Inferring Subjective Prior Knowledge: An Integrative Bayesian Approach

Sean Tauber (sean.tauber@uci.edu)
Mark Steyvers (mark.steyvers@uci.edu)
Department of Cognitive Sciences, University of California, Irvine
Irvine, CA 92697 USA

Abstract

The standard approach to Bayesian models of Cognition (also known as rational models) requires researchers to make strong assumptions about people's prior beliefs. For example, it is often assumed that people's subjective knowledge is best represented by "true" environmental data. We show that an integrative Bayesian approach—combining Bayesian cognitive models with Bayesian data analysis—allows us to relax this assumption. We demonstrate how this approach can be used to estimate people's subjective prior beliefs based on their responses in a prediction task.

Keywords: Bayesian modeling; rational analysis; cognitive models; Bayesian data analysis; Bayesian inference; knowledge representation; prior knowledge

Introduction

In the standard approach to Bayesian models of Cognition (also referred to as rational models), researchers make strong assumptions about people's prior beliefs in order to make predictions about their behavior. These models are used to simulate the expected behavior—such as decisions, judgments or predictions—of someone whose computational-level solution to a cognitive task is well described by the model. Analysis of Bayesian models of cognition usually involves a qualitative comparison between human responses and simulated model predictions. For an overview of Bayesian models of cognition see Oaksford and Chater (1998); but also see Mozer, Pashler, and Homaei (2008); and Jones and Love (2011) for a critique.

As an alternative to the standard approach, we present an integrative Bayesian approach that allows us to relax the assumptions about people's prior beliefs. This approach is motivated by previous efforts to infer subjective mental representations (Lewandowsky, Griffiths, & Kalish, 2009; Sanborn & Griffiths, 2008; Sanborn, Griffiths, & Shiffrin, 2010) and more specifically to combine Bayesian models of cognition and Bayesian data analysis (Huszar, Noppeney & Lengyel, 2010; Lee & Sarnecka, 2008). The integrative approach allows us to use people's responses on a cognitive task to infer posterior distributions over the psychological variables in a Bayesian model of cognition. It also allows us to estimate probabilistic representations of people's subjective prior beliefs.

We recently applied this approach to a Bayesian cognitive model of reconstructive memory (Hemmer, Tauber, & Steyvers, in prep). We estimated individuals' subjective prior beliefs about the distribution of people's heights based on their responses in a memory task. The technical requirements for integrated Bayesian inference were simplified because the posterior distribution, based on inference in the cognitive model, had a simple Gaussian form. This made it straight forward to define individuals' responses as Gaussian distributed random variables in an integrated Bayesian model.

In this study, we develop a method for applying integrated Bayesian inference that does not require the posterior of the cognitive model to have a simple parametric form. We apply this method to a Bayesian cognitive model for predictions that was developed by Griffiths and Tenenbaum (2006). Their Bayesian model of cognition was a computational-level description of how people combine prior knowledge with new information to make predictions about real-world phenomena. They asked participants to make a series of predictions about duration or extent that were similar to the following examples:

If you were assessing the prospects of a 60-year-old man, how much longer would you expect him to live?

If you were an executive evaluating the performance of a movie that had made $40 million at the box office so far, what would you estimate for its total gross?

All of the questions used by Griffiths and Tenenbaum (2006) were based on real-world phenomena such as, life spans, box office grosses for movies, movie runtimes, poem lengths and waiting times. Their assumption was that people make predictions about these phenomena based on prior beliefs that reflect their true extents or durations in the real world.

Although it is possible that people's beliefs about these phenomena are tuned to the environment, this assumption cannot be used to explain how people make similar sorts of predictions about counterfactual phenomena that have no true statistics in the environment. For example, consider the following question:

Suppose it is the year 2075 and medical science has advanced significantly. You meet a man that is 60 years old. To what age will this man live?

There is no "true" answer to this question and therefore no environmental data is available. This creates a problem for a Bayesian model of cognition that requires environmental data in order to make predictions.

Environmental Statistics as Prior Knowledge

Researchers can use Bayesian models of cognition to simulate the responses that people would make if their computational-level solution to the prediction problem is well described by the model. This process requires that the model includes representations of the prior knowledge people have about the phenomena being predicted. Researchers can represent prior knowledge in their models by collecting real-world environmental statistics and using them in their models as a stand-in for the subjective prior knowledge of individuals (Griffiths & Tenenbaum, 2006 Hemmer & Steyvers, 2009a; Hemmer & Steyvers, 2009b). Representing prior knowledge in this manner is based on the assumption that our knowledge and representations about real-world phenomena are based on actual exposure to these phenomena in the environment. A researcher's best guess at a participant's knowledge is that it reflects, on average, the actual statistics of that phenomenon in the environment.

