Title
Teacakes, Trains, Taxicabs and Toxins: A Bayesian Account of Predicting the Future

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 22(22)

Authors
Griffiths, Thomas L.
Tenenbaum, Joshua B.

Publication Date
2000

Peer reviewed

Teacakes, Trains, Taxicabs and Toxins: A Bayesian Account of Predicting the Future

Thomas L. Griffiths & Joshua B. Tenenbaum
Department of Psychology
Stanford University
Stanford, CA 94305-2130 USA
{gruffydd,jbt}@psych.stanford.edu

Abstract

This paper explores how people make predictions about the future. Statistical approaches to predicting the future are discussed, focussing on the method for predicting the future suggested by J. R. Gott (1993). A generalized Bayesian form of Gott's method is presented, and a specific psychological model suggested. Three experiments show that the predictions people make about the future are consistent with a Bayesian approach.

Despite the difficulty of predicting the future, people happily do it every day. We are confident about being able to predict the durations of events, how much time we will need to get home after work, and how long it will take to finish the shopping. In many cases we have a great deal of information guiding our judgments. However, sometimes we have to make predictions based upon much less evidence. When faced with new situations our decisions about how much longer we can expect events to last are based on whatever evidence is available. When the only information we possess concerns how long a particular event has lasted until now, predicting the future becomes a task of induction.

In this paper we explore the question of how people predict the future when told only about the past. We examine a simple statistical method of predicting the future, and consider how such a method could be made sufficiently flexible to be useful in everyday situations. The resulting Bayesian model makes strong predictions about the effects of providing further information, the symmetry of this form of reasoning, and how it should be affected by prior knowledge. We test these predictions empirically.

The Copernican Anthropic Principle

A simple solution to the problem of predicting the future was recently proposed by the cosmologist J. Richard Gott III (1993). Gott's method is founded upon what he calls the "Copernican anthropic principle", which holds that

> ...the location of your birth in space and time in the Universe is priveleged (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 (1993, p. 316)

Gott extends this principle to reasoning about our position in time – given no evidence to the contrary, we should not assume that we are in a "special" place in time. This means that the time at which an observer encounters a phenomenon should be randomly located in the total duration of that phenomenon.

Denoting the time since the start of a phenomenon t_{past}, and its total duration t_{total}, Gott forms what he terms the "delta t argument". Define the ratio

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

and assume that this is a random number between 0 and 1. It is possible to form probabilistic predictions about the value of r. For example, r will be between 0.025 and 0.975 with a probability $P = 0.95$, meaning that

$$\frac{1}{39}t_{past} < t_{future} < 39t_{past} \tag{2}$$

with 95% confidence, where $t_{future} = t_{total} - t_{past}$. Similarly, r will be less than 0.5 with probability $P = 0.5$, so $t_{past} < t_{future}$ with 50% confidence.

This method of reasoning has been used to predict a wide range of phenomena. Gott (1993) tells of his visit to the Berlin Wall in 1969 ($t_{past} = 8$ years). Assuming that his visit was randomly located in the period of the wall's existence, the 95% confidence interval for t_{future} would be 2.46 months to 312 years. The wall fell 20 years later, consistent with these predictions. Gott made similar calculations of t_{future} for Stonehenge, the journal *Nature*, the U.S.S.R., and even the human race. Subsequent targets of the principle have included Broadway musicals and the Conservative government in Britain (Landsberg, Dewynne, & Please, 1993).

What's Bayes got to do with it?

Gott's (1993) method for predicting the future yields interesting predictions in a wide range of situations. It is simple, but could prove useful in forming effective plans and expectations about future events. On this basis, it would be plausible for people to apply similar principles when making judgments concerning time.

Despite the attractiveness of this claim, there may be good reasons why Gott's (1993) method would not belong in our cognitive armory. One reason could be the restrictive assumptions of such an inference. In many cases in the real world where it might be desirable to predict the future, we know more than simply how long a process has been underway. In particular, our interaction with the world often gives us some prior expectations about the duration of an event. For example, meeting a 78 year-old man on the street, we are unlikely to think that there is a 50% chance that he will be alive at the age of 156.

