COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 38 (2014) 599–637
Copyright © 2014 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12101

One and Done? Optimal Decisions From Very Few Samples

Edward Vul,[a] Noah Goodman,[b] Thomas L. Griffiths,[c] Joshua B. Tenenbaum[d]

[a]*Department of Psychology, University of California, San Diego*
[b]*Department of Psychology, Stanford University*
[c]*Department of Psychology, University of California, Berkeley*
[d]*Brain and Cognitive Science, Massachusetts Institute of Technology*

Received 1 July 2011; received in revised form 29 March 2013; accepted 7 May 2013

Abstract

In many learning or inference tasks human behavior approximates that of a Bayesian ideal observer, suggesting that, at some level, cognition can be described as Bayesian inference. However, a number of findings have highlighted an intriguing mismatch between human behavior and standard assumptions about optimality: People often appear to make decisions based on just one or a few samples from the appropriate posterior probability distribution, rather than using the full distribution. Although sampling-based approximations are a common way to implement Bayesian inference, the very limited numbers of samples often used by humans seem insufficient to approximate the required probability distributions very accurately. Here, we consider this discrepancy in the broader framework of statistical decision theory, and ask: If people are making decisions based on samples—but as samples are costly—how many samples should people use to optimize their total expected or worst-case reward over a large number of decisions? We find that under reasonable assumptions about the time costs of sampling, making many quick but locally suboptimal decisions based on very few samples may be the globally optimal strategy over long periods. These results help to reconcile a large body of work showing sampling-based or probability matching behavior with the hypothesis that human cognition can be understood in Bayesian terms, and they suggest promising future directions for studies of resource-constrained cognition.

Keywords: Bayesian; Computational; Sampling; Inference; Bounded rationality

Correspondence should be sent to Edward Vul, Psychology, University of California, San Diego, 9500 Gilman Dr. 0109, La Jolla, CA 92093. E-mail: evul@ucsd.edu

1. Introduction

Across a wide range of learning, inference, and decision tasks, it has become increasingly common to analyze human behavior through the lens of optimal Bayesian models (in perception: Knill & Richards, 1996; motor control: Maloney, Trommershauser, & Landy, 2007; language: Chater & Manning, 2006; decision making: McKenzie, 1994; causal judgments: Griffiths & Tenenbaum, 2005; and concept learning: Goodman, Tenenbaum, Feldman, & Griffiths, 2008). However, despite the many observed parallels, the argument for understanding human cognition as a form of Bayesian inference remains far from complete. This study addresses two challenges. First, although human behavior often appears to be optimal when averaged over multiple trials and participants, it may not look that way within individual trials or participants. There will always be variance across trials and participants in any behavioral experiment, but the micro-level variation observed in many studies comparing human behavior with Bayesian models is not simply random noise around the model predictions. What kind of online processing is going on inside individual participants' minds that can appear so different at the local scale but approximate optimal behavior when averaged over many participants or many trials? Second, although ideal Bayesian computations are algorithmically straightforward in most small laboratory tasks, they are intractable for large-scale problems such as those that people face in the real world, or those that most Bayesian machine learning and artificial intelligence systems focus on. If human cognition is to be understood as a kind of Bayesian inference, we need an account of how the mind rapidly and effectively approximates these intractable calculations in the course of online processing.

Here, we argue that both of these challenges can be resolved by viewing cognitive processing in terms of stochastic sampling algorithms for approximate Bayesian inference, and analyzing the cost–benefit trade-off underlying the question of "How much to think?" Standard analyses of decision making as Bayesian inference assume that people should seek to maximize the expected utility (or minimize the expected cost) of their actions, relative to their posterior distribution over hypotheses. We show that in many settings, this ideal behavior can be approximated by an agent who considers only a small number of samples from the Bayesian posterior, and that the time cost to obtain more than a few samples outweighs the expected gain in decision accuracy they would provide. Hence, human cognition may approximate globally optimal behavior by making a sequence of noisy, locally suboptimal decisions—much as we see when we look closely at individual experimental participants and trials.

This first challenge—accounting for behavior within individual participants and trials—arises from the observation that while average behavior matches the ideal Bayesian agent, this is not true of individual trials and participants. Average judgments match the average of the Bayesian posterior distribution, but individual responses arise from the full range of the distribution with frequency proportional to the posterior probability. In the next section we will discuss two specific cases of this phenomenon in category learning (Goodman et al., 2008) and prediction (Griffiths & Tenenbaum, 2006), but such "sampling" behavior on individual trials appears to be ubiquitous. Sampling-based generalization has been found in word learning (Xu & Tenenbaum, 2007) and causal learning tasks (Sobel, Tenenbaum, & Gopnik, 2004; Xu & Tenenbaum, 2007), in both adults and children (Denison, Bonawitz, Gopnik, & Griffiths, 2013). Furthermore, studies in which individuals must produce more than one judgment on a given task have found that multiple guesses from one individual have independent errors, like independent samples from a probability distribution, in estimates of esoteric quantities in the world (Vul & Pashler, 2008), guesses about cued visual items (Vul, Hanus, & Kanwisher, 2009), and in illusory conjunctions in visual attention tasks (Vul & Rich, 2010). More broadly, models of category learning (Sanborn & Griffiths, 2008; Sanborn, Griffiths, & Navarro, 2006), change detection (Brown & Steyvers, 2008), associative learning (Daw & Courville, 2008), and language learning (Xu & Tenenbaum, 2007) have explicitly or implicitly relied on a sampling process like probability matching (Herrnstein, 1961), soft-max decision policies (Sutter & Barlo, 1998), or Luces choice rules (Luce, 1959) to link the ideal Bayesian posterior to participants' responses, indicating that in many cases when Bayesian models predict human behavior, they do so through the assumption that people sample instead of computing the response that will maximize expected utility under the full posterior distribution. Because the resulting variation in judgments across participants and trials suggests that each individual is guessing based on only a small number of samples, it might seem that it is a mistake to describe such behavior in terms of ideal Bayesian observers (Mozer, Pashler, & Homaei, 2008).

The second challenge—that Bayesian inference is intractable—arises from the difficulty of scaling probabilistic models to real-world problems. For problems involving discrete hypotheses about the processes that could have produced observed data, the computational cost of Bayesian inference increases linearly with the number of hypotheses considered. However, the number of hypotheses (possible generative processes) increases rapidly in common settings with combinatorial structure. For example, the number of causal structures relating a set of variables increases exponentially in the number of variables (with over 3 million possible structures for just six variables), and the number of clusterings of a set of objects increases similarly sharply (with over 100,000 partitions of just 10 objects). In other cases, possible generative processes are drawn from infinite discrete hypothesis spaces (such as when parsing with a recursive grammar), or continuous hypothesis spaces where there is no direct way to calculate the integrals required for Bayesian inference. The high computational cost that results from using probabilistic models has led computer scientists and statisticians to explore a variety of approximate algorithms, with exact computations being the exception rather than the rule in implementations of Bayesian inference (Gelman, Carlin, Stern, & Rubin, 2004).

Within cognitive science, these challenges are considered serious enough to question the whole program of Bayesian cognitive modeling. Mozer et al. (2008) argued that although many samples may adequately approximate Bayesian inference, behavior based on only a few samples is fundamentally inconsistent with the hypothesis that human cognition is Bayesian:

> If our investigations had found that. . . some sample-based model required, say, 20 samples per individual to match the data, we would not have considered the sampling account to be a qualitatively different story than the Bayesian account. However, when two samples per individual accounts for the data, our sense is that the [sample-based] and Bayesian accounts have to be viewed as qualitatively distinct. (p. 1146)

This perspective is consistent with commonly held heuristics in Bayesian statistics and artificial intelligence, where practitioners use hundreds or thousands of samples to approximate the relevant posterior distributions. For instance, Gelman et al. (2004) write that although 100 samples are often sufficient for reasonable posterior inference, 2,500 or more may often be required to avoid discounting rare events. Similarly, Gamerman and Lopes (2006) estimate that to obtain a 95% confidence interval with sufficient precision requires 600, or even 3,746, independent samples. Altogether, it is clear that just one or a few samples are far from sufficient to form an adequate approximation to the posterior distribution; therefore, if people make decisions with such poor approximations of the posterior, is it fruitful to describe their behavior as rational statistical inference? Others highlight the second challenge and argue that cognition cannot be Bayesian inference because exact Bayesian calculations are computationally intractable (Kwisthout, Wareham, & van Rooij, 2011), so the brain must rely on computationally efficient heuristics rather than Bayesian calculations (e.g., Gigerenzer, 2008). Kwisthout et al. (2011) argue that because Bayesian calculations are intractable, and even adequate approximate inference for these problems is computationally prohibitive, the enterprise of Bayesian modeling cannot ignore the algorithmic level of description. Similarly, Jones and Love (2011) stress the importance of integrating the computational and algorithmic levels of description for a cognitively and neurally plausible account of human cognition. Addressing these challenges is thus an important step toward addressing the psychological and empirical plausibility of probabilistic models as a framework for describing, modeling, and understanding human cognition.

In this study we will argue that acting based on a few samples can be easily reconciled with optimal Bayesian inference and may be the method by which people approximate otherwise intractable Bayesian calculations. Our argument has three central claims. First, that probability matching behavior can be understood in terms of sensible sampling-based approaches to approximating intractable inference problems of the kind used in Bayesian statistics and computer science. Second, that very few samples from the Bayesian posterior are often sufficient to obtain approximate predictions that are almost as good as predictions computed using the full posterior. And third, that under conservative assumptions about how much time it might cost to produce a sample from the posterior, making predictions based on very few samples (even just one) can actually be the globally optimal strategy.

2. Ideal aggregates from sampling behavior by individuals

We will begin by describing two characteristic cases that best highlight the apparent sampling phenomenon. Goodman et al. (2008) studied performance in classic categorization tasks, in which participants learn to discriminate two categories (A and B) by studying exemplars of each category, and are then asked to generalize the learned rules by categorizing new transfer items. Goodman et al. (2008) showed that the proportion of participants who classify transfer items as belonging to one of the trained categories fits almost perfectly with the probabilistic predictions of a Bayesian rule-learning model (Fig. 1-top). This model considers all possible logical rules for classification (expressed as disjunctions of conjunctions

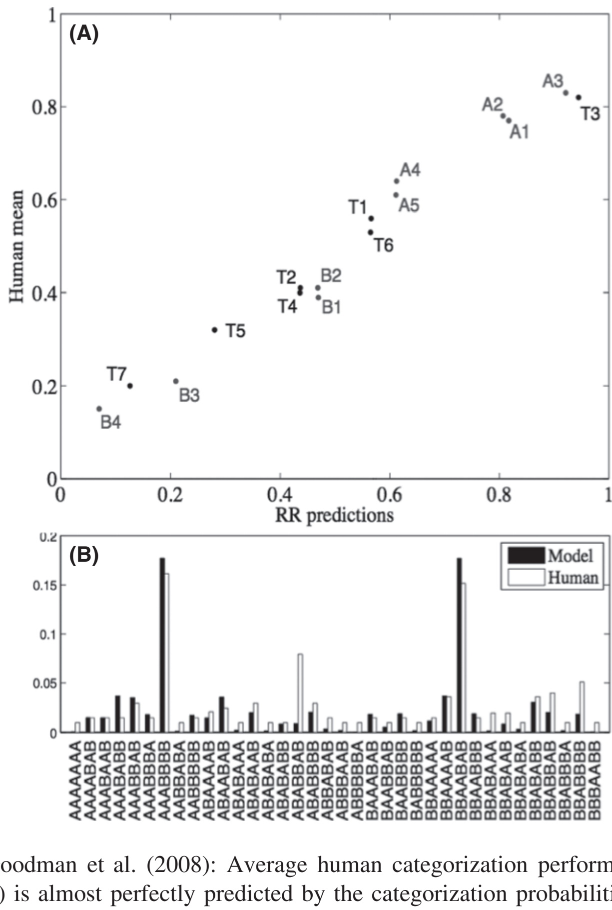

Fig. 1. (Top) From Goodman et al. (2008): Average human categorization performance (*y*-axis; proportion categorized as group A) is almost perfectly predicted by the categorization probabilities from an ideal Bayesian model (*x*-axis) that makes categorization predictions by integrating over the complete posterior probability distribution over categorization rules. Each point corresponds to previously seen objects from categories A and B (A^* and B^*, respectively) and new transfer items (T^*). (Bottom) A different picture emerges from the generalization patterns over seven test stimuli (*x*-axis; 'AAAAAAB,' for instance, corresponds to categorizing the first six test stimuli as category A, and the seventh as category B). The histogram of participants' generalization (white bars) does not match this ideal observer (gray bars). Instead, these patterns reflect much greater correlation of beliefs from one test probe to the next, consistent with individual participants adopting one or a few rules in proportion to their posterior probability, and making many generalization responses accordingly (black bars). Bayesian behavior emerges only on average, while individual participants seem to behave based on just a few samples from the posterior.

of Boolean features), computes a posterior probability for each rule given the training data, and then computes the probability that any item is a positive instance by averaging the decisions of all possible rules weighted by their posterior probabilities. Do individual participants compute this same average over all possible rules in their heads on any one trial? Not in this task. Goodman et al. (2008) analyzed the generalization patterns of more than 100 individual participants reported by Nosofsky, Palmeri, and McKinley (1994) and found that the response patterns across seven test exemplars were only poorly predicted by the Bayesian ideal, even allowing for random response noise on each trial. Rather than averaging over all rules, these generalization patterns were consistent with each participant classifying test items using only one or a few rules, although the rules considered vary across observers in proportion to the appropriate posterior probabilities (Fig. 1-bottom). Thus, it seems that individual human learners are somehow drawing one or a few samples from the posterior distribution over a complex hypothesis space of categorization rules, and the aggregate behavior that is consistent with integrating over the full posterior distribution emerges only in the average over many learners.