Standard Qualitative Analysis

In the standard approach to Bayesian cognitive modeling, researchers qualitatively compare model predictions to people's responses. The values of psychological parameters—which represent aspects of cognition that are "in people's heads"—are manually specified or estimated with non-Bayesian methods. For a critique of non-Bayesian analysis of Bayesian models, see Lee (2011). The researcher usually encodes subjective prior knowledge in the model using empirical priors (based on environmental data) or by specifying parametric priors with psychological parameters.

A limitation of this method is that researchers do not apply Bayesian inference techniques to participant response data, in order to make inferences about the prior knowledge and psychological parameters represented in the model. It does not allow for the possibility that participants' prior knowledge could be different from the form assumed by the researcher. Furthermore, a model that requires prior knowledge from real-world data cannot be used to generate predictions if the researcher is unable to encode this data in the model. For example, Griffiths' and Tenenbaum's (2006) model cannot be used to generate predictions for the counterfactual future life spans question; even though it involves the same sort of task as the factual prediction questions.

Quantitative Analysis: An Integrative Bayesian Approach

The limitations of the qualitative approach can be addressed by reframing a Bayesian model of cognition as a generative process for human response data. Researchers can then use an integrative Bayesian approach to make inferences about the subjective aspects of the cognitive model.

A Bayesian Model of Cognition for Predictions Griffiths and Tenenbaum (2006) had people make simple predictions about the duration or extent of real-world phenomena. For example, when told that a man was currently 60 years old, people had to predict the age to which he would live. We refer to the value that is presented in the question as t and to the person's prediction as t_{total}. So if a person predicted that the man would live to be 8o years old, then we would have $t = 60$ and $t_{total} = 80$.

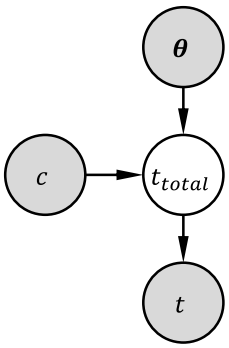

Figure 1. Graphical model (observer perspective)

The Bayesian model of cognition proposed by Griffiths and Tenenbaum used nonparametric environmental priors for t_{total}. We use a modified version of their model in which t_{total} has a parametric prior that is Normal, Erlang or Pareto distributed. We add a switch c that selects which parametric form is used for the prior.

Figure 1 is a graphical representation of our cognitive model for duration and extent from the perspective of the person making predictions (the observer). Shaded nodes represent variables that contain information that is known to the observer. Unshaded nodes contain information that is unknown to the observer.

The model depicts an observer's subjective model of the conditional dependencies between total duration/extent t_{total} of phenomena of different types c—which are determined by the form of the observer's prior knowledge for the domain. The vector $\boldsymbol{\theta}$ parameterizes prior distribution types such that $\theta_1, \theta_2, \theta_3$ parameterize Normal, Erlang and Pareto types, respectively. We specify the prior distribution t_{total} as:

$$t_{total} \sim \begin{cases} Norm(\theta_1) & , \quad c = 1 \\ Erlang(\theta_2) & , \quad c = 2 \\ Pareto(\theta_3) & , \quad c = 3 \end{cases} \qquad (1)$$

The time or duration t from which the observer must predict t_{total} is equally likely for all possible values $0 < t < t\_total$. We implemented this in the model by placing a uniform prior on t:

$$t \sim Unif(0, t_{total}) \qquad (2)$$

When presented with a prediction question with value t, we assume that observers access the relevant prior knowledge of t_total by determining the prior type c and the parameter

values θ_c and then infer a posterior distribution $P(t_{total}|t,c,\boldsymbol{\theta})$ that is described using Bayes' rule:

$$P(t_{total}|t,c,\boldsymbol{\theta}) \propto$$

$$\begin{cases} Unif(t|0,t_{total})f(t_{total}|\theta_c), & t \le t_{total} \\ 0 \quad , & t > t_{total} \end{cases} \tag{3}$$

where,

$$f(x|\theta_c) = \begin{cases} Norm(x|\theta_1) \quad , & c = 1 \\ Erlang(x|\theta_2) \ , & c = 2 \\ Pareto(x|\theta_3) \ , & c = 3 \end{cases} \tag{4}$$

Finally, the observer provides a prediction for the total extent or duration. This response is based on the posterior distribution $P(t_{total}|t,c,\boldsymbol{\theta})$, and could be related to the posterior in a number of ways. The response R could be a sample from the posterior,

$$R \sim P(t_{total}|t,c,\boldsymbol{\theta}) \tag{5}$$

or it could be a function of the posterior such as the median, mean or mode. Griffiths and Tenenbaum (2006) modeled predictions as the median of the posterior. We assume that each response is based on a single sample from the posterior. This assumption provides a technical simplification for modeling how people generate a response from the posterior distribution. We will not explore the theoretical implications of this assumption in depth; however, there is evidence supporting a response model that is based on limited samples from a posterior (Vul, Goodman, Griffiths & Tenenbaum, 2009).

