Research Article

Optimal Predictions in Everyday Cognition

Thomas L. Griffiths[1] and Joshua B. Tenenbaum[2]

[1]Department of Cognitive and Linguistic Sciences, Brown University, and [2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

ABSTRACT—***Human perception and memory are often explained as optimal statistical inferences that are informed by accurate prior probabilities. In contrast, cognitive judgments are usually viewed as following error-prone heuristics that are insensitive to priors. We examined the optimality of human cognition in a more realistic context than typical laboratory studies, asking people to make predictions about the duration or extent of everyday phenomena such as human life spans and the box-office take of movies. Our results suggest that everyday cognitive judgments follow the same optimal statistical principles as perception and memory, and reveal a close correspondence between people's implicit probabilistic models and the statistics of the world.***

If you were assessing the prospects of a 60-year-old man, how much longer would you expect him to live? If you were an executive evaluating the performance of a movie that had made $40 million at the box office so far, what would you estimate for its total gross? Everyday life routinely poses such challenges of prediction, situations in which the true answer cannot be determined on the basis of the limited data available, yet common sense suggests at least a reasonable guess. Analogous inductive problems—for example, identifying the three-dimensional structure underlying a two-dimensional image (Freeman, 1994; Knill & Richards, 1996) or judging when a particular fact is likely to be needed in the future (Anderson, 1990; Anderson & Milson, 1989)—arise in many domains of human psychology. Accounts of human perception and memory suggest that these systems effectively approximate optimal statistical inference, correctly combining new data with an accurate probabilistic model of the environment (Anderson, 1990; Anderson & Milson, 1989; Anderson & Schooler, 1991; Freeman, 1994; Geisler, Perry, Super, & Gallogly, 2001; Huber, Shiffrin, Lyle, & Ruys, 2001; Knill & Richards, 1996; Körding & Wolpert, 2004; Shiffrin & Steyvers, 1997; Simoncelli & Olshausen, 2001; Weiss, Simoncelli, & Adelson, 2002). In contrast—perhaps as a result of the great attention garnered by the work of Kahneman, Tversky, and their colleagues (e.g., Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974)—cognitive judgments under uncertainty are often characterized as the result of error-prone heuristics that are insensitive to prior probabilities. This view of cognition, based on laboratory studies, appears starkly at odds with the near-optimality of other human capacities, and with people's ability to make smart predictions from sparse data in the real world.

Address correspondence to Thomas Griffiths, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, e-mail: tom_griffiths@brown.edu.

To evaluate how cognitive judgments compare with optimal statistical inferences in real-world settings, we asked people to predict the duration or extent of everyday phenomena such as human life spans and the gross of movies. We varied the phenomena that were described and the amount of data available, and we compared the predictions of human participants with those of an optimal Bayesian model, described in detail in the appendix. Here, we illustrate the principles behind this Bayesian analysis by taking the example of trying to predict the total life span of a man we have just met, on the basis of the man's current age. If t_{total} indicates the total amount of time the man will live and t indicates his current age, the task is to estimate t_{total} from t. The Bayesian predictor computes a probability distribution over t_{total} given t, by applying Bayes's rule:

$$p(t_{\text{total}}|t) \propto p(t|t_{\text{total}})p(t_{\text{total}}) \tag{1}$$

The probability assigned to a particular value of t_{total} given t is proportional to the product of two factors: the likelihood $p(t|t_{\text{total}})$ and the prior probability $p(t_{\text{total}})$.

The likelihood is the probability of first encountering a man at age t given that his total life span is t_{total}. Assuming for simplicity that we are equally likely to meet a man at any point in his life, this probability is uniform, $p(t|t_{\text{total}}) = 1/t_{\text{total}}$, for all possible values of t between 0 and t_{total} (and 0 for values outside that range). This assumption of uniform random sampling is analogous to the Copernican anthropic principle in Bayesian cosmology (Buch, 1994; Caves, 2000; Garrett & Coles, 1993; Gott, 1993, 1994; Ledford, Marriott, & Crowder, 2001) and the generic-view principle in Bayesian models of visual perception (Freeman, 1994; Knill & Richards, 1996). The prior probability $p(t_{\text{total}})$ reflects our general expectations about the relevant class of events—in this case, about how likely it is that a man's life span will be t_{total}. Analysis of actuarial data shows that the distribution of life spans in our society is (ignoring infant mortality) approximately Gaussian—normally distributed—with a mean, μ, of about 75 years and a standard deviation, σ, of about 16 years.