Such probability matching behavior can also be seen when individuals make predictions from their own knowledge of the world. Griffiths and Tenenbaum (2006) studied participants' predictions about every day events, such as how long a cake will bake given that it has been in the oven for 45 min. By varying the time cutoff (e.g., 45 min) they showed that the median participants' judgments closely match the posterior medians of an optimal Bayesian predictor that knows the complete real-world distribution over these durations for a number of different event classes (cake baking times, movie run times, poem lengths, etc). However, again, it was not the case that individual participants made judgments in line with the average predictions: The variation in judgments across participants suggests that each individual is guessing based on only a small number of instances considered with probability proportional to the Bayesian posterior (Mozer et al., 2008). Moreover, the quantile–quantile comparison of the distribution of participants' responses and the Bayesian posterior distribution shows a near-perfect match, indicating that participants' guesses do not correspond to the single optimal choice under the full posterior (perturbed by random response noise), but instead correspond to samples from the posterior (for further details see Lewandowsky, Griffiths, & Kalish, 2009)). Fig. 2 shows the comparison of median human judgments with Bayesian posterior medians, along with the full quantile–quantile plots relating human and model predictions for seven different classes of everyday events, and an aggregate plot combining these data. Although there are some deviations in specific cases—such as a tendency to produce tighter predictions than the posterior for human life spans—the aggregate results show a close match between the two probability distributions, consistent with the idea that people are making predictions by sampling from the posterior distribution.

What mechanisms operating in individual participants would produce ideal Bayesian behavior in the aggregate via individual trials that appear to be probability matched to the posterior?

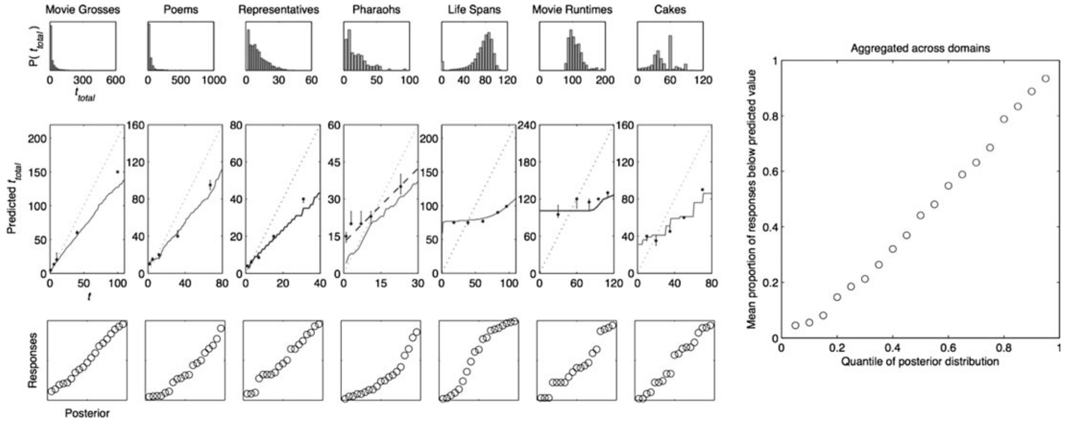

Fig. 2. Data from Griffiths and Tenenbaum (2006) showing optimal predictions for everyday quantities. (Left, top row) The real empirical distributions of quantities across a number of domains; from left to right: movie grosses, poem lengths, time served in the U.S. House of Representatives, the reigns of Egyptian pharaohs, human life spans, movie runtimes, and the time to bake a cake. (Left, middle row) When participants are asked to predict the total quantity based on a partial observation (e.g., what is the total baking time of a cake given that it has been baking for 45 min?), they make predictions that appear to match the Bayesian ideal observer that knows the real-world distribution. Thus, it would appear that in all these domains, people know and integrate over the full prior distribution of, for example, cake baking times when making one prediction. (Left, bottom row) However, the quantile–quantile plots comparing the distributions of human predictions with the corresponding posterior distributions reveal a different story. For each prediction, the quantiles of human response distributions were computed and then compared with the corresponding posterior distribution produced by using Bayesian inference with the appropriate prior (to produce each plot, quantiles were averaged across five predictions for each phenomenon). A match between the Bayesian posterior distribution and the distribution of people's responses corresponds to data points following along a diagonal line in these plots—where the quantiles of the two distributions are in direct correspondence. (Right) The correspondence between the posterior predictive and human responses is most pronounced when considering the quantile–quantile plot that reflects an aggregate over all seven individual quantities. Thus, people make guesses with frequency that matches the posterior probability of that answer, rather than maximizing and choosing the most likely alternative. This indicates that although participants *know* the distribution of cake baking times (as evidenced by the quantile–quantile match), they do not produce the optimal Bayesian response by integrating over this whole distribution, but instead respond based on only a small number of sampled baking times.

3. Approximating Bayesian inference by sampling

Bayesian probability theory prescribes a normative method for combining prior knowledge with observed data and making inferences about the world. However, the claim that human cognition can be described as Bayesian inference does not imply that people are doing exact Bayesian inference. Exact Bayesian inference amounts to fully enumerating hypothesis spaces every time beliefs are updated with new data. This is computationally intractable in any large-scale application, so inference must be approximate. As noted earlier, this is the case in Bayesian artificial intelligence and statistics, and is even more relevant to solving the kinds of problems we associate with human cognition, where the real-world inferences are vastly more complex and responses are time sensitive.

The need to approximate Bayesian inference leaves two important questions. For artificial intelligence and statistics: What kinds of approximation methods work best for Bayesian inference? For cognitive science and psychology: What kinds of approximation methods does the human mind use? In the tradition of rational analysis (Anderson, 1991), or analysis of cognition at Marr's (1982) computational level, one strategy for answering the psychological question begins with good answers to the engineering question. Thus, we will explore the hypothesis that the human mind approximates Bayesian inference with some version of the algorithmic strategies that have proven best in artificial intelligence and statistics, on the grounds of computational efficiency and accuracy.

In artificial intelligence and statistics, one of the most common methods for implementing Bayesian inference is with sample-based approximations. Inference by sampling rests on the ability to draw samples from an otherwise intractable probability distribution—that is, to arrive at a set of hypotheses which are distributed according to the target distribution, by using a simple algorithm (such as Markov Chain Monte Carlo [MCMC]; Robert & Casella, 2004; or particle filtering; Doucet, De Freitas, & Gordon, 2001). Samples may then be used to approximate expectations and predictions with respect to the target probability distribution, and as the number of samples grows, these approximations approach the exact quantities.[1] Sampling methods are typically used because they are applicable to a large range of computational models, are robust to increasing dimensionality, and degrade gracefully when computational resources limit the number of samples that can be drawn.

Computer scientists and statisticians use a wide range of sampling algorithms. Some of these algorithms have plausible cognitive interpretations, and specific algorithms have been proposed to account for aspects of human behavior (Brown & Steyvers, 2008; Gershman, Vul, & Tenenbaum, 2012; Levy, Reali, & Griffiths, 2009; Sanborn et al., 2006; Shi, Griffiths, Feldman, & Sanborn, 2010). For our purposes, we need only assume that a person has the ability to draw samples from the hypothesis space according to the posterior probability distribution. Thus, it is reasonable to suppose that people can approximate Bayesian inference via a sampling algorithm, and evidence that humans make decisions by sampling is not in conflict with the hypothesis that the computations they are carrying out are Bayesian.

However, using an approximation algorithm can often result in strong deviations from exact Bayesian inference. In particular, poor approximations can be produced when the number of samples is small. Recent empirical results suggest that if people are sampling from the posterior distribution, they base their decisions on very few samples (Goodman et al., 2008; Mozer et al., 2008; Vul & Pashler, 2008)—so few that any claims of convergence to the real probability distribution do not hold. Algorithms using only a few samples will have properties quite different from full Bayesian integration. This leaves us with the question: How bad are *decisions* based on few samples?

4. Two-alternative decisions

To address the quality of decisions based on few samples, we will consider performance of an ideal Bayesian agent (maximizing expected utility under the full posterior distribution over hypotheses) and a sample-based agent (maximizing expected utility under a small set of sampled hypotheses). We will start with the common scenario of choosing between two alternatives. Many experimental tasks in psychology are a variant of this problem: Given everything observed, make a two-alternative forced-choice (2AFC) response. Moreover, real-world tasks often collapse onto such simple 2AFC decisions; for instance, we must decide whether to drive to the airport via the bridge or the tunnel, depending on which route is likely to have least traffic. Although this decision will be informed by prior experiences that produced intricate cognitive representations of possible traffic flow, at the moment of decision these complex representations collapse onto a prediction about a binary variable: Is it best to turn left or right?

4.1. Bayesian and sample-based agents

Statistical decision theory (Berger, 1985) prescribes how information and beliefs about the world and possible rewards should be combined to define a probability distribution over possible payoffs for each available action (Kording, 2007; Maloney, 2002; Yuille & Bülthoff, 1996). An agent trying to maximize payoffs over many decisions should use these normative rules to determine the expected payoff of each action, and choose the action with the greatest expected payoff. Thus, the standard for decisions in statistical decision theory is to choose the action (A^*) that will maximize expected utility ($U(A;S)$) of taking an action under the posterior distribution over possible current world states (S) given observed data (D):

$$A^* = \arg\max_A \sum_S U(A;S)P(S|D). \tag{1}$$

To choose an action, the only property of world states we care about is the expected utility of possible actions given that state. Thus, if there are two possible actions (A_1 and A_2) and one action is "correct" for each world state (that is, there are two possible values for $U(A;S)$ and only one action for each state receives the higher value), then we may collapse the state space onto a binary alternative: Is A_1 correct or A_2?[2] Under this projection the posterior distribution becomes a Bernoulli distribution, where the posterior probability that A_1 is correct is p—this quantity fully parameterizes the problem, with respect to the 2AFC task. The ideal Bayesian agent who maximizes expected utility will then choose the action which is most likely to be correct (the *maximum a posteriori*, MAP, action, and will be correct p proportion of the time). (In what follows we assume p is between 0.5 and 1, without loss of generality.)

A sample-based agent samples possible world states (S_n) from the posterior distribution, uses those samples to estimate the expected utility of each action, and makes a decision based on that estimate:

$$A^* = \arg\max_A \sum_{i=1}^{k} U(A; S_i) \tag{2}$$
$$S_i \sim P(S|D).$$

Under the assumption that the utility has two values ("correct"/"incorrect"), the sample-based agent will thus choose the action which is most frequently correct in the set of sampled world states. In other words, the sample-based agent chooses the action which was best more than half of the time in the set of samples. As these samples favor the correct action with probability p, we can use the Binomial distribution to calculate the probability with which a given action appears more favorable (based on the set of k samples). Thus, a sample-based agent drawing k samples will choose action A_1 with probability:

$$q = 1 - \Theta_{CDF}\left(\lfloor\frac{k}{2}\rfloor, p, k\right), \tag{3}$$

where Θ_{CDF} is the binomial cumulative density function describing the probability that fewer than half ($\lfloor\frac{k}{2}\rfloor$) of k samples will suggest that the correct action is the best one, given that the posterior probability of the correct action is equal to p over the set of all possible samples. Thus, q is the probability that the majority of samples will point to the correct (MAP) action.[3] Therefore, the sample-based agent will be right with probability $qp + (1 - q)(1 - p)$.

Here, we discuss a sample-based agent specifying an a priori decision policy as a fixed number of samples, but other decision policies are possible. Particularly, it is worth considering (a) a sequential probability ratio test (SPRT) policy (Wald, 1947) that specifies a threshold amount of evidence required to make a choice (and can make a choice after a variable number of samples), and (b) an accumulator policy (following Vickers, 1979) that draws samples until a threshold number of them favor one of the two options. We will consider these policies in a later section (and in detail in Appendix A), but critically, the same conclusions are reached for all three decision policies for a sample-based agent. As such, we focus on a fixed-k (number of samples) policy, as it most clearly shows our primary conclusion: that very few samples are necessary.

4.2. Good decisions from few samples

So, how much worse will such 2AFC decisions be if they are based on a few samples rather than an inference computed by using the full posterior distribution? Bernoulli estimated that more than 25,000 samples are required for "moral certainty" about the true probability of a two-alternative event (Stigler, 1986).[4] Although Bernoulli's calculations were based on different derivations than those which are now accepted (Stigler, 1986), it is undeniable that *inference* based on a small number of samples differs from the exact Bayesian solution and will contain greater errors, but how bad are the *decisions* based on this inference?

