Performing Bayesian Inference with Exemplar Models

Lei Shi (lshi@berkeley.edu)
Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA 94720 USA

Naomi H. Feldman (naomi_feldman@brown.edu)
Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912 USA

Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California at Berkeley, Berkeley, CA 94720 USA

Abstract

Probabilistic models have recently received much attention as accounts of human cognition. However, previous work has focused on formulating the abstract problems behind cognitive tasks and their probabilistic solutions, rather than considering mechanisms that could implement these solutions. Exemplar models are a successful class of psychological process models that use an inventory of stored examples to solve problems such as identification, categorization and function learning. We show that exemplar models can be interpreted as a sophisticated form of Monte Carlo approximation known as importance sampling, and thus provide a way to perform approximate Bayesian inference. Simulations of Bayesian inference in speech perception and concept learning show that exemplar models can account for human performance with only a few exemplars, for both simple and relatively complex prior distributions. Thus, we show that exemplar models provide a possible mechanism for implementing Bayesian inference.

Keywords: Bayesian inference; exemplar models; speech perception; concept learning

Much of cognition and perception involves inference under uncertainty, using limited data from the world to evaluate underdetermined hypotheses. Probabilistic models provide a way to characterize the optimal solution to these problems, with probability distributions encoding the beliefs of agents and Bayesian inference updating those distributions as data become available. As a consequence, probabilistic models are becoming increasingly widespread in both cognitive science and neuroscience, providing explanations of behavior in domains as diverse as motor control (Körding & Wolpert, 2004), reasoning (Oaksford & Chater, 1994), memory (Anderson & Milson, 1989), and perception (Yuille & Kersten, 2006). However, these explanations are typically presented at Marr's (1982) *computational* level, focusing on the abstract problem being solved and the logic of that solution. Unlike many other formal approaches to cognition, probabilistic models are usually not intended to provide an account of the mechanisms underlying behavior – how people actually produce responses consistent with optimal statistical inference.

Understanding the mechanisms that could support Bayesian inference is particularly important since probabilistic computations can be extremely challenging. Representing and updating distributions over large numbers of hypotheses is computationally expensive, a fact that is often viewed as a limitation of "rational" models. The question of how people could perform Bayesian inference can be answered at at least two levels (as suggested by Marr, 1982). One kind of answer focuses on the neural level, exploring ways in which systems of neurons could perform probabilistic computations. The language of such answers is that of neurons, tuning curves, firing rates, and so forth (e.g., Ma, Beck, Latham, & Pouget, 2006). A second kind of answer is at the level of psychological processes – showing that the Bayesian inference can be performed using mechanisms that are used in psychological process models. The language of such answers is representations, similarity, activation, and so forth (e.g., Kruschke, 2006; Sanborn, Griffiths, & Navarro, 2006).

Our focus in this paper will be on a class of psychological process models known as exemplar models. These models assume that people store many instances ("exemplars") of events in memory, and evaluate new events by activating stored exemplars that are similar to those events (Medin & Schaffer, 1978; Nosofsky, 1986). It is well known that exemplar models of categorization can be analyzed in terms of nonparametric density estimation, and implement a Bayesian solution to this problem (Ashby & Alfonso-Reese, 1995). Here we show that exemplar models can be used to solve problems of Bayesian inference more generally, providing a way to approximate expectations of functions over posterior distributions. Our key result is that exemplar models can be interpreted as a sophisticated form of Monte Carlo approximation known as *importance sampling*. This result illustrates how Bayesian inference can be performed using a simple mechanism that is a common part of psychological process models.

Background

Exemplar models

Human knowledge is formed from examples. When we learned the concept "dog," we were not taught to remember the physiological and anatomical characteristics of dogs, but instead, saw examples of various dogs. Based on the large inventory of examples of dogs we have seen, we are able to reason about the properties of dogs, and make decisions about whether new objects we encounter are likely to be dogs. Exemplar models provide a simple explanation for how we do this, suggesting that we do not form abstract generalizations from experience, but rather store examples in memory and use those stored examples as the basis for future judgments (Medin & Schaffer, 1978; Nosofsky, 1986).

