Journal of Experimental Psychology: General
2011, Vol. 140, No. 4, 725–743

0096-3445/11/$12.00 DOI: 10.1037/a0024899

Predicting the Future as Bayesian Inference: People Combine Prior Knowledge With Observations When Estimating Duration and Extent

Thomas L. Griffiths
University of California, Berkeley

Joshua B. Tenenbaum
Massachusetts Institute of Technology

Predicting the future is a basic problem that people have to solve every day and a component of planning, decision making, memory, and causal reasoning. In this article, we present 5 experiments testing a Bayesian model of predicting the duration or extent of phenomena from their current state. This Bayesian model indicates how people should combine prior knowledge with observed data. Comparing this model with human judgments provides constraints on possible algorithms that people might use to predict the future. In the experiments, we examine the effects of multiple observations, the effects of prior knowledge, and the difference between independent and dependent observations, using both descriptions and direct experience of prediction problems. The results indicate that people integrate prior knowledge and observed data in a way that is consistent with our Bayesian model, ruling out some simple heuristics for predicting the future. We suggest some mechanisms that might lead to more complete algorithmic-level accounts.

Keywords: Bayesian inference, heuristics, predicting the future, mathematical modeling

> Making predictions is hard. Especially about the future.
> —Neils Bohr (or Yogi Berra)

Despite the difficulty of predicting the future, people do it effortlessly every day. They are confident about being able to predict the durations of events, how much time will be needed to get home after work, and how long it will take to finish the shopping. In many cases, people have a great deal of information guiding their judgments. However, sometimes they have to make predictions using much less evidence. When faced with new situations, decisions about how much longer events can be expected to last are based on whatever evidence is available. When the only information someone possesses concerns how long a particular event has lasted until now, predicting the future becomes a challenging inductive problem.

Being able to predict future events is an important component of many cognitive tasks. Expectations about the future are certainly fundamental to planning and decision making, but inferences about time also play a role in other aspects of cognition. Anderson's (Anderson, 1990; Anderson & Schooler, 1991) rational analysis of human memory takes the fundamental problem of memory to be predicting whether an item will be needed in the future, with retention favoring those items most likely to be in demand. Assessing the need for a particular item explicitly involves predicting the future, a problem that Anderson formulates as a Bayesian inference. Prediction is also intimately related to the discovery of causal relationships. The regularities induced by causal relationships lead people to make predictions about future events that can alter their perceptions (Eagleman & Holcombe, 2002) and result in surprises when their predictions are incorrect (Huettel, Mack, & McCarthy, 2002).

In this article, we test the predictions of a model of predicting the future, formulating the problem as one of Bayesian inference. Our Bayesian model indicates how people should combine their prior knowledge about a phenomenon with the information provided by observed data. Prior knowledge is expressed in terms of a probability distribution over the extent or duration of a phenomenon, whereas the effect of observations is incorporated via a statistical argument that is used in cosmology known as the *anthropic principle* (Gott, 1993). We explore the predictions of this model in depth and consider their implications for the psychological mechanisms that might guide human predictions.

In previous work, we showed that people are sensitive to the statistical properties of everyday quantities in the way that is consistent with our model (Griffiths & Tenenbaum, 2006). These results indicate that people use prior knowledge, but they do not imply that people are making predictions by performing the calculations indicated by Bayes' rule. Rather, they suggest that whatever algorithm people use for predicting the future, it is consistent with Bayesian inference in its sensitivity to prior knowledge. Many simple heuristics might have this property. For example, Mozer, Pashler, and Homaei (2008) argued that our results could be explained by people following a simple heuristic, in which they use only a small number of previous experiences to inform their

This article was published Online First August 29, 2011.

Thomas L. Griffiths, Department of Psychology, University of California, Berkeley; Joshua B. Tenenbaum, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

This research was supported in part by a grant from Mitsubishi Electronic Research Laboratories and a Hackett Studentship to Thomas L. Griffiths. We thank Mira Bernstein, Tania Lombrozo, Roger Shepard, and David Somers for valuable discussions. Experiments 1–3 were presented in July 2000 at the 22nd Annual Conference of the Cognitive Science Society, Philadelphia, Pennsylvania.

Correspondence concerning this article should be addressed to Thomas L. Griffiths, Department of Psychology, University of California, Berkeley, 3210 Tolman Hall 1650, Berkeley, CA 94720-1650. E-mail: tom_griffiths@berkeley.edu

judgments rather than having access to the whole probability distribution.

Our goal in this article is to gain a deeper understanding of the extent to which people's predictions are consistent with our Bayesian model and thus to obtain stronger constraints on the algorithms that they might be using in making these predictions. Our Bayesian model makes strong predictions about the effects of providing further observations, how prior knowledge should be combined with these observations, and the difference between independent and dependent observations. Each of these predictions provides an opportunity to identify a constraint that an algorithm would need to satisfy: If people behave in a way that is consistent with our model in each of these cases, then whatever algorithm they are using must also approximate Bayesian inference. We test these predictions empirically, using both cognitive and perceptual judgments about time, and consider their implications for simple heuristics that people might use in predicting the future.

Rational Models of Predicting the Future

We begin our analysis by describing the anthropic principle used in predicting the future. We then present the Bayesian generalization of this argument and consider the effects of manipulating the prior and the likelihood within this model. This leads us to the model predictions that we test in the remainder of the article.

Copernican Anthropic Principle

The cosmologist J. Richard Gott III (1993) proposed a simple heuristic for predicting the future, which was intended to provide insight into weighty matters like the prolonged existence of the human race but be just as valid when applied to matters arising in everyday life. Gott's heuristic is based on the Copernican anthropic principle, which holds that

> the location of your birth in space and time in the Universe is privileged (or special) only to the extent implied by the fact that you are an intelligent observer, that your location among intelligent observers is not special but rather picked at random. (Gott, 1993, p. 316)

Gott extends this principle to reasoning about a person's position in time: Given no evidence to the contrary, one should not assume that one is in a special place in time. If this principle is adopted, the time at which an observer encounters a phenomenon should be randomly located in the total duration of that phenomenon.[1]

This principle leads to a simple method for predicting the future: Discover how long a phenomenon has endured until the moment it was observed and predict that it should last that long into the future. This follows from the anthropic principle, because if it is assumed that you encounter a phenomenon at a random point, it is equally likely you observed it in the first or second half of its total duration. Denoting the time between the start of a phenomenon and its observation t_{past} and its total duration t_{total}, a good guess is $t_{total} = 2t_{past}$. The argument works just as well if you know how far a phenomenon will extend into the future (e.g., if you see a sign that is counting down to some event) but do not know how far it extends into the past. Again, you should guess $t_{total} = 2t_{future}$. For simplicity, we focus on predicting t_{total} from t_{past} for the remainder of this section.

More formally, Gott's (1993) rule can be justified by a probabilistic analysis that he called the "delta *t* argument" (p. 315). If we define the ratio

$$r = \frac{t_{past}}{t_{total}}, \tag{1}$$

the Copernican anthropic principle tells us that r should be uniformly distributed between 0 and 1. This lets us make probabilistic predictions about the value of r. In particular, the probability that $r < 0.5$ is 0.5, so there is a 50% chance that $t_{total} > 2t_{past}$. Likewise, there is a 50% chance that $t_{total} < 2t_{past}$, making $t_{total} = 2t_{past}$ a good guess. We can also use this argument to define confidence intervals over the durations of events by evaluating confidence intervals on r. For example, r will be between .025 and .975 with a probability of .95, meaning that with 95% confidence,

$$\frac{1}{39} t_{past} < t_{future} < 39 t_{past},$$

where $t_{future} = t_{total} - t_{past}$.

This method of reasoning has been used to predict a wide range of phenomena. Gott (1993) gave the example of the Berlin Wall, which he first encountered in 1969. At this point, the Berlin Wall had been in existence for 8 years, so t_{past} is 8 years. The 95% confidence interval on the future duration of the Berlin Wall, t_{past}, based on the assumption that Gott's visit was randomly located in the period of its existence, is 2.46 months to 312 years, firmly containing the actual t_{future} of 20 years. Gott made similar calculations of t_{future} for Stonehenge, the journal *Nature*, the U.S.S.R., and even the human race (the good news is that a 95% confidence interval gives humans at least 5,100 years, the bad news is that it also predicts less than 7.8 million). The principle has subsequently been applied to a surprisingly broad range of targets, including predictions of the runs of Broadway musicals (Landsberg, Dewynne, & Please, 1993).

A Bayesian Approach to Predicting the Future

Gott's (1993) Copernican anthropic principle suggests how we might formulate a rational statistical account of people's ability to predict the future, but the context in which people make daily predictions differs from that assumed by Gott in two important ways: prior knowledge and multiple observations. In many cases, in the real world, where it might be desirable to predict the future, people know more than simply how long a process has been underway. In particular, interaction with the world often gives people some prior expectations about the duration of an event. For example, if one meets a 78-year-old man on the street, one is unlikely to think that there is a 50% chance that he will be alive at the age of 156 (for a similar example, see Jaynes, 2003). Likewise, predictions are often facilitated by the availability of multiple observations of a phenomenon. For example, if one was attempting

[1] A more standard use of the anthropic principle in cosmology is as a way to explain otherwise improbable events, such as cosmological constants taking values that support a stable universe or the emergence of intelligent life. If we condition on there being an intelligent observer to observe such events, their probability is 1. A detailed discussion of uses of the anthropic principle is given in Bostrom (2002).

to determine the period that passes between subway trains arriving at a station, one would probably have several trips on which to base judgment. If, on the first trip, one discovered that a train had left the station 103 s ago, one might assume that trains run every few minutes. But, after three trips yield trains that have left 103, 34, and 72 s ago, this estimate might get closer to 103 s. After 10 trains, all leaving less than 103 s before we arrive, we might be inclined to accept a value very close to 103 s.

Gott's (1993) delta *t* argument does not incorporate the prior knowledge about durations that people bring to the problem of predicting the future or the possibility of multiple observations. However, it can be shown that the delta *t* argument is equivalent to a simple Bayesian analysis of the problem of predicting the future (Gott, 1994). Bayesian inference naturally combines prior knowledge with information from one or many observations, making it possible to extend Gott's argument to provide a more general account of predicting the future. Bayes' rule states that

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}, \tag{2}$$

where h is some hypothesis under consideration and d is the observed data. By convention, $P(h|d)$ is referred to as the posterior probability of the hypothesis, $P(h)$ the prior probability, and $P(d|h)$ the likelihood, giving the probability of the data under the hypothesis. The denominator $P(d)$ can be obtained by summing across $P(d|h)P(h)$ for all hypotheses, giving

$$P(h|d) = \frac{P(d|h)P(h)}{\int_H P(d|h)P(h)dh}, \tag{3}$$

where H is the set of all hypotheses.

In responding to a criticism offered by Buch (1994), Gott (1994) noted that his method for predicting the future could be expressed in Bayesian terms. In this setting, the data are the observation of the current duration of a phenomenon, t_{past}, and the hypotheses concern its total duration, t_{total}. Using the prior $P(t_{total}) \propto \frac{1}{t_{total}}$ and the likelihood $P(t_{past}|t_{total}) = \frac{1}{t_{total}}$ if $t_{total} \geq t_{past}$ (and 0 otherwise) yields the same results as his original formulation of the delta *t* argument.[2] This can be seen if Gott's choices of prior and likelihood are substituted into Equation 3:

$$P(t_{total}|t_{past}) = \frac{\frac{1}{t_{total}^2}}{\int_{t_{past}}^{\infty} \frac{1}{t_{total}^2} dt_{total}}, \tag{4}$$

which yields

$$P(t_{total}|t_{past}) = \frac{t_{past}}{t_{total}^2}. \tag{5}$$

The probability that an observed phenomenon will have $t_{total} > t$ for $t \geq t_{past}$ can be obtained by integrating this density over all values of $t_{total} > t$:

$$P(t_{total} > t|t_{past}) = \int_t^{\infty} \frac{t_{past}}{t_{total}^2} dt_{total} = \frac{t_{past}}{t}. \tag{6}$$

This is exactly the probability that would be obtained via the delta *t* argument: Asking whether $t_{total} > t$ is equivalent to asking whether $r < \frac{t_{past}}{t}$, and the result follows directly from the fact that r is uniformly distributed between 0 and 1.

