Optimal predictions in everyday cognition

Thomas L. Griffiths
Department of Cognitive and Linguistic Sciences
Brown University

Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Word count: 3714

Address for correspondence:
Thomas Griffiths
Department of Cognitive and Linguistic Sciences
Brown University, Box 1978
Providence, RI 02912
Phone: (401) 863 9563 E-mail: tom_griffiths@brown.edu
Abstract

Human perception and memory are often explained as an optimal statistical inference, informed by accurate prior probabilities. In contrast, cognitive judgments are usually viewed as following error-prone heuristics, 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 or 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.
Optimal predictions in everyday cognition

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 dollars at the box office so far, what would you estimate for its total gross? Everyday life routinely poses such challenges of prediction, where the true answer cannot be determined based on the limited data available, yet common sense suggests at least a reasonable guess. Analogous inductive problems arise in many domains of human psychology, such as 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). 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, cognitive judgments under uncertainty are typically characterized as the result of error-prone heuristics, insensitive to prior probabilities (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). 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.

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 or 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. To illustrate the principles behind this Bayesian analysis, imagine that we
want to predict the total life span of a man we have just met, based upon the man’s
current age. If $t_{\text{total}}$ indicates the total amount of time the man will live and $t$
indicates his current age, the task is to estimate $t_{\text{total}}$ from $t$. The Bayesian predictor computes a
probability distribution over $t_{\text{total}}$ given $t$, by applying Bayes’ rule:

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

(1)

The probability assigned to a particular value of $t_{\text{total}}$ 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 person at age $t$ given that
their total life span is $t_{\text{total}}$. Assuming for simplicity that we are equally likely to meet a
person 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 $t_{\text{total}}$ (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 person’s life span will be $t_{\text{total}}$. Analysis of actuarial
data shows that the distribution of life spans in our society is (ignoring infant
immortality) approximately Gaussian – normally distributed – with a mean of about 75
years and a standard deviation 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 $t_{\text{total}}$ for a man encountered at age $t$. A good guess for $t_{\text{total}}$ 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 \( t_{\text{total}} \) 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 \( t_{\text{total}} \) is approximately the mean; once \( t \) approaches the mean, the predicted value of \( t_{\text{total}} \) 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. Analysis of box office records shows that the total gross of movies follows a power-law distribution, a highly non-Gaussian shape, with most movies taking in only modest amounts but occasional blockbusters making huge amounts of money. In the Appendix, we prove that for power-law priors, the Bayesian prediction function picks a value for \( t_{\text{total}} \) that is a multiple of the observed sample \( t \). The exact multiple depends on the shape of the power-law prior. 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: if we
observe a movie has made $40 million to date, we should guess a total gross of around $60 million; if we had observed a current gross of only $6 million, we should guess about $9 million for the total. While 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 child and her 75-year-old grandfather, we would never predict that she will live a total of 15 years (1.5 × 10) and he will live to be 112 (1.5 × 75). Other classes of priors, such as the exponential-tailed Erlang distribution, are also associated with distinctive optimal prediction functions, 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 estimate the true prior distributions for several classes of events (details of the estimation procedure are given in the Appendix). For example, as shown in Figure 2, human life spans and the runtime of movies are approximately Gaussian, the gross of movies and the length of poems are approximately power-law distributed, and the number of years spent in office for members of the U.S. House of Representatives is approximately Erlang. The experiment examined how well people’s predictions corresponded to optimal statistical inference in these different settings.

Methods

Participants

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, Poems, Life Spans, Pharaohs, and Marriages. A second group of 142 undergraduates made predictions about Movie Runtimes, Representatives, Cakes, Waiting
Everyday predictions

Materials

Within each group, each participant made predictions about one phenomenon from each of the five different classes seen by that group. Predictions were based on one of five possible values of \( t \), varied randomly between subjects. These values were: 1, 6, 10, 40, and 100 million dollars for Movie Grosses; 2, 5, 12, 32, and 67 lines for Poems; 18, 39, 61, 83, and 96 years for Life Spans; 1, 3, 7, 11, and 23 years for Pharaohs; 1, 3, 7, 11, and 23 years for Marriages; 30, 60, 80, 95, and 110 minutes for Movie Runtimes; 1, 3, 7, 15 and 31 years for Representatives; 10, 20, 35, 50 and 70 minutes for Cakes; and 1, 3, 7, 11, and 23 minutes for Waiting Times. In each case, participants read several sentences establishing context and then were asked to predict \( t_{\text{total}} \) given \( t \).

