

Sampling in Human Cognition

Edward Vul

Massachusetts Institute of Technology; Dept. of Brain and Cognitive Sciences

Abstract

How do people reason from data to choose actions in novel situations? There is considerable flexibility in the language we can adopt to describe such acts of cognition, and different fields of cognitive science have used different levels of description to characterize subsets of cognitive phenomena. Recent developments in these fields have highlighted tensions between process-level descriptions from cognitive psychology and computational-level descriptions from computer science and machine learning: How can rational probabilistic inference be carried out in the human mind given its processing limitations? In my dissertation I argue that this tension can be resolved by explaining cognition and decision-making as sample-based approximate inference: instead of considering complete probability distributions, people entertain only a few hypotheses randomly considered with frequency proportional to their probability. I show in several cognitive and visual tasks that human perception and decisions follow the predictions of a sample-based approximate inference engine. Moreover, models of cognition based on specific sampling algorithms can describe previously elusive cognitive phenomena such as perceptual bistability and probability matching. Altogether, the sampling hypothesis unites the probabilistic modeling approaches from computer science with processing constraints from cognitive psychology and connects to several recently proposed neural implementations of complex reasoning.

A unified account of human cognition will yield understanding at least at three levels (Marr, 1982): *Computation* – what information is used to solve a problem and how is this information combined? *Process* – how is information represented and what procedures are used to combine the representations? *Implementation* – how are these representations and procedures implemented in the brain?

Subfields of cognitive science typically focus on a single level without considering connections to the others. Recent developments in computer science and artificial intelligence have produced impressive models for drawing sophisticated inferences from the sparse information available to humans, without considering how these computations can be carried out by the human brain. Cognitive psychologists, have characterized the failures of human cognition and defined the boundaries our processing capacities, without considering the computational goals of the algorithms or their neural constraints. Neuroscientists have probed in increasingly fine detail the implementation level – the physical hardware under-

lying cognition – without considering the processes and computations that the hardware has evolved to carry out. These fields have made tremendous progress without exploring interdisciplinary connections to bridge levels; however, recently progress has been hampered by the absence of interdisciplinary, cross-level research to connect the levels.

My dissertation takes on one such interdisciplinary challenge: connecting computational, probabilistic analyses of human cognition with the processing constraints known from cognitive psychology. I suggest that the seemingly intractable computational problems proposed by rational statistical analyses of cognition from computer science can be realistically approximated given processing constraints from cognitive psychology via *sampling* approximations. Testing this interdisciplinary theory requires methods from different subfields of cognitive science, so I have approached the problem from three angles: (1) Psychophysical experiments in visual perception (ch. 3-4), (2) Behavioral experiments on high level cognition (ch. 5), and (3) defining formal models of the connection between computation and process (ch. 2). The work in this dissertation paves the way for a Bayesian cognitive architecture, which synthesizes processing constraints from cognitive psychology with probabilistic computational considerations (Ch. 7).

The challenge

Bayesian inference and *decision theory* describe theoretically optimal computations for combining different sources of uncertain information to build structured models of the world and for using these models to plan actions and make decisions in novel situations. In recent years, this framework has become a popular and successful tool for describing the computations people must carry out to accomplish perceptual (Knill & Richards, 1996), motor (Maloney, Trommershauser, & Landy, 2007), memory (Anderson & Milson, 1989), and cognitive (Chater & Manning, 2006; McKenzie, 1994; Griffiths & Tenenbaum, 2005; Goodman, Tenenbaum, Feldman, & Griffiths, 2008) tasks both in the lab and in the real world; thus supporting the claim the Bayesian inference provides a promising description of *rational models* (Anderson, 1990) of cognition at the computational level. However, as Bayesian rational analysis gains ground as a computational description, several salient challenges have hampered progress, indicating that important constraints at the process level must be taken into account to accurately describe human cognition.

Challenges to a purely computational view

First, exact Bayesian calculations are practically (and often theoretically) impossible because computing the exact Bayesian answer frequently requires evaluating and integrating innumerable large hypothesis spaces. This is true of even small-scale inferences in artificial problems (for instance – possible sentence parses in a probabilistic context-free grammar (Charniak, 1995)). Thus, applications of Bayesian inference in machine learning have relied on approximate inference methods. Since exact Bayesian inference is usually intractable even for small artificial problems, the problem must be even more severe for the large-scale real-world problems that the human mind faces every day. How can the brain do Bayesian inference at a real-world scale?