Making a decision based on a limited number of samples is analogous to predicting the outcome of a single flip of a bent coin after flipping it a few times to gather information about its bias. In Fig. 3 we plot the difference in error rates between the sample-based and optimal agents as a function of the underlying probability (p) and number of samples (k). When p is near 0.5, there is no use in obtaining any samples because a perfectly informed decision will be as likely to be correct as a random guess; this is the case of a fair coin, even if we flip it many times to make sure of this fact, our prediction about the next coin flip will not be improved. When p is near 1, the first sample will indicate the (nearly, deterministically correct) answer, so there is much to be gained from one sample but subsequent samples are of little use; this is the case of a perfectly biased coin that always lands on heads—one flip yields all the information needed to inform a prediction. Most of the benefit of large numbers of samples occurs in interim probability values (around 0.7)—where the coin is substantially, but incompletely biased.

As the sample-based agent does not know what the true probability p may be for a particular decision, we can consider the scenarios such an agent should expect: the average scenario (expectation over p) and the worst-case scenario (maximization of the loss over p). These are displayed in Fig. 4A assuming a uniform probability distribution over p. The deviation from optimal performance decreases to negligible levels with very few samples, suggesting that the sample-based agent need not have more than a few samples to approximate ideal performance. We can go further to assess just how much is gained (in terms of decreased error rate) from an additional sample (Fig. 4B). Again, the vast majority of accuracy is gained with the first sample, and subsequent samples do very little to improve performance. Thus, even though few samples will not provide a very accurate

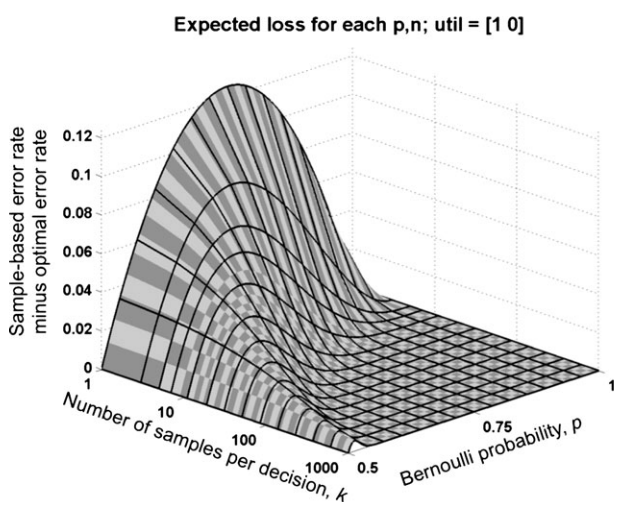

Fig. 3. Increased error rate for the sample-based agent over the optimal agent as a function of the probability that the first action is correct and the number of samples drawn for a decision (decisions based on zero sample not shown).

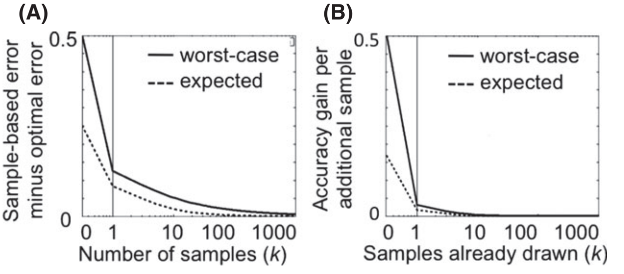

Fig. 4. Increased error rate for the sample-based agent in 2AFC decisions marginalizing over the Bernoulli probability (assuming a uniform distribution over p). (A) The maximum and expected increase in error for the sample-based agent compared with the optimal agent as a function of number of samples (see text). (B) Expected and maximum gain in accuracy from an additional sample as a function of the number of samples already obtained.

estimate of p—definitely not sufficient for "moral certainty"—they *are sufficient to choose an action*: We do not need moral certainty to act near optimally.

4.3. How many samples for a decision?

If people do make inferences based on samples—but as samples are costly—how many samples should people use before making a decision? For instance, how many possible arrangements of traffic across the city should we consider before deciding whether to turn left for the tunnel or right for the bridge? Considering one such possibility requires concerted thought and effort—it seems obvious that we should not pause at the intersection for several hours and enumerate all the possibilities. It also seems likely that we should not just turn left or right at random without any consideration. Therefore, how many samples should we take: How many times should we flip our bent coin to estimate its bias before making a guess about the outcome of the next flip? In other words, how hard should we think?

Determining an optimal answer to this meta-cognitive problem requires that we specify how much a sample may "cost." To be conservative (and for the sake of simplicity), we will assume that a sample can only cost time—it takes some amount of time to conjure up an alternate outcome, predict its value, and update a decision variable. Samples may have additional costs, not just time, but effort, metabolic costs, etc. (see Discussion). However, any additional costs will favor *fewer* samples, so for parsimony here we assume that additional samples pay only an opportunity cost of time.

If a given sample is free (costs 0 time), then we should take infinitely many samples, and make the best decision possible every time. If a sample costs 1 unit of time, and the *action time* (the time that it would take us to act once we have chosen to do so) is also 1 unit of time, then we should take zero samples; that is, we should guess randomly. To make this peculiar result intuitive, let us be concrete: If we have 100 s, and the action time is fixed to be 1 s, then we can make 100 random decisions, which will be right 50% of the time, thus giving us an expected reward of \$50 (assuming correct choices pay \$1, and incorrect choices are not penalized). If taking a single sample to improve our decision will cost an additional second per decision, then if we take one sample per decision, each decision will take 2 s, and we could make at most 50 of them. It is impossible for the expected reward from this strategy to be greater than guessing randomly, as even if 100% of the decisions are correct, only $50 will be gained. Moreover, as 100% accuracy based on one sample is extremely unlikely (this could only arise in a completely deterministic prediction task), substantially less reward should be expected. Thus, if obtaining a sample takes as long as the action, and we do not get punished for an incorrect answer, we should draw zero samples per decision and make as many random decisions as we can. More generally, we can parameterize how much a sample "costs" as the ratio between the time required to make an action and the time required to obtain one sample (action/sample ratio)—intuitively, a measure of how many samples it would take to double the time spent on a decision compared with making the decision using no samples.

The expected accuracy for a sample-based agent (previous section) gives us the expected utility per decision as a function of k (the number of samples) and p (the probability that the first action is correct; Fig. 6A), and the utility function. We consider two utility functions for the 2AFC case: *no punishment*—correct: gain 1; incorrect: lose 0; and *symmetric*—correct: gain 1; incorrect: lose 1. These two extremes are limits on the circumstances for which making a choice is always preferable to stalling,[5] and all other payoffs with positive expected return will fall somewhere between these bounds. Given one particular action/sample time ratio, we can compute the number of decisions made per unit time (Fig. 5B). Multiplying these two functions together yields the expected utility per unit time (Fig. 5C).

As p is unknown to the agent, an ideal k must be chosen by taking the expectation over p. This marginalization (assuming a uniform distribution over p) for many different action/sample time ratios is displayed in Fig. 6. It is clear that as samples become cheaper, one is best advised to take more of them—converging to the limit of infinitely many samples when the samples are free (the action/sample time ratio is infinity).

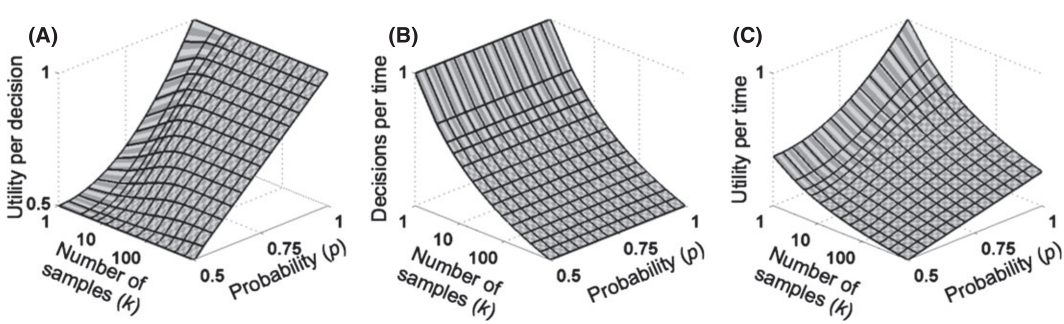

Fig. 5. Expected utility per decision, the number of decisions that can be made per unit time, and the expected rate of return (utility per unit time) as a function of the probability that the first action is correct and the number of samples (with an example action/sample cost ratio of 232, arbitrarily chosen from one of the logarithmically spaced cost ratios we evaluated).

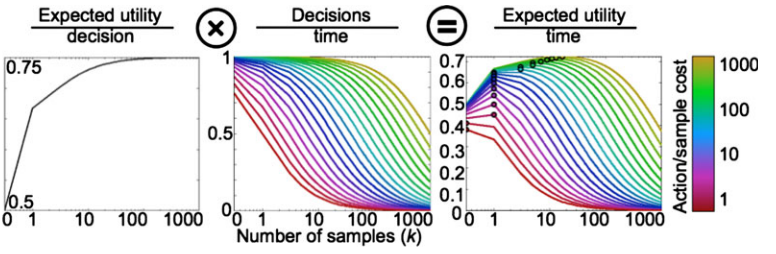

Fig. 6. Expected utility per decision (averaging over a uniform prior on p), number of decisions per unit time, and expected utility per unit time (rate of return) as a function of the number of samples and action/sample cost ratios. Action/sample cost ratios are logarithmically spaced between 1 (red) and 1,000 (yellow). In the last graph, the circles indicate the optimal number of samples at that action/sample cost ratio. (The utility function used here is +1 for a correct choice and 0 for incorrect.)

In Fig. 7 we plot the optimal number of samples as a function of the action/sample time ratio. Remarkably, for ratios less than 10, one is best advised to make decisions based on only one sample if the utility function is symmetric. Moreover, with no punishment for incorrect answers, the action/sample time ratio must be 2 or greater before taking any samples becomes a prudent course of action. Thus, under a wide range of assumptions about how much it costs to think, making guesses based on very few samples (e.g., one) is the best course of action: Making many locally suboptimal decisions quickly is the globally optimal strategy.

4.4. Optimal SPRT and accumulator decision policies for 2AFC

We have argued that the sample-based agent specifying a fixed number of samples a priori should, across a very wide range of sample costs, use just a few samples. Although

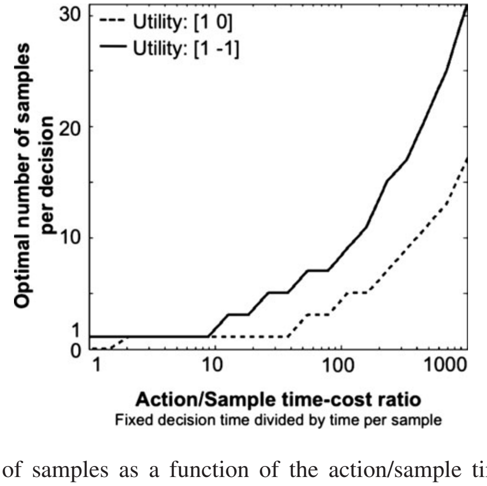

Fig. 7. The optimal number of samples as a function of the action/sample time-cost ratio for each of two utility functions (symmetric—correct: +1, incorrect: −1; and no punishment for incorrect answers—correct: +1, incorrect: 0).

this fixed-k policy highlights our claim that few samples are necessary, it is clearly not an ideal policy: If the agent decides to take five samples, and the first three all favor action A (*aaa*), there is no need to take another two samples; while if the first three are mixed (*aab*), then one or two further sample are worthwhile. Does our conclusion—that the optimal number of samples is very low—hold with decision policies that can choose to stop sampling more efficiently? We consider two-alternate decision policies for a sample-based agent: a SPRT policy (Wald, 1947), and an accumulator policy (Vickers, 1979) (illustrated in Fig. 8). We find that optimal stopping thresholds for these more efficient decision policies yield nearly identical expected numbers of samples per decision, and expected rates of return as a simple fixed-k policy.