The questions were presented in survey format. A paragraph began each survey 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.

Sample questions were:

```
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 that of box office intake for that movie?
```
Poems. If your friend read you her favourite 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?

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?

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 Runtimes. 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?

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?

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?
Procedure

Participants completed the surveys as part of a booklet of unrelated experiments.

Results

We first filtered out responses that could not be analyzed or indicated a misunderstanding of the task, removing predictions which did not correspond to numerical values or were less than $t_{total}$. Only a small minority of responses failed to meet this criterion for all stimuli except Marriages, yielding 174, 197, 197, 191, 136, 130, 126, and 158 for Movie Grosses, Poems, Life Spans, Pharaohs, Movie Runtimes, Representatives, Cakes, and Waiting Times respectively. The responses for the Marriages stimulus were problematic because the majority of participants (52%) indicated that marriages last “forever”. This accurately reflects the proportion of marriages that end in divorce (Kreider & Fields, 2002), but prevents the data from being analyzed using the methods described below, which are based upon median values. We thus did not analyze responses for the Marriages stimuli further.

People’s judgments for Life Spans, Movie Runtimes, Movie Grosses, Poems, and 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, or 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 conventional views of cognitive judgment, 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, $p(t_{total}) \propto 1/t_{total}$, regardless of
the phenomenon to be predicted (Gott, 1993, 1994; Jaynes, 2003; Jeffreys, 1961; Ledford et al., 2001).

A quantitative measure of goodness of fit confirmed that people’s predictions in each domain were best explained using a Bayesian prior of the appropriate parametric form (Gaussian, power-law, or Erlang). The consistency of different priors with human judgments was measured using the ratio of the variance of the median human judgments around the predicted value of $t_{total}$ to the variance about the overall median. This statistic is comparable to $1 - r^2$ in the context of linear regression, where $r$ is the correlation coefficient. Since, unlike regression models, these models have no parameters estimated from the data, the ratio of variances can take on values greater than 1, indicating that the overall median actually provides a better fit to the data. Smaller values reflect greater consistency between model predictions and human judgments. The fit values for the different families of priors are shown in Table 1.

The remaining stimuli – Cakes, Pharaohs, and Waiting Times – provided a more severe test of people’s ability to combine prior knowledge with new data. Making predictions in each of these cases requires using a more complex or impoverished form of prior knowledge than a simple parametric distribution. For instance, the Cakes stimuli ask about the duration a cake should spend in the oven. As shown in Figure 3, this quantity follows an irregular distribution with no simple parametric form, which could make it difficult to learn. However, people’s judgments were still close to the ideal Bayesian predictions, despite the complex form of the empirical prior distribution.

The Pharaohs stimuli posed a complementary challenge: the prior distribution followed a simple shape that would be intuitive to people based on general world knowledge, but most people should have neither concrete experience with any events of this type nor accurate knowledge about the mean or variance of the distribution. As shown in Figure 3, people’s predictions had a form consistent with the appropriate prior
Everyday predictions

(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 $\beta = 17.9$, as opposed to the true value of $\beta = 9.34$. Using this subjective Erlang prior produces a close correspondence to human judgments.

Finally, the Waiting Times stimuli required making predictions in a context where even extensive experience might not be sufficient to clearly determine the appropriate prior. 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 (Barabasi, 2005, provides a summary and explanation of these findings). Rather than seeking data on waiting times that could be used to assess people’s judgments, as we did for the other stimuli, we used people’s judgments to evaluate what the true distribution of waiting times might be. We fit prediction functions for Gaussian, power-law, and Erlang distributions to the human data, minimizing the ratio of variances described above. The power-law prior provided the best fit to human judgments, with $\gamma = 2.43$, producing the predictions shown in Figure 3.

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 serve as a guide to people’s implicit beliefs about the distributions of everyday quantities, and reveal that these beliefs are
Everyday predictions

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).