An exemplar model consists of stored exemplars $X^* =$

$\{x_1^*, x_2^*, \cdots, x_n^*\}$, and a similarity function $s(x, x^*)$, measuring how closely a new observation x is related to x^*. On observing x, all exemplars are activated in proportion to $s(x, x^*)$. The use of the exemplars depends on the task (Nosofsky, 1986). In an identification task, where the goal is to identify the x^* of which x is an instance, the probability of selecting x_i^* is

$$p_r(x_i^*|x) = \frac{s(x, x_i^*)}{\sum_{j=1}^n s(x, x_j^*)}, \tag{1}$$

where $p_r(\cdot)$ denotes the response distribution resulting from the exemplar model, and we assume that participants use the Luce choice rule (Luce, 1959) in selecting a response, with no biases towards particular exemplars. In a categorization task, where each exemplar x_i^* is associated with a category c_i, the probability that the new object x will be assigned to category c is given by

$$p_r(c|x) = \frac{\sum_{j|c_j=c} s(x, x_j^*)}{\sum_{j=1}^n s(x, x_j^*)}, \tag{2}$$

where again we assume a Luce choice rule without biases towards particular categories.

While exemplar models have been most prominent in the literature on categorization, the same basic principles have been used to define models of function learning (DeLosh, Busemeyer, & McDaniel, 1997), probabilistic reasoning (Juslin & Persson, 2002), and social judgment (Smith & Zarate, 1992). These models pursue a similar approach to models of categorization, but associate each exemplar with a quantity other than a category label. For example, in function learning each exemplar is associated with the value of a continuous variable rather than a discrete category index. The procedure for generating responses remains the same as that used in Equations 1 and 2: the associated information is averaged over exemplars, weighted by their similarity to the stimulus. Thus, the predicted value of some associated information f for a new stimulus x is

$$\hat{f} = \frac{\sum_{j=1}^n f_j s(x, x_j^*)}{\sum_{j=1}^n s(x, x_j^*)}, \tag{3}$$

where f_j denotes the information associated with the jth exemplar. The identification and categorization models can be viewed as special cases, corresponding to different ways of specifying f_j. Taking $f_j = 1$ for $j = i$ and 0 otherwise yields Equation 1, while taking $f_j = 1$ if $c_j = c$ and 0 otherwise yields Equation 2. Equation 3 thus provides the general formulation of an exemplar model that we will analyze.

Bayesian inference

Many cognitive problems can be formulated as evaluating a set of hypotheses about processes that could have produced observed data. Bayesian inference provides a solution to problems of this kind. Letting h denote a hypothesis and d the data, assume a learner encodes his or her degrees of belief regarding the hypotheses before seeing d using a probability distribution, $p(h)$, known as the *prior* distribution. Then, the degrees of belief after seeing d are given by the *posterior* distribution, $p(h|d)$, obtained from Bayes' rule

$$p(h|d) = \frac{p(d|h)p(h)}{\int_H p(d|h)p(h)\,dh} \tag{4}$$

where H is the set of hypotheses under consideration, and $p(d|h)$ is a distribution indicating the probability of seeing d if h were true, known as the *likelihood*.

While our analysis applies to Bayesian inference in the general case, we will focus on a specific example. Assume we observe a stimulus x, which we believe to be corrupted by noise and potentially missing some accompanying information, such as a category label. Let x^* denote the uncorrupted stimulus, and z denote the missing data. If there is no missing data (i.e. z is empty), then our goal is simply to reconstruct x, finding the x^* to which it corresponds. Otherwise, we seek to infer both x^* and the value of z which corresponds to x. We can perform both tasks using Bayesian inference.

The application of Bayes' rule is easier to illustrate in the case where there is no missing data z, and we simply wish to infer x^*. We will use the probability distribution $p(x|x^*)$ to characterize the noise process, indicating the probability with which the stimulus x^* is corrupted to x, and the probability distribution $p(x^*)$ to encode our a priori beliefs about the probability of seeing a given stimulus. We can then use Bayes' rule to compute the posterior distribution over the value of the uncorrupted stimulus, x^*, which might have generated the observation x, obtaining

$$p(x^*|x) = \frac{p(x|x^*)p(x^*)}{\int p(x|x^*)p(x^*)\,dx^*}, \tag{5}$$

where $p(x|x^*)$ is the likelihood and $p(x^*)$ is the prior.