Combining the prior with the likelihood according to Equation 1 yields a probability distribution $p(t_{\text{total}}|t)$ over all possible total life spans t_{total} for a man encountered at age t. A good guess for t_{total} is the median of this distribution—that is, the point at which it is equally likely that the true life span is longer or shorter. Taking the median of $p(t_{\text{total}}|t)$ defines a Bayesian *prediction function*, specifying a predicted value of t_{total} for each observed value of t. Prediction functions for events with Gaussian priors are nonlinear: For values of t much less than the mean of the prior, the predicted value of t_{total} is approximately the mean; once t approaches the mean, the predicted value of t_{total} increases slowly, converging to t as t increases but always remaining slightly higher, as shown in Figure 1. Although its mathematical form is complex, this prediction function makes intuitive sense for human life spans: A predicted life span of about 75 years would be reasonable for a man encountered at age 18, 39, or 51; if we met a man at age 75, we might be inclined to give him several more years at least; but if we met someone at age 96, we probably would not expect him to live much longer.

This approach to prediction is quite general, applicable to any problem that requires estimating the upper limit of a duration, extent, or other numerical quantity given a sample drawn from that interval (Buch, 1994; Caves, 2000; Garrett & Coles, 1993; Gott, 1993, 1994; Jaynes, 2003; Jeffreys, 1961; Ledford et al., 2001; Leslie, 1996; Maddox, 1994; Shepard, 1987; Tenenbaum & Griffiths, 2001). However, different priors will be appropriate for different kinds of phenomena, and the prediction function will vary substantially as a result. For example, imagine trying to predict the total box-office gross of a movie given its take so far. The total gross of movies follows a power-law distribution, with $p(t_{\text{total}}) \propto t_{\text{total}}^{-\gamma}$ for some $\gamma > 0$.[1] This distribution has a highly non-Gaussian shape (see Fig. 1), with most movies taking in only modest amounts, but occasional blockbusters making huge amounts of money. In the appendix, we show that for power-law priors, the Bayesian prediction function picks a value for t_{total} that is a multiple of the observed sample t. The exact multiple depends on the parameter γ. For the particular power law that best fits the actual distribution of movie grosses, an optimal Bayesian observer would estimate the total gross to be approximately 50% greater than the current gross: Thus, if we observe a movie has made \$40 million to date, we should guess a total gross of around \$60 million; if we observe a current gross of only \$6 million, we should guess about \$9 million for the total.

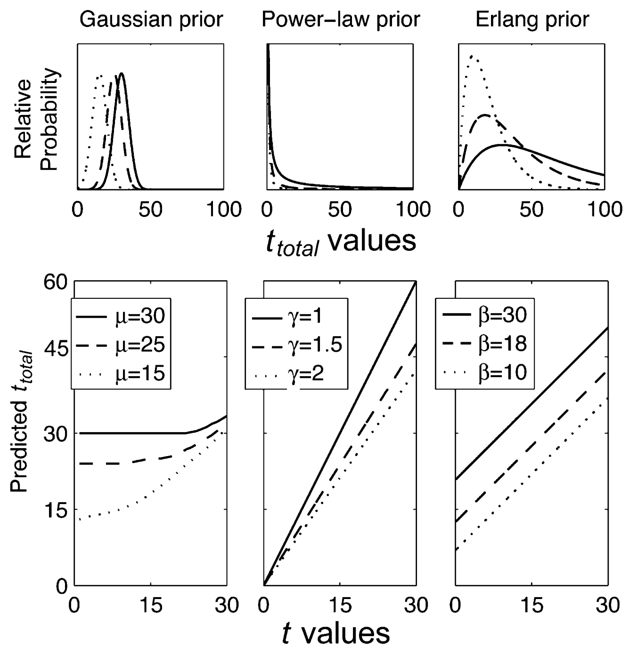

Fig. 1. Bayesian prediction functions and their associated prior distributions. The three columns represent qualitatively different statistical models appropriate for different kinds of events. The top row of plots shows three parametric families of prior distributions for the total duration or extent, t_{total}, that could describe events in a particular class. Lines of different styles represent different parameter values (e.g., different mean durations) within each family. The bottom row of plots shows the optimal predictions for t_{total} as a function of t, the observed duration or extent of an event so far, assuming the prior distributions shown in the top panel. For Gaussian priors (left column), the prediction function always has a slope less than 1 and an intercept near the mean μ: Predictions are never much smaller than the mean of the prior distribution, nor much larger than the observed duration. Power-law priors (middle column) result in linear prediction functions with variable slope and a zero intercept. Erlang priors (right column) yield a linear prediction function that always has a slope equal to 1 and a nonzero intercept.