Prior knowledge is not the only kind of information that Gott's (1993) method neglects. In some cases, our predictions are facilitated by the availability of multiple encounters with a phenomenon. For example, if we were attempting to determine the period that passes between subway trains arriving at a station, we would probably have several trips upon which to base our judgment. If on our first trip we discovered that a train had left the station 103 seconds ago, we might assume that trains run every few minutes. But, after three trips yield trains that have left 103, 34, and 72 seconds ago, this estimate might get closer to 103 seconds. And after ten trains, all leaving less than 103 seconds before we arrive, we might be inclined to accept a value very close to 103 seconds.

These limitations suggest that Gott's (1993) formalization lacks the flexibility that would be required of a method for predicting the future in real world situations. Such a method must allow for the influence of prior knowledge, and reflect the effect of multiple examples. Bayesian inference may provide a means of satisfying both of these requirements. Bayes Theorem states that

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)} \quad (3)$$

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 of the data given that hypothesis. In the case where the hypotheses are continuous, $P(d)$ can be obtained by summing across $P(d|h)P(h)$ for all hypotheses, giving

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{\int_{h \in H} P(d|h)P(h)dh} \quad (4)$$

where H is the set of all hypotheses.

Conveniently, some work extending Gott's (1993) method into a Bayesian framework already exists. 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. Using the prior $P(t_{total}) \propto \frac{1}{t_{total}}$ and the likelihood $P(t_{past}|t_{total}) = \frac{1}{t_{total}}$ yields the same results as his original formulation of the delta t argument.

These values are not chosen arbitrarily. The use of $\frac{1}{t_{total}}$ for the prior is motivated by sound statistical principles (Press, 1989), and provides a scale-invariant method for distributing probability over hypotheses. In many applications it is referred to as the uninformative prior, as it is appropriate when an inference is guided by no specific prior knowledge. The use of $\frac{1}{t_{total}}$ for the likelihood is also well motivated. The anthropic principle essentially states that t_{past} should be viewed as having been randomly sampled from t_{total}. Assuming a uniform distribution across values of t_{past}, the probability of any particular t_{past} will be $\frac{1}{t_{total}}$.

Crucially, both the priors and the likelihoods of this Bayesian framework can be modified to suit the situation at hand. A simple, flexible set of priors is provided by the Erlang distribution

$$P(t_{total}) = \frac{t_{total} e^{-t_{total}/\beta}}{2\beta^2} \quad (5)$$

where β is a free parameter. This distribution has a broad peak at $t_{total} = \beta$, and decays to zero at 0 and ∞. This parameterized peaked distribution provides a simple means to summarize many of the kinds of distributions that might be encountered across temporal domains. Similarly, the effect of multiple examples of t_{past} can be introduced by modifying the likelihoods. Extending the anthropic principle, we can assume that each example is drawn independently from t_{total}. The probability of observing a set of n examples will then be $(\frac{1}{t_{total}})^n$.

A Bayesian model

The specification of a Bayesian model requires identifying the distributions governing the prior probabilities and the likelihoods. For the case of predicting the future, Gott's (1993) method provides a good starting point. For each model considered in this section, we will assume that people's responses reflect the point t, such that $P(t < t_{total}) = 0.5$. This is essentially assuming that people adopt an unbiased criterion in making their judgments. Furthermore, we will examine the predictions of each model when provided with one, three, or ten examples of t_{past} ($n = 1, 3, 10$ respectively).

The simplest case of Gott's (1993) method is just the delta t argument, as presented in Equation 1. The predictions of the delta t argument are shown in the leftmost panel of Figure 1(a). The predictions are unaffected by n, and are thus constant at $t = 2t_{past}$. This seems to defy intuition, and is the weakest of the models we will consider.

At the next level of complexity is the introduction of $(\frac{1}{t_{total}})^n$ for the likelihood of a set of n examples and the uninformative prior $P(t_{total}) \propto \frac{1}{t_{total}}$, yielding the closed form prediction $t = 2^{1/n} t_{past}$. As shown in the second panel of Figure 1(a), the model shows an effect of the number of examples. The main problem with this model is that the prior does not make use of the flexibility provided by the inclusion of prior knowledge in inference. In particular, the uninformative prior makes scale invariant predictions about generalization, which means that t_{future} will be a constant proportion of t_{past}, whether predicting the future of the human race or a 78-year old man.