The SPRT policy (Wald, 1947) specifies a threshold for the amount of evidence required to make a choice, and it draws samples until that threshold is reached. Formally, the SPRT agent calculates $d_k = x_1 - x_2$: the difference in number of samples favoring choices A_1 and A_2. When $|d_k|$ reaches threshold T (i.e., when d_k reaches T or $-T$), the SPRT agent chooses the corresponding action. Thus, SPRT will choose action A_1 with probability $q = \frac{p^T}{p^T+(1-p)^T}$ (Navarro, 2007), after a variable number of samples k (where

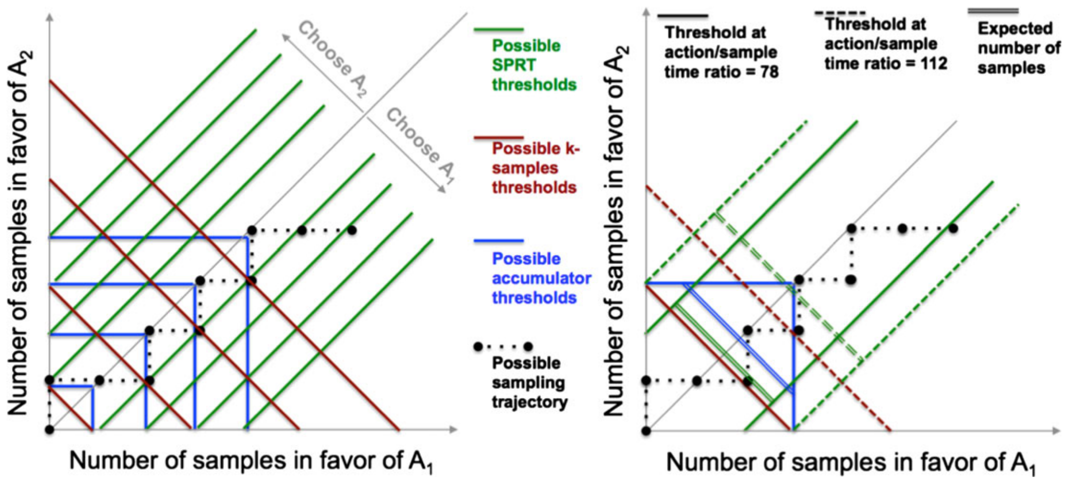

Fig. 8. Illustration of different decision policies for a sample-based agent. The accumulation of samples to support a 2AFC decision can be illustrated as a random walk in two dimensions: number of samples in support of each of the two alternatives. Each sample amounts to a step either rightward, or upward (the black dashed line shows one such possible random walk). (Left) Different decision policies draw stopping thresholds at different orientations in this plot. The fixed-k decision policy (red) amounts to picking a diagonal line corresponding to a constant number of samples. The SPRT policy thresholds (green) are pairs of orthogonal diagonals at an equal distance from the identity line (corresponding to different amounts of evidence in favor of an action). The accumulator policy (blue) chooses thresholds that correspond to a constant number of samples in favor of either action. (Right) Optimal thresholds for the three decision policies for two action/sample time ratios (solid lines: 78; dashed lines: 112; for the accumulator policy, both of these ratios yielded the same threshold $T = 3$). The diagonal double lines indicate the expected number of samples per decision for accumulator and SPRT policies.

$k \geq T$ and k has the same parity as T). The SPRT policy is optimal in the sense that it will attain a desired level of confidence faster than other possible policies.

The accumulator policy (Vickers, 1979) specifies a threshold on the number of samples obtained in favor of one of the two options. Formally, under the accumulator policy, the agent obtains samples until either x_1 or x_2 (the number of samples in favor of A_1 and A_2, respectively) reaches T, and then chooses the corresponding action. The accumulator policy guarantees neither a particular level of confidence, nor a fixed number of samples, but it does offer a more efficient policy than specifying a fixed number of samples. Requiring a fixed number of k samples for the fixed-k policy effectively specifies an Accumulator policy threshold $T = \lceil k/2 \rceil$ samples in favor of any one action (where $\lceil X \rceil$ denotes rounding X up to the nearest integer). The corresponding accumulator policy will take at most k samples, but it could reach a decision in as few as $\lceil k/2 \rceil$ samples, thus achieving the same end with fewer samples on easy trials (where p is close to 1).

We calculated optimal thresholds to maximize expected rate of return for SPRT and accumulator decision policies for various time costs of samples given that the sampling agent has a uniform prior over p (see Appendix A for calculations). Fig. 9 shows a comparison of the optimal thresholds, expected number of samples, and expected rate of return for the fixed-k, SPRT, and accumulator policies. There is very little variation in expected number of samples and rates of return across policies, and regardless of what decision policy the sample-based agent uses, the expected number of samples under the optimal threshold is very low: For the no punishment utility function, expected numbers of samples are 0 or 1 for action/sample time ratios below 25.

Our further analyses (of nAFC and continuous decision) will consider only the fixed-k policy because the fixed-k policy highlights our central point—that the optimal sampling agent uses very few samples—and this finding holds across other decision policies.

5. *N*-alternative decisions

So far we have only considered two-alternative decisions: In such cases, no matter how high dimensional the state of the world may be, the decision collapses onto one binary variable. It is likely that our analysis would produce different results when more than two alternatives are available (and thus, more information is required to choose among them). Therefore, we now ask the same questions of N-alternative forced-choice tasks, where N is 4, 8, 16, and 32: How bad are choices among many alternatives if such decisions are based on few samples? And how many samples *should* we use when we are faced with such a decision?

On the assumption that the utility functions for such N-AFC decisions are that one and only one of the N alternatives is "correct" and the others are "incorrect," the optimal agent (who knows the multinomial distribution describing the probability that any one choice is "correct") will always choose the alternative that has the highest probability (MAP), and the agent will be right with that probability—max$p$; thus, the performance of the optimal agent only depends on max$p$. The sample-based agent, just as in the 2AFC

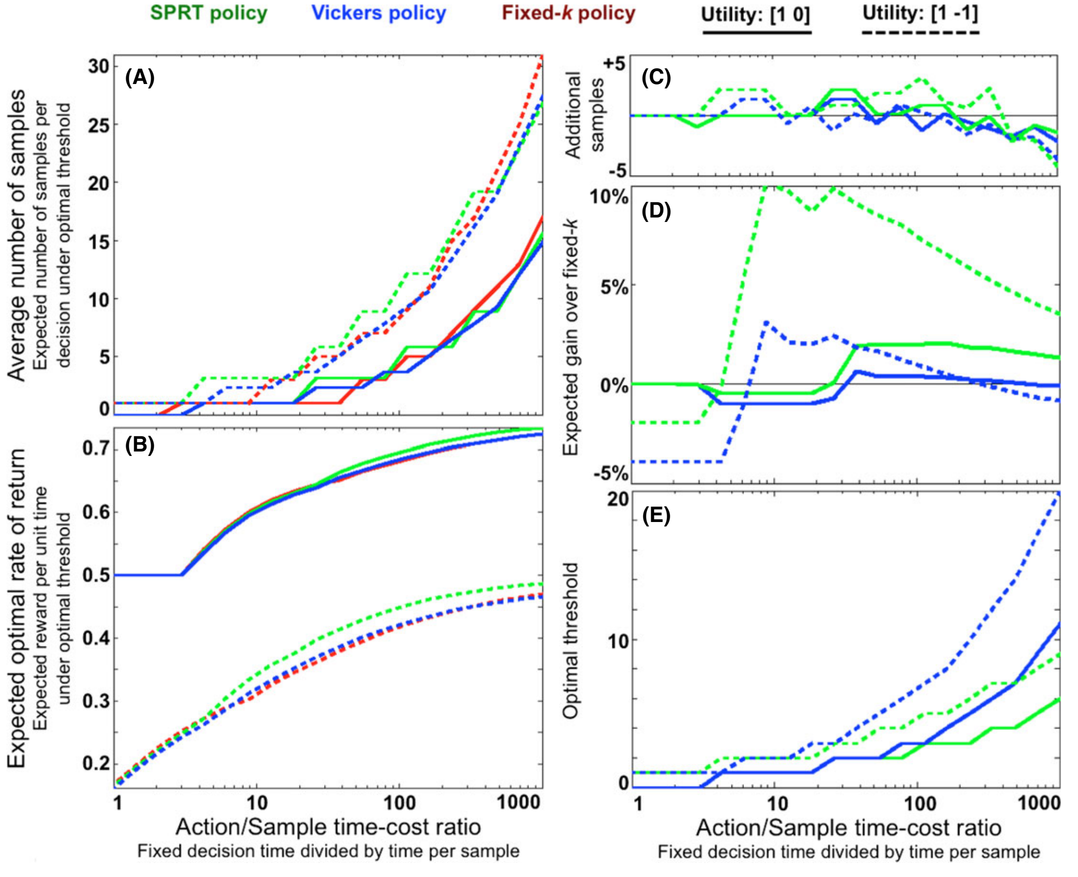

Fig. 9. Optimal fixed-*k* (red), SPRT (green), and accumulator (blue) policies as a function of action/sample time ratio for two utility functions: no punishment (solid; correct: +1; incorrect: 0), and symmetric (dashed; correct: +1; incorrect: −1). (A) Expected number of samples per decision under the optimal threshold (marginalizing over a uniform distribution on *p*): regardless of the decision policy, when samples are costly, the optimal threshold yields very few samples per decision. (B) Expected rate of return per unit time under the optimal threshold: of course, a no punishment utility function offers higher rates of return; rates of return increase as samples become cheaper; and policies that use samples more efficiently (like SPRT) show a slight advantage over fixed-*k*. (C) Difference in the expected number of samples per decision compared with the fixed-*k* policy: Although they use very different stopping criteria, optimal SPRT and accumulator policies end up using roughly the same number of samples at a given action/sample time ratio as the fixed-*k* policy—sometimes lower, sometimes higher. (D) Expected gain in rate of return over the fixed-*k* policy (as a percentage): Although they use roughly the same number of samples as the fixed-*k* policy, SPRT and accumulator decision policies are more efficient and thus can attain higher rates of return when samples are relatively cheap; however, none of these policies dominates the others, as there exists an action/sample time ratio and utility function under which each policy is optimal. (E) The optimal thresholds *T* for SPRT and accumulator policies (note that the optimal fixed-*k* thresholds are shown in plot (A)).

case, will choose the alternative that was "correct" under the most samples. Therefore, we show the inflation of error rates for the sample-based agent over the optimal agent as a function of the number of samples and the maximum probability of the multinomial

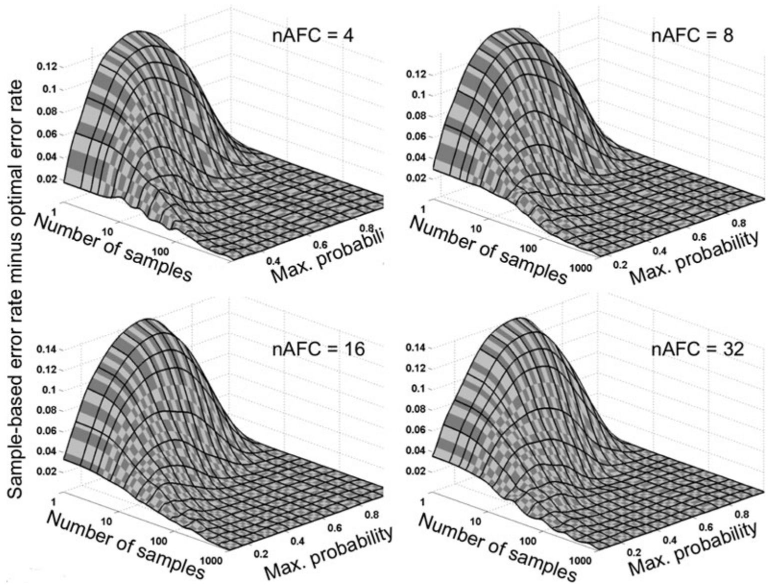

Fig. 10. Increased error rate for the sample-based agent over the optimal agent as a function of the number of alternatives in the decision (different panels), the number of samples, and the probability of the highest probability alternative. These values were produced by numerical simulation.

(Fig. 10). Just as in the 2AFC case, the optimal agent has the greatest advantage in problems of “interim” difficulty—when the maximum probability is neither too close to chance (where the sample-based agent and the optimal agent must both resort to random guessing) nor too close to certainty (when one sample will be sufficient for perfect accuracy for the sample-based agent). Again, just as in the 2AFC case, the advantage of many samples decreases quickly.

Once again, the relevant question is as follows: How much worse should the sample-based agent *expect* to fair, given that probabilities are unknown. Thus, we again marginalize over possible probability distributions over alternatives, assuming a uniform prior over multinomial distributions (a Dirichlet distribution with $\alpha = 1$), and obtain the expected additional error for the sample-based agent over the optimal agent as a function of the number of samples (Fig. 11A). And again, just as in the 2AFC case, we see that the expected additional error decreases quickly (albeit faster for choices with fewer alternatives).

Finally, we ask: How many samples *should* the sample-based observer take when faced with a choice among many alternatives? We take the same analysis strategy as in the 2AFC case: We assume that a given sample costs time, and thus slows down the decision, and that a rational sample-based agent is trying to maximize expected rate of return. As such, we can multiply the expected utility (here, we consider only the “no punishment” utility function[6]) by the number of decisions made per unit time, for each number of samples obtained. This interim calculation is shown in Fig. 11B.