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 (Shepard, 1987; Tenenbaum & Griffiths, 2001; Anderson, 1990) and discovering logical or causal relations (Anderson, 1990; Oaksford & Chater, 1994; Griffiths & Tenenbaum, in press). 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.
References


Appendix

The prediction problem

Assume that a point $t$ is sampled uniformly at random from the interval $[0, t_{\text{total}})$. What should we guess for the value of $t_{\text{total}}$? A Bayesian solution to this problem involves computing the posterior distribution over $t_{\text{total}}$ given $t$. Applying Bayes’ rule, this is

$$p(t_{\text{total}} | t) = \frac{p(t | t_{\text{total}}) p(t_{\text{total}})}{p(t)}$$  \hspace{1cm} (2)

where

$$p(t) = \int_{0}^{\infty} p(t | t_{\text{total}}) p(t_{\text{total}}) dt_{\text{total}}.$$  \hspace{1cm} (3)

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 3 thus simplifies to

$$p(t) = \int_{t}^{\infty} \frac{p(t_{\text{total}})}{t_{\text{total}}} dt_{\text{total}}.$$  \hspace{1cm} (4)

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 $\gamma \leq 1$, since the integral over $t_{\text{total}}$ diverges, but the posterior remains a proper probability distribution regardless. Applying Equation 4, we have

$$p(t) \propto \int_{t}^{\infty} t_{\text{total}}^{-(\gamma+1)} dt_{\text{total}}$$

$$= -\frac{1}{\gamma} t_{\text{total}}^{-\gamma} \bigg|_{t}^{\infty}$$

$$= \frac{1}{\gamma} t^{-\gamma},$$

where the constant of proportionality remains the same as in the original prior. We can
substitute this result into Bayes’ rule (Equation 2) to obtain

$$p(t_{total}|t) = \frac{t^{-(\gamma+1)}}{\frac{1}{\gamma} t^{-\gamma}} = \frac{\gamma t^\gamma}{t^{\gamma+1}},$$

(5)

for all values of $t_{total} \geq t$. Under the Erlang prior, $p(t_{total}) \propto t_{total} \exp\{-t_{total}/\beta\}$, we have

$$p(t) \propto \int_t^\infty \exp\{-t_{total}/\beta\}$$

$$= -\beta \exp\{-t_{total}/\beta\} \big|_t^\infty$$

$$= \beta \exp\{-t/\beta\}$$

where the constant of proportionality remains the same as in the original prior. Again, we can substitute this result into Bayes’ rule (Equation 2) to obtain

$$p(t_{total}|t) = \frac{\exp\{-t_{total}/\beta\}}{\beta \exp\{-t/\beta\}}$$

$$= \frac{1}{\beta} \exp\{-t_{total}/\beta\}\exp\{-t/\beta\}$$

(6)

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

**Predicting $t_{total}$**

We take the predicted value of $t_{total}$, which we will denote $t^*$, to be the posterior median. This is the point $t^*$ such that $P(t_{total} > t^*|t) = 0.5$: 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

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

(7)

We can derive $t^*$ analytically in the case of a power-law or Erlang prior. For the
Everyday predictions 18

power-law prior, we can use Equation 5 to rewrite Equation 7 as

\[
P(t_{total} > t^* | t) = \int_{t^*}^{\infty} \frac{\gamma t^{\gamma+1}}{t_{total}^{\gamma+1}} dt_{total}
\]

\[
= - \left( \frac{t}{t_{total}} \right)^{\gamma} |_{t^*}^{\infty}
\]

\[
= \left( \frac{t}{t^*} \right)^{\gamma}.
\]  

We can now solve for \( t^* \) such that \( P(t_{total} > t^* | t) = 0.5 \), obtaining \( t^* = 2^{1/\gamma} t \). For the Erlang prior, we can use Equation 6 to rewrite Equation 7 as

\[
P(t_{total} > t^* | t) = \int_{t^*}^{\infty} \frac{1}{\beta} \exp\{-(t_{total} - t)/\beta\} dt_{total}
\]

\[
= - \exp\{-(t_{total} - t)/\beta\} |_{t^*}^{\infty}
\]

\[
= \exp\{-(t^* - t)/\beta\}.
\]  

Again, we can solve for \( t^* \) such that \( P(t_{total} > t^* | t) = 0.5 \), obtaining \( t^* = t + \beta \log 2 \). For the Gaussian prior, we can find values of \( t^* \) by numerical integration and search.