Second, cognitive processing limitations pose an additional practical challenge to implementing approximate statistical inference in humans. Human cognition is limited in

memory (Wixted & Ebbesen, 1991; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), processing speed (Shepard & Metzler, 1971; Welford, 1952), and attention (Pashler, 1984; Broadbent, 1958; Treisman & Gelade, 1980; James, 1890), while people must make split-second decisions. Adequate approximations of Bayesian inference rely on millions of complex calculations (Robert & Casella, 2004). What procedures can people use to approximate statistical inferences in real-world decisions within a fraction of a second, despite their limited cognitive resources?

Third, although people seem to be Bayesian in many cognitive domains on the average over many trials or subjects, individuals on individual trials are often not optimal. Goodman et al. (2008) showed that optimal average Bayesian rule-learning behavior emerges from aggregating over many subjects, each of whom learns just one rule. Similarly, Griffiths and Tenenbaum (2006) demonstrated that on average, people know the distribution of quantities of the world, but individual responses reflect knowledge of only a small subset (Mozer, Pashler, & Homaei, 2008). These results suggest that average behavior reflects optimal actions that take into account complete probability distributions, while individual behavior does not reflect such fully probabilistic beliefs. What cognitive processes could produce optimal behavior on the average of many suboptimal trials?

Fourth, the characteristic dynamics of cognition have historically intrigued cognitive psychologists and highlighted the need for a process-level description: People forget what they have learned (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), they overweight initial training (Deese & Kaufman, 1957), they solve problems (Vul & Pashler, 2007) and rotate mental images (Shepard & Metzler, 1971) slowly, and they stochastically switch between interpretations when exposed to ambiguous stimuli (Alais & Blake, 2005). Such dynamics of human cognition are outside the scope of purely computational analysis, and require a process-level description.

The sampling hypothesis

In my dissertation, I suggest a resolution to all of these challenges: The mind approximates Bayesian inference by *sampling*.

Sampling algorithms represent probabilistic beliefs by considering small sets of hypotheses randomly selected with frequency proportional to their probability. Sampling can approximate probability distributions over large hypothesis spaces despite limited resources. Sampling predicts optimal behavior on average and deviations from optimality on individual decisions. Specific sampling algorithms have characteristic dynamics that may help explain the dynamics of human cognition. Altogether, sampling algorithms are formal process-level descriptions of Bayesian inference that could bridge the gap between ideal-observer analyses from machine learning and known resource constraints from cognitive psychology.

To elucidate the sampling hypothesis, it should be contrasted with two alternate theories of representations: *Boolean point estimates* and *probability distributions*.

Boolean-valued point estimates

Classical accounts of neural and psychological representation assume that beliefs are noisy, Boolean-valued point-estimates. Boolean-valued belief representations contain single estimates of a belief: In choices from multiple discrete options, one or more options may be

deemed true, and the others false. An object either belongs to a category, or it does not. A signal has either passed threshold, or it has not. In choices along continuously valued dimensions (e.g., brightness), all-or-none representations take the form of point-estimates (e.g., 11.3 Trolands). Although the content of a point-estimate is continuous (11.3), its truth value is Boolean. Such Boolean accounts of mental representation have been postulated for signal detection (point estimates corrupted by noise; e.g., Green & Swets, 1966), memory (memory traces as point estimates; e.g., Kinchla & Smyzer, 1967), concepts and knowledge (as logical rules and Boolean valued propositions; e.g., Bruner, Goodnow, & Austin, 1956).

However, to produce behavior consistent with ideal Bayesian observers, as they have been shown to do in signal detection (Whitely & Sahani, 2008), memory (Steyvers, Griffiths, & Dennis, 2006), categorization (Tenenbaum, 1999), and knowledge (Shafto, Kemp, Bonawitz, Coley, & Tenenbaum, 2008), people must represent *uncertainty* about their beliefs. They must know how much they believe different uncertain alternatives. Unfortunately, Boolean-valued belief representations fail to represent uncertainty, and as such, cannot support the Bayesian probabilistic computations that describe human behavior in these same domains.