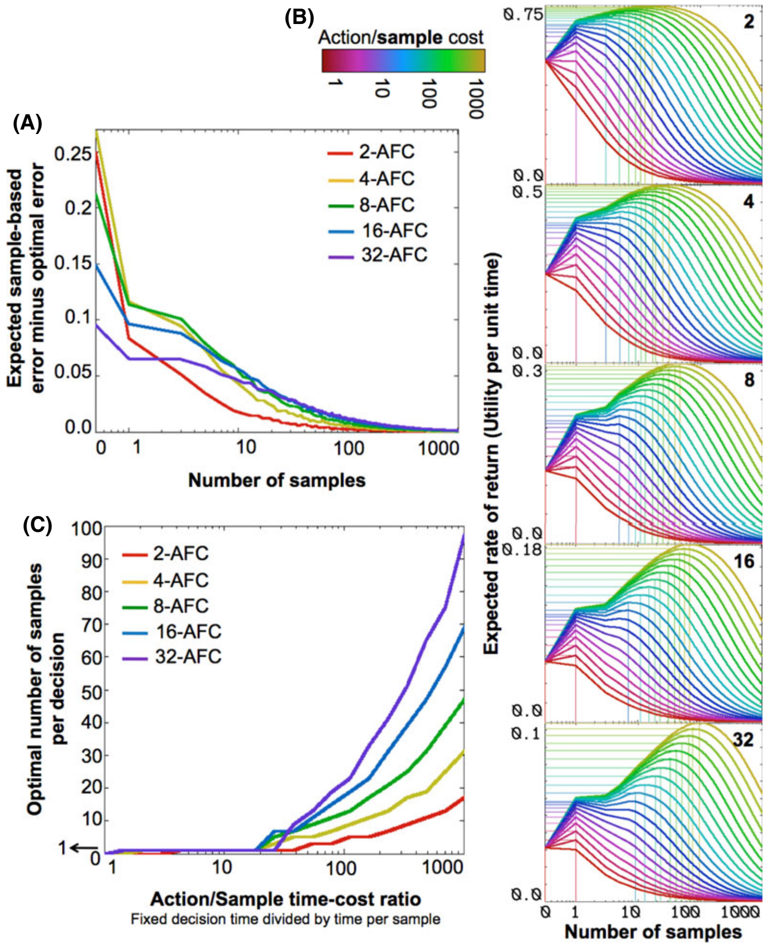

Fig. 11. (A) Expected increased error rate for the sample-based agent over the optimal agent as a function of the number of alternatives in the decision (different lines) and the number of samples (horizontal axis). (B) Expected rate of return for the sample-based agent as a function of number of alternatives in a decision (different panels), the number of samples used per decision (horizontal axis), and the action/sample cost ratio (different lines). (C) Optimal number of samples for the sample-based agent as a function of the action/sample cost ratio (horizontal axis) and the number of alternatives in the decision being made (different lines). All these plots marginalize over possible multinomial probability distributions describing which alternative is correct, using a uniform Dirichlet prior.

From the calculation in Fig. 11B, we can then plot the optimal number of samples given a particular sample cost, for decisions with different numbers of alternatives. This is displayed in Fig. 11C. Just as in the 2AFC case, a large regime of possible sample costs result in 1 as the optimal number of samples. However, the more alternatives there are, the faster the optimal number of samples rises as a function of decreasing sample cost, reaching an optimal calculation of as many as 75 samples within our tested range. Nonetheless, again, we see that in a large range of possible sample costs, making very

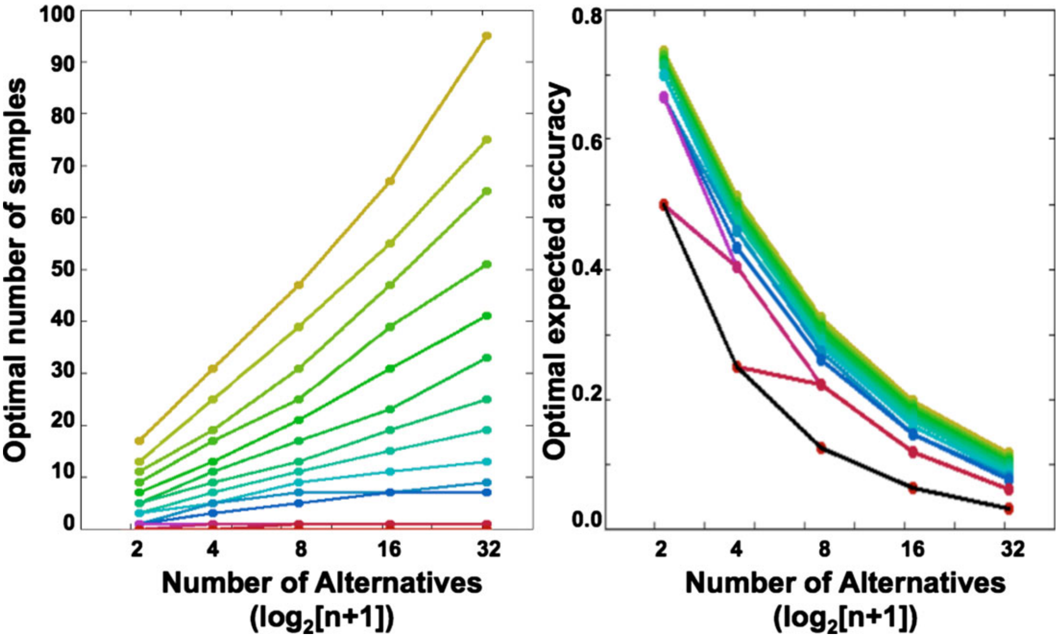

Fig. 12. Hick's law from a sample-based agent. (Left) Optimal number of samples (vertical axis) for the sample-based agent as a function of the number of alternatives ($\log_2(n + 1)$, horizontal axis) and the cost per sample (different lines). (Right) Expected accuracy given the optimal number of samples. The optimal sample-based agent shows linear scaling of response times with the logarithm of the number of alternatives, while allowing accuracy to drop as the number of alternatives increases (as people do; Brown et al., 2009).

quick, suboptimal decisions is the best policy, even when choosing among as many as 32 alternatives.

These results dovetail with the literature on the Hick–Hyman law (Hick, 1952; Hyman, 1953), which describes the near-ubiquitous log-linear relationship between the number of alternatives (n) in a decision problem, and the average response time on the problem as $RT = a + b \log_2(n + 1)$, where a is the fixed time of a movement. Previous work has shown that an agent making nAFC decisions based on racing diffusion/accumulator models, who adjusts thresholds to reach constant high accuracy rates, will exhibit Hick's law reaction times (Usher, Olami, & McClelland, 2002), and that this feature arises from basic statistical requirements for obtaining sufficient information to attain a certain level of accuracy as the number of alternatives increases (Usher & McClelland, 2001). Our result shows that the Hick–Hyman relationship falls out of our analyses as the optimal policy, without presuming a requirement to maintain constant accuracy (Fig. 12). Indeed, as the number of alternatives increases, the optimal sample-based policy shows a marked drop in accuracy (while the number of samples increased proportionally to $\log(n)$). This is exactly the behavior that real subjects exhibit when their response times and accuracies are not constrained by the experiment (Brown, Steyvers, & Wagenmakers, 2009). Thus, the number of samples a sample-based agent should take before reaching a decision for nAFC tasks, given a fixed action/

sample cost ratio follows the Hick–Hyman law. On our analysis, the slope, b, of this linear relationship is determined by the time cost, or difficulty, of obtaining a single sample for the decision, or as intuitively described by Hick (1952), the inverse of the processing speed.

6. Continuous decisions

Thus far we have shown that for choices among 2, 4, 8, 16, and 32 alternatives, a sample-based agent is often best advised to make decisions based on very few samples—thus, it should not be surprising that people are often observed to make decisions as though they are taking only a few samples in such scenarios. However, many human actions are defined over continuous variables: where to move an arm, how long to wait for a tardy friend, etc. These are all "continuous" decisions, rather than choices among discrete alternatives. These scenarios never have a single, explicitly correct answer, but rather, are often rewarded for precision—the closer to the optimal answer, the higher the reward. We will now consider actions defined by continuous variables, and again ask: How bad are decisions based on few samples, and how many samples should a sample-based agent use?

Just as in the binomial (2AFC) and multinomial (nAFC) cases, we assume that for continuous choices the "correct" answer on a given trial is drawn from the posterior probability distribution that the optimal observer has access to (and which the sample-based agent is approximating with samples). For simplicity, we will assume that here this posterior probability distribution takes the form of a Gaussian with standard deviation σ_P.

6.1. Making continuously valued decisions

Determining choices among a set of discrete alternatives under a correct/incorrect payoff scheme for optimal and sample-based observers was straightforward: choose the alternative with highest probability (or the most samples). However, when choosing along a continuous dimension, this formalization no longer makes sense. The task would be hopeless if the reward structure were a delta function—that is, if only one of infinitely many continuously varying possibilities was deemed "correct" and rewarded.

Therefore, instead of structuring rewards as a delta function, it is common practice to define a reward function for continuous decisions that decreases as a function of distance from the correct answer. For instance, the target for archery competitions is composed of many concentric circles and archers attempt to get an arrow as close as possible to the center because the inner circles are worth more points. Typical reward functions for such games are different from the loss functions considered in statistics, which are commonly unbounded (for instance, L2: loss that increases with the square of the distance from the target): ranging from zero for a perfect answer to infinite loss. However, in the games we consider, and arguably in the real world, the loss function drops off until it reaches some bound—if one were to miss the archery target altogether, one gets zero points, regardless of how badly the target was missed. A variant of such a utility function has been characterized mathematically as a "maximum local mass" loss function: essentially, a utility function that is shaped like a Gaussian probability distribution peaking at the correct answer and dropping off to zero with distance (Brainard & Freeman, 1997). Thus, for continuous choice decisions we use the maximum local mass utility function, which captures the idea that there is one best answer and many equally wrong answers, but also avoids the impossible pitfalls of assuming a delta function as the utility structure of the task.

Given the maximum local mass utility function, the optimal observer should choose the mean of the Gaussian probability distribution describing her uncertainty; and the sample-based agent should choose the mean of the obtained set of samples (this holds in so far as the utility function and posterior are unidimensional, unimodal, and symmetric; for multi-dimensional problems, see Brainard & Freeman, 1997 for approximation algorithms).

6.2. How many samples should the sample-based agent use?

Error rates as quantified by the squared distance from the target are quite meaningless, as these do not take into account a meaningful utility function.[7] Thus, we skip directly to an analysis of how many samples will maximize the rate of return for the sample-based agent.

The optimal number of samples for a continuous decision will depend on two factors we had not previously considered. First, the breadth of the distribution predicting the target location, parameterized by its standard deviation, σ_P. Second, the breadth of the utility function, parameterized also by its standard deviation σ_U: how close to the target center does our response have to be to be rewarded. The optimal number of samples turns out to be a function of the ratio between these two standard deviations.

When σ_P is much larger than σ_U, then no matter how many samples we take (to obtain an accurate estimate of the mean of the predictive distribution) our prediction will still be so uncertain that the correct answer is unlikely to be close enough to the mean to be rewarded. Asking how many samples we should take in this case is like asking, "How carefully should we aim when throwing a crumpled piece of paper from the Empire State building into a trash can on the ground?" Obviously, we should not aim very carefully because no matter how carefully we aim our success will be left to chance. For exactly the same reasons, in such circumstances we should make decisions based on very few samples, as additional samples will be of no use.

When σ_U is much larger than σ_P, then it also does not make sense to take many samples. When this is the case, if we take one, or infinitely many samples, our guess is still guaranteed to be near the peak of the utility function, and we will obtain similar rewards. This case is analogous to throwing a piece of crumpled paper into a large trash can situated an inch under your hand; in this case, it makes no sense to spend time aiming because there simply is no way you could miss.

However, in an intermediate range, when the relationship between σ_U and σ_P is just right, then we should obtain many samples to improve performance. This is the scenario when we must throw our paper ball into a trash can from across the room—it is doable, but it is not so easy that we should not aim—it is *just the right level of difficulty*. In this case we would spend the time to take careful aim and delicately arc the toss. Similarly, in this case, we should take many samples when trying to make a decision.

Fig. 13 shows the optimal number of samples as a function of sample cost and the log of the ratio between σ_U and σ_P. We see exactly the effects described above—at interim ratios, when samples are cheap enough, we should take many of them (within our tested range, as many as 70). However, when decisions are too hard, or too easy, or the sample cost is not low (when it would take at most 10 samples to double our time per decision), we are best off taking just one sample, and making a guess accordingly. Thus, again, when making continuous decisions, it seems that often the best course of action is to make many quick, imperfect decisions to maximize long-run rewards.[8]

7. Strategic adjustment of sampling precision

Thus far, we have shown that under some assumptions, in cases when people try to maximize their expected rate of return, making decisions based on very few samples is actually optimal. However, based on our analysis, we expect that people would use more samples for decision that have higher stakes or are allotted more time—do people make these predicted, optimal adjustments?