This analysis is straightforward to generalize to the case where z contains important missing data, such as the category from which x was generated. In this case, we need to define our prior as a distribution over both x^* and z, $p(x^*, z)$. We can then use Bayes' rule to compute the posterior distribution over the uncorrupted stimulus, x^*, and missing data, z, which might have generated the observation x, obtaining

$$p(x^*, z|x) = \frac{p(x|x^*)p(x^*, z)}{\int\int p(x|x^*)p(x^*, z)\,dx^*\,dz} \tag{6}$$

where we also assume that the probability of observing x is independent of z given x^*, so $p(x|x^*, z) = p(x|x^*)$.

Evaluating expectations by Monte Carlo

Posterior distributions on hypotheses given data can be used to answer a variety of questions. To return to the example above, a posterior distribution on x^* and z can be used to evaluate the properties of x^* and z given x. For any function $f(x^*, z)$, the *expectation* of that function given x is

$$E\left[f(x^*, z)|x\right] = \int\int f(x^*, z)p(x^*, z|x)\,dx^*\,dz \tag{7}$$

being the average of $f(x^*, z)$ over the posterior distribution. Since $f(x^*, z)$ can pick out any property of x^* and z that might be of interest, many problems of reasoning under uncertainty can be expressed in terms of expectations. However, evaluating expectations over the posterior distribution can be challenging: it requires computing a posterior distribution, which is a hard problem in itself, and the integrals in Equation 7 can range over many values for x^* and z. Consequently, Monte Carlo methods are often used to approximate expectations.

The Monte Carlo method approximates the expectation of a function with respect to a probability distribution with the average of that function at points drawn from the distribution. Assume we want to evaluate the expectation of a function $g(y)$ over the distribution $p(y)$, $E_p[g(y)]$. Let μ denote the value of this expectation. The law of large numbers justifies

$$\mu = E_p[g(y)] = \int g(y)p(y)\,dy \approx \frac{1}{m}\sum_{j=1}^{m} g(y_j) = \hat{\mu}_{MC} \tag{8}$$

where the y_j are all drawn from the distribution $p(y)$.

Using the Monte Carlo method requires that we are able to generate samples from the distribution $p(y)$. However, this is often not the case: it is quite common to encounter problems where $p(y)$ is known at all points y but hard to sample from. If another distribution $q(y)$ is close to $p(y)$ but easy to sample from, a form of Monte Carlo called *importance sampling* can be applied (see Neal, 1993, for a detailed introduction). Manipulating the expression for the expectation of g gives

$$\int g(y)p(y)\,dy = \frac{\int g(y)p(y)\,dy}{\int p(y)\,dy} = \frac{\int g(y)\frac{p(y)}{q(y)}q(y)\,dy}{\int \frac{p(y)}{q(y)}q(y)\,dy} \tag{9}$$

The numerator and denominator of this expression are each expectations with respect to $q(y)$. Applying simple Monte Carlo (with the same set of samples from $q(y)$) to both,

$$\mu = E_p[g(y)] \approx \frac{\sum_{j=1}^{m} g(y_j)\frac{p(y_j)}{q(y_j)}}{\sum_{j=1}^{m}\frac{p(y_j)}{q(y_j)}} = \hat{\mu}_{IS} \tag{10}$$

where each y_j is drawn from $q(y)$. The ratios $\frac{p(y_j)}{q(y_j)}$ can be viewed as "weights" on the samples y_j, correcting for having sampled from $q(y)$ rather than $p(y)$. Samples with higher probability under $p(y)$ than $q(y)$ occur less often than if we were sampling from $p(y)$, but receive greater weight.

Both simple Monte Carlo and importance sampling can be applied to the problem of evaluating the expectation of a function $f(x^*, z)$ over a posterior distribution on x^* and z with which we began this section. Simple Monte Carlo would draw values of x^* and z from the posterior distribution $p(x^*, z|x)$ directly. Importance sampling would generate from another distribution, $q(x^*, z)$, and then reweight those samples. One simple choice of $q(x^*, z)$ is the prior, $p(x^*, z)$. If we sample from the prior, the weight assigned to each sample is the ratio of the posterior to the prior

$$\frac{p(x^*, z|x)}{p(x^*, z)} = \frac{p(x|x^*)}{\int\int p(x|x^*)p(x^*, z)\,dx^*\,dz} \tag{11}$$

where we use the assumption that $p(x|x^*, z) = p(x|x^*)$. Substituting these weights into Equation 10 , we obtain

$$E[f(x^*, z)|x] \approx \frac{\sum_{j=1}^{m} f(x_j^*, z_j)p(x|x_j^*)}{\sum_{j=1}^{m} p(x|x_j^*)} \tag{12}$$

where we assume that x_j^* and z_j are drawn from $p(x^*, z)$.