These choices for the likelihood and the prior are not arbitrary. The Copernican anthropic principle determines the likelihood: If a phenomenon is encountered at a random point in its duration, then its current duration, t_{past}, is uniformly distributed between zero and its total duration, t_{total}. Consequently, $P(t_{past}|t_{total}) = \frac{1}{t_{total}}$ for all t_{past} less than t_{total}. The prior reflects the minimal possible state of knowledge about t_{total}. Using $\frac{1}{t_{total}}$ for the prior avoids building any sense of scale into the predictions: It gives the same amount of probability to a region even if the scale is transformed multiplicatively. For example, it gives the same probability to t_{total} taking a value between 1 and 2 minutes, 1 and 2 hours, and 1 and 2 years. This kind of prior, which gives very little information about the value of t_{total}, is known as an *uninformative* prior (Press, 1989; Jeffreys, 1961). This choice of prior leads to the scale-free character of the predictions made by Gott's rule: t_{total} is predicted to be a constant multiple of t_{past}, regardless of the value that t_{past} takes on. By not imposing a natural scale, the same simple rule can produce predictions for phenomena that have values of t_{past} ranging from a few seconds to hundreds of thousands of years, but it also leads to counterintuitive conclusions in contexts where a natural scale does seem to apply, as with the human life span.

Changing the Prior and Likelihood

Crucially, both the prior and the likelihood used in this Bayesian analysis can be altered to accommodate situations in which one has more knowledge, either as a result of previous experiences with similar phenomena or having the opportunity to make multiple observations. Griffiths and Tenenbaum (2006) explored the effects that prior knowledge of the duration and extent of everyday quantities has on predicting the future. Figure 1 shows the distributions of five everyday quantities: human life spans, the runtimes of movies, the box office gross of movies, the length of poems, and the time served by members of the U,S. House of Representatives. These quantities have different distributions, which translate into different predictions. Our Bayesian analysis can be applied in all of these cases, with the modification of t_{past} and t_{total} to refer to the observed and total extent of phenomena, respectively (e.g., an amount of money or the number of lines of text), not just their duration.

We can examine the consequences of using different priors by looking at how the posterior median—the value *t* such that it is equally likely that the total duration is greater than or less than *t*, with $P(t_{total} > t|t_{past}) = P(t_{total} < t|t_{past}) = 0.5$—changes as a function of the current duration t_{past}. This *prediction function*

[2] It should be noted that this is an improper prior, because it does not integrate to 1 over all values of t_{total} from 0 to ∞. Such priors can be used in Bayesian inference in cases where the resulting posterior distribution still integrates to 1, as it does in this case. For details, see Jaynes (2003) or Bernardo and Smith (1994).

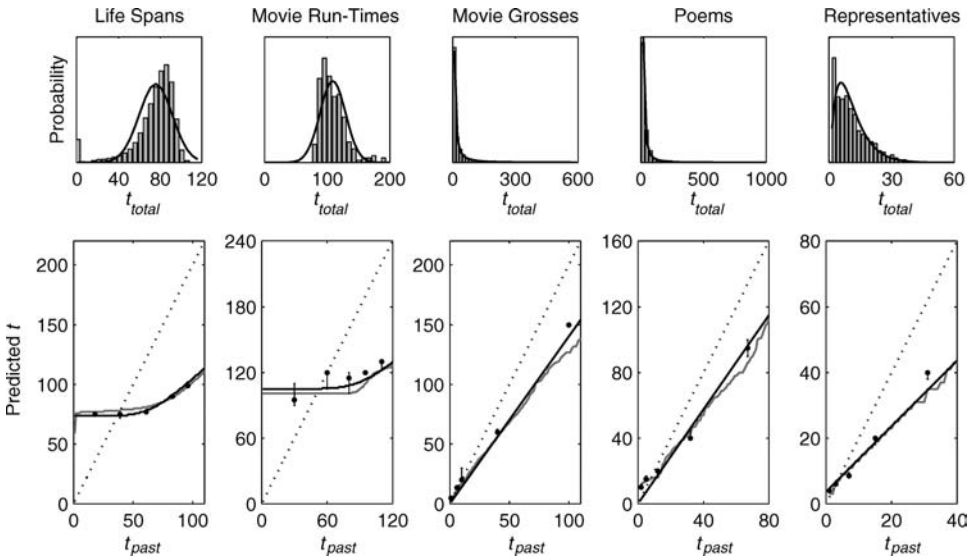

Figure 1. Effects of prior knowledge on predicting the future. The upper panels show the distributions associated with five everyday quantities: human life spans (in years), the run-times of movies (in minutes), the box office gross of movies (in millions of United States dollars), the length of poems (in lines), and the time served by members of the U.S. House of Representatives (in years). The histograms show the actual distributions, and the black curves show approximations from Gaussian (life spans and movie run-times), power-law (movie grosses and length of poems), and Erlang (time in the House of Representatives) distributions. The lower panels show optimal Bayesian predictions for these quantities, corresponding to the posterior median of t_{total} given t_{past}. Gray lines use the histogram as a prior and black lines use the approximating distributions. The dotted lines show Gott's (1993) rule, with t_{total} estimated as twice t_{past}, which seems appropriate in some cases (e.g., movie grosses) but not in others (e.g., life spans). The black dots indicate the median of human predictions for these quantities, with error bars corresponding to a 68% confidence interval estimated by bootstrap (see Griffiths & Tenenbaum, 2006, for details).

takes on a different shape for different priors. Human life spans and movie run-times roughly follow Gaussian distributions, with

$$P(t_{total}) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(t_{total}-\mu)^2/2\sigma^2}, \quad (7)$$

where μ and σ are the mean and standard deviation of the distribution, respectively. The resulting prediction function picks values of t_{total} close to the mean of the prior when t_{past} is small and then predicts a value slightly greater than t_{past} as t_{past} becomes large. Movie grosses and the length of poems follow power-law distributions, with

$$P(t_{total}) \propto t_{total}^{-\gamma}, \quad (8)$$

where γ is a parameter of the distribution. The resulting prediction function suggests that t_{total} will be $2^{1/\gamma} t_{past}$ (see Griffiths & Tenenbaum, 2006, for derivations of the results in this paragraph). The predicted value is thus simply a multiple of t_{past}, as in Gott's rule, although the multiplier depends on the shape of the distribution. The time spent in the U.S. House of Representatives roughly follows an Erlang distribution:

$$P(t_{total}) = \frac{t_{total} e^{-t_{total}/\beta}}{\beta^2}, \quad (9)$$

where β is a parameter. This distribution has a broad peak at $t_{total} = \beta$ and decays to zero at 0 and ∞. The resulting prediction function indicates that t_{total} will be slightly larger than t_{past}, with the predicted value being $t_{past} + \beta \log 2$.

Figure 1 also shows the predictions made by human participants in an experiment in which each participant was asked to make a prediction based on a single value of t_{past} for a subset of these phenomena. In aggregate, these predictions were in remarkably close correspondence with the predictions made by our Bayesian model, with the median human judgment being close to the median of the posterior distribution. However, Mozer et al. (2008) argued that these results do not provide definitive evidence that people had detailed knowledge of the underlying prior distribution. Mozer et al. showed that a heuristic in which each participant only has access to *k* samples from the prior and then makes a prediction corresponding to the smallest of these samples that is greater than t_{past}—the Min*k* heuristic—can produce similarly good approximations to our Bayesian model when aggregated across many simulated participants. If people are using this heuristic, they may have less detailed knowledge of the distribution of events than our original results suggested. In addition, although the Min*k* heuristic approximates our Bayesian model in this particular case, it does not incorporate assumptions about how t_{past} is sampled or indicate how observations and prior knowledge should be combined more generally, so it will differ from our Bayesian model in other settings. We have argued elsewhere that people's prior knowledge of the distributions of everyday quantities goes beyond a few samples (Lewandowsky, Griffiths, & Kalish, 2009) and shown that

there are circumstances under which approximating Bayesian computations with small samples can be rational (Shi, Griffiths, Feldman, & Sanborn, 2010; Vul, Goodman, Griffiths, & Tenenbaum, 2009). In this article, we turn to the question of whether people's judgments remain consistent with our Bayesian model for more complex prediction problems.

Bayesian models indicate not just how prior knowledge should be used but also how this knowledge should be integrated with observed data. The experiments we present in this article thus focus on examining how people combine prior knowledge with observations when predicting the future. One property of our Bayesian model, but not of heuristic accounts such as Gott's (1993) delta *t* argument or the Min*k* heuristic proposed by Mozer et al. (2008), is that the number of observations of the current duration of a phenomenon people see should affect their predictions. This is most applicable for cyclic phenomena, where t_{total} represents the interval between instances of a recurring event. For example, we might try to estimate how often trains run on a particular subway line on the basis of having waited for trains on multiple occasions. The Copernican anthropic principle determines the probability of a set of observations of this kind in exactly the same way that it determines the probability of a single observation. We use T to denote a set of n times drawn independently and uniformly at random from the total duration of a phenomenon, as might occur when waiting for a train n different times on the same subway line. Then, taking t_{past} to be the largest value in T, we have

$$P(T|t_{total}) = \begin{cases} \left(\frac{1}{t_{total}}\right)^n & t_{total} \geq t_{past} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where n appears in the exponent because the probabilities of independent events multiply and each observation has probability $\frac{1}{t_{total}}$. This equation says that for any total duration t_{total}, any set of observations less than t_{total} is equally likely, but the probability of the set of observations decreases as t_{total} increases, favoring values of t_{total} that are closer to t_{past}. As the number of observations increases, so does the preference for smaller values of t_{total}.

Changing the prior and likelihood used in our Bayesian model has important implications for the resulting predictions. We can see some of these implications by examining the effects of these changes on the posterior median *t* when we manipulate t_{past} and the number of examples ($n = 1, 3, 10$). The predictions of Gott's (1993) delta *t* argument, presented in Equation 1, are shown in Figure 2A. The predictions are unaffected by n and are thus constant at $t = 2t_{past}$, a consequence of the fact that the simple form of the delta *t* argument is not designed to take into account situations in which multiple observations are available. The predictions that result from following this heuristic are also insensitive to prior information, offering a single prediction for all situations. This seems to defy intuition and is the weakest of the models we consider.

At the next level of complexity is a Bayesian model with $\frac{1}{t_{total}}$ for the likelihood of a single example and the uninformative prior $P(t_{total}) \propto \frac{1}{t_{total}}$. As shown in Figure 2B, the model shows an effect of the number of examples, resulting from the fact that the likelihood of n independent examples is $\left(\frac{1}{t_{total}}\right)^n$. The exponent in the likelihood gives smaller durations more weight as the number of examples increases, reducing the estimate of *t*. The curve shown in the figure corresponds to the prediction $t = 2^{1/n}t_{past}$, approaching t_{past} as n becomes large. This reduction in *t* makes intuitive sense: Consider the problem of predicting the time that passes between trains discussed at the start of this section: An initial t_{past} of 103 s suggests t_{total} is a few minutes, but seeing more examples, each less than 103 s, brings our estimate much closer to t_{past}. The main problem with this model is that it does not make use of the flexibility provided by the inclusion of prior knowledge in inference.