Although such constant-multiple prediction rules are optimal for event classes that follow power-law priors, they are clearly inappropriate for predicting life spans or other kinds of events with Gaussian priors. For instance, upon meeting a 10-year-old girl and her 75-year-old grandfather, we would never predict that the girl will live a total of 15 years (1.5×10) and the grandfather will live to be 112 (1.5×75). Other classes of priors, such as the exponential-tailed Erlang distribution, $p(t_{\text{total}}) \propto t_{\text{total}}\exp(-t_{\text{total}}/\beta)$ for $\beta > 0$,[2] are also associated with distinctive optimal prediction functions. For the Erlang distribution, the

[1] When $\gamma > 1$, a power-law distribution is often referred to in statistics and economics as a Pareto distribution.

[2] The Erlang distribution is a special case of the gamma distribution. The gamma distribution is $p(t_{\text{total}}) \propto t_{\text{total}}^{k-1}\exp(-t_{\text{total}}/\beta)$, where $k > 0$ and $\beta > 0$ are real numbers. The Erlang distribution assumes that k is an integer. Following Shepard (1987), we use a one-parameter Erlang distribution, fixing k at 2.

TABLE 1

Sources of Data for Estimating Prior Distributions

Data set	Source (number of data points)
Movie grosses	http://www.worldwideboxoffice.com/ (5,302)
Poem lengths	http://www.emule.com/ (1,000)
Life spans	http://www.demog.berkeley.edu/wilmoth/mortality/states.html (complete life table)
Movie run times	http://www.imdb.com/charts/usboxarchive/ (233 top-10 movies from 1998 through 2003)
U.S. representatives' terms	http://www.bioguide.congress.gov/ (2,150 members since 1945)
Cake baking times	http://www.allrecipes.com/ (619)
Pharaohs' reigns	http://www.touregypt.com/ (126)

Note. Data were collected from these Web sites between July and December 2003.

best guess of t_{total} is simply t plus a constant determined by the parameter β, as shown in the appendix and illustrated in Figure 1.

Our experiment compared these ideal Bayesian analyses with the judgments of a large sample of human participants, examining whether people's predictions were sensitive to the distributions of different quantities that arise in everyday contexts. We used publicly available data to identify the true prior distributions for several classes of events (the sources of these data are given in Table 1). For example, as shown in Figure 2, human life spans and the run time of movies are approximately Gaussian, the gross of movies and the length of poems are approximately power-law distributed, and the distributions of the number of years in office for members of the U.S. House of Representatives and of the length of the reigns of pharaohs are approximately Erlang. The experiment examined how well people's predictions corresponded to optimal statistical inference in these different settings.

METHOD

Participants and Procedure

Participants were tested in two groups, with each group making predictions about five different phenomena. One group of 208 undergraduates made predictions about movie grosses, poem lengths, life spans, reigns of pharaohs, and lengths of marriages. A second group of 142 undergraduates made predictions about movie run times, terms of U.S. representatives, baking times for cakes, waiting times, and lengths of marriages. The surveys were