Exemplar models as importance samplers

Inspection of Equations 3 and 12 yields our main result: that exemplar models can be viewed as implementing a form of importance sampling. More formally, assume X^* is a set of m exemplars x^* and associated information z drawn from the probability distribution $p(x^*, z)$, and $f_j = f(x_j^*, z_j)$ for some function $f(x^*, z)$. Then the output of Equation 3 for an exemplar model with exemplars X^* and similarity function $s(x, x^*)$ is an importance sampling approximation to the expectation of $f(x^*, z)$ over the posterior distribution on x^* and z, as given in Equation 6, for the Bayesian model with prior $p(x^*, z)$ and likelihood $p(x|x^*) \propto s(x, x^*)$.

This connection between exemplar models and importance sampling provides an alternative rational justification for exemplar models of categorization, as well as a more general motivation for these models. The justification for exemplar models in terms of nonparametric density estimation (Ashby & Alfonso-Reese, 1995) provides a clear account of their relevance to categorization, but does not explain why they are appropriate in other contexts, such as identification (Equation 1) or the general response rule given in Equation 3. In contrast, we can use importance sampling to provide a single explanation for identification, categorization, and other uses of exemplar models, viewing each as the result of approximating an expectation of a particular function $f(x^*, z)$ over the posterior distribution $p(x^*, z|x)$. For identification, z is empty and $f(x^*, z) = 1$ for all x^* within a small range ε of a specific value x_i^* and 0 otherwise. For categorization, z contains the category label, and $f(x^*, z) = 1$ for all $z = c$ and 0 otherwise. For function learning, z contains the value of the continuous variable associated with x^*, and $f(x^*, z) = z$. Similar analyses apply in other cases, with exemplar models providing a rational method for answering questions expressed as an expectation of a function of x^* and z.

Simulations

The success of importance sampling as a scheme for approximating expectations justifies using exemplar models as an approximation to Bayesian inference. In this section, we evaluate exemplar models as a scheme for approximating Bayesian inference in two tasks, examining the effect of number of exemplars on performance in order to evaluate the consequences of biological and psychological constraints.

The perceptual magnet effect

The perceptual magnet effect is a categorical effect in speech perception in which discriminability of speech sounds is reduced near phonetic category prototypes and enhanced near category boundaries, presumably due to a perceptual bias toward phonetic category centers (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Feldman and Griffiths (2007) argued that this effect can be characterized as Bayesian inference if one assumes that listeners are using their knowledge of phonetic categories to optimally recover the phonetic detail of a speaker's target production through a noisy speech signal. Here we demonstrate than an exemplar model derived through importance sampling can provide a psychologically plausible implementation of this Bayesian model, mirroring human performance with a reasonable number of exemplars.

The Bayesian model assumes that a speaker's target production T is sampled from a Gaussian phonetic category c with category mean μ_c and category variance σ_c^2 and that listeners hear a speech sound S, perturbed by articulatory and acoustic noise, that is normally distributed around the target production T with noise variance σ_S^2. The prior on target productions is therefore a mixture of Gaussians representing a language's phonetic categories,

$$p(T) = \sum_c N(\mu_c, \sigma_c^2) p(c) \tag{13}$$

and the likelihood function is a Gaussian whose variance is determined by the speech signal noise,

$$p(S|T) = N(T, \sigma_S^2) \tag{14}$$

Listeners hear the speech sound S and use Bayes' rule to compute the expectation $E[T|S]$ and optimally recover the phonetic detail of a speaker's target production.

To perform this computation using importance sampling, listeners need only store exemplars of previously encountered speech sounds, giving them a sample from $p(T)$, the prior on target productions (Equation 13)[1] . Upon hearing a new speech sound, they weight each stored exemplar by its likelihood $p(S|T)$ (Equation 14) and take the weighted average of these exemplars to approximate the posterior mean

$$E[T|S] \approx \frac{\sum_{j=1}^{m} T_j p(S|T_j)}{\sum_{j=1}^{m} p(S|T_j)} \tag{15}$$

where T_j denotes the phonetic detail (e.g. formant value) of a stored target production.