Substituting the Erlang distribution for the uninformative prior renders Equation 4 into

$$P(t_{total}|T) = \frac{(t_{total})^{1-n} e^{-t_{total}/\beta}}{\int_{t_{past}}^{\infty} (t_{total})^{1-n} e^{-t_{total}/\beta} dt_{total}} \quad (6)$$

where T is the set of n examples of t_{past}. The third panel of Figure 1(a) shows the predictions of this model for $n = 1, 3, 10$ with β ranging from 0.5 to 4.5 in unit increments. The model shows reduced predictions with more examples, and increased predictions with larger values of β.

This third model is what will be tested against people's predictions about the future. The use of a parameterized prior and specification of the likelihoods makes it appear somewhat more complex than Gott's (1993) original prescription. As parsimony contributed to the psychological plausibility of the approach, this additional complexity needs to be justified. One source of justification is the success of the Bayesian framework in describing behavior in other domains. Tenenbaum (1999) discussed several examples of tasks that map naturally onto the temporal problems addressed by Gott.

The tasks considered by Tenenbaum (1999) were cases of inductive concept learning, where people learn about a concept through the provision of positive examples. An example of this kind of task would be predicting healthy levels of imaginary toxins. People would be given a number that they are told is a healthy level of a particular toxin, then asked to guess 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, a person is given a number that is assumed to be randomly sampled from the set of all numbers satisfying a particular criterion, and asked to judge the nature of this criterion. Since both duration and toxin levels are numbers required to be between 0 and some maximum number, this judgement requires the estimation of the maximum number (t_{total} in the case of predicting the future). Tenenbaum (1999) found that a Bayesian framework gave a good account of people's performance on this kind of task.

Model predictions

The Bayesian model outlined above, and depicted in the third panel of Figure 1(a), has some obvious implications. Most central is how the provision of further information should affect predictions. The tightening of the range of acceptable values of t_{total} corresponds to an important component of Tenenbaum's (1999) account of concept learning. As people are given more examples of a concept, they become less inclined to generalize beyond the properties of those examples. In the case of predicting the future, where all hypotheses have a value of t_{total} as their sole property, this manifests as a tendency to accept the smallest possible value of t_{total} that includes all observed values of t_{past}.

The Bayesian model also clearly defines the phenomena of prediction to be symmetric. In forming a judgment about t_{total}, knowledge of t_{past} and knowledge of t_{future} are equally informative. Given one of these pieces of information, it is possible to calculate a range of acceptable values for t_{total}. If people apply similar methods in making judgments about time the effects should be maintained regardless of which of t_{past} and t_{future} are provided. This is not an absolute symmetry, however: if scenarios in which t_{future} and t_{past} are provided differ in the distribution of prior probability, then the predicted values of t_{total} may also differ.

One further implication of the model is 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 will bias inferences towards hypotheses around that value. If people employ approximately Bayesian methods in forming their judgments, introducing information that biases the prior probability distribution in this way should result in higher predictions, especially when those predictions are based upon few observations.

The Bayesian framework for predicting the future has three clear implications for the kind of judgments that it will produce. The effects of further information, symmetry of predictions, and effects of prior probabilities are all important properties of how these judgments are made. Experiments 1, 2, and 3 examine these predictions in turn.

Experiment 1: New information

Method

Participants Participants were 81 undergraduates from Stanford University, participating for partial course credit. The participants were randomly assigned to four groups.

Materials Four simple scenarios were developed for exploring the predictions of the Bayesian framework. The first scenario described a coffee shop that had recently started selling teacakes. This scenario is given below. Participants were shown t_{past}, and asked to predict t_{total}. The second scenario told participants that they were visiting a foreign country in which trains ran precisely to schedule. The schedule was set up so that exactly the same amount of time passed between successive trains. On the platform was a clock showing how long it had been since the last train arrived. Participants were told the value on the clock when they reached the station, t_{future}, and asked to predict t_{total}.

