized by free energy minimization with respect to actions. However, the authors build on this foundation to assert a much more provocative concept: that for biologically evolved organisms, the desired equilibrium is by definition just the species' evolved equilibrium state distribution, and so minimizing free-energy with respect to actions is, in effect, equivalent to maximizing expected utility. What makes this claim provocative is that

it rejects decision theory's core distinction between a state's likelihood and its utility: nowhere in the definition of free-energy is utility mentioned. The mathematical equivalence rests on the evolutionary argument that hidden states with high prior probability also tend to have high utility. This situation arises through a combination of evolution and ontogenetic development, whereby the brain is immersed in a "statistical bath" that prescribes the landscape of its prior distribution. Because agents who find themselves more often in congenial states are more likely to survive, they inherit (or develop) priors with modes located at the states of highest congeniality. Conversely, states that are surprising given your evolutionary niche — like being out of water, for a fish — are maladaptive and should be avoided.

Although the free energy principle appears at the least to be a very useful formulation for exposing the computational parallelism between perceptual and decision problems, the more radical maneuver of treating them both as literally optimizing a single objective function is harder to swallow. A state's equilibrium likelihood and its utility are, on the classical view, not the same thing; rare events might be either unusually bad (being out of water, for a fish), good (being elected president, for an African American), or indeed neither. The idea that the two are united within a biological niche seems in one sense to appeal to evolutionary considerations to the end of substituting equilibrium for evolution, and risks precluding adaption or learning. Should the first amphibian out of water dive back in? If a wolf eats deer not because he is hungry, but because he is attracted to the equilibrium state of his ancestors, would a sudden bonanza of deer inspire him to eat only the amount to which he is accustomed? Should a person immersed in the "statistical bath" of poverty her entire life refuse a winning lottery ticket, since this would necessitate transitioning from a state of high equilibrium probability to a rare one? In all these cases, the possibility of upward mobility, within the individual, seems to rest on at least some role for traditional notions of utility or fitness in guiding their decisions, though the idea remains intriguing that in the ethological setting these have more in common with probability than a decision theorist might expect.

5.2   The costs of representation and computation

Probabilistic computations make exorbitant demands on a limited resource, and in a real physiological and psychological sense, these demands incur a cost that debits the utility of action. As shown in recent experiments by Kool et al. (2010), humans are "cognitive misers" who seek to avoid effortful thought at every opportunity, and this effort diminishes the same neural signals that are excited by reward (Botvinick et al., 2009).

One example of this issue arises in representing probability distributions over future states. We have mentioned already that predominant accounts of the dopamine system suggest that this system learns utilites in expectation over stochastic future states, rather than adopting the full decision-theoretically motivated representation of learning the probabilities of outcomes, and their utilites, separately, and computing utilities by integrating over them. (In reinforcement learning, the latter approach is known as "model-based.") These representations can actually be distinguished by clever experiments. For instance, one can study whether a rat who has learned to leverpress for food while hungry will continue to do so when full; a full probabilistic representation over outcomes will adjust its expected utilities to the changed outcome value, whereas representing utilities only in expectation can preclude this and so predicts hapless working for unwanted food. The upshot

of many such experiments is that the brain adopts both approaches, depending on circumstances. As suggested by Daw et al. (2005) and Keramati et al. (2011), which circumstances elicit which approach can be explained by a sort of meta-optimization over the costs (e.g. extra computation) of maintaining the full representation relative to its benefits (better statistical accuracy).

Thus, whether to represent a probability distribution over states, altogether, may itself be subject to (meta) decision theoretic analysis. Assuming the agent does represent probabilities, a finer question is how these should be approximated. Making the natural assumption that more accurate approximations of the posterior incur larger costs, it would make sense for the brain to seek an approximation that balanced the costs of computation against the utility of accuracy. However, this principle appears to make a strict segregation of probability and reward not only impossible but foolhardy.

Consider a simple Monte Carlo scheme for approximating expected utilities by drawing samples from the state distribution. Vul et al. (2009) analyzed the effects of imposing sampling costs on an such an agent. They showed, using a simple two-alternative forced-choice example and levels of sampling cost, that maximal total expected utility (gains minus costs) is achieved with remarkably few samples (on the order of 1-50), which was argued to capture phenomena such as probability matching observed in humans playing similar tasks. A key feature of this cost-benefit analysis of approximation is that it crosscuts probability and utility: the costs of adopting more or less faithful approximation to the probability must be weighed against the utility foregone by making worse decisions.