**Estimating prior distributions**

The distributions appearing in Figures 2 and 3 were estimated from information available on the internet (Table A1). The parameters of different families of priors were estimated from these data (Table A2). Power law parameters (\( \gamma \)) were estimated using linear regression in log-log coordinates, considering only the values of \( t_{total} \) greater than the mean of the distribution to ensure that the tail of the distribution was modeled. For two distributions with exponential tails – Movie Runtimes and Pharaohs – there was insufficient data to estimate \( \gamma \). Parameters for the Erlang (\( \beta \)) and Gaussian (\( \mu, \sigma \)) were found using maximum-likelihood estimation. The results shown in the figures use Gaussian distributions for Life Spans and Movie Runtimes (\( \mu = 75.74 \) and 108.87, \( \sigma = 15.92 \) and 20.46), power-law distributions for Movie Grosses and Poems (\( \gamma = 2.04 \) and 1.91), and Erlang distributions for Representatives and Pharaohs (\( \beta = 5.28 \) and 9.34).
The subjective prior distributions for Pharaohs and Waiting Times were estimated from human data, as described in the main text.
Author Note

We thank Liz Baraff and Onny Chatterjee for their assistance in running the experiments, and Mira Bernstein, Daniel Casasanto, Peter Dayan, Konrad Körding, Tania Lombrozo, Rebecca Saxe, and Marty Tenenbaum for comments on the manuscript. The second author was supported by the Paul E. Newton Chair.
Footnotes

1Standard correlation coefficients allow a linear transformation of the regressor, which would remove the differences between the prediction functions for power-law and Erlang priors. Our measure is essentially a mean squared error, normalized by the total variance to facilitate comparison across stimuli.
Table 1. Fit to Human Predictions for Bayesian Model with Priors of Different Families

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Family</th>
<th>Uninformative</th>
<th>Power-law</th>
<th>Erlang</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie Grosses</td>
<td>Power-law</td>
<td>0.199</td>
<td>0.012</td>
<td>0.132</td>
<td>0.116</td>
</tr>
<tr>
<td>Poems</td>
<td>Power-law</td>
<td>0.397</td>
<td>0.029</td>
<td>0.064</td>
<td>0.263</td>
</tr>
<tr>
<td>Representatives</td>
<td>Erlang</td>
<td>0.586</td>
<td>0.091</td>
<td>0.033</td>
<td>0.074</td>
</tr>
<tr>
<td>Life Spans</td>
<td>Gaussian</td>
<td>34.576</td>
<td>7.799</td>
<td>3.965</td>
<td>0.025</td>
</tr>
<tr>
<td>Movie Runtimes</td>
<td>Gaussian</td>
<td>21.667</td>
<td>–</td>
<td>2.298</td>
<td>0.593</td>
</tr>
<tr>
<td>Cakes</td>
<td>-</td>
<td>2.607</td>
<td>0.611</td>
<td>0.142</td>
<td>0.166</td>
</tr>
<tr>
<td>Pharaohs</td>
<td>Erlang</td>
<td>2.059</td>
<td>–</td>
<td>0.097</td>
<td>0.657</td>
</tr>
<tr>
<td>Waiting Times</td>
<td>Power-law</td>
<td>0.647</td>
<td>0.010</td>
<td>0.034</td>
<td>0.100</td>
</tr>
</tbody>
</table>

Note: The values given for Pharaohs and Waiting Times correspond to subjective priors in the appropriate family. For Pharaohs, objective priors gave fits of 1.066 (Erlang) and 0.653 (Gaussian).
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Source (number of datapoints)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move Grosses</td>
<td><a href="http://www.worldwideboxoffice.com/">http://www.worldwideboxoffice.com/</a> (5302)</td>
</tr>
<tr>
<td>Poems</td>
<td><a href="http://www.emule.com">http://www.emule.com</a> (1000)</td>
</tr>
<tr>
<td>Life Spans</td>
<td><a href="http://www.demog.berkeley.edu/wilmoth/mortality/states.html">http://www.demog.berkeley.edu/wilmoth/mortality/states.html</a> (complete lifetable)</td>
</tr>
<tr>
<td>Representatives</td>
<td><a href="http://bioguide.congress.gov">http://bioguide.congress.gov</a> (2150 members since 1945)</td>
</tr>
<tr>
<td>Cakes</td>
<td><a href="http://www.allrecipes.com/">http://www.allrecipes.com/</a> (619)</td>
</tr>
<tr>
<td>Pharaohs</td>
<td><a href="http://www.touregypt.com/">http://www.touregypt.com/</a> (126)</td>
</tr>
</tbody>
</table>
Table A2: Parameters of Distributions of $t_{total}$ for Everyday Phenomena