Full probability distributions

A strictly computational account of Bayesian cognition would suggest that cognition represents exact probability distributions. A probability distribution may be exactly represented in two ways. First, analytically: as a mathematical function that codifies the probability of any possible hypothesis. It seems cognitively and neurally implausible for mental representations to be, literally, mathematical functions. Second, probability distributions may be represented as fully enumerated weighted lists: a paired list of every hypothesis along with its associated probability (e.g., probabilistic population codes; Ma, Beck, Latham, & Pouget, 2006). While weighted lists may be plausible representations for cases with fairly simple inferences, they break down in large-scale combinatoric problems, where the number of hypotheses grows exponentially to potentially infinite length. In these cases, a weighted list would need to be impossibly, or at least implausibly, long given constraints on human cognition.

Sample-based representations

Since, neither Boolean representations nor full probability distributions seem adequate, I propose *sampling* as an alternative representation. According to this sampling hypothesis, people represent probability distributions as sample-generating procedures, and as sets of samples that have been generated from these procedures. Inference by sampling rests on the ability to draw samples from an otherwise intractable probability distribution: to arrive at a set of hypotheses which are distributed according to the target distribution. Of course, producing exact samples that are distributed according to the correct probability distribution is hard in itself; thus, a number of algorithms have been developed for solving precisely this problem, for instance, Markov chain Monte Carlo; 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 distributions of interest. As a physical example, consider the “plinko” machine (Figure ??, Galton,

1889) – this device represents a Gaussian distribution in so far as additional balls dropped in can generate additional (approximately) Gaussian samples. Representations via samples and sample-generating procedures can represent uncertainty as the variation of the set of samples (in contrast to single point-estimates). Moreover, in contrast to weighted lists, sample-based representations may be truncated at a short, finite length without introducing systematic error.

Theoretical considerations

Sampling is commonly used in engineering to approximate inference for the same reasons that make sampling an appealing process-model for cognitive science. First, sampling algorithms are applicable to most inference problems used across cognitive domains. Second, sampling algorithms can plausibly implement real-world reasoning because they scale efficiently to high-dimensional problems. Third, sampling is a “just-in-time” algorithm that smoothly trades off precision, speed and computational load; thus sampling algorithms can be used when time or cognitive resources are limited, while also allowing precise inferences when resources allow.

In Chapter 2, I capitalize on the graceful degradation of sampling with limited cognitive resources to explore the meta-cognitive question: how many samples should people use to make a decision? In Bayesian statistics and machine learning, accurate *inference* requires thousands of samples, each of which is effortful to produce. But how many samples are necessary to make a sufficiently accurate *decisions*? Surprisingly, across a large range of tasks, using few samples often yields decisions that are not much worse than those based on more precise inferences (Vul, Goodman, Griffiths, & Tenenbaum, 2009). Moreover, on the assumption that sampling is effortful and takes time, I found that using just one sample for decisions often maximizes expected reward: making quick, suboptimal decisions is often the globally optimal policy.

Relationship between sampling and classical theories

Not only is the sampling hypothesis is a novel process-level description that connects computational Bayesian models to processing constraints from cognitive psychology, but it is also closely related to several classical laws and theories of cognition.

Probability matching, Luce choice, and soft-max decisions

Probability matching (Herrnstein, 1961; Vulkan, 2000) refers to the characteristic behavior of people and animals when faced with a risky choice among several alternatives. In a typical probability matching experiment, an observer is faced with a choice between two options, one rewarded with probability p , the other with probability $1-p$. The optimal choice is to *maximize* and always choose the option with the greater probability of reward; however, instead, people choose the two alternatives with frequency proportional to the probability of reward; thus *matching* the reward probability. Luce (1959) generalized the matching rule to describe the gamut of responses that people make between probability matching and maximizing. Although people do not always strictly *match* probabilities, deviations from probability matching follow a law-like regularity, where responses are chosen according to an exponentiated probability matching law – when the exponent is 1, people

match probabilities, and as the exponent increases, people are closer and closer to maximizing. The Luce choice rule also generalizes beyond probabilities based on previous rewards, to probabilities derived from any set of beliefs or representations. For this reason, the Luce choice rule has been applied as the function linking optimal probabilistic models of cognition to suboptimal human behavior, via a “soft-max” decision rule Frank, Goodman, & Tenenbaum, 2009; Goodman et al., 2008.

