aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
21(4) 263–268

DOI:10.1177/0963721412447619
http://cdps.sagepub.com
SAGE

Bridging Levels of Analysis for Probabilistic Models of Cognition

Thomas L. Griffiths[1], Edward Vul[2], and Adam N. Sanborn[3]
[1]University of California, Berkeley; [2]University of California, San Diego; and [3]University of Warwick

Abstract

Probabilistic models of cognition characterize the abstract computational problems underlying inductive inferences and identify their ideal solutions. This approach differs from traditional methods of investigating human cognition, which focus on identifying the cognitive or neural processes that underlie behavior and therefore concern alternative levels of analysis. To evaluate the theoretical implications of probabilistic models and increase their predictive power, we must understand the relationships between theories at these different levels of analysis. One strategy for bridging levels of analysis is to explore cognitive processes that have a direct link to probabilistic inference. Recent research employing this strategy has focused on the possibility that the Monte Carlo principle—which concerns sampling from probability distributions in order to perform computations—provides a way to link probabilistic models of cognition to more concrete cognitive and neural processes.

Keywords

cognitive modeling, levels of analysis, probabilistic models of cognition, rational process models

From language learning and categorization to visual perception and motor control, the tasks that the brain must solve require the use of noisy, incomplete information about the world to make generalizations and future decisions. How do people make inductive inferences and guide their behavior on the basis of such limited data? The computational challenge in understanding human behavior in these complex tasks is fundamentally statistical; therefore, over the past two decades, researchers have started to answer this question by using probabilistic models of cognition to analyze human inferences in such tasks (Anderson, 1990; Ashby & Alfonso-Reese, 1995; Kersten, Mamassian, & Yuille, 2004; Levy, Reali, & Griffiths, 2009; Sanborn, Griffiths, & Navarro, 2010; Trommershäuser, Maloney, & Landy, 2008).

Probabilistic models of cognition are cast at what Marr (1982) called the *computational* level. They specify the ideal solution to an abstract statistical problem that people must solve: Given the decision that must be made, how *should* people use the limited available information? (See Table 1.) Researchers compare human behavior with candidate ideal solutions to characterize the assumptions that guide human inductive inference. This approach is quite different from traditional methods used to study the mind. Historically, cognitive psychologists have defined models at Marr's *algorithmic* level, focusing on identifying the cognitive processes involved in representing and manipulating information. Neuroscience has added analyses at Marr's *implementation* level, examining how these cognitive processes might be realized in the brain. This range of approaches raises a basic question: How are insights at these different levels of analysis connected?

The Importance of Bridging Levels of Analysis

Understanding the relationships among results from these different levels of analysis is central to evaluating the contributions that probabilistic models of cognition can make to psychology. In the 30 years since Marr described the computational level, the strategy of seeking computational-level explanations has grown in popularity. As probabilistic models have been applied to more aspects of cognition, it has become increasingly important to understand the implications of such computational-level analyses for algorithmic- and implementation-level analyses. Elucidating this relationship can inform both predictions across tasks and the theoretical constraints that hold between levels of analysis.

On the empirical side, a fruitful strategy for identifying cognitive processes is to look for ways in which human behavior deviates from ideal solutions obtained from computational

Corresponding Author:
Thomas L. Griffiths, Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, MC 1650, Berkeley, CA 94720-1650
E-mail: tom_griffiths@berkeley.edu

Table 1. Levels of Analysis Identified by Marr (1982)

Level of analysis	Goal
Computational	Identify the abstract computational problem to be solved and its ideal solution
Algorithmic	Identify the algorithm and representation used in executing (or approximating) the solution
Implementation	Identify the physical process underlying the algorithm

analyses. The heuristics and biases research program (Tversky & Kahneman, 1974) provides one of the best examples of this approach. However, this strategy requires an understanding of when such deviations reflect mistaken assumptions about ideal solutions on the part of the researcher (often a result of misunderstanding the problem people are solving) and when they provide clues about the cognitive and neural processes by which people approximate those ideal solutions. Probabilistic models of cognition at the computational level can make this strategy for identifying cognitive processes applicable in a wider range of domains by expanding the set of problems for which we know the ideal solutions.