A large prior literature on "probability matching" (Herrnstein, 1961; Vulkan, 2000) has studied a very similar phenomenon in a simpler task. In probability matching, participants predict the outcome of a trial based on the relative frequencies with which that outcome has been observed in the past. Thus, participants have direct evidence of the probability that

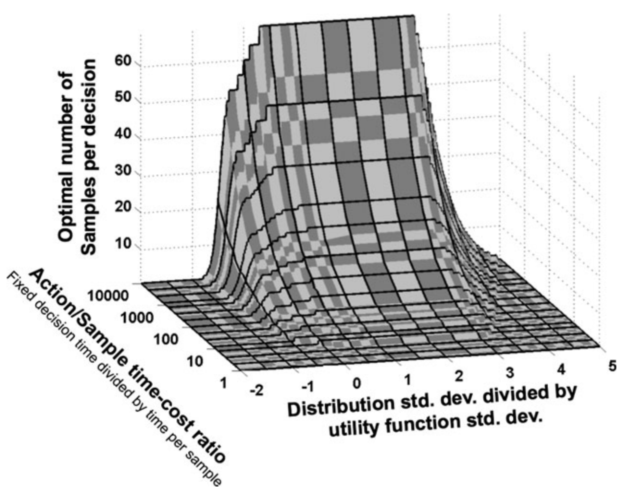

Fig. 13. Optimal number of samples for a sample-based agent as a function of the time cost of each additional sample and the ratio of the breadths of the utility function and of the posterior predictive distribution. There is a narrow range when the continuous decision is just difficult enough so as to warrant taking multiple samples (as many as 40 or 60 when samples are very quick, requiring 1,000 or 10,000 to halve the rate of decision). However, when samples are costly (requiring fewer than 10 to halve the decision rate), or in the infinitely wide range of decisions where the intrinsic uncertainty of the task is mismatched to the reward structure, making decisions based on just one sample is optimal.

lever A or lever B should be pulled, but they do not seem to maximize; instead, they "probability match" and choose levers with a frequency proportional to the probability of reward. On our account, this literal "probability matching" behavior amounts to making decisions based on one sample, whereas decisions based on more samples would correspond to Luces choice decisions (Luce, 1959) with an exponent greater than 1.

As probability matching contains a large body of experimental work, we can use this literature for a preliminary evaluation of a key question: Do people adjust the number of samples they use as key parameters of the decision process change? Shanks, Tunney, and McCarthy (2002) concluded that this is the case from a finding indicating that people tend to adopt an ideal maximizing strategy as more training and reward are provided. We can further test the effect of higher stakes on the apparent number of samples used to make a decision in a more graded fashion within the set of experimental findings reviewed by Vulkan (2000). Specifically, we computed the average stakes of the decisions and an estimate of the number of samples participants used to make those decisions for each of the studies reviewed in Vulkan (2000).

We measure the stakes of decisions as the difference in expected reward (in cents) between a probability matching decision and a maximizing decision. These studies vary in the probability of the alternative most likely to be "correct," p, the reward for a correct response, $u^{(+)}$, and the utility for an incorrect response, $u^{(-)}$. The expected *maximizing* reward for these studies is thus $U_* = pu^{(+)} + (1 - p)u^{(-)}$, and the expected *probability matching* reward is $U_m = (p^2 + (1 - p)^2)u^{(+)} + 2p(1 - p)u^{(-)}$. The quantity we are interested in—what we refer to as the stakes of the decision—is the advantage of maximizing over probability matching, or $U_\delta = U_* - U_m$: For studies where this number is higher, there is more to be gained by taking more samples.

The Luces choice rule describes the relationship between the probabilities of reward associated with various actions and the frequency with which agents choose these alternatives (see Eq. 4; Luce, 1959). If the frequency with which participants choose the option most likely to contain the reward is p_s, and the probability that the most likely option is rewarded is p_e, then the Luces choice odds ratio can be described as

$$\left(\frac{p_e}{1-p_e}\right)^L = \frac{p_s}{1-p_s}, \tag{4}$$

where L is the Luces choice exponent. Solving for L, we get:

$$L = \log\left(\frac{p_s}{1-p_s}\right) / \log\left(\frac{p_e}{1-p_e}\right). \tag{5}$$

With this expression we can measure the Luces choice exponent, L. On the assumption that agents make decisions by sampling, the Luces choice exponent yields a proxy for the number of samples used in a decision. If the agent uses a SPRT decision policy (see Appendix A), the Luces choice exponent corresponds to the SPRT threshold, and the expected number of samples per decision for that threshold is given by Feller (1966) as

$$\mathbb{E}[k] = \frac{L}{1-2p} - \frac{2L}{1-2p}\frac{1-(1-p/p)^T}{1-(1-p)/p)^{2T}}. \tag{6}$$

Thus, we computed the stakes and a proxy for the number of samples used in each of the 25 studies reviewed by Vulkan (2000) that tested probability matching with utility functions defined in terms of correct and incorrect answers, and we can assess how the effective number of samples per decision ($\mathbb{E}[k]$) varies as a function of the stakes of a decision (U_δ). We measure the correlation between the logarithms of these two quantities[9] in Fig. 14. Our prediction is that when the stakes are higher (that is, when the difference in expected rewards between the maximizing and probability matching response strategies is large), participants would use more samples for each decision, and thus would show a higher Luces choice exponent. This is precisely what we find—the stakes and the effective number of samples are positively correlated: $r(25) = .61$, $p = .0013$, 95% confidence interval on the correlation coefficient: (.28, .81). Thus, despite all the other variation across studies, laboratories, and so on, when stakes are higher, people are closer to maximizing—they seem to use more samples per decision when it matters more. We suspect that using rational algorithmic analyses of meta-cognitive decisions in this manner may be a fruitful way to systematically derive satisficing heuristics—decision policies that, although explicitly suboptimal, yield near-optimal behavior under cognitive processing constraints, thus offering a formal approach to "bounded rationality" (Simon, 1956).

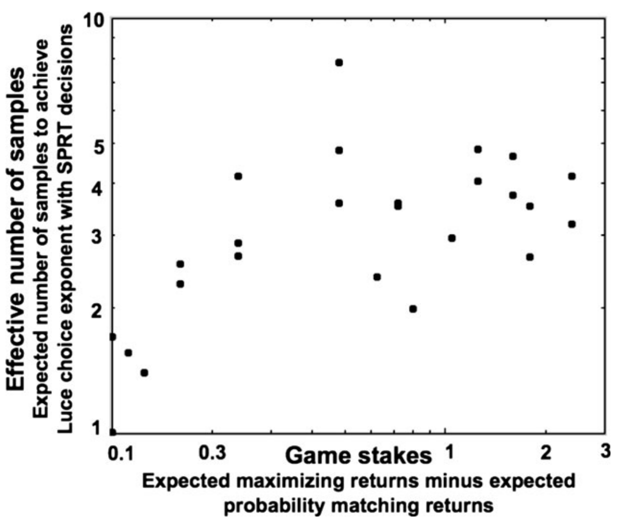

Fig. 14. Effective number of samples (based on the Luces choice exponent evident in human choices) as a function of the reward structure (the expected reward from *maximizing* decisions minus the expected reward (in U.S. cents) from *probability matching* decisions. Because both quantities are bounded at 0, we plot their logarithms against each other. Each data point corresponds to one study as surveyed by Vulkan (2000)—despite all the extraneous variation between studies, there is a significant correlation: $r(25) = .61, p = .0013$.

8. Discussion

We began with the observation that, on average, people tend to act consistently with ideal Bayesian inference, integrating information to optimally build models of the world; however, locally, they appear to be systematically suboptimal, acting based on a very limited number of samples. This has been used to argue that people are not exactly Bayesian (Mozer et al., 2008). Instead, we have argued that sample-based approximations are a powerful method for implementing approximate Bayesian inference. Although with few samples, sample-based inferences will deviate from *exact* Bayesian inference, we showed that for choices among 2, 4, 8, 16, and 32 discrete alternatives and for unidimensional continuous choices, a decision based on a very small set of samples is nearly as good as an optimal decision based on a full probability distribution. Moreover, we showed that given reasonable assumptions about the time it takes to produce an exact sample, a policy of making decisions based on very few samples (even just one) is globally optimal, maximizing long-run utility for choices among discrete alternatives as well as choices along continuous variables. Furthermore, our analysis predicts that when the stakes are higher, participants should use more samples for a decision, and we found evidence of such optimal meta-cognition in a meta-analysis of the probability matching literature.

8.1. Related arguments

Other authors have invoked various kinds of sampling as a way to explain human decision making. Stewart, Chater, and Brown (2006) suggested that a policy of making decisions through binary preference judgments among alternatives sampled from memory can account for an assortment of human judgment and decision-making errors. Schneider, Oppenheimer, and Detre (2007) suggest that votes from sampled orientations in multidimensional preference space can account for violations of coherent normative utility judgments. A promising direction for future research would be to relate models like these, based on samples drawn from memory or over preferences, to models like those we have described in our study, in which samples are drawn from probability distributions reflecting ideal inferences about the world.

Another related argument comes from proponents of "bounded rationality" (Gigerenzer, 2008; Simon, 1956): Given cognitive limitations, strict optimization is impractical if not impossible, and instead people may adopt heuristics that are sufficient for good performance. On this view, people do not optimize their behavior given the presented reward structure, but instead they "satisfice" to make approximate decisions that respect cognitive constraints. Making decisions based on one or just a few samples is an example of such a satisficing heuristic—although under such a decision policy the agent does not optimize any one decision, the policy yields behavior that is actually optimal in the context of the limitations of the agent. We are optimistic that by explicitly considering meta-cognitive optimization, one might find such heuristics that are not optimally rational, but are optimal under the constraints of bounded rationality.

8.2. *Internal samples versus external information gathering and decisions from experience*

It is important to distinguish between internal sampling from the posterior, and external sampling of data from the world. Formalizing external information gathering as sampling data from the world is a useful tool for analyzing how much information people gather from the world before making a decision. In contrast, internal sampling, as we have described here, is a strategy for approximate inference and a candidate general description of effortful "thought" to solve a problem, with the goal of asking how much do people "think" before making a decision. Although these accounts intend to explain different bodies of work, they share a close formal relationship, so similar optimality considerations ought to apply.

Whether or not people are exactly optimal, one thing is clear: When sampling data from the world, people are inclined to gather little evidence, and make decisions quickly, rather than spend a lot of time "sampling" evidence (Hertwig & Erev, 2009; Hertwig & Pleskac, 2010). Thus, it seems that whether decisions are being informed by external information gathering or internal deliberation, the trade-off between making quick, less-informed decisions and slow, more informed decisions is similar. In both cases, people seem to choose a globally optimal policy of using few samples to make quick decisions.

8.3. *Optimality in sequential sampling models*

Sequential sampling models (Ratcliff & Smith, 2004) based on the SPRT (Wald, 1947), drift diffusion (Gold & Shadlen, 2000; Ratcliff, 1978), or accumulators (Brown & Heathcote, 2008; Vickers, 1979) have been applied not only to sampling data from the environment but also to more internal processes, such as memory (e.g., Ratcliff, 1978). Although these models have not been formulated as an approximate inference strategy for carrying out Bayesian inference, they share a number of similarities without analysis (as described in Figs. 8 and 9 and Appendix A).

Models based on the SPRT are optimal in that they can achieve a given level of confidence using the least amount of data—in other words, using fewest samples. Sequential sampling models have mostly focused on characterizing the diffusion/accumulation process that best captures human response time distributions. Our analysis focuses on the choice of optimal thresholds for such models (see also Bogacz, 2007), and the effective number of samples that they use under an optimal policy. While the emphasis of our study differs from that of the bulk of the sequential sampling literature, our results are fundamentally consistent with those modeling approaches. Indeed, when using sequential sampling models to describe how people gather information about the world, it has been suggested that people adopt nearly optimal decision criteria (so as to maximize the rate of return Bogacz, 2007; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Gold &

Shadlen, 2002)—decision criteria that amount to using very few samples, and making many quick, locally suboptimal decisions.

8.4. What is a sample?

What is a sample as we consider it, and what does it take to produce such a sample? There are two important points to be made here. First, to pre-empt an objection from experts in sampling who know that one typically needs many samples for inference, we want to clarify that a sample, as we consider it, is an *exact, independent* sample—that is, a sample from the actual posterior probability distribution. Various approximate sampling schemes, such as *Gibbs* sampling (Geman & Geman, 1984), MCMC (Robert & Casella, 2004), importance sampling (Srinivasan, 2002), or particle filters (Doucet et al., 2001), produce correlated samples. For instance, to produce an exact sample from a MCMC algorithm one must run the algorithm for a fairly long "burn-in" period—in effect, what we consider one sample would require many MCMC iterations—and even after the burn-in period, subsequent samples are still correlated. All these approximate sampling methods are associated with schemes for estimating the effective sample size. In MCMC, this amounts to the auto-correlation of the sampling chain; in importance sampling and particle filters, this is computed from the variance of the importance weights. We expect that these schemes for estimating the effective sample size will yield numbers that can link these more sophisticated sampling algorithms to the analyses we present in this study. Indeed, recent arguments suggest that when there is a cost to each MCMC sample, very few samples (yielding correlation and bias) are optimal (Lieder, Griffiths, & Goodman, 2012).

Second, where does deductive, procedural thought come into play if we cast thinking in terms of sampling? Here, we want to clarify that the process for producing one exact sample from a complex, structured model of the world will often require much deductive, procedural calculation. Thus, although we cast the output of this processing as a "sample," the processing itself will often contain many deterministic calculations. For these reasons—that a sample is actually the output of potentially very complicated calculations—we believe that the process of producing one such exact sample is likely to be a rather slow process.