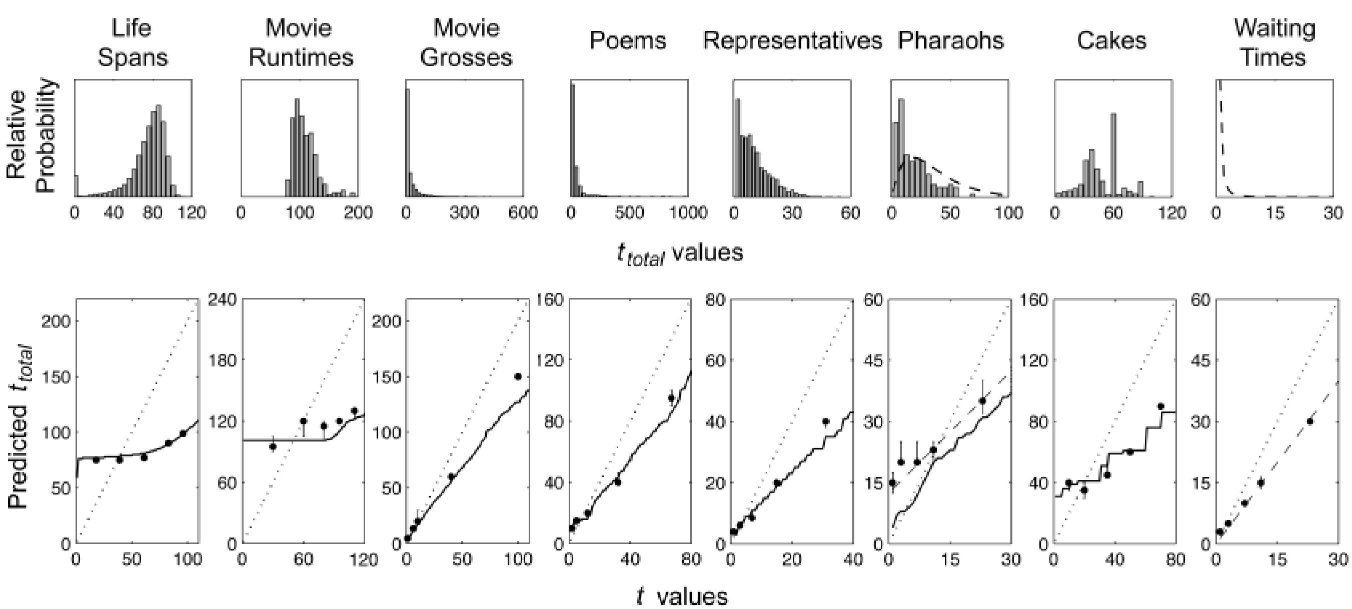

Fig. 2. People's predictions for various everyday phenomena. The top row of plots shows the empirical distributions of the total duration or extent, t_{total}, for each of these phenomena. The first two distributions are approximately Gaussian, the third and fourth are approximately power-law, and the fifth and sixth are approximately Erlang. The bottom row shows participants' predicted values of t_{total} for a single observed sample t of a duration or extent for each phenomenon. Black dots show the participants' median predictions of t_{total}. Error bars indicate 68% confidence intervals (estimated by a 1,000-sample bootstrap). Solid lines show the optimal Bayesian predictions based on the empirical prior distributions shown above. Dashed lines show predictions made by estimating a subjective prior, for the pharaohs and waiting-times stimuli, as explained in the main text. Dotted lines show predictions based on a fixed uninformative prior (Gott, 1993).

included in a booklet that participants completed for a set of unrelated experiments.

Materials

Each participant made a prediction about one instance from each of the five different classes seen by his or her group. Each prediction was based on one of five possible values of t, varied randomly between subjects. These values were \$1, \$6, \$10, \$40, and \$100 million for movie grosses; 2, 5, 12, 32, and 67 lines for poem lengths; 18, 39, 61, 83, and 96 years for life spans; 1, 3, 7, 11, and 23 years for reigns of pharaohs; 1, 3, 7, 11, and 23 years for lengths of marriages; 30, 60, 80, 95, and 110 min for movie run times; 1, 3, 7, 15, and 31 years for terms of U.S. representatives; 10, 20, 35, 50, and 70 min for baking times for cakes; and 1, 3, 7, 11, and 23 min for waiting times. In each case, participants read several sentences establishing context and then were asked to predict t_{total} given t.

The questions were presented in survey format. Each survey began as follows:

> Each of the questions below asks you to predict something—either a duration or a quantity—based on a single piece of information. Please read each question and write your prediction on the line below it. We're interested in your intuitions, so please don't make complicated calculations—just tell us what you think!

Each question was then introduced with a couple of sentences to provide a context. Following are sample questions:

> *Movie grosses:* Imagine you hear about a movie that has taken in 10 million dollars at the box office, but don't know how long it has been running. What would you predict for the total amount of box office intake for that movie?

> *Poem lengths:* If your friend read you her favorite line of poetry, and told you it was line 5 of a poem, what would you predict for the total length of the poem?

> *Life spans:* Insurance agencies employ actuaries to make predictions about people's life spans—the age at which they will die—based upon demographic information. If you were assessing an insurance case for an 18-year-old man, what would you predict for his life span?