The sampling hypothesis naturally captures this gamut of decision-making behaviors. Probability matching can be described as sampling from prior experience: randomly selecting previously experienced outcomes and choosing the option that was rewarded most often in the set of sampled outcomes. When only one previous trial is considered, this procedure yields probability matching, and if more trials are considered then behavior will vary between probability matching and maximizing. When more than one trial is considered, decisions deviate from probability matching, and follow the general Luce choice/soft-max decision rule. Moreover, the sampling hypothesis is not restricted to sampling directly previously experienced outcomes: hypotheses may be sampled not only from direct experience but also from internal beliefs inferred indirectly from observed data. The sampling hypothesis thus allows provides an explanation for the general use of a soft-max decision rule connecting probabilistic models to human behavior.

Point-estimates, noise, and drift-diffusion models

Point-estimate based representations explain variation across trials and individuals as point-estimates corrupted by noise. Because the structure of the noise defines a probability distribution over possible states, the variation in responses across trials predicted by sampling and predicted by the noise accounts align under these special circumstances. Crucially, however, as I will describe below, predictions of these two accounts diverge when considering the relationship between the errors contained in different guesses.

A specific case of the noisy point-estimate account – the drift-diffusion model used to describe decision-making over time in cognitive psychology and neuroscience (Ratcliff, 1978; Gold & Shadlen, 2000) – allows for quantitative assessment of speed-accuracy tradeoffs on the tacit assumption that people aggregate noisy samples of input data until they reach some decision criterion. These cases may be construed as obtaining “samples” from the external world when used to account for perceptual decisions (Gold & Shadlen, 2000), but when applied to cognitive decisions, such as memory retrieval (Ratcliff, 1978), the samples must be internally generated. In cases where drift-diffusion models are applied to memory, they are superficially isomorphic to sample-based inference about internal beliefs.

Thus, the sampling hypothesis unifies internal noise, drift diffusion models, and soft-max probability matching behavior under one general framework that describes how people can approximate optimal Bayesian inference in situations without direct prior experience about the task at hand and must make decisions and inferences based solely on pure reasoning.

Evidence for sampling in human cognition

Theories postulating representation via point-estimates, probability distributions, or samples, postulate different causes for response variability, and thus make different predictions about the information contained in multiple guesses. If internal representations are

noisy point-estimates, then variation across responses must arise from the accumulation of noise on the point-estimates – thus, errors will be cumulative and correlated. If internal representations are complete probability-distributions, then there should be no response variation (variation in responses can only reflect variation in inputs or utility functions). In contrast, if internal representations are sampling processes, then variation in responses arises from the randomness of sampling – thus, multiple responses will have independent error. Crucially, this results in novel and unique predictions about the information contained in response errors, which we can look for in human behavior.

Although sampling and point-estimate representations may predict similar distributions of errors, they differ in their predictions about the relationship between multiple guesses. According to internal noise models, a single point-estimate is corrupted by noise, and this noise will therefore be shared by a number of guesses based on the same point estimate. In contrast, sampling based predict independent error between multiple guesses based on the same stimulus information. In my dissertation I test this prediction in the case of visual attention – a domain which facilitates precise quantitative measurements – as well as cognitive knowledge estimation tasks.

In Chapter 3, we investigate sampling in simple visual attention tasks (Vul, Hanus, & Kanwisher, 2009). We asked subjects to report one letter from a circular array cued by a line, but we asked them for multiple guesses on each trial. In these circumstances, subjects often report nearby items instead of the target, but we asked whether two guesses share a source of spatial error. If errors in this task reflect noise corrupting the position of the cue, then there would be a correlation in errors between the two guesses: if the first guess contained an error clockwise from the cue, the second guess should as well. However, if these errors arise from sampling given uncertainty about the spatial co-occurrence of cue and target, then the errors should be independent. We confirmed the sampling hypothesis prediction and found that two guesses independent and identically distributed. This result was replicated for the case of visual selection in time, rather than space (Vul et al., 2009). Together, these results indicate that when subjects are asked to select items under spatiotemporal uncertainty, subjects make guesses by independently sampling alternatives from a probability distribution over space and time.