8.5. Sample cost

How much might a sample "cost"? In our analyses the relevant measure of sample cost in multiple-trial experiments is the ratio between the time it takes to make an action and go on to the next trial and the time required to draw a sample to inform a decision about that action—a measure of how much a sample will slow down the rate of decision making. Ratios near 10 seem quite reasonable: Most experimental trials last a few seconds, and it can arguably cost a few hundred milliseconds to consider a hypothesis. This is speculation; however, it seems to us that in most experimental tasks, the benefits gained from a better decision are relatively small compared with the costs of spending a very long time thinking. So, if thinking amounts to sampling possible alternatives before making a decision, it should not be surprising that people regularly seem to use so few samples.

For the sake of parsimony we have only focused on the time cost of samples, and we have assumed that each sample is an exact sample from the appropriate posterior. However, these assumptions may not be valid, people may (and are even likely to) face costs over and above opportunity, to generating a sample, for instance, energy, metabolism, or even "effort."

Moreover, the value of a sample may be substantially smaller than we estimated: The world changes, and our model of the world may be only approximate. If the world changes while we sample for a decision (again, on the assumption that obtaining a sample takes some time), then not only do additional samples slow the rate of decision making but they also render the eventual decision progressively less matched to the ongoing state of the world. Furthermore, even if the world does not change, but we assume that our model of the world is not perfectly calibrated, then the asymptotic performance (achieved with infinitely many samples) will be worse than we assumed. Insofar as either of these conditions applies, the value of a sample for a decision will be substantially lower than we had assumed in our calculations.

We have ignored these considerations from this article because both greater costs and smaller gains from a sample would produce incentives to use *fewer* samples for a decision. As we have argued that using just a few samples for each of many decisions may be an optimal policy, this conclusion holds with even more strength if the expected gains from a sample are lower.

8.6. *Reusing samples for multiple decisions*

Thus far, we have considered scenarios where a sample-based agent takes a new set of samples for each decision being made. This makes sense in many laboratory tasks where each trial involves a new set of information on which to base a decision; however, many real-world tasks require making many decisions based on the same information. In these scenarios, it makes sense for an agent to cache and reuse samples for several decisions. When samples correspond to a more abstract hypothesis space (e.g., a general categorization rule), those samples offer greater opportunities for reuse, whereas specific samples (e.g., a particular world state or action) cannot be reused. Thus, when more abstract hypothesis spaces are considered, the time cost per sample is effectively reused because many more decisions can use each sample obtained. We imagine that in such cases people would use more samples than they would if only one choice was contingent on their deliberation. Such scenarios result in an inherent difficulty in measuring how many samples went into a decision, and whether people are indeed sampling from the appropriate posterior.

First, if people reuse samples across a number of decisions, then their decisions will be correlated, and will reflect a narrower range of alternatives than sampling from the appropriate posterior will predict. This dependency sometimes yields crucial evidence that people are using a few samples for a decision, as in Goodman et al. (2008), where generalization behavior across multiple trials is most consistent with just a few sampled rules being reused on every trial (see Fig. 1). However, sometimes this dependency will yield ambiguities about whether individual participants make judgments based on the full posterior by reusing a few samples, or whether individual participants have large idiosyncratic biases about the posterior in question (in the limiting case, basing predictions of cake baking times based on only a few experienced cakes; Mozer et al., 2008). This phenomenon can be well illustrated by human judgments of the prices of individual goods: We asked people to make several estimates about the price of product categories (e.g., a diamond ring); while the across-subject dispersion in mean estimates tracked the dispersion of prices in the real world (estimated from Google products), the dispersion of guesses within an individual was substantially lower (Fig. 15). In such data, it is impossible to assess whether individual participants simply produce correlated guesses because they are reusing samples or because they have biased, impoverished beliefs.

There is some reason to suspect that in many cases these correlations arise from participants reusing samples, or obtaining correlated samples, rather than from individual

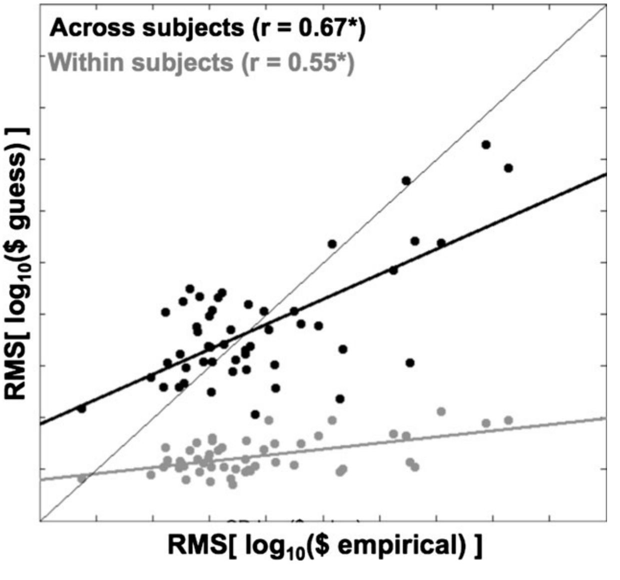

Fig. 15. Human judgments of prices of various products—each participant made multiple guesses about the price of a product category (e.g., diamond rings; snowboards, etc.). For each product we plot the within-subject dispersion of the price setting (gray; the average within-subject standard deviation) and the across-subject dispersion (black; standard deviation across subjects of the average guessed price for each participant). Although both within- and across-subject dispersion are well correlated with the empirical dispersion of product prices, within-subject dispersion is objectively much smaller than the empirical dispersion, whereas across-subject dispersion tracks empirical dispersion more closely. These results could be attributed either to systematic reuse of sampled prices by individuals across multiple trials, or to systematic biases in beliefs across participants (or some combination). We argue based on other work that systematic reuse of samples is likely to play a larger role.

biases. First, correlations between two guesses from one individual decrease as more time passes between guesses (Vul & Pashler, 2008), suggesting that individual biases dissipate as participants forget their previous estimates or samples. Second, even when an additional guess about the same problem is obtained without an additional delay, the two guesses are less correlated for individuals with lower working memory capacity, suggesting that the dependency between multiple guesses does not arise from individual biases, but from an ability to remember and reuse multiple samples across decisions (Hourihan & Benjamin, 2010). Finally, when task constraints prevent participants from reusing samples across multiple guesses, those guesses appear to be independently sampled from an identical distribution (Vul et al., 2009). Together, these results suggest that stable individual variability across judgments arises not from idiosyncratic deviations in the posterior, but from systematic reuse or correlation of samples from the posterior across a number of trials, when such reuse is possible.

8.7. *Assumption of a uniform prior over* p

Our analyses, particularly in the alternate forced-choice domains, have assumed that the sample-based agent assumes an uninformative prior about the problem—that is, the sample-based agent assumes a uniform prior over p in the Bernoulli case. On this assumption, our calculations of the optimal number of samples seem most robust and general; however, the optimal number of samples will vary if the structure of the problem confers a more informative prior. If we assume that we are in a nearly deterministic setting where p tends to extreme values (0 or 1), then the optimal number of samples will change: These cases guarantee that the first sample will be informative and the second sample will be redundant. On the other hand, if we assume we are dealing with a very random, unconstrained problem, where p tends to be around 0.5, then we know that all samples will be uninformative. If we think that p tends to be around 0.7—a regime where more samples payoff—then we would assume that we should use more samples. As assumptions about p will vary, the shape of the optimal number of samples as a function of sample cost will vary; however, only under very constrained conditions will the optimal number of samples be much higher than our analyses have described.

8.8. *Black Swans and variable utility functions*

What happens with variable utility functions? In our analysis we have assumed that utility functions are constant in nAFC decisions—one reward is assigned for a "correct" answer, and another for an "incorrect" answer. This assumption holds for the bulk of psychological experiments, and even for most methods of evaluating machine learning algorithms; however, it does not apply universally in the real world, where some outcomes are better than other positive outcomes. Although little about our results will change when such variation in utilities is small, it poses an interesting problem when this variation is large.

Take, for instance, the game of "Russian roulette," in which one bullet is placed within a six-shot revolver, and the drum is randomly spun; the player then aims the revolver at his or her head and pulls the trigger. In this game, there is a five in six chance that the current chamber does not contain the bullet, and pulling the trigger will cause no harm, and will confer a slight reward (the game is said to have been played by 19th-century Russian military officers to demonstrate their bravado to others). However, there is a one in six chance that the chamber does contain the bullet, in which case the loss is catastrophic (death). In these cases, and others, the sample-based agent that relies on few samples might not consider the state of the world in which the bullet is in the current chamber. Thus, the agent will not consider the relatively low probability of an extreme outcome. This is referred to as the "Black Swan" problem (Taleb, 2008): Ignoring very important but low probability events leads to substantial biases and irrationalities, which Taleb (2008) argues exist in finance.

One possible method that sample-based agents may adopt to avoid the Black Swan problem is increasing the sampling rate of high-stakes scenarios. For instance, instead of sampling from just the posterior probability distribution over possible world states, one might weight the samples by the variance of possible outcomes in that state. Using this modified sampling strategy, world states in which decisions are particularly high stakes will be over-represented relative to their probability but will allow the agent to compute the expected utility of a particular action in a more useful manner. Such a sampling scheme predicts some forms of availability effects (Tversky & Kahneman, 1974)—mental over-representation of the possibility of events with extreme outcomes. It will be an interesting direction for future research to assess how availability may be used to overcome the Black Swan problem for sample-based agents and whether this sampling strategy underlies human decision-making biases.

8.9. Limitations

We should emphasize that we are not arguing that all human actions and decisions are based on very few samples. The evidence for sampling-based decisions arises in high-level cognition when people make a decision or a choice based on what they think is likely to be true (Which example is in the concept? How long will this event last? How many airports are there in the United States?). In other situations people appear to integrate over the posterior, or to take many more samples, such as when people make graded inductive judgments (How similar is A to B? How likely is it that X has property P given that Y does? How likely do you think that F causes G?). Moreover, in low-level sensory and motor tasks, decisions often seem to be much closer to ideal Bayesian performance, rather than decisions based on few samples, as seen in cognition (Trommershauser, Maloney, & Landy, 2003, although see Battaglia & Schrater, 2007). It is interesting to consider why there may be a difference between these sorts of decisions and tasks.

8.10. Conclusion

Under reasonable discrete and continuous choice scenarios, people are best advised to make decisions based on few samples. This captures a very sensible intuition: When we are deciding whether to turn left or right at an intersection, we should not enumerate every possible map of the world. We do not need “moral certainty” about the probability that left or right will lead to the fastest route to our destination—we just need to make a decision. We must implicitly weigh the benefits of improving our decision by thinking for a longer period of time against the cost of spending more time and effort deliberating. Intuition suggests that we do this in the real world: We think harder before deciding whether to go north or south on an interstate (where a wrong decision can lead to a detour of many miles), than when we are looking for a house (where the wrong decision will have minimal cost). Indeed, empirical evidence confirms this: when the stakes are high, people start maximizing instead of “probability matching” (Shanks et al., 2002), and we show that they do so in a graded fashion as stakes increase. Nonetheless, it seems that in simple circumstances, deliberating is rarely the prudent course of action—for the most part, making quick, locally suboptimal, decisions is the globally optimal policy: one (or a few) and done.

Acknowledgments

This work was supported by an ONR MURI grant N00014-07-1-0937 (JBT, EV); BIAL, NSF Dissertation grants, and NDSEG fellowship (EV); and grant FA9550-07-1-0351 from the Air Force Office of Scientific Research (TLG). EV and JBT were supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract D10PC20023. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government. We thank Frederick Eberhardt, David Danks, and Vikash Mansinghka for helpful discussions. A preliminary version of this work was presented at the 31st annual Cognitive Science Society meeting (Vul, Goodman, Griffiths, & Tenenbaum, 2009).

Notes

1. The Monte Carlo theorem states that the expectation over a probability distribution can be approximated from samples:

$$E_{P(S)}[f(S)] \simeq \frac{1}{k}\sum_{i=1}^{k} f(S_i), \quad \text{when} S_i \sim P(S). \tag{7}$$

2. The analysis becomes more subtle when the utility structure is more complex. We return to this point later.
3. With an even number of samples, there is the possibility of a tie, which this equation does not handle. A more general equation for q is as follows:

$q = \sum_{x=0}^{k} C(x) \times \Theta(x, p, k)$ where x is the number of samples favoring the optimal choice; $C(x)$ is the choice function (0 when $x < 0.5*k$, 1 when $x > 0.4*k$, and 0.5 when $x = 0.5*k$); and $\Theta()$ is the Binomial probability mass function. This more general equation appropriately handles ties, but it yields the same answer for odd numbers of samples as the simpler equation in the main text. Critically, using this more general formulation does not change any of our results—even numbers of samples are never optimal; thus, we only consider odd numbers of samples here and use the simple equation.