Although we should certainly use precise information about the distribution of t_{total} if available (as was done in Griffiths & Tenenbaum, 2006), in general, we can use a simpler prior to impose some sense of scale on the predictions. In the remainder of the article, we use an Erlang distribution as a prior (see Equation 9). This parameterized peaked distribution provides a simple way to summarize many of the kinds of distributions that might be encountered across temporal domains and to explore the results of manipulating different aspects of the model. Of the everyday quantities considered by Griffiths and Tenenbaum (2006), two were well-described by an Erlang distribution (length of terms in the U.S. House of Representatives and the durations of reigns of pharaohs in Ancient Egypt). The use of an informative prior imposes a natural scale on a phenomenon, meaning that people's predictions about t_{total} are no longer simply a fixed multiple of t_{past}. Figure 2C shows the predicted values of t_{total} for one, three, and 10 observations using an Erlang prior where the maximum value of t_{past} is varied but β is held constant. As t_{past} increases, the ratio of the predicted value of t_{total} to t_{past} decreases, providing a clear pattern that indicates the use of an informative prior.

Model Predictions

Our Bayesian model of predicting the future has several properties that discriminate it from existing heuristic accounts, such as the rule proposed by Gott (1993) or the Min*k* heuristic proposed by Mozer et al. (2008). Exploring the extent to which people's judgments are consistent with these properties provides constraints on algorithms that could be used to explain how people predict the future. Each point of consistency highlights a property that a heuristic account must possess and makes it more challenging to define a single simple heuristic that might explain people's behavior.

The most important property of our Bayesian model is its treatment of multiple observations. The tightening of predictions about t_{past} is a direct consequence of the Copernican anthropic principle and plays an important role in models of human generalization and word learning in the form of the *size principle* (Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007a, 2007b). As more observations less than t_{past} are obtained, predictions should converge toward the smallest possible value of t_{total} that is greater than t_{past}. The equivalent statistical structure of predicting the future and generalization also suggest that this phenomenon should be observed whether a judgment is about time or some other quantity for which the same set of hypotheses (i.e., intervals from 0 to some upper limit) is appropriate, such as healthy levels

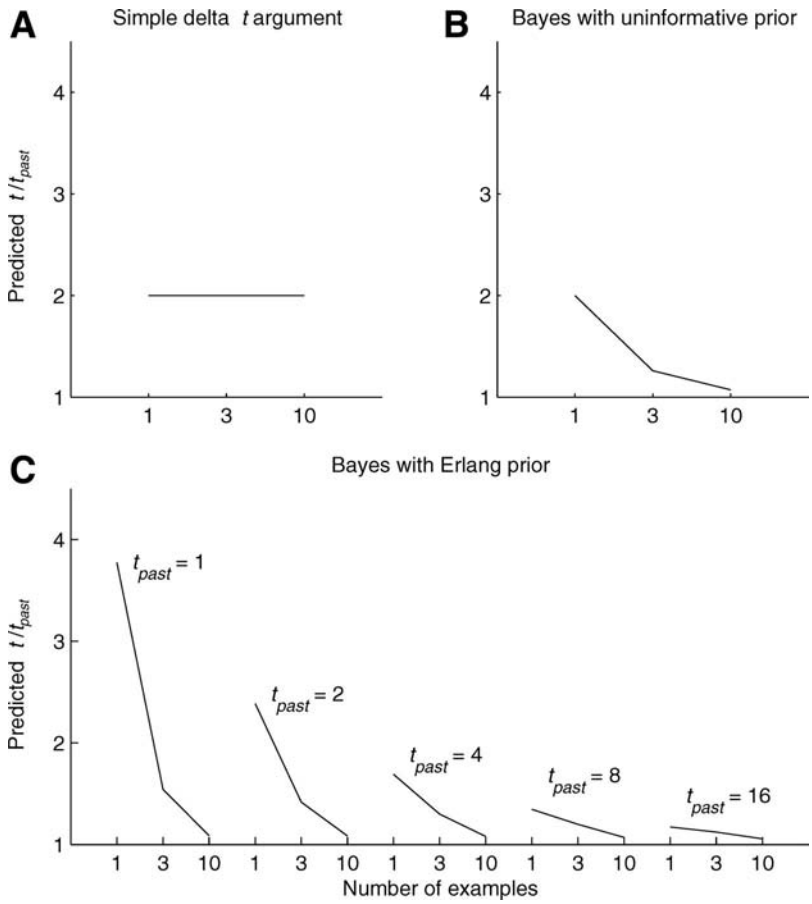

Figure 2. Predictions of the various models, depicting the posterior median—the point t at which $P(t_{\text{total}} < t|T) = 0.5$—for a set T containing one, three, and 10 observations. The predictions of each model are shown as a line connecting three points: The left point is the prediction for a single example, the center point shows the prediction for three observations, and the right point is the prediction for 10 observations. On all graphs, the vertical axis shows the predicted value of t in proportion to t_{past}, the largest value in the set of observations T. A: The predictions produced by the simple delta t argument. B: The predictions of a Bayesian model with uninformative prior but a likelihood that accommodates multiple observations. C: The predictions of a Bayesian model with an Erlang prior of $\beta = 4$ given sets of observations varying in the maximum value of t_{past}.

of a toxin (where small amounts are presumably not dangerous but it becomes, at some point, a risk to health).

Our Bayesian model also asserts that manipulating the prior probability distribution across the hypothesis space will produce a general change in predictions, at least until the effect of the priors is overwhelmed by the likelihoods. In particular, inducing a prior preference for a relatively high value of t_{total} will bias inferences toward hypotheses around that value. Griffiths and Tenenbaum (2006) showed that people are sensitive to the distribution of everyday quantities when they make predictions, meaning that they make use of prior knowledge. However, our Bayesian model also makes clear quantitative predictions about how this prior knowledge of durations should be combined with the information provided by multiple observations: As the number of observations increases, the influence of prior knowledge should decrease, with judgments becoming more similar regardless of priors. These predictions are not consistent with simple heuristic accounts. Use of Gott's (1993) rule cannot predict any effect of prior knowledge. If the biasing information is not provided in terms of samples from the prior, it is hard to see how it could be explained by the Min*k* heuristic. Because neither of these heuristics can handle multiple observations, they also cannot predict the diminishing effect of prior knowledge as the number of observations increases.

Finally, the effect of multiple observations under our Bayesian model should be sensitive to the process by which the observations were generated. In particular, we should only expect a rapid reduction in the extent of generalization if a set of observations are probabilistically independent. If the observations are dependent, then successive pieces of data will not reduce our predictions about t_{total}. For example, imagine we were trying to predict how long Stonehenge would continue to exist. The first time we visit provides an observation of t_{past} from which we can extrapolate t_{total}. Each subsequent visit gives us a new value for t_{past}, but these visits do not provide independent observations: If Stonehenge existed at

the most recent value of t_{past}, it must also have existed at all smaller values. This corresponds to using the likelihood function

$$P(T|t_{total}) = \begin{cases} \left(\frac{1}{t_{total}}\right) & t_{total} \geq t_{past} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

for a set of observations T with maximum value t_{past}. Using this likelihood function, we should not reduce our estimate of t_{total} each time we visit Stonehenge, as we would if we were using the likelihood given in Equation 10, which assumes independent observations. This prediction falls out of the statistical analysis of the prediction problem and is hard to explain unless people are sensitive to the way in which the data they observe are sampled. It is interesting that previous research has provided both positive and negative results concerning people's sensitivity to dependence between observations: Xu and Tenenbaum (2007a) found evidence that children are sensitive to the way in which examples are sampled in learning words, whereas Barsalou, Huttenlocher, and Lamberts (1998) found that adults can fail to use such information in learning categories.

Our Bayesian model has three clear implications for the kind of judgments that people should produce when reasoning about the duration of events. The effect of multiple observations, integration of prior probabilities, and impact of independent observations are all important properties of how these judgments are made. These predictions provide an opportunity to identify constraints on algorithms that people could be using in predicting the future and are inconsistent with existing heuristic accounts. In Experiments 1–3, we examine these predictions in turn, using a task in which people read about the durations of events and then form predictions. Experiments 4 and 5 supplement these results with a task in which people actually experience events of different durations and form their predictions on the basis of their perceptual experiences.

Experiment 1: Multiple Observations

In the first experiment, we examined the effect of introducing multiple observations into problems of predicting the future and compared the resulting predictions with judgments made in nontemporal domains. The aim was to evaluate whether our Bayesian model could provide a good account of these judgments and to examine the correspondence between predicting the future and other problems of generalization.

Method

Participants. Participants were 81 undergraduates participating for partial course credit. The participants were randomly assigned to four conditions, with 21 participants in the *teacake* condition, 21 in the *train* condition, 16 in the *toxin* condition, and 23 in the *taxicab* condition.

Stimuli. A simple questionnaire format was used, examining four scenarios in which inductive decisions were required, two of which involved temporal judgments. Each questionnaire had three sections. The first section outlined the situation and gave a single number on which judgments were to be based. This summary was intended to establish a context for the judgment and to provide some information relevant to forming reasonable prior expectations. The second and third sections added further information, giving a total of three numbers and 10 numbers, respectively.

The first scenario, the teacake scenario, described a coffee shop that had recently started selling teacakes. The situation was described as follows:

> Each day, on your way to class, you walk past a coffee shop. The shop has recently started a new advertising campaign: they bake fresh teacakes regularly throughout the day, and have a clock outside that shows how long it has been since the teacakes were taken out of the oven. You are interested in buying a teacake as soon as it is removed from the oven, and wonder how often batches of teacakes are baked. Today, the clock shows that it has been 34 minutes since the last batch of teacakes was removed from the oven.

The story gives t_{past}, and participants were asked to predict t_{total}. Specifically, they were asked the following:

> Please write down your best guess of how much time elapses between batches of teacakes, in minutes. Try to make a guess, even if you feel like you don't have enough information to make a decision—just go with your gut feeling. You may assume that the batches of teacakes are always separated by the same amount of time.

The second and third sections added further times. The second section read as follows:

> Suppose that you check the clock on your way to class for the next two days. Each time, you note how long it has been since the teacakes came out of the oven. Including the first day, the times you have seen are 34, 8, and 21 minutes.

Participants were then asked exactly the same question about the time between batches of teacakes. Finally, the third section read as follows:

> Suppose that you note the time on the clock every time you walk past the coffee shop in the next week, checking how long it has been since the teacakes came out of the oven. Including the three days from the previous week, the times you have seen are 34, 8, 21, 18, 2, 5, 27, 22, 10, and 14 minutes.

This was followed by the same question about the time between batches of teacakes.

The second scenario, the train scenario, outlined a similar situation in the domain of public transport:

> Imagine that you are visiting a foreign country, and want to catch a subway train. Your guidebook says that the subway service is very efficient, with trains running to fixed schedules, guaranteeing that you won't have to wait very long at the station. The book doesn't mention how much time passes between trains, so you will have to try to guess. When you get to the station, you see a clock on the wall indicating that the next train will arrive in 103 seconds.