> *Reigns of pharaohs:* If you opened a book about the history of ancient Egypt to a page listing the reigns of the pharaohs, and noticed that at 4000 BC a particular pharaoh had been ruling for 11 years, what would you predict for the total duration of his reign?

> *Lengths of marriages:* A friend is telling you about an acquaintance whom you do not know. In passing, he happens to mention that this person has been married for 23 years. How long do you think this person's marriage will last?

> *Movie run times:* If you made a surprise visit to a friend, and found that they had been watching a movie for 30 minutes, what would you predict for the length of the movie?

> *Terms of U.S. representatives:* If you heard a member of the House of Representatives had served for 15 years, what would you predict his total term in the House would be?

> *Baking times for cakes:* Imagine you are in somebody's kitchen and notice that a cake is in the oven. The timer shows that it has been baking for 35 minutes. What would you predict for the total amount of time the cake needs to bake?

> *Waiting times:* If you were calling a telephone box office to book tickets and had been on hold for 3 minutes, what would you predict for the total time you would be on hold?

RESULTS

We first filtered out responses that could not be analyzed or that indicated a misunderstanding of the task, removing predictions that did not correspond to numerical values or were less than t_{total}. Only a small minority of responses failed to meet these criteria, except in the case of the marriage predictions. The total number of responses analyzed was 174 for movie grosses, 197 for poem lengths, 197 for life spans, 191 for reigns of pharaohs, 136 for movie run times, 130 for terms of U.S. representatives, 126 for baking times for cakes, and 158 for waiting times. The responses for the marriage stimuli were problematic because the majority of participants (52%) indicated that marriages last "forever." This accurately reflects the proportion of marriages that do not end in divorce (Kreider & Fields, 2002), but prevented us from analyzing the data using methods based on median values. We therefore did not analyze responses for the marriage stimuli further.

People's judgments for life spans, movie run times, movie grosses, poem lengths, and terms of U.S. representatives were indistinguishable from optimal Bayesian predictions based on the empirical prior distributions, as shown in Figure 2. People's prediction functions took on very different shapes in domains characterized by Gaussian, power-law, and Erlang priors, just as expected under the ideal Bayesian analysis. Notably, the model predictions shown in Figure 2 have no free parameters tuned specifically to fit the human data, but are simply the optimal functions prescribed by Bayesian inference given the relevant world statistics. These results are inconsistent with claims that cognitive judgments are based on non-Bayesian heuristics that are insensitive to priors (Kahneman et al., 1982; Tversky & Kahneman, 1974). The results are also inconsistent with simpler Bayesian prediction models that adopt a single uninformative prior, $p(t_{\text{total}}) \propto 1/t_{\text{total}}$, regardless of the phenomenon to be predicted (Gott, 1993, 1994; Jaynes, 2003; Jeffreys, 1961; Ledford et al., 2001).

Examining the results for the remaining stimuli—reigns of pharaohs, baking times for cakes, and waiting times—provides an opportunity to learn about the limits of people's capacity for prediction. As shown in Figure 2, people's predictions about the reigns of pharaohs had a form consistent with the appropriate prior (an Erlang distribution), but were slightly too high. We established people's subjective priors for the reigns of pharaohs in a follow-up experiment, asking 35 undergraduates to state the typical duration of a pharaoh's reign. The median response was 30 years, which corresponds to an Erlang prior on t_{total} with parameter β equal to 17.9, as opposed to the true value of approximately 9.34. Using this subjective Erlang prior produces a close correspondence to the human judgments.

The pharaohs stimuli provide an instance of a situation in which people make inaccurate predictions: when they know the appropriate form for the prior, but not the details of its parameters. In contrast, responses to the cakes stimuli reveal that people can make accurate predictions even in contexts in which priors lack a simple form. The duration a cake should spend in the oven is a quantity that follows a rather irregular distribution, as shown in Figure 2. However, people's judgments were still close to the ideal Bayesian predictions, despite the complex form of the empirical prior distribution.