On the theoretical side, we need to know how particular probabilistic models constrain and are constrained by particular theories at the algorithmic or implementation level—or whether these accounts are independent. For example, the recent debate between proponents of probabilistic and connectionist models (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; McClelland et al., 2010) emphasized that many probabilistic models are defined in terms of structured, discrete representations, such as rules and grammars, whereas connectionist models use continuous, graded representations that can mimic discrete structures when appropriate. However, because these models are cast at different levels of analysis, it is not clear whether this representational discrepancy reflects a fundamental incompatibility. In general, we cannot know whether a given probabilistic model is inconsistent with particular cognitive or neural processes unless we identify connections between computational-level models and models at the algorithmic and implementation levels.

A Strategy for Bridging Levels of Analysis

When he proposed the idea of different levels of analysis for information-processing systems, Marr (1982) expected that there would be constraints that hold between levels. Successful computational-level analyses impose a strong constraint on analyses at the algorithmic and implementation levels: Whatever form those cognitive and neural processes take, they need to approximate the solution to the computational problem—after all, people somehow make sensible decisions and inferences on the basis of limited available data much of the time. Similarly, algorithmic- and implementation-level analyses constrain computational-level analyses: The information available for analysis at the computational level is determined by limitations identified at the algorithmic and implementation levels. To take a simple example, people must perceive the world through biological sensors and make decisions using their brains. Just as the structure of organisms constrains evolutionary solutions to the problems posed by their environments, the structure of the human mind and brain should constrain people's solutions to the computational problems they consider.

In line with this logic, a persistent match between human behavior and the predictions of a probabilistic model suggests that the cognitive and neural processes producing this behavior are somehow approximating probabilistic inference. This observation leads to a strategy for bridging levels of analysis: Consider the best algorithms for approximating probabilistic inference in computer science and statistics as candidate models of cognitive and neural processes. Such *rational process models* (Sanborn et al., 2010; Shi, Griffiths, Feldman, & Sanborn, 2010) may elucidate the processes operating at the algorithmic and implementation levels but also provide a direct link to the computational level.

Rational process models differ from traditional process models in cognitive psychology, which postulate a set of psychological mechanisms and examine how those mechanisms can be combined to model behavior. In contrast, proposing a rational process model involves identifying an algorithm for approximating probabilistic inference, determining whether the components of the algorithm are consistent with what we know about cognitive processes, and then examining how well the model fits behavior. The resulting models can approximate probabilistic inference arbitrarily well for situations in which sufficient time and memory are available, but they deviate from ideal solutions in systematic ways that can capture human behavior for situations in which information-processing resources are limited.

Developing rational process models bridges the computational and algorithmic levels in two ways. First, it involves considering a continuum of models: At one end is ideal performance; at the other, a systematic pattern of deviations from this ideal that depends on the particular algorithm being used. These models are thus directly connected to and constrained by the computational-level analysis, with the strength of the constraint being reduced as the limitations on time and space imposed on the algorithm increase. Second, the way in which these models deviate from the ideal solution is not arbitrary. The algorithms that are used are the best solutions that computer scientists and statisticians have developed for solving problems when time and space are limited. As a consequence, rational process models can be viewed as pushing the principle of optimization that underlies computational-level accounts down to the algorithmic level, representing our best guess at a strategy that human minds could use to solve a computational problem given particular information-processing constraints.

Monte Carlo as a Psychological Mechanism

The common thread running through all probabilistic models of cognition is the use of Bayesian inference to combine available data with prior beliefs in order to form new beliefs that can be used to guide behavior. Bayesian inference provides a computational-level account of belief revision (and, more generally, of learning and inductive inference) by describing beliefs as probability distributions over possibilities. Assume that a learner entertains a set of hypotheses about something (say, which of a number of possible acquaintances is calling her cell phone) and represents her degree of belief in each hypothesis h with a probability $p(h)$. The resulting distribution is referred to as the *prior distribution* because it indicates the learner's degree of belief in the likelihood of each hypothesis before any data are observed (in our example, this might reflect the frequency with which the learner's acquaintances call her, as well as anticipated calls). After observing some data d (say, the area code of the caller), the learner needs to revise these degrees of belief to obtain a *posterior distribution* $p(h \mid d)$. This is done by applying Bayes' rule,

$$p(h \mid d) = \frac{p(d \mid h)p(h)}{\sum_{h'} p(d \mid h')p(h')} \quad (1)$$

where $p(d \mid h)$, the likelihood, indicates the probability of d if h is true (how likely each possible caller is to have a phone number with a particular area code), and the sum in the denominator ranges over all hypotheses. In the context of real cognitive tasks, the actual use of Bayes' rule to calculate the result quickly becomes intractable because a very large number of hypotheses must be compared. Bayesian inference therefore requires some algorithm to allow for approximation by human minds and brains.