Here, the clock provides t_{future}, and participants are asked to predict t_{total}. As noted above, our Bayesian model does not differentiate between t_{past} and t_{future}, as they are both observations of a period less than t_{total} assumed to be uniformly distributed between 0 and t_{total}. We thus expect these predictions to be captured by our model in the same way as the predictions based on t_{past} from the first scenario are captured. Participants were then asked to write down their best guess of how much time passes between successive trains. The second section introduced additional observations of 34

and 72 s, then the third section added 61, 17, 29, 101, 97, 42, and 52 s to the times.

These scenarios were complemented with two analogous situations that made no reference to time. The first asked people to estimate what levels of a toxin might be healthy. The description of the toxin scenario was as follows:

> On a visit to the doctor, you undergo a range of tests. In one case, the doctor informs you that he is testing for levels of a certain environmental toxin in your blood. He has a chart that indicates healthy levels of environmental toxins, but you can't quite read the numbers indicating the threshold between healthy and unhealthy amounts. The doctor tells you that you need only get another appointment if the test suggests that you have unhealthy levels of toxin in your blood. When you go in to pick up your results, you find a stack of reports that have been classified as "healthy." You are relieved to find your report in the "healthy" stack. You see that the concentration listed on your report is 34 ng/mL.

Participants were then asked to give their best guess of the highest concentration of the toxin that would be considered healthy. Subsequent information was introduced in the next two sections, with additional observations of 8 and 21 ng/mL, then 18, 2, 5, 27, 22, 10, and 14 ng/mL.

The second comparison scenario was a version of the Jeffreys (1961) tramcar problem: Participants were told the serial number of a taxicab (as well as being given the information that all cabs are given a unique number between 1 and the total number of cabs in the company) and asked to guess the number of cabs in the company (the presentation of the problem in terms of taxicabs was inspired by Jaynes, 1994). The taxicab scenario gave participants the information shown below.

> Imagine that your business requires you to travel by train to a certain town or city for ten weeks. When the train pulls up at the station, you get into the first taxicab you find. As you get into the cab, you notice a serial number on the rear fender of your cab. You are curious about this, and ask the driver how the cabs are labeled. He tells you that each cab is given a unique number between 1 and the total number of cabs in the company, and that each number in that range corresponds to a cab. The serial number of this particular cab is 103.

Participants were asked to estimate the total number of cabs in the company. Additional observations were introduced in the next two sections, with numbers 34 and 72, then 61, 17, 29, 101, 97, 42, and 52.

For each of the scenarios, the first number given was the largest, meaning that further observations would only tighten the range of generalization. The largest examples were 34 min and 34 ng/mL for the teacake and toxin scenarios, respectively, and 103 s and 103 cabs for the train and taxicab scenarios, respectively. These values were selected on the basis of an expectation that they would be broadly consistent with people's expectations for these quantities. As can be seen above, the sets of numbers given were identical for the teacake and toxin scenarios and the train and taxicab scenarios. The numbers provided after the first number in each scenario were approximately uniformly distributed and were selected to suggest a random sampling process.

Procedure. Each participant received a questionnaire containing all three sections on a single sheet. At the top of the sheet was the statement "Please answer each question before going on to the next, and do not go back and change your answers" in boldface. These instructions were the same for all four conditions.

Results and Discussion

Plausible responses to these problems are constrained to be greater than the largest example provided, so participants who produced responses less than t_{past} were viewed as not having understood the task and omitted from the analysis. This criterion eliminated approximately 15% of the participants in each of Experiments 1–3. Inspection of participants who were rejected suggested that they were giving estimates either of the spacing between observations or of the central tendency of the observations, consistent with two possible misunderstandings of the instructions. Responses more than 3 standard deviations from the mean were considered outliers and also excluded. Only one outlier was identified in the course of Experiments 1–3. Responses were transformed to $\frac{t}{t_{past}}$, where t is the raw score and t_{past} is the largest example.

A one-way within-subjects analysis of variance (ANOVA) showed a statistically significant effect of the number of examples for each scenario: For the teacake scenario, $F(2, 30) = 9.71$, mean square error (MSE) = 0.23; for the train scenario, $F(2, 32) = 18.00$, $MSE = 2.39$; for the toxin scenario, $F(2, 30) = 15.05$, $MSE = 0.017$; and for the taxicab scenario, $F(2, 44) = 9.57$, $MSE = 3.69$, all $ps < .001$. Means and standard errors are shown in Figure 3, which demonstrates that all four scenarios show a similar effect of new information. The figure also shows fits generated by our Bayesian model, using the Erlang prior with β estimated for each scenario by minimizing the sum of the normalized squared errors to the human data (i.e., taking the difference between the model prediction and the mean of the human data for each point, dividing by the standard error for that mean, then summing over points). The parameterization of the distribution reflects the different priors that might exist across different scenarios, relative to the scale of the examples selected. The values of β for the teacake, train, toxin, and taxicab scenarios were 1.6, 5.4, 0.7, and 4.4 times the maximum of t_{past}, respectively. The peak of the Erlang prior is at $t_{total} = \beta$, yielding values of 54 min between batches of teacakes, 9 min 21 s between trains, 24 ng/mL of toxin, and 460 taxicabs, all of which seem appropriate.

The results are consistent with the predicted effect of multiple observations: Smaller predictions were made as more observations were provided. This effect appeared in both the temporal and the nontemporal tasks, as might be expected if all four tasks are viewed as instances of Bayesian inference. This pattern of behavior indicates that people update their beliefs as they receive additional observations, something that is not accommodated by existing heuristics. The variation in the absolute magnitude of responses across different tasks is also consistent with using an informative prior. Under the prior used by Gott (1994), we would not expect to see such variation. In Experiment 2, we explore the issue of priors in more detail, examining whether priors can also be manipulated within scenarios.

Experiment 2: Priors

Griffiths and Tenenbaum (2006) explored the basic prediction that results from taking a Bayesian approach to predicting the

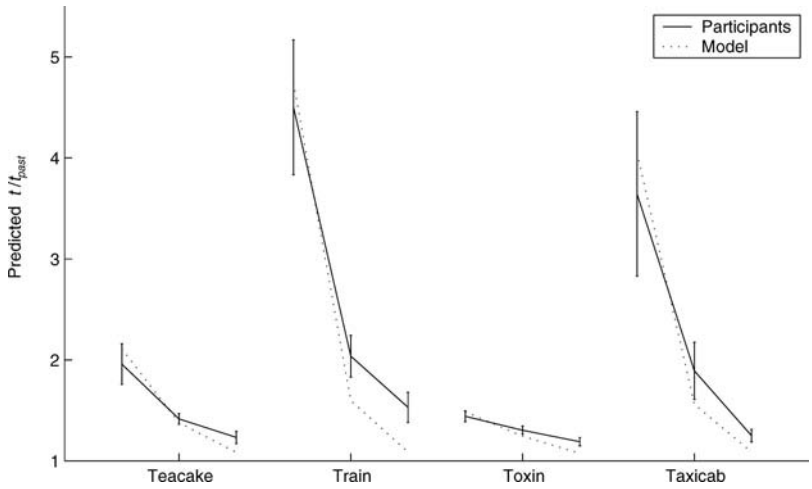

Figure 3. Results of Experiment 1. The mean responses for each condition are displayed in the same format as Figure 2, with one, three, and 10 examples corresponding to the left, center, and right points on each plotted line. Error bars show one standard error. All four conditions show a clear effect of number of examples, irrespective of whether they are judgments involving time, as well as variation in the overall range of generalization corresponding to differences in priors.

future, showing that prior knowledge about the distribution of durations influences people's predictions. Experiment 1 provided further support for this idea, showing that there was variation in people's predictions across scenarios, which can be captured in the different distributions assumed for t_{total} in our model. However, there are still two important questions to address related to the use of prior knowledge in predicting the future, concerning how such knowledge is represented and how it is combined with multiple observations. In Experiment 2, we explore these questions.

The question of how prior knowledge of durations is represented is relevant to understanding what kinds of algorithms might allow people to predict future events. The Min*k* heuristic proposed by Mozer et al. (2008) assumes that people have access to samples from the probability distribution $P(t_{total})$ associated with the quantity to be predicted. Although this may be a reasonable assumption for some of the everyday prediction problems considered by Griffiths and Tenenbaum (2006), it is clear that people can also make predictions in contexts where they might not have direct experience. In Experiment 2, we use a simple manipulation of prior knowledge, providing instructions that modify people's expectations, which is harder to account for in terms of drawing on some stored set of samples of t_{total}.

The question of how prior knowledge is combined with multiple observations provides a further opportunity to examine whether people form predictions in a way that is consistent with Bayesian inference. The experiment conducted by Griffiths and Tenenbaum (2006) asked participants to make a prediction on the basis of only one observation. The results of Experiment 1 demonstrate that people can take into account information from multiple observations and seem to do so in a way that is consistent with Bayesian inference, reducing their predictions toward a value that is the largest observation so far as the number of observations increases. This is the result of the likelihood function (Equation 10) asserting a stronger preference for smaller values of t_{total} as n increases and should thus appear independently of manipulation of the prior. In Experiment 2, we test this by providing multiple observations, in addition to manipulating the prior.

The basic strategy we adopted for exploring manipulations of people's priors was to provide our participants with information that we expected to affect their priors, obtain predictions with different numbers of observations, and then fit our Bayesian model to these results to estimate a prior distribution. We could then check whether the change in the inferred prior distributions was consistent with the change that should have been produced by the information that we provided to our participants.

Method

Participants. Participants were 78 undergraduates participating for partial course credit. The participants were randomly assigned to four conditions. The *teacake no prior* condition had 17 participants, *teacake prior* had 16, *train no prior* had 22, and *train prior* had 23.

Stimuli. The teacake and train scenarios from Experiment 1 were used, together with two new scenarios. The new scenarios gave participants information that was designed to alter their prior before they were given specific numbers on which a prediction could be based. The sentence "A friend who you are walking with says that he worked in a coffee shop in the same chain last year, and that shops usually bake every two hours, although it varies from shop to shop" was added to the teacake scenario, and "In the course of your travels, you have noticed that most subway trains in this country run approximately every seven minutes, although it varies from place to place" to the train scenario. Both pieces of information were intended to increase the durations people would expect for these scenarios, being greater than the estimated mode in Experiment 1. These new scenarios were termed *teacake prior* and *train prior*, whereas the original scenarios were called *teacake no prior* and *train no prior*. All scenarios asked for predictions

with one, three, and 10 examples, using the same numbers as those presented in the stimuli for Experiment 1.

Procedure. The procedure was the same as for Experiment 1.

Results and Discussion

Responses were screened using the same procedure as in Experiment 1. The scenarios were sorted into teacake and train groups and examined for the effect of number of examples and manipulating priors using two-way within-between ANOVAs. The teacake scenarios showed an effect of the number of examples, $F(2, 42) = 25.87$, $MSE = 0.27$, $p < .001$, and manipulating priors, $F(1, 21) = 4.70$, $MSE = 1.72$, $p < .05$, as well as an interaction between the two, $F(2, 42) = 3.80$, $p < .05$. Similar results were shown for the train scenarios. There was a statistically significant effect of the number of examples, $F(2, 68) = 50.31$, $MSE = 0.62$, $p < .001$, as well as an effect of manipulating priors, $F(1, 34) = 5.85$, $MSE = 1.58$, $p < .05$). In both groups, the effect of the number of examples replicates the results of Experiment 1, and the higher means for the group given the raised prior is consistent with the predictions of our Bayesian model.