A more general analysis of this problem was recently undertaken by Gershman and Wilson (2010), exploring the idea that the brain might adopt approximate forms of the probability distribution over states, and treat its choice of which approximation to use as a "meta-decision." They incorporated a neurally-motivated approximation cost into the utility function, and then derived a variational lower bound on the log expected utility (closely related to the free energy bound on the log marginal likelihood). By maximizing this bound with respect to the choice of approximate distribution, the agent can near-optimally balance the costs and benefits of accurately approximating the state distribution. They showed that this model could account for the finding by Machens et al. (2005) that auditory representations seem to over-represent behaviorally relevant signals (like mating calls). The basic idea is that approximate distributions that represent the entire auditory spectrum with high fidelity are metabolically costly, and that accurately representing low-utility auditory signals is metabolically wasteful. Thus, approximations that concentrate their density in regions of high utility will achieve a higher utility lower bound. Finally, then, choosing appropriate approximations at the sensory level pushes utility considerations back to this level, perhaps explaining why the brain seems not to respect this basic decision theoretic distinction.

6   Conclusions

This chapter has attempted to demonstrate that while statistical decision theory provides a tantalizingly simple framework for decision-making, the neural reality is not so simple: perception, action and utility are ensnared in a tangled skein. To unravel this skein — or at least to motivate its tangle — we have pointed to new ideas that reconfigure the relationships between these variables,

and which suggest novel organizations for the underlying neural systems. These novel organizations involve a richer ensemble of dynamical interactions between perceptual and motivational systems than that which is anticipated by statistical decision theory.

The ideas discussed in this chapter borrow from recent as well as old computational concepts from artificial intelligence, engineering, statistics and physics. There still remain many ideas from these sources that have yet to percolate into the consciousness of neuroscientists. For example, an extremely rich research tradition in artificial intelligence has examined how to incorporate computational costs into decision-making systems (e.g., Horvitz, 1988; Russell and Wefald, 1991; Zilberstein, 1995). We hope that contact with these ideas will reinvigorate thinking about the organizational principles of the brain.

Acknowledgements

References

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700–765.

Bossaerts, P., Preuschoff, K., and Hsu, M. (2008). The neurobiological foundations of valuation in human decision making under uncertainty. In P.W. Glimcher, C.F. Camerer, E. F. and Poldrack, R., editors, *Neuroeconomics: Decision Making and the Brain.* Elsevier Press.

Botvinick, M. and An, J. (2009). Goal-directed decision making in prefrontal cortex: a computational framework. *Advances in Neural Information Processing Systems*, pages 169–176.

Botvinick, M., Huffstetler, S., and McGuire, J. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, 9(1):16.

Braun, D., Mehring, C., and Wolpert, D. (2010). Structure learning in action. *Behavioural Brain Research*, 206(2):157–165.

Camerer, C. (1998). Bounded rationality in individual decision making. *Experimental Economics*, 1(2):163–183.

Cooper, G. (1988). A method for using belief networks as influence diagrams. In *Workshop on Uncertainty in Artificial Intelligence*, pages 55–63.

Daw, N. and Courville, A. (2008). The pigeon as particle filter. *Advances in Neural Information Processing Systems*, 20:369–376.

Daw, N., Courville, A., and Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. In Chater, N. and Oaksford, M., editors, *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, pages 431–452.

Daw, N., Courville, A., and Touretzky, D. (2006). Representation and timing in theories of the dopamine system. *Neural Computation*, 18(7):1637–1677.

Daw, N., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711.

Dayan, P. and Daw, N. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4):429–453.

Dayan, P. and Hinton, G. (1997). Using expectation-maximization for reinforcement learning. *Neural Computation*, 9(2):271–278.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):815.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Friston, K., Daunizeau, J., and Kiebel, S. (2009). Reinforcement learning or active inference. *PloS One*, 4(7):e6421.

Friston, K., Daunizeau, J., Kilner, J., and Kiebel, S. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–260.

Gallistel, C., Fairhurst, S., and Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences*, 101(36):13124.

Gershman, S. and Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology*, 20(2):251–256.

Gershman, S., Vul, E., and Tenenbaum, J. (2009). Perceptual multistability as Markov chain Monte Carlo inference. *Advances in Neural Information Processing Systems*.

Gershman, S. and Wilson, R. (2010). The neural costs of optimal control. *Advances in Neural Information Processing Systems*.

Glimcher, P. (2003). *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. The MIT Press.