What cognitive and neural processes are candidates for approximating the computational-level solutions identified by Bayes' rule? One highly successful strategy for approximating probabilistic inference in computer science and statistics is the Monte Carlo principle: Instead of using the probability distribution itself to perform computations, use a number of samples from that distribution, each randomly selected with a frequency proportional to its probability in the full distribution. Different Monte Carlo algorithms are used to approximate probabilistic inference across a range of circumstances, and these algorithms provide a rich source of hypotheses about possible rational process models (see Table 2).

The Monte Carlo algorithms we consider all have an important property: If they use enough samples, their answers will be exactly the same as those of computations performed with the entire probability distribution. This means that the ideal behavior indicated by Bayes' rule can be achieved by these algorithms. However, these algorithms can also be used to generate answers when the resources for solving a problem are limited. To do this, we just use a small number of samples. Reducing the number of samples can introduce systematic deviations from Bayesian inference, which we can look for in human behavior as a strategy for identifying the algorithms that people might be using (just as particular biases were used as clues about the heuristics that guide people's judgments by Tversky & Kahneman, 1974).

The idea that people might approximate probabilistic inference by sampling connects to a long literature in cognitive psychology. Sampling appears as a basic component of a variety of psychological theories of choice and decision making (Busemeyer, 1985; Luce, 1959; Stewart, Chater, & Brown, 2006). Moreover, sampling is often implicitly used in probabilistic models of cognition as a result of the assumption that people make judgments with frequency proportional to their probability (a strategy known as *probability matching*), which is consistent with using only a few Monte Carlo samples to make a judgment.

Although a few samples are insufficient to adequately approximate a probability distribution, decisions based on even one sample can be almost as good as the ideal calculation, and they may even be optimal in the long run if obtaining additional samples is cognitively demanding (Vul, Goodman, Griffiths, & Tenenbaum, 2009). Behavior consistent with the

Table 2. Sophisticated Monte Carlo Methods for Approximating Bayesian Inference

Algorithm	Purpose	Example behaviors captured	Example applications
Importance sampling	Approximating the posterior distribution using only samples from the prior distribution	Extrapolations from training trials	Modeling reproductions from memory of perceptual stimuli (Shi, Griffiths, Feldman, & Sanborn, 2010)
Particle filter	Updating a posterior distribution over hypotheses as more data are observed	Order effects when new information is encountered	Learning categories as exemplars are revealed sequentially (Sanborn, Griffiths, & Navarro, 2010) and understanding sentences as words are heard (Levy, Reali, & Griffiths, 2009)
Markov chain Monte Carlo	Exploring a space of hypotheses while maintaining a single hypothesis at a time	Change of beliefs over time without new information	Bistable perception, whereby people move stochastically between interpretations (Gershman, Vul, & Tenenbaum, 2012)

use of a few samples to make judgments occurs in experiments where participants make multiple guesses using the same impoverished knowledge. For instance, multiple guesses about obscure facts contain independent error, so that the average of two guesses from one person is more accurate than either guess alone (Vul & Pashler, 2008). In the remainder of this section, we review some preliminary results that illustrate how Monte Carlo methods might provide insight into how people approximate probabilistic computations.

One intuitive method for approximating probabilistic inference is to retrieve memories of past events that were similar to a current event. This algorithm turns out to be a specific form of a Monte Carlo method known as *importance sampling*, whereby samples are drawn from a distribution other than the target distribution and then reweighted to approximate a sample from the target distribution (see Robert & Casella, 2004). A simple importance-sampling algorithm for Bayesian inference involves sampling hypotheses h from the prior distribution $p(h)$ and then weighting those hypotheses by the likelihood function $p(d \mid h)$ on observing data d to obtain an approximation to the posterior distribution $p(h \mid d)$. The events remembered from the past act as samples from the prior distribution, and the similarity function corresponds to the likelihood. This algorithm may sound familiar, as it can be shown to be formally equivalent to classical exemplar process models. This simple rational process model predicts human behavior in a variety of tasks that have been analyzed using probabilistic models of cognition (Shi et al., 2010).