Means and standard errors are shown in Figure 4, together with the model fits, obtained via the same method as was used in Experiment 1. The β values used in fitting the data were 1.6, 3.3, 3.25, and 3.85 for the teacake no prior, teacake prior, train no prior, and train prior conditions, respectively. The β values are greater in the conditions where the prior was raised and give peak values of 1 hr 52 min for the teacakes and 6 min 40 s for the trains. It is notable that these values are within 10% of the priors supplied in the experimental materials, supporting the efficacy of the manipulation and the appropriateness of the model.

Experiment 3: Independence

Experiments 1 and 2 required participants to respond to a sequence of observations in which they see the largest value, make a response, see two smaller values, make another response, then see another seven values smaller than the first before making a third response. Although we interpret the decreasing trends observed in all three experiments as a rational statistical inference, a possible confound in these experiments is that the largest observation was always presented first. This suggests that participants might be applying an *anchoring and adjustment* heuristic (Tversky & Kahneman, 1974). Specifically, participants form an estimate after the first section of the questionnaire, use this estimate as an anchor, and then adjust downward as more information becomes available.

Experiment 3 was designed to obtain further constraints on heuristics that people might be using to make predictions. The experiment had two differences from Experiments 1 and 2. First, it used a between-subjects design in which each participant saw a single set of numbers (containing one, three, or 10 examples). The between-subjects design was intended to remove the opportunity for people to adjust their judgments between responses: The largest observation was always presented last and each participant made only a single response, so the task demands no longer made an explicit strategy of this kind attractive. Second, we manipulated the dependency structure of the samples. This provided a way to test whether the statistical structure of the problem affected people's predictions, which would be another aspect of human behavior that would need to be incorporated into heuristic accounts.

Formally, the treatment of dependency is as follows. The likelihood given in Equation 10 only applies if the *n* examples of the set are independent. This would be the case if a phenomenon was observed on several different occasions, where each occasion corresponds to a different instance of the phenomenon (as in our examples in Experiments 1 and 2, where different episodes of baking or different trains are observed). If the observations are completely dependent, as would be the case if they were all observations of a single instance of the phenomenon, the likelihood is that given in Equation 11, and new observations less than

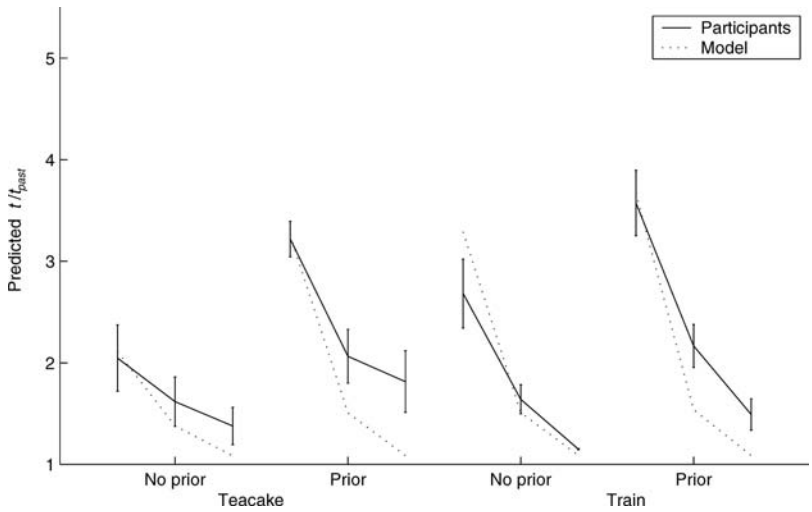

Figure 4. Results of Experiment 2. The mean responses for each condition are displayed in the same format as Figure 3, with one, three, and 10 examples corresponding to the left, center, and right points on each plotted line. Error bars show one standard error. The results are consistent with the manipulation of the priors described in the text.

the maximum value of t_{past} will have no effect on predictions about t_{total}. An example of this would be standing outside the coffee shop and waiting, noting down the time on the clock every minute. Only the largest observed time should be used in evaluating how often teacakes are baked, as all of the other observations are rendered irrelevant by this (if the clock says 54 min, it would have said 53 min a minute ago, etc.). If people are using values of t_{past} to make a statistical inference about t_{total}, they should expect to see an effect of the number of observations when those examples are independent but not if they are dependent.

Method

Participants. Participants were 220 undergraduates participating for partial course credit. Participants were randomly assigned to five conditions: 42 participants saw only a single example, whereas 37 saw three independent examples, 52 saw ten independent examples, 45 saw three dependent examples, and 44 saw ten dependent examples.

Stimuli. Five questionnaires were used, representing two scenarios and three different numbers of examples. The original teacake scenario was used, together with a second scenario in which the successive observations of the clock outside the shop were rendered dependent. This was made apparent through a change in instructions. The questionnaire for three dependent examples read as follows:

> Suppose that one day you decide to wait outside the shop until a fresh batch of teacakes is baked. You look up several times, and see that the clock reads 8, then 21, then 34 minutes, all without a new batch of teacakes being made.

Because the dependent times were listed from smallest to largest, the independent times on the original teacake scenario were presented in the same order. Each participant saw either one, three, or 10 examples. Because the notion of independence is irrelevant for a single example, only five conditions were needed.

Procedure. The procedure was the same as that used in Experiments 1 and 2.

Results and Discussion

Responses were screened as in Experiment 1. The effect of examples was examined separately for the independent condition and the dependent condition, because the single example cell was used in both analyses. A one-way between-subjects ANOVA revealed that varying the number of examples had a statistically significant effect on predictions when the examples were independent, $F(2, 102) = 4.10$, $MSE = 0.58$, $p < .05$, but not when they were dependent, $F(2, 115) = 0.55$, $MSE = 1.52$, $p = .58$, as predicted by our Bayesian account. The means for the different conditions are displayed in Figure 5.

The results of this experiment suggest that people show a sensitivity to the statistical properties of the information provided when making judgments about time and cast doubt on explanations of the results in terms of a simple heuristic like anchoring and adjustment. First, the between-subjects design meant that people saw only one set of numbers and were thus not given an opportunity to explicitly anchor a judgment and then revise it incrementally in the face of further smaller durations. Second, the results provide no evidence for a tightening of generalization when the observations are dependent. A sensitivity to dependency needs to be part of an account of predicting the future: If people are using a simple heuristic to address the task, they are still making an inference about whether the statistical structure of the problem warrants the application of that heuristic.

Complete dependence between observations is obviously an extreme case, but it provides the strongest manipulation for testing whether people are sensitive to the statistical structure of their

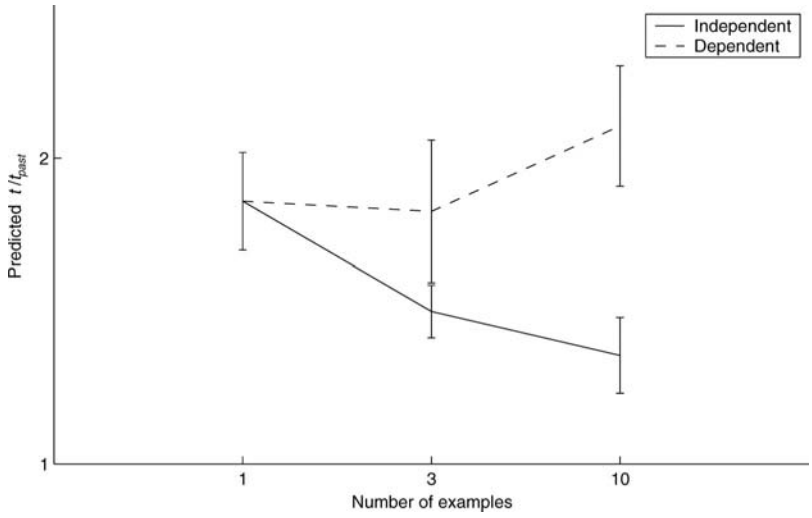

Figure 5. Results of Experiment 3. The mean responses for each condition are displayed in the same format as Figure 3, with one, three, and 10 examples corresponding to the left, center, and right points on each plotted line. Error bars show one standard error. An effect of multiple observations is only observed when the examples are independent.

observations. We used this strong manipulation in part because a similar strategy has been used in exploring children's sensitivity to the way in which examples are sampled in word learning (Xu & Tenenbaum, 2007a). However, our results raise the question of how sensitive people are to this kind of statistical property of the input, which is something that can be explored in future work by providing vignettes that interpolate between complete independence and complete dependence. Understanding how people reason about dependencies between observations is particularly interesting given other work showing that people fail to appropriately account for biases in the way that their observations are sampled (Fiedler & Juslin, 2006; Hertwig, Barron, Weber, & Erev, 2004; Ross & Nisbett, 1991).

Experiment 4: Predictions From Experienced Durations

Experiments 1—3 suggest that people's temporal predictions correspond well to the qualitative properties of our Bayesian account. However, these experiments all involve cognitive judgments about time: People receive a set of numbers that correspond to durations and are asked to produce a response in similar format. In our remaining experiments, we examined the effects of multiple observations and prior beliefs in a set of tasks where both the stimuli and the responses were actual durations. These experiments involved learning about *Goodman gems*, named for Goodman (1955). These gems have the unique property that after a fixed amount of time, they change from one color to another. Each type of gem has two characteristic colors, spending the same amount of time displaying one color before switching to the other, and all gems of the same type take the same amount of time to change. People presumably expect gems not to change in color, so we used these stimuli as a way to introduce the task in a way where people would have relatively little prior knowledge about durations. In each experiment, the participants were presented with sets of gems under different viewing conditions and asked to judge, using a timer, how long it takes those gems to change color. In Experiment 4, we tested the basic predictions of our Bayesian model, exploring how people's judgments vary as the value of t_{past} and the number of observations change. This was thus a replication and extension of Experiment 1 with perceptual rather than conceptual presentation of durations.

Method

Participants. Participants were 27 undergraduates participating for partial course credit.

Stimuli. The gems were presented as two-dimensional diamonds on a computer screen, and participants directly observed how long it took them to change color. There was no need to consider the psychophysical scaling of the stimuli, because the relationship between actual and perceived time is approximately linear for the intervals used in this experiment (Allan, 1979).

Procedure. In an initial instruction and practice phase, participants were told about the properties of Goodman gems and then watched seven gems go through their full color change cycle, starting in one color, changing to a second, and then changing back to the first. The time taken for these gems to change color was uniformly spaced on a logarithmic scale, from 0.5 to 32 s. Participants saw these gems in random order and then learned how to click *Start* and *Stop* buttons with the mouse to record how long they thought the gems took to change color, on average. The initial instructions read as follows:

> This is an experiment about a special kind of gemstone, called a Goodman gem. Goodman gems have the unique property of cyclically changing color. Each gem has two different colors. After a fixed amount of time, it will suddenly change from one color to the other. Some gems change color rapidly, while others take longer.
>
> In this experiment, you will see Goodman gems of many different kinds. Your job is to estimate the how much time passes between color changes for each kind of Goodman gem.
>
> Every Goodman gem of a particular kind changes color after exactly the same amount of time. You can tell that two gems are of the same kind because they have the same colors. This means that you can use your experiences with one gem of a particular kind to predict how much time should pass between changing colors for another gem of that kind.
>
> Clicking on the button below with the mouse will show you seven samples of different kinds of Goodman gems changing color, one after the other. Each gem starts having just changed color, and you will see it change to its other color and then back again.