These results suggest that people's predictions can also be used to identify the prior beliefs that inform them. The waiting-times stimuli provide an opportunity to explore this possibility. The true distribution of waiting times in queues is currently a controversial question in operations research. Traditional models, based on the Poisson process, assume that waiting times follow a distribution with exponential tails (e.g., Hillier & Lieberman, 2001). However, several recent analyses suggest that in many cases, waiting times may be better approximated by a power-law distribution (Barabási, 2005, provides a summary and explanation of these findings). Hence, it is not clear what the objective distribution of durations should be for these stimuli. Rather than using objective statistics on real-world durations to assess the optimality of people's judgments, as we did for the other stimulus classes, we used people's judgments for these stimuli to assess which distributional form they assumed the phenomenon would follow. We fit prediction functions for Gaussian, power-law, and Erlang distributions to the behavioral data, attempting to minimize the sum of the squared differences between the median human judgments and the predicted values of t_{total}. The power-law prior with $\gamma = 2.43$ provided the best fit to the human judgments, producing the predictions shown in Figure 2. Assuming that people's predictions are near-optimal with respect to the true distribution of durations, these results are qualitatively consistent with recent power-law models for waiting-time distributions (Barabási, 2005).

DISCUSSION

The results of our experiment reveal a far closer correspondence between optimal statistical inference and everyday cognition than suggested by previous research. People's judgments were close to the optimal predictions produced by our Bayesian model across a wide range of settings. These judgments also served as a guide to people's implicit beliefs about the distributions of everyday quantities, and revealed that these beliefs are surprisingly consistent with the statistics of the world. This finding parallels formal analyses of perception and memory, in which accurate probabilistic models of the environment play a key role in the solution of inductive problems (Anderson, 1990; Anderson & Milson, 1989; Anderson & Schooler, 1991; Freeman, 1994; Geisler et al., 2001; Huber et al., 2001; Knill & Richards, 1996; Körding & Wolpert, 2004; Shiffrin & Steyvers, 1997; Simoncelli & Olshausen, 2001; Weiss et al., 2002).

Although people's predictions about everyday events were on the whole extremely accurate, the cases in which their predictions deviated from optimality may help to shed light on the implicit assumptions and strategies that make these intuitive judgments so successful most of the time in the real world. One interesting hypothesis concerning such strategies is suggested by the pattern of people's errors in predicting the reigns of pharaohs. Both the magnitude of errors and the variance in judgments across participants were substantially greater for this question than for our other questions. This should not be surprising, as most participants probably had far less direct experience with the reigns of pharaohs than with the other kinds of scenarios we presented. Despite this lack of direct experience, people's predictions were not completely off the mark: Their judgments were consistent with having implicit knowledge of the correct form of the underlying distribution but making incorrect assumptions about how this form should be parameterized (i.e., its mean value).

The predictions for the reigns of pharaohs suggest a general strategy people might employ to make predictions about unfamiliar kinds of events, which is surely an important prediction problem faced in everyday life. Given an unfamiliar prediction task, people might be able to identify the appropriate form of the distribution by making an analogy to more familiar phenomena in the same broad class, even if they do not have sufficient direct experience to set the parameters of that distribution accurately. For instance, participants might have been familiar with the length of time that various modern monarchs have spent in their positions, as well as with the causes (e.g., succession, death) responsible for curtailing those times, and it is not unreasonable to think that analogous mechanisms could have governed the durations of pharaohs' reigns in ancient Egypt. Yet most people might not be aware of (or might not remember) just how short life spans typically were in ancient Egypt compared with modern expectations, even if they know life spans were somewhat shorter. If participants predicted the reign of the pharaoh by drawing an analogy to modern monarchs and adjusting the mean reign duration downward by some uncertain but insufficient factor, that would be entirely consistent with the pattern of errors we observed. Such a strategy of prediction by analogy could be an adaptive way of making judgments that would otherwise lie beyond people's limited base of knowledge and experience.

The finding of optimal statistical inference in an important class of cognitive judgments resonates with a number of recent suggestions that Bayesian statistics may provide a general framework for analyzing human inductive inferences. Bayesian models require making the assumptions of a learner explicit. By exploring the implications of different assumptions, it becomes possible to explain many of the interesting and apparently inexplicable aspects of human reasoning (e.g., McKenzie, 2003). The ability to combine accurate background knowledge about the world with rational statistical updating is critical in many aspects of higher-level cognition. Bayesian models have been proposed for learning words and concepts (Tenenbaum, 1999), forming generalizations about the properties of objects (Anderson, 1990; Shepard, 1987; Tenenbaum & Griffiths, 2001), and discovering logical or causal relations (Anderson, 1990; Griffiths & Tenenbaum, 2006; Oaksford & Chater, 1994). However, these modeling efforts have not typically attempted to establish optimality in real-world environments. Our results demonstrate that, at least for a range of everyday prediction tasks, people effectively adopt prior distributions that are accurately calibrated to the statistics of relevant events in the world. Assessing the scope and depth of the correspondence between probabilities in the mind and those in the world presents a fundamental challenge for future work.