These scenarios were compared with two analogous situations that made no reference to time. One of the comparison scenarios was the healthy levels of toxin experiment described above. The second 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.

Each scenario had three sections. The first section outlined the situation and give a single number on which judgments were to be based. The second and third sections added further information, giving a total of three numbers and ten numbers respectively. The first number given was the largest, meaning that further observations would only tighten the range of generalization. The sets of numbers given were identical for the teacake and toxin scenarios and the train and taxicab scenarios, and were approximately uniformly distributed. The largest example was 34 minutes (ng/mL) for the teacake (toxin) scenario, and 103 seconds (cabs) for the train (taxicab) scenario.

For example, the first section of the teacake scenario was

> 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.
>
> 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 section gave the additional times of 21 and 8 minutes, and the third section gave further times of 18, 2, 5, 27, 22, 10 and 14 minutes.

Procedure The procedure used in all three experiments was identical: Each participant received a sheet providing general instructions about the task, and a questionnaire of one of four kinds.

Results and Discussion

Plausible responses to these problems are constrained to be greater than the largest example provided. Responses were transformed such that $t = \frac{x}{t_{past}}$, where x is the raw score and t_{past} is the largest example, and participants who gave $t < 1$ were excluded from the analysis, eliminating approximately 15% of the participants in each experiment. Responses more than three standard deviations from the mean were considered outliers, and were also excluded. Only one outlier was identified in the course of all three experiments.

A one-way within-subjects ANOVA showed a statistically significant effect of the number of examples for each scenario ($F(2,30) = 9.71$, $F(2,32) = 18.00$, $F(2,44) = 9.57$, $F(2,30) = 15.05$, for the teacake, train, taxicab, and toxin respectively, all $p < .001$). Means and standard errors are shown in Figure 1(b), which demonstrate that the two temporal tasks show a similar effect of new information to the tasks of analogous statistical structure. The Figure also shows predictions generated by the Bayesian model, using the Erlang prior with β set independently for each scenario. 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 respectively. The peak of the Erlang prior is at $t_{total} = \beta$, yielding values of 54 minutes between batches of teacakes, 9 minutes 21 seconds between trains, 24 ng/mL of toxin, and 460 taxicabs, all of which seem appropriate.

Experiment 2: Symmetry of effects

Method

Participants Participants were another 77 undergraduates from Stanford University, participating for partial course credit. The participants were randomly assigned to four groups.

Materials Again, four scenarios were used. The teacake and train scenarios from Experiment 1 were applied to a different set of participants, and modified versions of these scenarios were generated to make it possible to test the symmetry of predictions. The new scenarios were identical to the original teacake and train scenarios, except for the provision of t_{future} instead of t_{past} - the participants were informed how long it would be before the next batch of teacakes were removed from the oven, or the next train would arrive, and were asked to predict the period that went between these events. All scenarios asked for predictions with 1, 3, and 10 examples, replicating Experiment 1.

Results and Discussion

Responses were screened using the same procedure as in Experiment 1. Each pair of putatively symmetric scenarios was subjected to a two-way within-between ANOVA, examining the effects of number of examples and temporal direction. The train scenario showed a statistically significant effect of number of examples ($F(2,56) = 17.40$, $p < .001$), and no evidence of an effect of temporal direction ($F(1,28) = 0.50$, $p = 0.49$) or an interaction between the factors ($F(2,56) = 0.08$, $p = 0.93$). The effect of new information reproduces that of Experiment 1, and the non-significant result for temporal direction implies that any asymmetry in prediction is too weak to be detected by the present experiment.