Clicking a button then led to the initial examples and training on the use of the *Start* and *Stop* buttons, with the following instructions:

> Because this experiment involves making judgments about how much time it takes for a particular kind of Goodman gem to change color, you are going to use a timer to record your responses.
>
> At several points in the experiment, you will be asked a question about time. For each question, the procedure will be exactly the same. You will see a "Start Timer" button, like the one below. When you are ready to record your response, click on the "Start Timer" button with the mouse.
>
> When you click on the button, it will disappear and you will be told that the timer is running. A "Stop timer" button will also appear. You should then wait until the time since you pressed the "Start Timer" button is the same as the amount of time that corresponds to the answer to the question, then click on the "Stop Timer" button.

Following this initial phase, participants received further instructions outlining a story about miners and scientists, intended to establish the nature of the experimental task:

> The seven kinds of Goodman gems you have seen so far were once thought to be the only Goodman gems in existence. However, it has recently been discovered that a group of miners in a faraway country has found fifteen other kinds of Goodman gems, and is trying to sell them to the highest bidder. As a consequence, the gems will probably be sold to private collectors. You represent a group of scientists who are interested in knowing how long it takes these particular Goodman gems to change color. Unfortunately the miners realize that once the scientists have this knowledge, they are less likely to try to buy the gems. Consequently, they won't let you see any of their new kinds of gems go through a complete color change cycle. However, since the miners need to demonstrate that these are actually Goodman gems, you have the opportunity to see each of their gems change color once.

> There are fifteen new kinds of gems, and the miners are selling ten gems of each kind. You are only allowed to view one gem at a time.
>
> Each gem is kept in a different room, and you enter the room at a random point in the color change cycle. Once the gem changes color, the miners lead you back out of the room.
>
> The scientists with whom you are working want you to use the information you obtain from your observation of the gems to decide how much time it takes for gems of each kind to change color.

The instructions were discussed with each participant, ensuring that the nature of the sampling procedure was clear and emphasizing that they would have incomplete information for making these judgments.

In the main part of the experiment, participants saw 15 sets of new Goodman gems. The sets of gems were chosen in a 5×3 factorial array, crossing the maximum time taken to change color (1, 2, 4, 8, or 16 s) with the number of examples (one, three, or 10). All of the gems of a single kind were seen one after the other, and the participants were presented with a display showing up to 10 rooms that they could choose to enter by clicking on different buttons. An example set of instructions at this point was as follows:

> You are about to see a new kind of Goodman gem.
>
> The miners have ten gems of this kind, each housed in a different room. On the next screen, you will see a set of buttons corresponding to the different rooms you can enter. To enter a room, click on the appropriate button. When you enter a room, you will see the color of the gem and will wait until it changes color.
>
> All of the Goodman gems you see will be of the same kind, although they will be at a random point in their color change cycle when you encounter them.

In each set, one gem took the maximum time to change color, while the remaining durations were random values below this maximum, reflecting the fact that the participant had entered the room at a later point in the gem's color change cycle. The gems were shown in random order, so the point at which the gem taking the maximum time to change color appeared was randomized across sets and subjects. After viewing each set of gems, participants made a judgment about how long it took that type of gem to change color, using the *Start* and *Stop* buttons. For these judgments, they were given the following instructions:

> Now make a guess about how much time passes before this particular kind of Goodman gem changes color. Since you haven't seen any gems change color and then change back again, you can't know for certain how much time it takes. Try to use the information you have obtained so far to make an informed guess.

Results and Discussion

The design provides the same statistical structure as used in Experiments 1–3, although the effect of multiple observations is manipulated across sets of gems rather than within sets of gems. We should thus expect participants to produce smaller predictions for sets containing more examples. As in the previous experiments, responses shorter than the longest example indicate a misunderstanding of the requirements of the task. However, because actual duration estimates tend to be less precise than written responses, we relaxed our criterion to reject only those participants who produced responses that were less than half as long as the longest example. This procedure reduced the number of participants contributing data to 23. We used the same transformation as in the previous experiments for all analyses, converting responses t to $\frac{t}{t_{\text{past}}}$.

The results of the experiment are shown in Figure 6, with the characteristic pattern of reduced generalization with more examples found in Experiments 1–3. A two-way within-subjects ANOVA revealed that there was a significant effect of number of examples, $F(2, 44) = 35.87$, $MSE = 0.69$, $p < .001$, and stimulus duration, $F(4, 88) = 43.04$, $MSE = 0.48$, $p < .001$, as well as a significant interaction between the two, $F(8, 176) = 15.16$, $MSE = 0.37$, $p < .001$. This pattern of results is consistent with our Bayesian model, as illustrated by the predictions resulting from using a prior in which β is chosen so that the peak is at 4 s, the geometric mean of the values given as practice stimuli. These predictions are also shown in Figure 6A.

The main effect of number of examples supports our argument, showing the characteristic pattern predicted by our Bayesian model as a result of the independent sampling assumption made in the likelihood. The main effect of stimulus duration also provides strong support for another aspect of our model, reflecting the combination of observed data with a prior distribution. Figure 6A shows that this main effect results from participants producing smaller predictions relative to the maximal value of t_{past} for stimuli of longer duration. That is, people tended to produce predictions for the total duration that were closer to the maximum observed duration as the maximum observed duration increased. Such an effect is impossible to explain with the uninformative prior used by Gott (1994). Under Gott's analysis, we would expect to see the same pattern of predictions for all five durations. The smaller predictions with longer durations indicate that people have expectations about how long the gems should take to change color. When the observed durations are shorter than expected, they make a larger prediction. As the observed durations increase, the extent to which they are willing to go beyond the maximum observed value decreases. The ability of our model to capture this pattern with a single choice of β equal to the geometric mean of the practice data suggests that people have estimated the typical time scale for this scenario and use that as an ingredient in a rational statistical prediction of duration.

Experiment 5: Irrelevance of Order

Experiment 4 showed that the magnitude of people's predictions relative to the maximum value of t_{past} decreases as a function of the number of examples, consistent with the predictions of our Bayesian model. An alternative explanation for these results could be formed on the basis of the fact that as the number of examples increases, so does the elapsed time between observing the maximum value of t_{past} and making a prediction. With one observation, the maximum value has to be the last thing the participant saw. With three observations, there will typically be one other observation between the maximum value and the time when participants make a judgment. With 10 observations, there will typically be four or five observations between the maximum value and the time

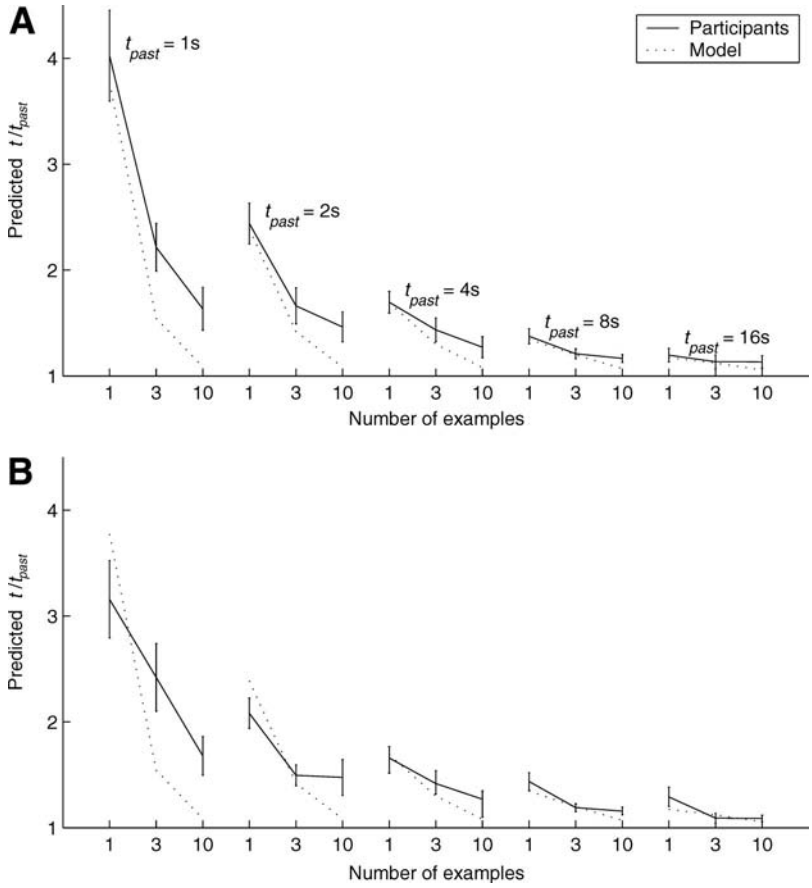

Figure 6. A: Results of Experiment 4. B: Results of Experiment 5. The mean responses for each condition are displayed in the same format as Figure 3, with one, three, and 10 examples corresponding to the left, center, and right points on each plotted line. Error bars show one standard error. The effect of multiple observations persists with perceptual stimuli and the variation in the size of predictions can be accounted for by assuming a single prior distribution over total duration.

when participants make a judgment. This increase in elapsed time could account for our results if people forget earlier observations or if people put greater weight on more recent trials (a recency effect).

To address this alternative explanation, we replicated Experiment 4 but fixed the order of examples so that the largest value of t_{past} always appeared last. Our Bayesian model is insensitive to the order of the observations and thus the basic pattern of results should be similar to those of Experiment 4. Recent work has pointed out that order effects pose a challenge for Bayesian models that assume observations are independent and identically distributed (e.g., Kruschke, 2006). Our goal here is not to exhaustively rule out order effects in the predicting the future task but to examine whether this specific aspect of order—the location of the maximal observation—could account for our findings in Experiment 4. If the distance between this maximum value and the prediction is responsible for the observed effect, then the effect should be removed by adopting this presentation order. If people are instead sensitive just to the statistical structure of the task, we would expect to see an effect of multiple observations despite the maximum value of t_{past} appearing last.

Method

Participants. Participants were 21 undergraduates participating for course credit.

Stimuli. Stimuli were the same as for Experiment 4.

Procedure. Procedure was the same as for Experiment 4, except the largest value of t_{past} was always the last to be presented.

Results and Discussion

Two participants produced values of t less than half as long as t_{past} and were excluded from the analysis. The same transformation as in Experiment 4 was used for all analyses. The results are shown in Figure 6Bb. The results reproduced the pattern seen in Experiment 4, with a significant effect of number of examples, $F(2, 38) = 31.50$, $MSE = 0.27$, $p < .001$, and stimulus duration, $F(4, 76) = 20.94$, $MSE = 0.69$, $p < .001$, and a significant interaction between the two, $F(8, 152) = 5.71$, $MSE = 0.24$, $p < .001$. Again, the results were consistent with our Bayesian account, as can be seen from the predictions of the model with the same value of β as in Experiment 4, also shown in Figure 6B.

Our results indicate that forgetting previous observations or placing greater weight on more recent observations cannot account for the decrease in the magnitude of people's predictions relative to the maximal value of t_{past} as they obtain more observations. Even when the maximal value was the most recent, we saw a decrease in predictions consistent with our Bayesian model. However, this experiment examined only one manipulation of the order of observations and does not rule out the possibility of order effects in this task. Because our Bayesian model predicts that the order of observations should be irrelevant to people's predictions, a more comprehensive investigation of order effects would provide a strong test of the model.

Taken together, these two experiments suggest that the effect of multiple examples observed in Experiments 1–3 extends beyond cognitive stimuli dealing with time, affecting explicitly temporal judgments in which both stimuli and responses are actual durations. The results also illustrate that the effect of multiple observations can be obtained in situations where judgments are not a direct result of the revision of beliefs in the face of new information. Because the experiment used different stimuli for sets of different sizes, the resulting judgments did not reflect the gradual accumulation of information about a single temporal quantity but several decisions about different quantities. People showed the predicted effect of multiple observations in making these decisions. This outcome is similar to that obtained in Experiment 3 but extends the result to perceptual stimuli and a situation in which a single individual makes multiple judgments.