Acknowledgments—We thank Liz Baraff and Onny Chatterjee for their assistance in running the experiments, and Mira Bernstein, Daniel Casasanto, Nick Chater, David Danks, Peter Dayan, Reid Hastie, Konrad Körding, Tania Lombrozo, Rebecca Saxe, Marty Tenenbaum, and an anonymous reviewer for comments on the manuscript. The second author was supported by the Paul E. Newton Chair.

REFERENCES

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

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

Anderson, J.R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396–408.

Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*, 207–211.

Buch, P. (1994). Future prospects discussed. *Nature*, *368*, 107–108.

Caves, C.M. (2000). Predicting future duration from present age: A critical assessment. *Contemporary Physics*, *41*, 143–153.

Freeman, W.T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, *368*, 542–545.

Garrett, A.J.M., & Coles, P. (1993). Bayesian inductive inference and the anthropic principles. *Comments on Astrophysics*, *17*, 23–47.

Geisler, W.S., Perry, J.S., Super, B.J., & Gallogly, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Gott, J.R. (1993). Implications of the Copernican principle for our future prospects. *Nature*, *363*, 315–319.

Gott, J.R. (1994). Future prospects discussed. *Nature*, *368*, 108.

Griffiths, T.L., & Tenenbaum, J.B. (2006). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354–384.

Hillier, F.S., & Lieberman, G.J. (2001). *Introduction to operations research* (7th ed.). New York: McGraw Hill.

Huber, D.E., Shiffrin, R.M., Lyle, K.B., & Ruys, K.I. (2001). Perception and preference in short-term word priming. *Psychological Review*, *108*, 149–182.

Jaynes, E.T. (2003). *Probability theory: The logic of science*. Cambridge, England: Cambridge University Press.

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

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

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

Kreider, R.M., & Fields, J.M. (2002). *Number, timing, and duration of marriages and divorces: 1996* (Current Population Reports P70-80). Washington, DC: U.S. Census Bureau.

Ledford, A., Marriott, P., & Crowder, M. (2001). Lifetime prediction from only present age: Fact or fiction? *Physics Letters A*, *280*, 309–311.

Leslie, J. (1996). *The end of the world: The ethics and science of human extinction*. London: Routledge.

Maddox, J. (1994). Star masses and Bayesian probability. *Nature*, *371*, 649.

McKenzie, C.R.M. (2003). Rational models as theories—not standards—of behavior. *Trends in Cognitive Sciences*, *7*, 403–406.

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

Shepard, R.N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

Simoncelli, E.P., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Tenenbaum, J.B. (1999). Bayesian modeling of human concept learning. In M.S. Kearns, S.A. Solla, & D.A. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 59–65). Cambridge, MA: MIT Press.

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

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

Weiss, Y., Simoncelli, E.P., & Adelson, E.H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*, 598–604.

(RECEIVED 5/31/05; REVISION ACCEPTED 12/7/05; FINAL MATERIALS RECEIVED 1/30/06)

APPENDIX

The Prediction Problem

Assume that a point t is sampled uniformly at random from the interval $[0, t_{total}]$. What should we guess for the value of t_{total}? A

Bayesian solution to this problem involves computing the posterior distribution over t_{total} given t. Applying Bayes's rule, this posterior distribution is

$$p(t_{\text{total}}|t) = \frac{p(t|t_{\text{total}})p(t_{\text{total}})}{p(t)} \tag{A1}$$

where

$$p(t) = \int_0^\infty p(t|t_{\text{total}})p(t_{\text{total}})dt_{\text{total}}. \tag{A2}$$

By the assumption that t is sampled uniformly at random, $p(t|t_{\text{total}}) = 1/t_{\text{total}}$ for $t_{\text{total}} \geq t$ and 0 otherwise. Equation A2 thus simplifies to

$$p(t) = \int_t^\infty \frac{p(t_{\text{total}})}{t_{\text{total}}} dt_{\text{total}} \tag{A3}$$

The form of the posterior distribution for any given value of t is thus determined entirely by the prior, $p(t_{\text{total}})$.