Figure 1 compares the performance of this exemplar model to multidimensional scaling data from Iverson and Kuhl (1995), who used an AX discrimination task to generate a perceptual map of thirteen equally spaced stimuli in the /i/ and /e/ categories. Model parameters are the same as those used by Feldman and Griffiths (2007). The figure shows the nonlinear mapping between psychoacoustic and perceptual space that is characteristic of the perceptual magnet effect: stimuli near the /i/ and /e/ category means are clustered together

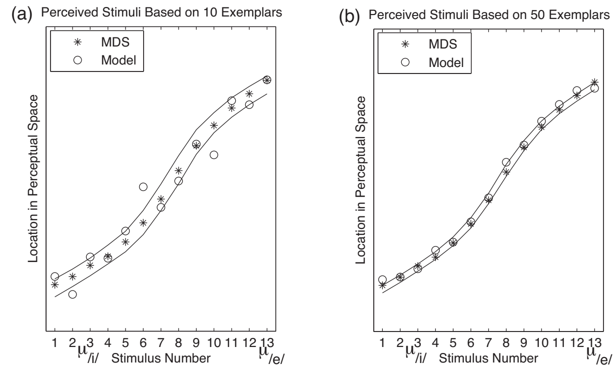

Figure 1: Locations of stimuli in perceptual space from Iverson and Kuhl's (1995) multidimensional scaling data and from a single hypothetical subject (open circles) and the middle 50% of hypothetical subjects (solid lines) using an exemplar model in which perception is based on (a) ten and (b) fifty exemplars. The labels $\mu_{/i/}$ and $\mu_{/e/}$ show the locations of category means in the model.

in perceptual space in both data and model. The simulations suggest that a relatively small number of exemplars suffices to capture human performance in this perceptual task. Model performance using ten exemplars already demonstrates the desired effect, and with fifty exemplars, the model gives a precise approximation that closely mirrors the combined performance of the 18 subjects in Iverson and Kuhl's multidimensional scaling experiment.

The exemplar model provides several advantages over the original Bayesian formulation. It allows listeners to compute speakers' target productions without explicit knowledge of phonetic categories, thereby giving a more plausible account of how six-month-olds might acquire enough information to show the perceptual magnet effect (Kuhl et al., 1992). Listeners can still compute category membership based on labeled exemplars using Equation 2, but labeled exemplars are not required in order to show perceptual warping. Furthermore, parametric knowledge of category structure is not required for either computation: Equation 15 generalizes easily to the case of non-Gaussian categories, allowing listeners to perform optimally for a range of category structures. Finally, similar exemplar-based mechanisms have previously been proposed by Guenther and Gjaja (1996) and Pierrehumbert (2001) to create a bias toward category centers, and importance sampling provides a way of integrating the Bayesian model with these exemplar-based approaches.

The number game

While the perceptual magnet effect is an example where the exemplar model is applied in a space of continuous variables (frequency in acoustic space), exemplars can also be hypotheses over a discrete space. The "number game" of Tenenbaum (1999; Tenenbaum & Griffiths, 2001) is a good example. This game is formulated as follows: given natural numbers from

[1]Exemplars in a continuous space that are acquired by sensory experience may be corrupted by noise and thus are not perfect samples from the prior. However, often exemplars still closely follow the prior distribution since such noise can be significantly reduced by averaging over repetitive identical observations and/or weighting over cues from multiple sensory modalities.

1 to 100, if number x belongs to an unknown set C (e.g., $\{59,60,61,62\}$), what is the probability that y also belongs to the same set?Here, the exemplars of interest are not numbers themselves, but sets of numbers following rules, such as squares ($\{1,4,9,16,...\}$) or natural numbers between 89 and 91 ($\{89,90,91\}$).