General Discussion

Predicting the future is a difficult inductive problem but one that people often solve effortlessly. Our results suggest that people's predictions about temporal events are consistent with a rational statistical account of predicting the future, which uses Bayesian inference to combine prior knowledge of durations with information from multiple examples. Experiment 1 showed that the effect of providing further examples conformed to the predictions of our Bayesian model: More examples promoted a reduction in the scope of generalization, with predictions becoming closer to the largest example provided. Experiment 2 showed that people's predictions could be affected by the manipulation of their prior expectations and that this effect was consistent with the interaction of priors and likelihoods in Bayesian inference. Experiment 3 showed that the revision of beliefs as a result of further information was appropriately sensitive to the dependency structure of the observations. Finally, Experiments 4 and 5 demonstrated that the effect of multiple observations extended to explicitly temporal judgments with perceptual stimuli and situations in which judgments did not require the direct revision of beliefs.

These results provide constraints on possible algorithms that people could be using to predict future events. People's predictions are inconsistent with existing simple heuristics, such as the multiplicative rule proposed by Gott (1993) and the Min*k* heuristic proposed by Mozer et al. (2008). In particular, neither of these heuristics provides an account of how multiple observations should influence predictions, making it difficult to explain the results of Experiments 1, 4, and 5. The sensitivity to instructional manipulation of priors in Experiment 2 and statistical dependency structure observed in Experiment 3 also suggests that people are doing something more sophisticated than naively applying a simple heuristic, because they seem to take into account prior knowledge of durations in a form other than samples from the distribution and the sampling process by which the observations were generated. This kind of sensitivity is consistent with treating the problem of forming predictions as one of statistical inference.

In the remainder of the article, we explore in detail several issues that are raised by our results. First, we discuss how accounts of predicting the future at different levels of analysis can be mutually informative. Our analysis is at Marr's (1982) computational level, defining the computational problem presented by predicting the future and considering how it might be solved. Characteristically for this kind of analysis, we have been concerned with why people make the judgments they do rather than how they make those judgments. Further insights into the algorithms behind everyday prediction of the future can be gained from a study by Sternberg and Kalmar (1997), and our computational account suggests some simple heuristics in addition to those considered so far that may guide people's predictions. Having discussed these issues, we briefly consider how predicting the future relates to other cognitive problems.

Levels of Analysis

In presenting a Bayesian account of predicting the future, we are not claiming that people explicitly perform Bayesian calculations whenever they pass a coffee shop or arrive at a train station. Rather, we are claiming that a probabilistic analysis of the problem of predicting the future can provide insight into the human ability to solve inductive problems. This account is complementary to analyses of the cognitive processes involved in forming predictions (e.g., Sternberg & Kalmar, 1997) and can be conducted in parallel with investigations of whether people's judgments can be explained in terms of simple heuristics.

Computation and algorithm. Marr (1982) identified three levels at which theories of cognition can provide explanations: computational, algorithmic, and implementational. In analyzing the problem of predicting the future, we aim to provide an explanation at the computational level, addressing the question "What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out? " (Marr, 1982, p. 25). We have made no commitments to the algorithms by which people solve this computational problem or how they might be implemented. Theories at these different levels have the potential to provide complementary insights into people's ability to predict the future. Although neuroscientists have begun to investigate the neural basis of prediction (Huettel, Mack, & McCarthy, 2002) and the perception of time (Eagleman & Holcombe, 2002), we still lack a clear account of how the ability to predict the future might be implemented in people's brains. However, a detailed algorithmic-level analysis of everyday induction was conducted by Sternberg and Kalmar (1997), providing some results that parallel our findings.

Sternberg and Kalmar (1997) had participants make a series of judgments about events either in the future or in the past (they termed these tasks *prediction* and *postdiction*, respectively) and then used psychometric techniques to establish which properties of the stimuli affected the time taken to form these judgments. Their results were interpreted in terms of a 17-step information-

processing model, where each step identified a computational operation involved in encoding the stimuli, accessing memory, forming a prediction, and making a response. The variables found to affect response time included the amount of time between the present and the past or future date, whether the judgment was prediction or postdiction, and how much additional information might be needed to make a prediction. On the basis of these results, Sternberg and Kalmar concluded that the key steps in the algorithm were those that involved computing the difference between the current and the past or future time and developing a schema describing the event.

Sternberg and Kalmar's (1997) results suggest that the strategies people use to make predictions about the past and the future are highly similar, with overlapping (but nonidentical) factors influencing response latencies in prediction and postdiction. This provides a nice algorithmic analogue to our analysis, in which values of t_{past} and t_{future} are equally informative about t_{total} (as discussed briefly in Experiment 1). Their findings also illustrate the important role that prior knowledge plays in everyday inductive inferences. Although our analysis assumes a simple prior, in which domain knowledge is reflected only in the choice of β, the retrieval and processing of information relevant to a problem will be an important step in being able to reach solutions that reflect both appropriate knowledge about the domain and the effect of the observations.

Statistics and heuristics. The consistency of human behavior with our Bayesian model suggests that people's intuitions about predicting the future can be understood in terms of rational statistical inference but does not imply that they work through Bayes' rule whenever they need to make predictions. Presumably, people are following some cognitive algorithm when solving this problem. An analysis of a problem at the computational level can provide informative constraints on possible algorithmic-level theories. In this case, behavior consistent with our Bayesian model provides constraints on the heuristics that people might be using in solving the problem of predicting the future. The standard use of heuristics in psychological theory is as a direct alternative to applying any form of Bayesian reasoning to a problem. In contrast, we view heuristics as providing simple means of achieving solutions similar to the Bayesian solution, without the computational overhead. A probabilistic analysis of a cognitive problem is complementary to identifying a heuristic that seems consistent with human behavior: The probabilistic analysis can explain why a particular heuristic is useful, while the heuristic answers the question of how people are able to perform a task.

Our experiments illustrate that people seem to combine observations with prior knowledge in a way that is consistent with Bayesian inference and can take into account information from multiple observations. A heuristic account needs to be able to accommodate both of these findings. The Min*k* heuristic introduced by Mozer et al. (2008) illustrates how the effects of prior knowledge on predictions could be implemented without requiring sophisticated probabilistic computations. Mozer et al. used this heuristic to show that results consistent with our Bayesian analysis could be produced by aggregating across a population in which each participant has access to only *k* samples from the prior. Subsequent work using a within-subjects version of the predicting the future task provides evidence against this account (Lewandowsky et al., 2009). However, sampling from the prior still provides a good way to approximate Bayesian inference and is consistent with performance on this task if participants draw a new set of *k* samples from the prior each time they have to make an inference (Shi et al., 2010). Indeed, approximations based on a small number of samples can represent an optimal trade-off between accuracy and computation time (Vul et al., 2009).

The Min*k* heuristic, in its current form, is not capable of accounting for how people incorporate information from multiple observations (as illustrated in Experiments 1, 4, and 5), how they adjust their priors on the basis of verbal instructions (as illustrated in Experiment 2), or why they are sensitive to the statistical dependency structure of their observations (as illustrated in Experiment 3). Each of these results thus poses a challenge to this account. However, this does not mean that it will not be possible to construct a heuristic that satisfies these constraints. For the purpose of illustration, we show how simple heuristics might be able to account for the effect of multiple observations.

We consider two variants on a simple heuristic, making use of the intuition that as the density of observations in an interval increases, the range of generalization should decrease. The simplest form of this density heuristic is the rule "The distance to extrapolate is the range of scores divided by the number of examples." Using this heuristic, the predicted value of t_{total} is

$$t = t_{\text{past}} + \frac{t_{\text{past}}}{n}. \tag{12}$$

This heuristic derives its pedigree from frequentist statistics, where statistical problems equivalent to our problem of predicting the future are addressed by using t as an unbiased estimator of t_{total}. This simple density heuristic is insensitive to scale and thus cannot account for the variation in judgments seen in Experiments 1 and 2. In fact, it is closely related to our Bayesian model with the scale-free prior: A first-order Taylor series expansion of $P(t_{\text{total}} > t|T)$ around the maximum value of t_{past} gives the prediction $t = t_{\text{past}} + \frac{t_{\text{past}}}{2n}$. Using this heuristic might thus make sense if one has very little knowledge about the domain, but it would be inappropriate if one has a sense of the scale of predictions.

To accommodate the results of our experiments, a heuristic needs to reflect the natural scale of durations in a particular domain. Observing that the Erlang prior results in the prediction $t = t_{\text{past}} + \beta\log 2$ with a single observation, we might define a density heuristic for a domain with a natural scale to be

$$t = t_{\text{past}} + \frac{\beta}{n}. \tag{13}$$

Here, the density of the observations is evaluated not with respect to t_{past} but with respect to the assumed scale of the problem. This heuristic gives a reasonably good account of our data from Experiments 4 and 5 with $\beta = 2.7$ s, as can be seen in Figure 7. The full Bayesian model with an Erlang prior gives correlations of .933 and .920 for Experiments 4 and 5, respectively, while this density heuristic gives correlations of .906 and .895, respectively. Just as the Erlang prior is not suitable for all circumstances (e.g., predicting human life spans and the grosses of movies, as discussed above and in Griffiths & Tenenbaum, 2006), this particular heuristic might be expected to have limited application. However, heuristics based on density should be able to capture one of the central results

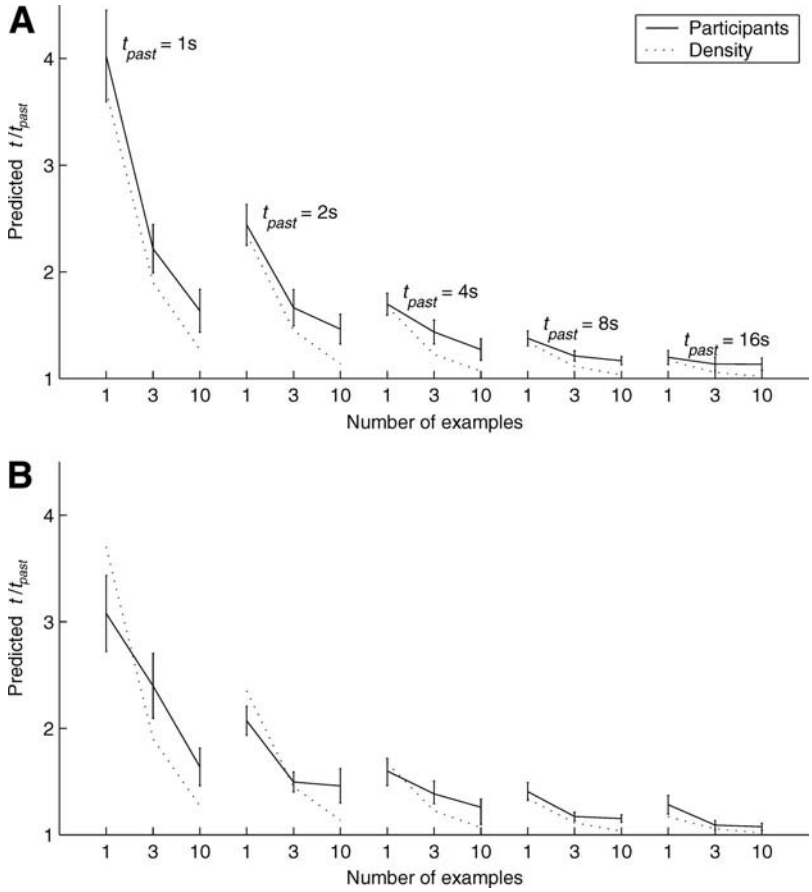

Figure 7. A: Results of Experiment 4. B: Results of Experiment 5. The dotted lines indicate the predictions of a version of the density heuristic that fixes a natural scale, as described in the text.

of our Bayesian analysis and our experiments: the decrease in the extent of generalization as a function of the number of examples.