Gold, J. and Shadlen, M. (2002). Banburismus and the Brain:: Decoding the Relationship between Sensory Stimuli, Decisions, and Reward. *Neuron*, 36(2):299–308.

Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. Wiley New York.

Hoffman, M., de Freitas, N., Doucet, A., and Peters, J. (2009). An expectation maximization algorithm for continuous markov decision processes with arbitrary rewards. In *Twelfth International Conference on Artificial Intelligence and Statistics*. Citeseer.

Horvitz, E. (1988). Reasoning about beliefs and actions under computational resource constraints. *International Journal Approximate Reasoning*, 2(3):337–338.

Jordan, M., Ghahramani, Z., Jaakkola, T., and Saul, L. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291.

Keramati, M., Dezfouli, A., and Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology*, 7(5):e1002055.

Knill, D. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., Kaga, K., and Ono, T. (2001). Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature*, 412(6846):546–549.

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

Kording, K. and Wolpert, D. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10(7):319–326.

Kruschke, J. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 22(1):3–26.

Kruschke, J. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review*, 113(4):677–698.

Kruschke, J. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning & Behavior*, 36(3):210.

Lee, T. and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSA A*, 20(7):1434–1448.

Loewenstein, G., Rick, S., and Cohen, J. (2008). Neuroeconomics. *Annual Review of Psychology*, 59:647–672.

Machens, C., Gollisch, T., Kolesnikova, O., and Herz, A. (2005). Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron*, 47(3):447–456.

Machina, M. and Schmeidler, D. (1992). A more robust definition of subjective probability. *Econometrica*, 60(4):745–780.

Maloney, L. and Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research*.

McNamara, J. and Houston, A. (1980). The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, 85(4):673–690.

Medin, D. and Bettger, J. (1991). Sensitivity to changes in base-rate information. *The American Journal of Psychology*, 104(3):311–332.

Montague, P., King-Casas, B., and Cohen, J. (2006). Imaging valuation models in human choice. *Annual Review of Neuroscience*, 29:417–448.

Newsome, W., Britten, K., and Movshon, J. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341(6237):52–54.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154.

O'Connor, D., Fukui, M., Pinsk, M., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, 5(11):1203–1209.

Platt, M. and Glimcher, P. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–238.

Rao, R. (2010). Decision making under uncertainty: A neural model based on partially observable Markov decision processes. *Frontiers in Computational Neuroscience*, 4.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Verlag.

Russell, S. and Wefald, E. (1991). Principles of metareasoning. *Artificial Intelligence*, 49(1-3):361–395.

Salzman, C., Britten, K., and Newsome, W. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346(6280):174–177.

Sanborn, A., Griffiths, T., Navarro, D., To, S., and Sanborn, A. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117(4):1144–1167.

Sanborn, A. and Silva, R. (2009). Belief Propagation and Locally Bayesian Learning. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Savage, L. (1954). *The Foundations of Statistics*. John Wiley and Sons.

Schultz, W., Dayan, P., and Montague, P. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593.

Serences, J. (2008). Value-based modulations in human visual cortex. *Neuron*, 60(6):1169–1181.

Shadlen, M. and Newsome, W. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4):1916.

Shuler, M. and Bear, M. (2006). Reward timing in the primary visual cortex. *Science*, 311(5767):1606.

Stephens, D. and Krebs, J. (1986). *Foraging Theory*. Princeton University Press.

Sugrue, L., Corrado, G., and Newsome, W. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, 304(5678):1782.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Tatman, J. and Shachter, R. (1990). Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man and Cybernetics*, 20(2):365–379.

Theodorou, E., Buchli, J., and Schaal, S. (2010). Learning policy improvements with path integrals. In *International Conference on Artificial Intelligence and Statistics*.

Toussaint, M. and Storkey, A. (2006). Probabilistic inference for solving discrete and continuous state Markov Decision Processes. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 945–952. ACM.

Trommershäuser, J., Maloney, L., and Landy, M. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences*, 12(8):291–297.

Vlassis, N. and Toussaint, M. (2009). Model-free reinforcement learning as mixture learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1081–1088. ACM.

Von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press.

Vul, E., Goodman, N., Griffiths, T., and Tenenbaum, J. (2009). One and done? Optimal decisions from very few samples. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 66–72.

Zacksenhouse, M., Bogacz, R., and Holmes, P. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, 54(2):230–246.

Zilberstein, S. (1995). Operational rationality through compilation of anytime algorithms. *AI Magazine*, 16(2):79.