In Chapter 4, we further tested the sampling hypothesis in “binding” tasks: subjects had to report two features of the target letter to assess whether illusory conjunctions (Treisman & Schmidt, 1982), or misbinding errors, also arise from sampling under spatiotemporal uncertainty. Vul and Rich (in press) presented subjects with arrays and RSVP streams of colored letters and asked subjects to report both the color and the letter. Given the permuted arrangement of colors and letters, these two dimensions yielded two independent estimates of the reported spatial positions. Again, the correlation between these reports could be used to evaluate the independence of the two guesses. In this case as well, errors were independent indicating that different features are independently sampled and that illusory conjunctions and binding errors arise from spatiotemporal uncertainty rather than noise.

In Chapter 5 we test a predicted consequence of the sampling hypothesis in high level knowledge. If different responses from one person correspond to independent samples from their uncertain knowledge, we should see a “wisdom of crowds” (Surowiecki, 2004) within one individual. Galton (1907) demonstrated that averaging guesses from multiple individ-

uals yields a more accurate answer than can be obtained from one individual alone because the independent error across individuals averages out to yield a more accurate estimate. If multiple guesses from one individual are independent samples, they should also contain independent error, and then the average of multiple guesses from one individual should also yield a similar “wisdom of crowds” benefit, where the crowd is within one individual. Vul and Pashler (2008) tested this prediction by asking subjects to guess numerical trivia (e.g., what proportion of the world’s airports are in the United States?). After subjects made one guess for each question, they were asked to provide a second guess for each. The average of two guesses was more accurate than either guess alone, suggesting that the two guesses contained some independent error, as predicted under the sampling hypothesis.

The above cases verify the predictions of the sampling account in cases where the implicit computational model is not specified, but does the sampling hypothesis yield additional predictive power in cases with a concrete computational model? Goodman et al. (2008) investigated human decisions in a category-learning task, where subjects see several examples of a category, and are then asked to decide whether new items belong to the category or not. Goodman et al. (2008) found that the average frequency with which subjects classify new items fits almost perfectly with the probabilistic predictions of a Bayesian rule-learning model. The model considers all possible classification rules, computes a posterior probability for each rule given the training data, and then computes the probability that any item belongs to the category by averaging the decisions of all possible rules weighted by their posterior probabilities. Is this fully Bayesian inference what individual subjects do on any one trial? Not in this task. Goodman et al. (2008) analyzed the generalization patterns of individual subjects reported by (Nosofsky, Palmeri, & McKinley, 1994) and found that response patterns across seven test exemplars were only poorly predicted by the Bayesian ideal. Rather than averaging over all rules, these generalization patterns were instead consistent with each participant classifying test items using only one or a few rules; while the particular rules considered vary across observers according to the appropriate posterior probabilities. Thus, it seems that individual human learners are somehow drawing one or a few samples from the posterior distribution over possible rules, and behavior that is consistent with integrating over the full posterior distribution emerges only in the average over many learners. Similar sampling-based generalization behavior has been found in word learning (Xu & Tenenbaum, 2007) and causal learning tasks (Sobel, Tenenbaum, & Gopnik, 2004), in both adults and children.

Specific sampling algorithms for specific tasks

Although we confirmed the predictions of sample-based inference in several cognitive domains, the fact remains that producing a sample from the appropriate posterior distribution is not trivial. In computer science and statistics there are many algorithms available for doing Monte Carlo inference. Simple sample-generating algorithms, like rejection sampling, tend to be slow, inefficient, and computationally expensive. In practice, different sampling algorithms are chosen for particular problems where they may be most appropriate. Therefore, while “sampling” may capture some cognitive phenomena at a coarse grain, the exact sampling algorithms used may vary across domains, and may provide more accurate descriptions of specific behavioral phenomena and the dynamics of cognition.

Most real-world domains offer only small amounts of training data which must then

support a number of future inferences and generalizations. Shi, Griffiths, Feldman, and Sanborn (in press) showed that in such domains, exemplar models (Medin & Shaffer, 1978) using only a few examples can support Bayesian inference as an importance sampler (Ripley, 1987). This can be achieved using an intuitive psychological process of storing just a small set of exemplars and evaluating the posterior distribution by weighting those samples by their probability – a process known as “importance sampling”. Shi et al. (in press) argued that such an importance sampler accounts for typicality effects in speech perception (Liberman, KS, Hoffman, & Griffith, 1957), generalization gradients in category learning (Shepard, 1987), optimal estimation in everyday predictions (Griffiths & Tenenbaum, 2006), and reconstructive memory (Huttenlocher, Hedges, & Vevea, 2000).