Taken together, using samples from the prior and paying attention to the density of observations might provide a way to approximate our Bayesian model. However, adopting the appropriate heuristic for producing such an approximation might be a challenging task in itself. Although assessing the density of a set of observations does not require any statistical sophistication, people still need to be sufficiently aware of the statistical structure of a problem to know that applying a particular heuristic is justified: The results of Experiment 3 indicate that if a set of observations has the wrong dependency structure, people do not apply this kind of heuristic. Judging whether a simple heuristic is applicable can require a complex inference that relies on subtle details of a problem and is a current topic of investigation in research on heuristics (Gigerenzer & Brighton, 2009).

Predicting the Future and Other Cognitive Problems

We have considered predicting the future to be a desirable goal in itself, but being able to form such judgments also has important implications for other areas of cognition. In this section, we consider how our analysis relates to work on memory and generalization.

Memory and the future. Our analysis of predicting the future is based on the problem of predicting the duration of events, but a similar statistical analysis can be provided for other kinds of prediction tasks. In any such task, the key to making predictions will be specifying a statistical model that captures the quantities to be predicted and the relationship between past and future observations. Anderson (1990) argued that a number of phenomena in human memory can be understood by considering the goal of memory to be the efficient storage of facts that are likely to be needed in the future. Anderson's analysis of this problem views the memory system as assuming a simple statistical model of when facts are needed, essentially a nonhomogeneous Poisson process, and estimating the properties of each fact according to this model. While the specifics of his account differ from ours, the two approaches embody a similar philosophy in providing a computational level analysis of the problem of predicting the future. Anderson (1990) suggested that these inferences occur unconsciously and are an important part of the human memory. One attractive component of future research is thus exploring the extent to which unconscious temporal judgments more broadly reflect simple statistical principles.

Predicting the future as generalization through time. Our Bayesian approach to predicting the future is closely related to

Shepard's (1987) analysis of the problem of generalization. Given two objects x and y, represented as points in a psychological space, and the knowledge that x has a particular property (equivalent to belonging to a consequential region in psychological space C, containing all objects with that property), the problem of generalization involves computing the probability that y also has that property, $P(y \in C|x)$. Shepard (1987) used a simple Bayesian argument to show that this probability decreases as an exponential function of the distance between x and y.

Tenenbaum and Griffiths (2001) extended Shepard's (1987) analysis of the problem of generalization to cover different kinds of consequential regions and to allow for situations in which not just a single object x but a set of objects X were offered as examples drawn from C. The generalization function $P(y \in C|X)$ can be evaluated via Bayesian inference, computing

$$P(y \in C|X) = \frac{\int_{h \in H|\{X,y\} \subseteq h} P(X|h)P(h)}{\int_{h \in H|X \subseteq h} P(X|h)P(h)}, \quad (14)$$

where h is a hypothetical consequential region, H is the set of all such hypotheses, $P(X|h)$ is the probability of the set of objects X being sampled from the consequential region h, and $P(h)$ is a prior on consequential regions. Tenenbaum and Griffiths (2001) defined the likelihood $P(X|h)$ to be

$$P(X|h) = \begin{cases} \left(\frac{1}{|h|}\right)^n & X \in h \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where $|h|$ is the size of h, being the number of objects in a discrete set or the measure of a continuous hypothesis, and n is the number of objects in X. The likelihood reflects the assumption that the objects were independently sampled uniformly at random from h and implies that smaller, more specific hypotheses will receive higher probabilities than larger, more general hypotheses, even when both are equally consistent with the set of observations. Tenenbaum and Griffiths (2001; Tenenbaum, 1999) called this the *size principle*.

The method for predicting the future outlined above can be expressed in terms of Equation 14. H is the set of all possible durations of a phenomenon, with each h an interval $[0,t_{\text{total}}]$ for $t_{\text{total}} \in [0,\infty)$. Instead of X, we observe T, a set of n observations with maximum value t_{past}, and it is easily seen that Equation 15 becomes Equation 10: The Copernican anthropic principle is equivalent to the size principle. Instead of y, we have a new time t, and thus Equation 14 becomes

$$P(t_{\text{total}} > t|T) = \frac{\int_t^\infty P(T|t_{\text{total}})P(t_{\text{total}})dt_{\text{total}}}{\int_{t_{\text{past}}}^\infty P(T|t_{\text{total}})P(t_{\text{total}})dt_{\text{total}}},$$

which is exactly the same computation as Equation 6.

The problem of predicting the duration of a phenomenon is equivalent to Shepard's (1987) problem of generalization. Although Shepard's original formulation involved generalization through psychological space, predicting the future involves generalization through time. We should thus expect to find similar principles guiding people's generalizations regardless of whether they involve time or some other quantity, as long as they embody the same statistical structure. As mentioned earlier in the article, one problem that has the same structure as predicting the future is predicting healthy levels of imaginary toxins: given a number that is a healthy level of a particular toxin, guessing the highest level of the toxin that would be considered healthy. This situation is exactly analogous to predicting the total duration of an event from the amount of elapsed time since the start of the event. In both cases, one is given a number that is assumed to be randomly sampled from the set of all numbers satisfying a particular criterion and then asked to judge the nature of this criterion. Because both duration and toxin levels are numbers required to be between 0 and some maximum number, this judgment requires the estimation of the maximum number (t_{total} in the case of predicting the future). The results of Experiment 1, in which judgments for healthy levels of toxins were similar to predictions of t_{total}, provides support for the idea that predicting the future may simply be an aspect of the human ability to form meaningful generalizations.

Other problems of estimation and prediction. Our analysis of predicting the future involves a very specific situation in which people need to form an estimate of the total extent or duration of a phenomenon and that quantity is constant across multiple observations. This limits the range of problems to which the simple solution that we present here should be applied. One restrictive assumption is that there is no variability in the total extent or duration across instances of a phenomenon. For example, in our teacake scenario, the coffee shop baked to a fixed but unknown schedule. However, most coffee shops probably bake on a variable schedule throughout the day, meaning that t_{total} is not a single fixed value but something that varies from batch to batch. This can be incorporated into our Bayesian model by making our hypotheses represent different forms for the *distribution* of t_{total}, rather than a single value, but this adds a significant amount of complexity.

This example should make it clear that our Bayesian model does not provide a general solution to problems of estimation and prediction. The Bayesian framework, in which prior knowledge is combined with the information provided by a set of observations, can be used to solve these problems in general, but the nature of the hypotheses under consideration will vary from task to task. The analysis given in the previous section should make it clear that another Bayesian model—Shepard's (1987) account of generalization—corresponds to assuming a slightly more general class of hypotheses. Bayesian models used in other domains, such as models of category learning (e.g., Anderson, 1991), provide strategies for estimating probability distributions over observations, taking different forms for those distributions as the hypotheses that they evaluate. Forming good predictions requires using a set of hypotheses that matches the structure of the problem to be solved, as well as prior knowledge appropriate to that domain.

Conclusion

The ability to effortlessly solve inductive problems is one of the most compelling aspects of human cognition. Bayesian inference provides a natural framework for studying everyday induction, being a formal method for combining rational statistical inferences from observations with rich prior knowledge. These two components are central to understanding how it is that people can solve difficult inductive problems. In this article, we have shown how the problem of predicting the future can be addressed within this framework and examined how well a Bayesian model of predicting the future accounts for people's judgments. The results of our

experiments indicate that people behave in a way that is consistent with our Bayesian model, incorporating information from multiple observations, using prior knowledge, and being sensitive to the statistical structure of the observed data. These results provide constraints on algorithms that might explain how people are able to solve the challenging problem of predicting the future.

References

Allan, L. G. (1979). The perception of time. *Perception & Psychophysics, 26,* 340–354. doi:10.3758/BF03204158

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

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429. doi:10.1037/0033-295X.98.3.409

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408. doi:10.1111/j.1467-9280.1991.tb00174.x

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology, 36,* 203–272. doi:10.1006/cogp.1998.0687

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* New York, NY: Wiley.

Bostrom, N. (2002). *Anthropic bias: Observation selection effects in science and philosophy.* New York, NY: Routledge.

Buch, P. (1994, March 10). Future prospects discussed. *Nature, 368,* 107–108. doi:10.1038/368107b0

Eagleman, D. M., & Holcombe, A. O. (2002). Causality and the perception of time. *Trends in Cognitive Sciences, 6,* 323–325. doi:10.1016/S1364-6613(02)01945-9

Fiedler, K., & Juslin, P. (Eds.). (2006). *Information sampling and adaptive cognition.* Cambridge, England: Cambridge University Press.

Gigerenzer, G., & Brighton, H. (2009). *Homo heuristicus*: Why biased minds make better inferences. *Topics in Cognitive Science, 1,* 107–143. doi:10.1111/j.1756-8765.2008.01006.x

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.

Gott, J. R., III. (1993, May 27). Implications of the Copernican principle for our future prospects. *Nature, 363,* 315–319. doi:10.1038/363315a0

Gott, J. R., III. (1994, March 10). Future prospects discussed. *Nature, 368,* 108. doi:10.1038/368108a0

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15,* 534–539. doi:10.1111/j.0956-7976.2004.00715.x

Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience, 5,* 485–490.

Jaynes, E. T. (1994). *Probability theory: The logic of science* (Fragmentary ed.). Available from http://omega.math.albany.edu:8008/JaynesBook.html

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511790423

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

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113,* 677–699. doi:10.1037/0033-295X.113.4.677

Landsberg, P. T., Dewynne, J. N., & Please, C. P. (1993, September 30). Rise and fall. *Nature, 365,* 384. doi:10.1038/365384e0

Lewandowsky, S., Griffiths, T. L., & Kalish, M. L. (2009). The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. *Cognitive Science: A Multidisciplinary Journal, 33,* 969–998. doi:10.1111/j.1551-6709.2009.01045.x

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

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science: A Multidisciplinary Journal, 32,* 1133–1147. doi:10.1080/03640210802353016

Press, S. J. (1989). *Bayesian statistics: Principles, models, and applications.* New York, NY: Wiley.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation.* New York, NY: McGraw-Hill.

Shepard, R. N. (1987, September 11). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. doi:10.1126/science.3629243

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review, 17,* 443–464. doi:10.3758/PBR.17.4.443

Sternberg, R. J., & Kalmar, D. A. (1997). When will the milk spoil? Everyday induction in human intelligence. *Intelligence, 25,* 185–203. doi:10.1016/S0160-2896(97)90042-8

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). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 59–65). Cambridge, MA: MIT Press.

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. doi:10.1126/science.185.4157.1124

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? Optimal decisions from very few samples. In N. Taatgen & H. van Rijn (Eds.), *CogSci 2009 Proceedings* (pp. 148–153). Wheat Ridge, CO: Cognitive Science Society. Retrieved from http://csjarchive.cogsci.rpi.edu/Proceedings/2009/papers/28/paper28.pdf

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10,* 288–297. doi:10.1111/j.1467-7687.2007.00590.x

Xu, F., & Tenenbaum, J. B. (2007b). Word learning as Bayesian inference. *Psychological Review, 114,* 245–272. doi:10.1037/0033-295X.114.2.245

Received April 15, 2010
Revision received June 13, 2011
Accepted June 17, 2011 ■