Applying Bayesian Data Analysis to the Bayesian Model of Cognition The goal of the researcher is to apply Bayesian data analysis to the Bayesian model of cognition in order to infer the values of θ and c given t and observer predictions R about t_{total}. This requires an integrative application of Bayesian inference from the perspective of the researcher. Each and every value of θ and c for which the researcher wishes to evaluate the posterior likelihood requires Bayesian inference of the posterior likelihood of the observer's response in the rational model given the values of θ and c.

From the perspective of the researcher, the responses provided by an observer are the result of a generative process that encapsulates an application of Bayesian inference to a Bayesian model of cognition (fig. 1) resulting in a posterior distribution (Eq. 3) from which the result is sampled. We call this generative process a *Bayesian Inference and Response Process* (BIRP) and define it as a probability distribution with likelihood function:

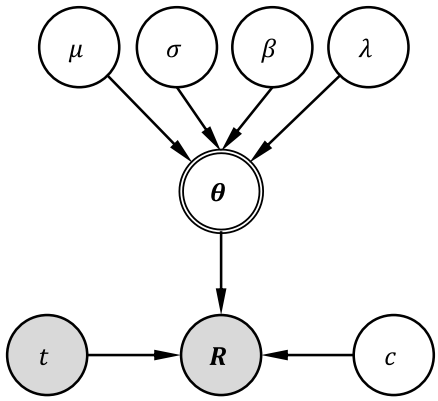

Figure 2. Graphical model (researcher perspective)

$$BIRP(R|t,\theta,c) \propto \begin{cases} Unif(t|0,R)f(R|\theta_c), & t \le R \\ 0 \quad , & t > R \end{cases} \tag{6}$$

Figure 2 shows a graphical model from the perspective of the researcher that incorporates a BIRP. In this model the original stimulus t and the observer responses R are data that is known to the researcher. The form of the prior distribution used by the observer is indexed by c, and the parameters for the observer's possible prior distributions are all latent (unobserved) variables for which posterior distributions will be inferred. Observer responses R are generated as samples from the BIRP:

$$R \sim BIRP(t,\theta,c) \tag{7}$$

The researcher must place suitable hyper priors on the latent prior type c and latent parameters for the observer prior distributions μ, σ, β and λ. We define the deterministic vector $\boldsymbol{\theta} = \langle(\mu,\sigma),\beta,\lambda\rangle$ for the purpose of notational compactness.

Experiment

We described an integrative Bayesian approach that allows us to make inferences about people's subjective beliefs based on their responses in a prediction task. We ran an experiment in order to collect people's predictions for several of the same questions used by Griffiths and Tenenbaum (2006). We also collected predictions for the counterfactual lifespans question.

Method

Participants

A total of 25 undergraduates from the University of California, Irvine participated in the study and were compensated with partial course credit.

Materials

Prediction questions were presented to participants through a web-based survey. There were 8 different question types