Particle filtering is a candidate sampling algorithm for domains where inference must be carried out online as data are coming in, such as sentence processing (Levy, Reali, & Griffiths, 2009), object tracking (Vul, Frank, Alvarez, & Tenenbaum, 2010), or change-point detection (Brown & Steyvers, 2008). Particle filters track a sampled subset of hypotheses as they unfold over time; at each point when additional data are observed, the current set of hypothesized states are weighted based on their consistency with the new data, and resampled accordingly – as a consequence, this inference algorithm produces a bias against initially implausible hypotheses. Levy et al. (2009) showed how this bias can account for garden-path effects in sentence processing: when the start of the sentence suggests one interpretation, but the end of the sentence disambiguates the interpretation in favor of a less likely alternative, people are substantially slowed as they search for the correct parse. This difficulty arises in particle filtering because of the challenge of resampling/updating when the correct hypothesis is not within the currently entertained set of particles. Similar arguments have been used to explain individual differences in change detection (Brown & Steyvers, 2008), and performance while tracking objects (Vul et al., 2010), and might be fruitfully applied to describe other classic sequential inference biases (e.g., Bruner & Potter, 1964).

In some real-world and laboratory tasks, the observer sees all the relevant data and must make sense of it over a period of time. For instance, when looking at a 2D projection of a wireframe cube (Necker, 1832), observers are provided with all of the relevant data at once, but must then come up with a consistent interpretation of the data. In cases where two equally likely interpretations of the stimulus are available, the perceived interpretation changes stochastically over time, jumping between two modal interpretations. Sundareswara and Schrater (2007) demonstrated that the dynamics of such rivalry in the case of a Necker cube arises from approximate inference via Markov Chain Monte Carlo (MCMC; Robert & Casella, 2004). Gershman, Vul, and Tenenbaum (2010) elaborated on this argument by showing that MCMC in a coupled markov random field – like those typically used as computational models of low-level vision – not only produces bistability and binocular rivalry, but also produces the characteristic traveling wave dynamics of rivalry transitions (Gershman et al., 2010; Wilson, Blake, & Lee, 2001).

Conclusion

I started with a set of challenging question for cognitive science: How can optimal probabilistic models of human cognition be reconciled with the cognitive processing constraints documented in cognitive psychology? How can people approximate ideal statistical

inference despite their limited cognitive resources? How can we account for the dynamics of human cognition along with the associated errors and variability of human decision-making? I proposed the sampling hypothesis as a means to reconcile theories and data from different cognitive science disciplines: Instead of representing complete probability distributions, or single point-estimates, people represent their beliefs as sample-generating processes. The unique predictions of the sampling hypothesis are confirmed across several visual and cognitive tasks. Moreover, sampling-based approximate inference algorithms provide an account of the process-level dynamics of human cognition as well as variation in responses. Altogether, the sampling hypothesis provides a necessary link between computational models of cognition from computer science and process-level limitations from cognitive psychology.

References

- Alais, B., & Blake, R. (2005). *Binocular rivalry*. Cambridge: MIT Press.
- Anderson, J. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*.
- Broadbent, D. (1958). *Perception and communication*. London: Pergamon Press.
- Brown, S., & Steyvers, M. (2008). Detecting and predicting changes. *Cognitive Psychology*, *58*, 49–67.
- Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York, NY: Wiley.
- Bruner, J., & Potter, M. (1964). Interference in visual recognition. *Science*, *144*, 424-425.
- Cepeda, N., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distribution of practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380.
- Cepeda, N., Vul, E., Rohrer, D., Wixted, J., & Pashler, H. (2008). Spacing effects in learning: A temporal ridge of optimal retention. *Psychological Science*, *19*, 1095-1102.
- Charniak, E. (1995). *Parsing with context-free grammars and word statistics* (Tech. Rep.). Providence, RI: Brown university.
- Chater, N., & Manning, C. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344.
- Deese, J., & Kaufman, R. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, *54*(3), 180-187.
- Doucet, A., De Freitas, N., & Gordon, N. (2001). *Sequential monte carlo methods*. New York: Springer.
- Frank, M., Goodman, N., & Tenenbaum, J. (2009). Using speakers' referential intentions to model early cross-situational world learning. *Psychological Science*.
- Galton, F. (1889). *Natural inheritance*. London: MacMillan.
- Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 450-451.
- Gershman, S., Vul, E., & Tenenbaum, J. (2010). Perceptual multistability as markov chain monte carlo inference. *Advances in Neural Information Processing Systems*.
- Gold, J., & Shadlen, M. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390-394.
- Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.
- Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- 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.
- Huttenlocher, J., Hedges, L., & Vevea, J. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220-241.