This problem can be addressed by Bayesian inference. Our data are the knowledge that x belongs to the set C, and our hypotheses concern the nature of C. Since C is unknown, we should sum over all possible hypotheses $h \in H$ when evaluating whether y belongs to C,

$$p(y \in C|x) = \sum_{h \in H} p(y \in C|h)p(h|x) = \sum_{h \in H} \mathbf{1}(y \in h)p(h|x) \quad (16)$$

where $\mathbf{1}(y \in h)$ is the indicator function of the statement $y \in h$, taking value 1 if this is true and 0 otherwise. In the analysis presented by Tenenbaum (1999; Tenenbaum & Griffiths, 2001), the likelihood $p(x|h)$ is proportional to the inverse of the size of h (the "size principle") being $1/|h|$ if $x \in h$ and 0 otherwise. A broad range of hypotheses were used, including intervals of numbers spanning a certain range, even numbers, odd numbers, primes, and cubes.

The number game is challenging because any given number (say $x = 8$) is consistent with many hypotheses (not only intervals containing 8, but also hypotheses such as even numbers, cube numbers, number with ending digit 8, etc.). Interestingly, the responses of human participants can be captured quite accurately with this Bayesian model (Figure 2 (a)). However, this involves instantiating all 6,412 hypotheses, calculating the likelihood for each rule and integrating over the product of the prior and likelihood. Human subjects are not likely to perform such computations given limitations on memory capacity and computational power, so a mechanism that approximates the exact solution is desirable.

Performing the computations involved in the number game requires extending our analysis of exemplar models to the general case of Bayesian inference. We can do this by replacing the role of exemplars in the preceding analysis with hypotheses sampled from the prior $p(h)$. These hypotheses are activated in response to how well they explain the data, with activation proportional to $p(x|h)$. Averaging any function of h over the distribution defined by normalizing the activations will be an importance sampler for the expectation of that function over the posterior, $p(h|c)$. Thus, storing a few hypotheses in memory and activating those hypotheses in response to data provides a psychologically plausible mechanism for performing Bayesian inference.

We can now apply this framework to the number game. Equation 16 is an expectation of an indicator function over the posterior distribution $p(h|x)$. This expectation can be approximated using a set of m hypotheses $h_1, \ldots, h_m$ sampled from the prior and activated in proportion to the likelihood,

$$p(y \in C|x) \approx \frac{\sum_j \mathbf{1}(y \in h_j)p(x|h_j)}{\sum_j p(x|h_j)} \quad (17)$$

meaning that $p(y \in C|x)$ is just the ratio of the summed likelihoods of the hypotheses stored in memory that generate y to the summed likelihoods of all hypotheses stored in memory. Considering limitations in memory capacity and computational power, we conducted two sets of simulations. In the computation-limited case, the bottleneck is the number of exemplars that can be processed simultaneously, but not the supply of qualified hypotheses, being those hypotheses such that $x \in h$. In contrast, the memory-limited case assumes that only a limited number of hypotheses are stored in memory and those exemplars are not necessarily qualified. When the right hypothesis is missing (say cubes for $\{1,8,27,64\}$), the exemplar model gives incorrect results, as when a person fails to recognize the underlying rule. Our simulations use the same parameters as those in Tenenbaum (1999) except that the likelihood function assigns a small non-zero probability to all natural numbers from 1 to 100 for every hypothesis to ensure numerical stability.

Figure 2 (b) and (c) show a single hypothetical subject's responses to the number game. The results suggest a small number of exemplars (20 and 50) is sufficient to account for human performance. The memory limited case needs more exemplars because not all exemplars are qualified hypothesis. Therefore, the effective number of exemplars, which determines the computational load, is small. The consistency of these results with the human judgments indicates that this kind of generalized exemplar model provides a plausible mechanism for performing Bayesian inference that relies on reasonable memory and computational resources and can be used with highly structured hypothesis spaces.

Conclusion

Our theoretical results indicate that exemplar models can be interpreted as a form of importance sampling, and thus provide a simple psychological mechanism for producing behavior consistent with Bayesian inference. Our simulations demonstrate that this approach produces predictions that correspond reasonably well with human behavior and that relatively few exemplars are needed to provide a good approximation to the true Bayesian solution to a simple problem. These simulations also highlight the flexibility of this approach, since exactly the same model can be used to make predictions regardless of the form of the prior.

The approach that we have taken in this paper represents one way of addressing questions about the mechanisms that could support probabilistic inference. Our results suggest that exemplar models are not simply process models, but a kind of "rational process model" – an effective and psychologically plausible scheme for approximating statistical inference. This approach pushes the principle of optimality that underlies probabilistic models down to the level of mechanism, and suggests a general strategy for explaining how people perform Bayesian inference: look for connections between psychological process models and approximate inference algorithms developed in computer science and statistics.