The teacake scenario likewise showed a statistically significant effect of number of examples ($F(2,60) = 22.18$, $p < .001$). However, the results also indicated a significant effect of temporal direction ($F(1.30) = 4.46$, $p < .05$) and an interaction between the two factors ($F(2,60) = 3.667$, $p < .05$). This difference between the two scenarios may be a result of the way that the asymmetry introduces new information about the teacakes. Changing "the clock shows that it has been 34 minutes since the last batch of teacakes was removed from the oven" to "the clock shows that it will be 34 minutes until the next batch of teacakes is removed from the oven" provides the implication that the next batch of teacakes is currently in the oven. The time between batches of teacakes can be divided into time in the oven and time left waiting. Of these, the time the teacakes spend in the oven is less flexible. Applying the anthropic principle, the observation that the teacakes are currently in the

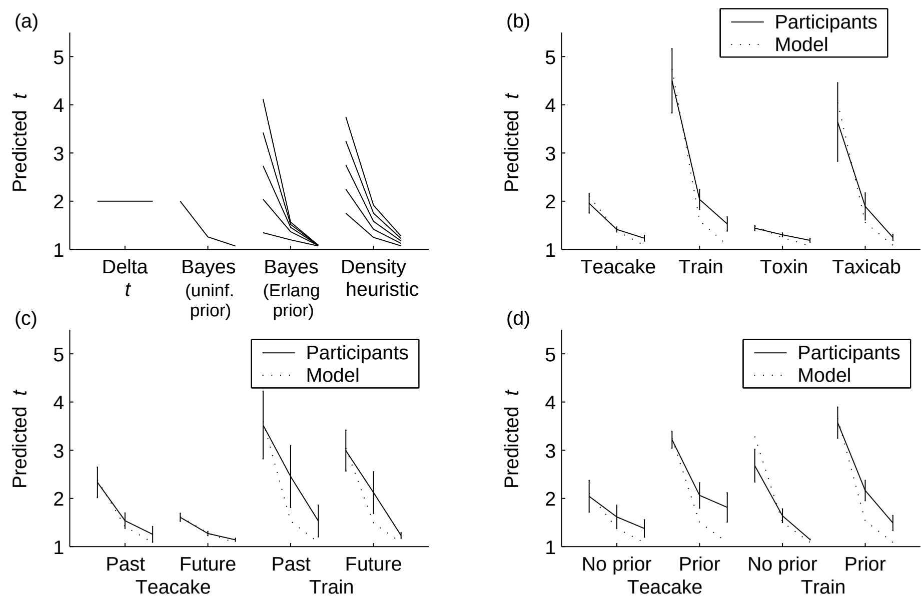

Figure 1: (a) Predictions of the various models, depicting point at which $P(t < t_{total}) = 0.5$ for 1, 3 and 10 examples. On all graphs, the vertical axis shows the predicted value of t in proportion to t_{past}. (b) Results for Experiment 1. The solid line shows means (with one standard error). The dotted line shows the predictions of the Bayesian model with an Erlang prior. (c) Results for Experiment 2. (d) Results for Experiment 3.

oven suggests that the teacakes spend more time in the oven than waiting to be purchased. This correspondingly reduces the total amount of time that might be expected to pass between successive batches of teacakes.

Figure 1(c) shows the means and standard errors, which are reminiscent of those found in Experiment 1. The predictions of the model were made with β values of 2, 0.91, 3.75, and 2.95 for teacake-past, teacake-future, train-past, and train-future respectively. These values approximate those used in fitting the results of Experiment 1.

Experiment 3: Manipulating priors

Method

Participants Participants were another 78 undergraduates from Stanford University, participating for partial course credit. The participants were randomly assigned to four groups.

Materials The teacake scenario from Experiment 1 and the train scenario from Experiment 2 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 upon 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.

All scenarios asked for predictions with 1, 3, and 10 examples, replicating Experiment 1.

Results and Discussion

Responses were screened using the same procedure as in Experiment 1. The scenarios were grouped into teacakes and trains, 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 ($F = 25.86, p < .001$) and manipulating priors ($F = 4.70$, $p < .05$), as well as an interaction between the two ($F = 3.80, p < .05$). Similar results were shown for the train scenarios. There was a statistically significant effect of the number of examples ($F = 50.31$, $p < .001$), as well as an effect of manipulating priors ($F = 5.85$, $p < .05$). 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 the Bayesian model.

Means and standard errors are shown in Figure 1 (d), together with the model predictions. The β values used in fitting the data were 1.6, 3.3, 3.25, and 3.85 for the teacake, teacake-prior, train, and train-prior conditions respectively. The β values are greater in the conditions where the prior was raised, and give peak values of 1 hour, 52 minutes for the teacakes and 6 minutes 40 seconds 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.