- James, W. (1890). *The principles of psychology*. Dover publications.
- Kinchla, R., & Smyzer, F. (1967). A diffusion model fo perceptual memory. *Perception and psychophysics*, 2(219-229).
- Knill, D., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
- Levy, R., Reali, F., & Griffiths, T. (2009). Modeling the effects of memory on human online sentence processing with particle filters. In *Advances in neural information processing systems* (Vol. 21). Whistler, Canada.
- Liberman, A., KS, H., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358-368.
- Luce, R. (1959). *Individual choice behavior*. New York, NY: Wiley.
- Ma, W., Beck, J., Latham, P., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432-1438.
- 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: MIT Press.
- McKenzie, C. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and bayesian inference. *Cognitive Psychology*, 26, 209–239.
- Medin, D., & Shaffer, M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Mozer, M., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science*, 32, 1133-1147.
- Necker, L. (1832). Observations on some remarkable phenomenon which occurs in viewing a figure of a crystal or geometrical solid. *Phil. Mag. J. Sci*, 3, 329-337.
- Nosofsky, R., Palmeri, T., & McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53-79.
- Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 358-377.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ripley, B. (1987). *Stochastic simulation*. New York: Wiley.
- Robert, C., & Casella, G. (2004). *Monte carlo statistical methods*. New York: Springer.
- Shafto, P., Kemp, C., Bonawitz, E., Coley, J., & Tenenbaum, J. (2008). Inductive reasoning about causally transmitted properties. *Cognition*, 109(2), 175-192.
- Shepard, R. (1987). Towards a universal law of generalization for psychological science. *Science*, 237, 1317-1323.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701-703.
- Shi, L., Griffiths, T., Feldman, N., & Sanborn, A. (in press). Exemplar models as a mechanism for performing bayesian inference. *Psychonomic Bulletin and Review*.

- 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.
- Steyvers, M., Griffiths, T., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, *10*(7), 327-334.
- Sundareswara, R., & Schrater, P. (2007). Dynamics of perceptual bistability. *Journal of Vision*, *8*(5).
- Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Random House.
- Tenenbaum, J. (1999). Bayesian modeling of human concept learning. *Advances in Neural Information Processing Systems*.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107-141.
- Vul, E., Frank, M., Alvarez, G., & Tenenbaum, J. (2010). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*.
- Vul, E., Goodman, N., Griffiths, T., & Tenenbaum, J. (2009). One and done? optimal decisions from very few samples. *Proceedings for the 31st Annual Meeting of the Cognitive Science Society*.
- Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all or none samples. *Journal of Experimental Psychology: General*, *4*, 546-560.
- Vul, E., & Pashler, H. (2007). Incubation is helpful only when people are misled. *Memory and Cognition*, *35*, 701-710.
- Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645-647.
- Vul, E., & Rich, A. (in press). Independent sampling of features enables conscious perception of bound objects. *Psychological Science*.
- Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*(1), 101-118.
- Welford, A. (1952). The 'psychological refractory period' and the timing of high-speed performance – a review and a theory. *British Journal of Psychology*, *43*, 2-19.
- Whitely, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty is used to guide statistically optimal decisions. *Journal of Vision*, *8*(1-15).
- Wilson, H., Blake, R., & Lee, S. (2001). Dynamics of travelling waves in visual perception. *Nature*, *412*, 907-910.
- Wixted, J., & Ebbesen, E. (1991). On the form of forgetting. *Psychological Science*, *2*(6), 409-415.
- Xu, F., & Tenenbaum, J. (2007). Sensitivity to sampling in bayesian word learning. *Developmental Science*, *10*(3), 288-297.