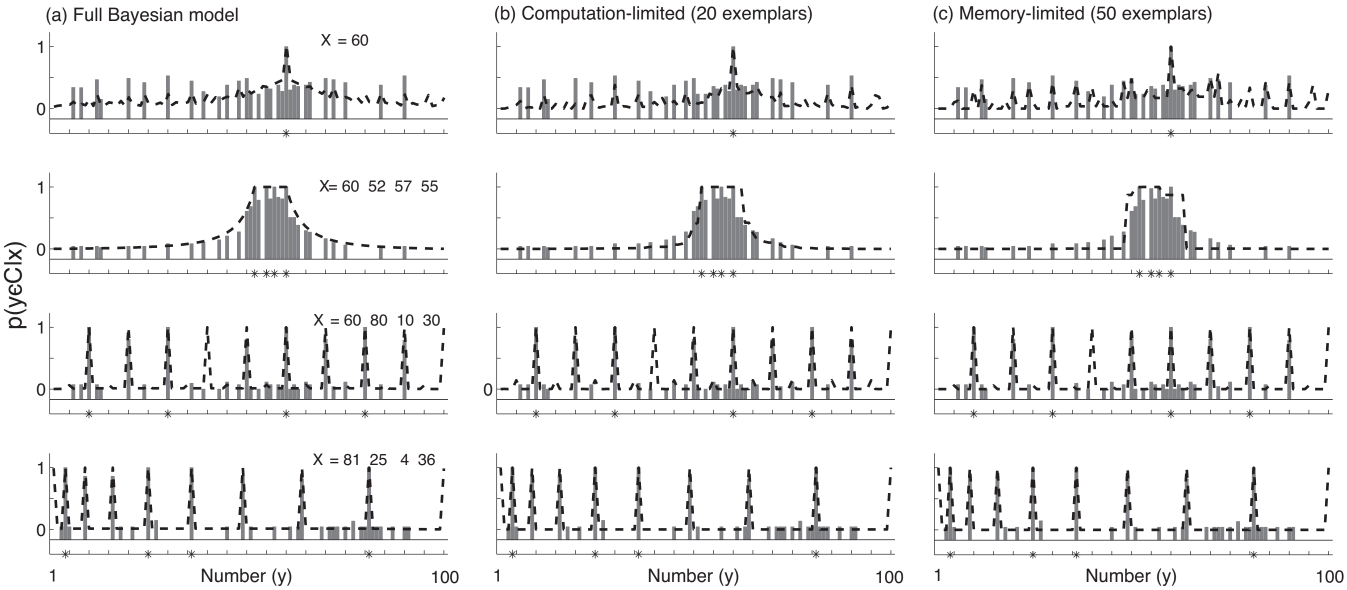

Figure 2: Simulations (dashed line) and behavioral data from Tenenbaum (1999) (gray bars) for the number game. The full Bayesian model uses 6412 hypotheses. Results of computation-limited and memory-limited exemplar models are based on a single hypothetical subject with a single set of hypotheses (exemplars) sampled from the prior. Models are tested under conditions suggesting single point generalization $x = 60$, a consecutive interval $x = \{60, 52, 57, 55\}$, multiples of 10 $x = \{60, 80, 10, 30\}$ and squares $x = \{81, 25, 4, 36\}$.

Acknowledgments. This research was supported by grant FA9550-07-1-0351 from the Air Force Office of Scientific Research (LS,TLG) and grant HD32005 from the National Institute of Health (NHF).

References

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703-719.

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

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non of abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 968-986.

Feldman, N. H., & Griffiths, T. L. (2007). A rational account of the perceptual magnet effect. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (p. 257-262). Austin, TX: Cognitive Science Society.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*(2), 1111-1121.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553-562.

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): a lazy algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563-607.

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

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

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

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

Ma, W. J., Beck, J., Latham, P., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432-1438.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods* (Tech. Rep. No. CRG-TR-93-1). University of Toronto: University of Toronto.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608-631.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*, 3-21.

Tenenbaum, J. B. (1999). *A Bayesian framework for concept learning*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629-641.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*, 301-308.