When people must update their beliefs over time as new information arrives, they often deviate from ideal behavior in systematic ways. One example is *order effects*: The order in which people receive information affects their judgments, even in cases where computational-level analyses say it should not (see Sanborn et al., 2010, for examples). These effects can be captured by Monte Carlo algorithms known as *particle filters*, which represent a posterior distribution with a set of samples that is updated as new data become available (see Robert & Casella, 2004). Because the representation of the learner's beliefs is reduced to a few hypotheses sampled from the posterior distribution, it becomes easy for hypotheses that were initially supported by the data to dominate even when they are no longer justified. Particle filters have been used to explain order effects in category learning that are problematic for computational-level models to account for (Sanborn et al., 2010), as well as "garden path" effects in sentence processing (Levy et al., 2009) and the detection of changes in the distribution from which events are drawn (Brown & Steyvers, 2009).

In other situations, people deviate from purely computational-level analyses because they have to *think*: Despite seeing all the data, people do not immediately know a solution, but they slowly come up with one over time, perhaps changing their mind several times in doing so. Perceptual bistability provides an interesting example of a case in which beliefs change over time without the addition of new information as people switch between possible interpretations of a visual object. Markov chain Monte Carlo (MCMC; see Robert & Casella, 2004) algorithms provide a way to understand how people might change their beliefs even without observing new data by exploring a complex space of hypotheses while considering only one hypothesis at a time. The most common class of MCMC algorithms is based on the idea of taking a random walk through hypotheses, proposing local changes to the current hypothesis h, and accepting a proposed variant h' on the basis of its relative posterior probability $p(h' \mid d)/p(h \mid d)$. Hypotheses that better explain the data are more likely to be accepted, and, in the long run, the proportion of times the random walk visits a given hypothesis h converges to its posterior probability $p(h \mid d)$. When such random-walk MCMC algorithms are applied to probabilistic models of vision to infer the latent cause underlying a presentation of different images to the left and right eyes, the result exhibits the dynamics of binocular rivalry, including the distribution of switching times, patchy perception, and traveling waves (Gershman, Vul, & Tenenbaum, 2012). Although a purely computational-level account can explain why there are two stable percepts in bistable phenomena (i.e., the two percepts correspond to two hypotheses that have high probability in the posterior distribution), these results represent promising initial steps toward the idea that a rational process model based on MCMC can also capture the rich dynamics of bistable perception.

Prospects and Challenges for Rational Process Models

There are several appealing features of process models based on approximate inference algorithms for computational-level models. First, and most important, these process models make an explicit connection to the computational level, thus bridging levels of analyses. For example, the identification of exemplar models as a form of importance sampling strengthened the connection between existing process and computational-level models (Shi et al., 2010). Second, while forming such a bridge, the computational and algorithmic levels remain partitioned, thus allowing for greater modularity of theories in the creation of novel models of new tasks. For example, particle filters can be used to explain phenomena in category learning (Sanborn et al., 2010), sentence processing (Levy et al., 2009), and change-point detection (Brown & Steyvers, 2009), with a single algorithm that is applicable across different computational-level models. Third, the explicit constraints from the computational level tend to increase the parsimony of the resulting process model because the few free parameters that remain must produce ideal behavior in the limiting case of infinitely many samples. Consequently, the whole range of parameter settings will often produce reasonable behavior, which can then be used to explicate individual differences (e.g., Brown & Steyvers, 2009). Finally, even with parameter settings that yield notable deviations from optimal solutions, process models based on Monte Carlo principles can often produce average behavior that matches the ideal solution. This

explains how individuals might use approximation algorithms in each specific computation but the average behavior of an individual across multiple settings or of many individuals in one setting ends up resembling the ideal solution (Shi et al., 2010). The explicit approximation of probabilistic inference can thus potentially yield a better match to human behavior than an arbitrary process model, with greater parsimony and generality across tasks.