General Discussion

Predicting the future is a difficult task, particularly when the predictions are formed on the basis of very little information. Gott (1993) suggested a simple method for making predictions about the future. Gott's method lacks the flexibility to be useful in a wide range of real-world situations, but the same principles allow the construction of a more general Bayesian model. This model shows many similarities to Tenenbaum's (1999) Bayesian framework for inductive concept learning, which has proven successful in other domains. Experiments 1, 2, and 3 explored whether the Bayesian model provided a reasonable account of the effects of new information, the symmetry of predictions, and the effects of prior probabilities upon people's judgments about the future.

Experiment 1 showed that the effect of providing further examples conformed to the predictions of the 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 these predictions were symmetric in time, taking the same form regardless of whether the judgment concerned the past or the future. Experiment 3 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.

Psychological plausibility

It seems unlikely that the participants in this experiment were consciously performing Bayesian inference. A more probable explanation is that the problem can be solved by the application of a simple heuristic, which more closely resembles the cognitive process by which answers can be reached (cf. Gigerenzer, 1999).

One simple heuristic that produces results consistent with both the data and the normative Bayesian model is the rule "The distance to extrapolate is the range of scores divided by the number of examples", which is an unbiased frequentist estimator of t_{total}. The predictions of this 'density' heuristic are shown in the fourth panel of Figure 1(a). Priors are implemented by taking the average of t_{past} and β before dividing by n, and the curves shown illustrate β ranging between 0.5 and 4.5 in unit increments. Note that the extent of generalization decreases with the number of examples, as in the present experiments.

The existence of such a heuristic does not affect the claim that people's predictions about temporal events are consistent with a Bayesian framework. The algorithmic properties of the Bayesian model and the heuristic may differ, but their computational properties are similar. In fact, producing the response $t = t_{past} + \frac{t_{past}}{n}$, where n is the number of examples, serves as a first order approximation of the Bayesian model with prior $P(t_{total}) \propto \frac{1}{t_{total}}$.

Predicting future research

The present results provide support for the claim that a process consistent with Bayesian inference underlies people's judgments on these tasks. However, the generality of the findings needs to be extended. In particular further scenarios need to be investigated. Demonstrating the consistency of the results across a range of contexts and set of numerical examples will increase the strength of the findings.

The model-fitting presented in the preceding experiments helps to show that the results are consistent with a Bayesian model, but one of the most important outcome of the model-fitting is that the peak values of the prior distributions are in appropriate ranges. This provides a further avenue for future research: empirically estimating the parameters of the prior distribution, and using these results to predict the effects of providing examples. This would involve administering the scenarios without giving a time displayed on the clock, and asking people to estimate t_{total}. The resulting distribution will give a prior, $P(t_{total})$, that can be used to predict further responses.

Finally, it is interesting to note that the ability to predict the future may be important to domains other than conscious planning. For instance, Anderson (1990) argued that memory may display similar temporal sensitivities. The major challenge for any memory system is to index entries in a fashion whereby those that are needed will be readily available. This requires predicting the future: given an event, the expected future occurrence of the event must be inferred if it is to be indexed appropriately. Anderson (1990) suggests 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 reflect Bayesian principles.

Acknowledgements

This research was supported in part by a grant from MERL, and a Hackett Studentship to the first author. For valuable discussions, we thank Mira Bernstein, Roger Shepard and David Somers.

References

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Lawrence Erlbaum, Hillsdale, NJ.

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

Gigerenzer, G. (1999). *Simple heuristics that make us smart*. Oxford University Press, Oxford.

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

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

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford.

Landsberg, P., Dewynne, J., and Please, C. (1993). Rise and fall. *Nature*, 365:384.

Press, S. J. (1989). *Bayesian statistics: Principles, Models, and Applications*. Wiley, New York.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In Kearns, M. S., Solla, S. A., and Cohn, D. A., editors, *Advances in Neural information Processing Systems 11*, pages 59–65. MIT Press, Cambridge, MA.