4. Bernoulli considered *moral certainty* to be at least 1,000:1 odds that the true ratio will be within $\frac{1}{50}$ of the measured ratio.
5. As an example, consider a situation with a payoff structure outside these bounds, where a correct answer yields a gain of 1, and an incorrect answer yields a loss of -2. In this case, if p is between 0.5 and $\frac{2}{3}$, the expected return per choice is negative, even for the ideal observer; thus, the ideal observer would try to *minimize* the number of choices made per unit time—the ideal observer should stall.
6. With punishment for an incorrect decision among many alternatives, in many situations the expected reward may be negative, rather than positive, in which case the optimal agent will try to *minimize* the number of decisions made per second—we avoid this degenerate scenario by not considering punishment.
7. Squared distance is relevant for an L2 loss function, but such loss functions are rarely (if ever) relevant for real-world decisions.
8. Here, we consider an infinite possible range of the continuous decision variable—for our purposes, what is required is that the range of the variable is considerably larger than σ_U and σ_P. In cases where the range is close to either of these standard deviations, optimal policies will employ fewer samples (perhaps 0).
9. We used logarithms because both quantities are effectively bounded at zero and are not normally distributed otherwise.
10. In practice, the approximation

$$\mathbb{E}[U/t|T,c] = \int_{0.5}^{1} \mathrm{d}p\mathcal{P}(p)\frac{\mathbb{E}[U|T,p]}{\mathbb{E}[k|T,p]/c+1} \tag{8}$$

yields very similar answers, and identical optimal thresholds within the set of cs we considered.

References

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Battaglia, P., & Schrater, P. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *Journal of Neuroscience*, *27*, 6984–6994.

Berger, J. (1985). *Statistical decision theory and bayesian analysis*. New York: Springer.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behavior. *Trends in Cognitive Sciences*, *11*, 118–125.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & 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*, 700–765.

Brainard, D., & Freeman, W. (1997). Bayesian color constancy. *Journal of the Optical Society of America*, *14*(7), 1393–1411.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Brown, S., & Steyvers, M. (2008). Detecting and predicting changes. *Cognitive Psychology*, *58*, 49–67.

Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462.

Chater, N., & Manning, C. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344.

Daw, N., & Courville, A. (2008). The rat as particle filter. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*. Vol. *20* (pp. 369*–376). Cambridge, MA: MIT Press.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*(3), 285–399.

Doucet, A., De Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods*. New York: Springer.

Feller, (1966). *An introduction to probability theory and its applications*. New York: Wiley.

Gamerman, D., & Lopes, H.(2006). *Markov chain Monte Carlo*. New York: Chapman and Hall.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. New York: Chapman and Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Gershman, S., Vul, E., & Tenenbaum, J. (2012). Multistability and perceptual inference. *Neural Computation*, *24*(1), 1–24.

Gigerenzer, G. (2008). Rationality for mortals: How people cope with uncertainty. Oxford University Press, USA.

Gold, J., & Shadlen, M. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390–394.

Gold, J., & Shadlen, M. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*, 299–308.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.

Griffiths, T., & Tenenbaum, J. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354–384.

Griffiths, T., & Tenenbaum, J. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*, 767–773.

Herrnstein, R. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Science*, *13*(12), 517–523.

Hertwig, R., & Pleskac, T. (2010). Decisions from experience: Why small samples? *Cognition, 115,* 225–237.

Hick, W. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.

Hourihan, K., & Benjamin, A. (2010). Smaller is better (when sampling from the crowd within): Low memory-span individuals benefit more from multiple opportunities for estimation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 1068–1074.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*, 188–196.

Jones, M., & Love, B. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188.

Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

Kording, K. (2007). Decision theory: What should the nervous system do? *Science, 318*(5850), 606.

Kwisthout, J., Wareham, T., & van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, *1*, 1–6.

Levy, R., Reali, F., & Griffiths, T. (2009). Modeling the effects of memory on human online sentence processing with particle filters. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*, Vol. *21* (pp. 937–944). Cambridge, MA: MIT Press.

Lewandowsky, S., Griffiths, T., & Kalish, M. (2009). The wisdom of individuals: Exploring peoples knowledge about everyday events using iterated learning. *Cognitive Science*, *33*, 969–998.

Lieder, F., Griffiths, T. L., & Goodman, N. D. (2012). Burn-in, bias, and the rationality of anchoring. In P. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*. Vol. *25* (pp. 2699–2707). Cambridge, MA: MIT Press.

Luce, R. (1959). *Individual choice behavior*. New York: Wiley.

Maloney, L. (2002). Statistical decision theory and biological vision. In D. Heyer & R. Mausfield (Eds.), *Perception and the physical world* (pp. 145–189). New York, NY: Wiley.

Maloney, L., Trommershauser, J., & Landy, M. (2007). Questions without words: A comparison between decision making under risk and movement planning under risk. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 297–313). New York, NY: Oxford University Press.

Marr, D. (1982). *Vision.* Cambridge, MA: MIT Press.

McKenzie, C. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, *26*, 209–239. Available at: http://psy.ucsd.edu/mckenzie/McKenzie1994CogPsych.pdf

Mozer, M., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science, 32,* 1133–1147.

Navarro, D. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, *51*, 85–98.

Nosofsky, R., Palmeri, T., & McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.

Ratcliff, & Smith. (2004). A comparison of sequential-sampling models for two choice reaction time. *Psychological Review*, *111*, 333–367.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Robert, C., & Casella, G. (2004). *Monte Carlo statistical methods*. New York: Springer.

Sanborn, A. N., & Griffiths, T. L. (2008). Markov chain Monte Carlo with people. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*. Vol. *20* (pp. 1265–1272). Cambridge, MA: MIT Press.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. (2006). A more rational model of categorization. In R. Sun, & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 726–731). Vancouver, BC.

Schneider, A., Oppenheimer, D., & Detre, G. (2007). Application of voting geometry to multialternative choice. In D. McNamara, & G. Trafton (Eds.),*Proceedings of the 29th annual conference of the cognitive science society* (pp. 635–640). Nashville, TN.

Shanks, D., Tunney, R., & McCarthy, J. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*, 233–250.

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, *17*(4), 443–464.

Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Sobel, D., Tenenbaum, J., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303–333.

Srinivasan, R. (2002). *Importance sampling - applications in communications and detection*. Berlin: Springer-Verlag.

Stewart, N., Chater, N., & Brown, G. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26.

Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.

Sutter, R., & Barlo, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Taleb, N. (2008). *The black swan*. New York: Random House.

Trommershauser, J., Maloney, L., & Landy, M. (2003). Statistical decision theory and rapid, goal-directed movements. *Journal of the Optical Society A*, 1419–1433.

Tversky, A., & Kahneman, D. (1974). Judgments under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Usher, M., & McClelland, J. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Usher, M., Olami, & McClelland, J. (2002). Hick's law in a stochastic race model with speed-accuracy tradeoff. *Journal of Mathematical Psychology*, *46*, 704–715.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Vul, E., Goodman, N., Griffiths, T., & Tenenbaum, J. (2009). One and done? Optimal decisions from few samples. In N. Taatgen, H. van Rijn, L. Schomaker, & J. Nerbonne (Eds.), *Proceedings of the 31st annual conference of the cognitive science society*. Amsterdam, the Netherlands.

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all or none samples. *Journal of Experimental Psychology: General*, *138*(4), 546–560.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.

Vul, E., & Rich, A. (2010). Independent sampling of features enables conscious perception of bound objects. *Psychological Science*, *21*(8), 1168–1175.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*(1), 101–118.

Wald. (1947). *Sequential analysis*. New York: Wiley.

Xu, F., & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288–297.

Yuille, A., & Bülthoff, H. (1996). Bayesian decision theory and psychophysics. In D. Knill & W. RIchards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge, MA: MIT Press.

Appendix A: SPRT and accumulator agents for 2AFC decisions

A.1. Optimal thresholds for an SPRT policy

Under the sequential probability ratio test (SPRT) policy (Wald, 1947), the sample-based agent calculates $d_k = x_1 - x_2$: The difference in number of samples favoring choice A_1 and A_2, and choose the corresponding action when $|d_k|$ reaches threshold T. SPRT will choose action A_1 with probability

$$q_{T,\,p} = \frac{p^T}{p^T + (1-p)^T} \tag{9}$$

after a variable number of samples k, yielding probability of correct choice $w^{(+)}_{T,p}+ = pq_{T,p} + (1-p)(1-q_{T,p})$. The expected utility per decision is thus

$$\mathbb{E}[U|T,\,p] = (w^{(+)}_{T,p})u^{(+)} + (1 - w^{(+)}_{T,p})u^{(-)}, \tag{10}$$

where $u^{(+)}$ is the reward for a correct answer and $u^{(-)}$ is the reward (more likely punishment) for an incorrect answer.

The probability distribution of the number of samples (k) for an SPRT policy to reach a decision given threshold T and p is given by (Feller, 1966, ch. XIV, eq. 5.7):

$$\mathcal{P}(k|T,p) = \left(\frac{2^k}{2T}\right)p^{\frac{k-T}{2}}(1-p)^{\frac{k+T}{2}}\sum_{v=1}^{2T-1}\cos\left(\frac{v\pi}{2T}\right)^{k-1}\sin\left(\frac{v\pi}{2T}\right)\sin\left(\frac{Tv\pi}{2T}\right) \tag{11}$$

for all k that are of the same parity as T and $k \geq T$ (otherwise, $\mathcal{P}(k|T,p) = 0$); $\mathcal{P}(k = 1|T = 1,p) = 1$, and $\mathcal{P}(k = 0|T = 0,p) = 1$. The average number of samples can be calculated more easily via (Feller, 1966, ch. XIV, eq. 3.4):

$$\mathbb{E}[k|T,p] = \frac{T}{1-2p} - \left(\frac{2T}{1-2p}\right)\frac{1-\left(\frac{1-p}{p}\right)^T}{1-\left(\frac{1-p}{p}\right)^{2T}}. \tag{12}$$

The expected rate of return (in arbitrary time units) for a particular threshold (T) given a utility function (specifying $u^{(+)}$ and $u^{(-)}$) and a prior on p ($\mathcal{P}(p)$), and an action/sample cost ratio (c) is

$$\mathbb{E}[U/t|T,c] = \int_{0.5}^{1} \mathrm{d}p\mathcal{P}(p)\sum_{x=0}^{\infty}\frac{\mathbb{E}[U|T,p]}{(2x+T)/c+1}\mathcal{P}(2x+T|T,p) \tag{13}$$

(note that we need only integrate over $p \in [0.5,1.0]$ due to symmetry).[10] Thus, the optimal threshold (T^*_c) that maximizes the rate of return for a given action/sample time cost (c) is

$$T^*_c = \arg\max_T \mathbb{E}[U/t|T,c]. \tag{14}$$

The optimal thresholds, expected rates of return ($\mathbb{E}[U/t|T^*_c,c]$) and expected number of samples per decision ($\mathbb{E}[k|T^*_C,c] = \int_{0.5}^{1}\mathrm{d}p\mathcal{P}(p)\mathbb{E}[k|T^*_c,p]$) are shown in Fig. 9.

A.2. Optimal thresholds for an accumulator policy

Under the accumulator policy (Vickers, 1979), a sample-based agent obtains samples until either x_1 or x_2 (the number of samples in favor of A_1 and A_2, respectively) reaches T, and then chooses the corresponding action. The accumulator policy guarantees neither a fixed level of confidence per threshold nor a fixed number of samples. The probability distribution of the number of samples (k) before a decision is made under an accumulator policy with threshold T and probability p is

$$\mathcal{P}(k|T,p) = \left[\binom{k}{T} - \binom{k-1}{T}\right]\left[p^T(1-p)^{k-T} + (1-p)^T p^{k-T}\right], \tag{15}$$

for $T \leq k < 2T$, where $\binom{n}{m}$ is the binomial coefficient, defined to be 0 when $n < m$.

The probability of choosing option A_1 given T, p, and k is

$$q_{k,T,p} = \frac{p^T(1-p)^{k-T}}{p^T(1-p)^{k-T} + (1-p)^T p^{k-T}} \tag{16}$$

for $k \geq T > 0$, and $q_{k,T,p} = 0.5$ if $T = 0$. From this we can calculate the probability of correct choice as: $w^{(+)}_{k,T,p} + = pq_{k,T,p} + (1-p)(1-q_{k,T,p})$, and the expected utility per decision as $\mathbb{E}[U|k,T,p] = (w^{(+)}_{k,T,p})u^{(+)} + (1 - w^{(+)}_{k,T,p})u^{(-)}$.

The expected rate of return is therefore

$$\mathbb{E}[U/t|T,c] = \int_{0.5}^{1} \mathcal{P}(p)\mathrm{d}p \sum_{k=T}^{2T-1} \frac{\mathbb{E}[U|k,T,p]}{k/c+1}\mathcal{P}(k|T,p). \tag{17}$$

Yielding the optimal accumulator threshold $T^*_c = \arg\max_T \mathbb{E}[U/t|T,c]$. The optimal thresholds, expected rates of return ($\mathbb{E}[U/t|T^*_c,c]$), and expected number of samples per decision ($\mathbb{E}[k|T^*_C,c] = \int_{0.5}^{1}\mathcal{P}(p)\mathrm{d}p\sum_{k=T}^{2T-1} k\mathcal{P}(k|T,p)$) are shown in Fig. 9.