Despite these potential advantages of rational process models, building such models presents some considerable challenges. First, there is often uncertainty about a given person's overall goal in a task and about the structure of the computational-level model; thus, there is ambiguity about how to interpret specific deviations from our best guess about ideal behavior. Second, there are many Monte Carlo methods that can be used to approximate Bayesian inference, each with different behavioral deviations; therefore, the idea of exploring these methods only weakly constrains the set of models we might consider.

Identifying candidates for rational process models can potentially be facilitated by the fact that Monte Carlo methods have been developed to be computationally efficient for particular kinds of problems. We outlined some of the methods designed to solve specific kinds of problems—such as the use of particle filters for sequentially updating beliefs—in the previous section. A strategy is therefore to investigate algorithms that are suitable for solving the problem that people face in a given task and then to determine whether the algorithm's signature weaknesses (e.g., primacy effects from particle filters) are present in people's behavior. Likewise, we can investigate whether the algorithms that people use to solve particular problems are well-matched to the statistical structure of those problems—a kind of rational metacognition.

Although our focus in this article has been on cognitive processes suggested by Monte Carlo methods, a similar approach might also shed light on the neural processes that support probabilistic inference. Recent papers have shown how simple neural circuits can implement algorithms such as importance sampling (Shi & Griffiths, 2009) and how variability in neural responding might be interpreted in terms of sampling (Fiser, Berkes, Orban, & Lengyel, 2010). Since different sampling methods are better suited for different problems and provide good models of different psychological phenomena, we might expect that the brain would employ not just one mechanism for probabilistic inference, but many. Investigating this possibility is an exciting direction for future research.

Conclusion

Determining the empirical and theoretical implications of probabilistic models of cognition requires understanding how different levels of analysis are related. Rational process models provide one strategy for constructing a bridge between levels of analysis, using the constraints provided by computational-level theories while incorporating ideas about cognitive and neural processes. Monte Carlo methods, which are based on the principle of sampling from a probability distribution, have proved a rich source of such models so far. Other strategies for bridging levels of analysis exist, such as starting with process models and determining what computational-level problem they appear to solve, as Ashby and Alfonso-Reese (1995) did for categorization. Ultimately, we hope that the use of these different strategies will lead to a more complete understanding of how people perform the amazing feats of learning and inference that characterize human cognition, all the way from abstract computational problems and their solutions to concrete neural processes.

Recommended Reading

Anderson, J. R. (1990). (See References). A pioneering work applying computational-level analyses to human cognition that includes a nuanced discussion of the relationships between levels of analysis.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer. An accessible introduction to probabilistic models and inference algorithms in the context of machine learning.

Ashby, F. G., & Alfonso-Reese, L. (1995). (See References). Explores a different strategy for linking levels of analysis, focusing on the problem of categorization.

Fiser, J., Berkes, P., Orban, G., & Lengyel, M. (2010). (See References). A provocative analysis of the relationship between Monte Carlo strategies and variability in neural systems.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113*, 677–699. Develops an alternative approach to connecting process models and probabilistic inference.

Declaration of Conflicting Interests

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

Funding

This work was supported by National Science Foundation Grant IIS-1018733 and Air Force Office of Scientific Research Grant FA-9550-10-1-0232 to T. L. G. and by Office of Naval Research Grant N00014-07-1-0937 to E. V.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Alfonso-Reese, L. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39*, 216–233.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology, 58*, 49–67.

Busemeyer, J. R. (1985). Decision making under uncertainty: Simple scalability, fixed sample, and sequential sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 538–564.

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

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

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14*, 357–364.

Kersten, D., Mamassian, P., & Yuille, A. L. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271–304.

Levy, R., Reali, F., & Griffiths, T. L. (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). La Jolla, CA: NIPS Foundation.

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

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to understanding cognition. *Trends in Cognitive Sciences, 14*, 348–356.

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

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review, 117*, 1144–1167.

Shi, L., & Griffiths, T. L. (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems* (Vol. 22, pp. 1669–1677). La Jolla, CA: NIPS Foundation.

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*, 443–464.

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

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

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

Vul, E., Goodman, N., Griffiths, T., & Tenenbaum, J. (2009). One and done? Optimal decisions from few samples. In N. Taatgen and H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 66–72). Austin, TX: Cognitive Science Society.

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