We can derive an analytic form for the posterior distribution obtained with power-law and Erlang priors. The posterior distribution resulting from the Gaussian prior has no simple analytic form. With the power-law prior, $p(t_{\text{total}}) \propto t_{\text{total}}{}^{-\gamma}$ for $\gamma > 0$. This prior is improper if $\gamma \leq 1$, because the integral over t_{total} diverges, but the posterior remains a proper probability distribution regardless. Applying Equation A3, we have

$$\begin{aligned} p(t) &\propto \int_t^\infty t_{\text{total}}^{-(\gamma+1)} dt_{\text{total}} \\ &= -\frac{1}{\gamma} t_{\text{total}}^{-\gamma}\Big|_t^\infty \\ &= \frac{1}{\gamma} t^{-\gamma}, \end{aligned}$$

where the constant of proportionality remains the same as in the original prior. We can substitute this result into Bayes's rule (Equation A1) to obtain

$$\begin{aligned} p(t_{\text{total}}|t) &= \frac{t_{\text{total}}^{-(\gamma+1)}}{\frac{1}{\gamma} t^{-\gamma}} \\ &= \frac{\gamma t^\gamma}{t_{\text{total}}^{\gamma+1}}, \end{aligned} \tag{A4}$$

for all values of $t_{\text{total}} \geq t$.

Under the Erlang prior, $p(t_{\text{total}}) \propto t_{\text{total}} \exp(-t_{\text{total}}/\beta)$, we have

$$\begin{aligned} p(t) &\propto \int_0^\infty \exp\{-t_{\text{total}}/\beta\}\, dt_{\text{total}} \\ &= -\beta \exp\{-t_{\text{total}}/\beta\})|_t^\infty \\ &= \beta \exp\{-t/\beta\} \end{aligned}$$

where the constant of proportionality remains the same as in the original prior. Again, we can substitute this result into Bayes's rule (Equation A1) to obtain

$$\begin{aligned} p(t_{\text{total}}|t) &= \frac{\exp\{-t_{\text{total}}/\beta\}}{\beta \exp\{-t/\beta\}} \\ &= \frac{1}{\beta} \exp\{-(t_{\text{total}} - t)/\beta\}, \end{aligned} \tag{A5}$$

for all values of $t_{\text{total}} \geq t$.

Predicting t_{total}

We take the predicted value of t_{total}, which we denote t^*, to be the posterior median. This is the point t^* such that $P(t_{\text{total}} > t^*|t) = .5$: A Bayesian predictor believes that there is a 50% chance that the true value of t_{total} is greater than t^* and a 50% chance that the true value of t_{total} is less than t^*. This point can be computed from the posterior, using the fact that

$$P(t_{\text{total}} > t^*|t) = \int_{t^*}^\infty p(t_{\text{total}}|t)dt_{\text{total}}. \tag{A6}$$

We can derive t^* analytically in the case of a power-law or Erlang prior. For the power-law prior, we can use Equation A4 to rewrite Equation A6 as

$$\begin{aligned} P(t_{\text{total}} > t^*|t) &= \int_{t^*}^\infty \frac{\gamma t^\gamma}{t_{\text{total}}^{\gamma+1}} dt_{\text{total}} \\ &= -\left(\frac{t}{t_{\text{total}}}\right)^\gamma \Big|_{t^*}^\infty \\ &= \left(\frac{t}{t^*}\right)^\gamma. \end{aligned} \tag{A7}$$

We can now solve for t^* such that $P(t_{\text{total}} > t^*|t) = .5$, obtaining $t^* = 2^{1/\gamma}t$. For the Erlang prior, we can use Equation A5 to rewrite Equation A6 as

$$\begin{aligned} P(t_{\text{total}} > t^*|t) &= \int_{t^*}^\infty \frac{1}{\beta} \exp\{-(t_{\text{total}} - t)/\beta\} dt_{\text{total}} \\ &= -\exp\{-(t_{\text{total}} - t)/\beta\})|_{t^*}^\infty \\ &= \exp\{-(t^* - t)/\beta\}. \end{aligned} \tag{A8}$$

Again, we can solve for t^* such that $P(t_{\text{total}} > t^*|t) = .5$, obtaining $t^* = t + \beta\log 2$. For the Gaussian prior, we can find values of t^* by numerical integration and optimization.