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Power law ($\gamma$)</th>
<th>Erlang ($\beta$)</th>
<th>Gaussian ($\mu$)</th>
<th>Gaussian ($\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie Grosses</td>
<td>2.04</td>
<td>11.62</td>
<td>23.22</td>
<td>39.31</td>
</tr>
<tr>
<td>Poems</td>
<td>1.91</td>
<td>22.00</td>
<td>44.01</td>
<td>90.94</td>
</tr>
<tr>
<td>Life Spans</td>
<td>4.83</td>
<td>38.16</td>
<td>75.74</td>
<td>15.92</td>
</tr>
<tr>
<td>Movie Runtimes</td>
<td>–</td>
<td>54.44</td>
<td>108.87</td>
<td>20.46</td>
</tr>
<tr>
<td>Representatives</td>
<td>16.92</td>
<td>5.28</td>
<td>10.57</td>
<td>8.05</td>
</tr>
<tr>
<td>Cakes</td>
<td>1.67</td>
<td>24.96</td>
<td>49.92</td>
<td>32.22</td>
</tr>
<tr>
<td>Pharaohs</td>
<td>–</td>
<td>9.34</td>
<td>17.69</td>
<td>11.06</td>
</tr>
<tr>
<td>Waiting Times</td>
<td>2.43</td>
<td>5.19</td>
<td>3.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: Parameters for Waiting Times were estimated from human predictions. All other parameters were estimated from publically available data indicating values of $t_{total}$. 
Figure Captions

Figure 1. Optimal Bayesian prediction functions depend on the form of the prior distribution. 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, $t_{\text{total}}$, that could describe events in a particular class. Lines of different colours represent different parameter values (e.g., different mean durations) within each family. The bottom row of plots shows the optimal predictions for $t_{\text{total}}$ 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 (column 1) the prediction function always has slope less than one and an intercept near the mean $\mu$; predictions are never much smaller than the mean of the prior distribution, nor much larger than the observed duration. Power-law priors (column 2) result in linear prediction functions with variable slope and a zero intercept. Erlang priors (column 3) yield a linear prediction function that always has slope equal to one and a nonzero intercept.

Figure 2. Results for Life Spans, Movie Run times, Movie Grosses, Poems, and Representatives. The top row of plots shows the empirical distributions of the total duration or extent, $t_{\text{total}}$, for each of these everyday phenomena. The first two distributions are approximately Gaussian, the next two approximately power-law, and the last is approximately Erlang. Best-fitting distributions from the appropriate parametric families are plotted in red. The bottom row shows predicted values of $t_{\text{total}}$ for a single observed sample $t$ of a duration or extent for each phenomenon. Black dots show median predictions of $t_{\text{total}}$ for human participants. Error bars indicate 68% confidence intervals (estimated by a 1000-sample bootstrap). Black lines show the optimal Bayesian predictions based on the empirical prior distributions shown above, red lines show predictions based on the parametric priors that best fit the empirical distributions, and
dotted lines show predictions based on a fixed uninformative prior.

*Figure 3.* Results for Cakes, Pharaohs, and Waiting Times. The top row of plots shows empirical prior distributions for the baking times of cakes and the length of reigns of Egyptian pharaohs, as well as a prior estimated from people's predictions regarding waiting times. The baking times of cakes follow no simple parametric form. The durations of pharaohs' reigns follow an Erlang distribution (\(\beta = 9.34\), red curve), but people's judgments are consistent with using a subjective prior with a higher mean value (\(\beta = 17.9\), blue curve). The bottom row of plots shows median predictions about these three phenomena from human participants, depicted using the same conventions as in Figure 2. Black curves represent the optimal Bayesian prediction functions based on the empirical prior distributions shown above. Red curves indicate predictions under objective parametric priors, while blue curves indicate predictions under subjective priors estimated from human data (see main text for details).
Everyday predictions, Figure 1
Everyday predictions, Figure 2

![Graphs showing distributions and predicted values for different datasets: Life Spans, Movie Runtimes, Movie Grosses, Poems, and Representatives. The graphs display probability distributions and predicted values over time.](image-url)
Everyday predictions, Figure 3