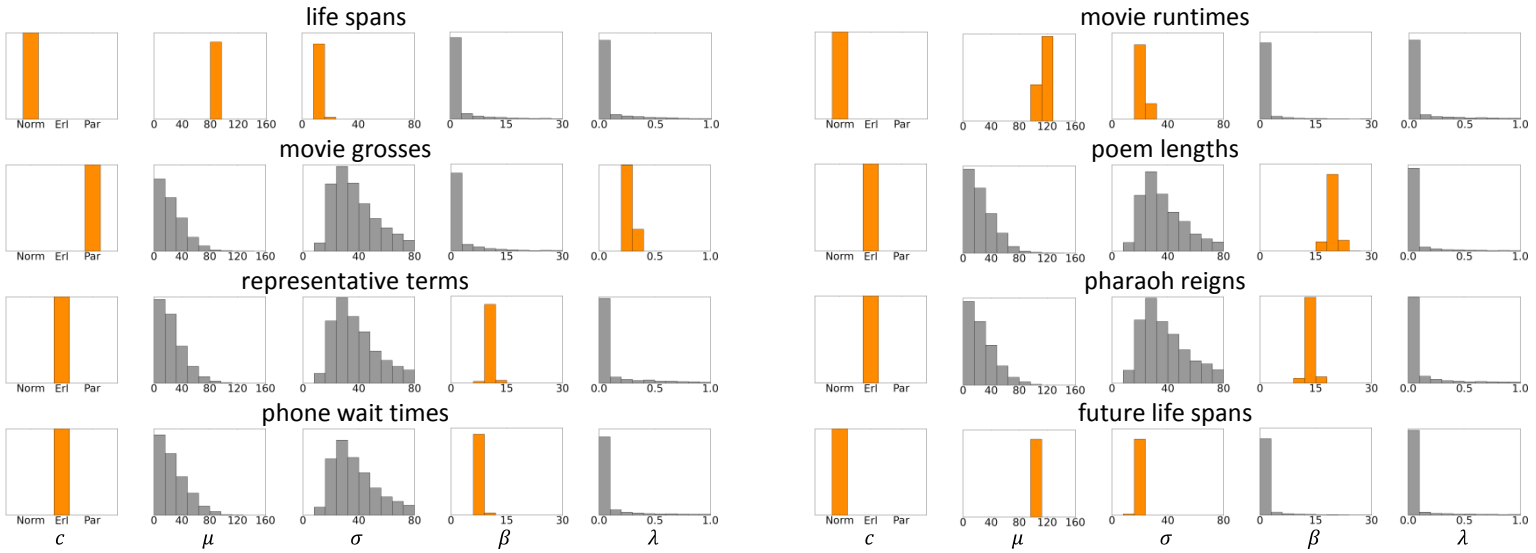

Figure 3. Posterior distributions of people's subjective prior types and parameter values from the researcher's perspective. For each of the eight question types the subplot for the indicator variable c shows the relative posterior probability for each of the prior types (normal, Erlang, or Pareto). The remaining subplots show the posterior distributions of the parameters for these prior types. Parameters that correspond to prior types with zero posterior probability are shown in gray.

and 5 variations of each question. Each variation corresponded to 1 of 5 possible values of t.The survey instructions and 7 of the questions were identical to those used by Griffiths and Tenenbaum (2006). For the unabbreviated questions and survey instructions, refer to Griffiths and Tenenbaum (2006). Below are abbreviated examples of each of the questions with all 5 of the possible t values included: (1*) Predict the age a man will live to if he is currently (18, 39, 61, 83, 96) years old*; (2) *Predict what the total box-office intake for a movie that has taken in* ($1, $6, $10, $40, $100) *so far*; (3) *Predict the length of a movie that has already been playing for (30, 60, 80, 95, 110) minutes*; (4) *Predict the total length of a poem from which you were just quoted line (2, 5, 12, 32, 67)*; (5) *Predict the total time a pharaoh will be in power if he had already reigned for (1, 3, 7,11, 23) years in 4000 BC*; (6) *Predict the total years that a (1, 3, 7, 15, 31) year member of the U.S. House serve*; (7) *Predict how long you will be on hold if you have already been holding on the phone for (1, 3, 7, 11, 23) minutes*. There was an eighth question that was not part of the Griffiths and Tenenbaum study: *Suppose it is the year 2075 and medical science has advanced significantly. You meet a man that is (18, 39, 61, 83, 96) years old. To what age will this man live?*

Procedure

Each participant made a prediction about all 5 instances of the 8 different types of phenomena for a total of 40 questions. Each prediction was based on one of the five possible values of t. The questions were presented in a different random order for each participant. Only one question was presented on-screen at a time and participants entered their answer in a text-entry box before moving to the next question.

Inference and Data Analysis

Responses from each participant were considered for exclusion on a per question-type basis. If any of a participant's five responses for one of the eight question-types were below the value of t that was presented in the question, then all five of that participant's responses for that question-type were excluded for analysis but their responses for other question-types were still included—as long as they passed the inclusion requirement above. The number of participants that were included in the analysis for each question-type was: 24 for life spans; 23 for box office intake; 23 for movie durations; 25 for poem lengths; 24 for pharaoh reigns; 20 for U.S. representative terms; and 25 for lifespans in the future.

We aggregated participant responses for each question such that each response provided an additional data point for Bayesian analysis. We implemented a customized Markov-chain Monte Carlo (MCMC) sampler to perform Bayesian inference using the researcher model. To complete the model, we used the following priors:

$$c \sim Cat\left(\tfrac{1}{3}.\tfrac{1}{3},\tfrac{1}{3}\right) \qquad \mu \sim HalfNorm(prec = .001)$$
$$\rho \sim Ga(1{,}1000) \qquad \sigma = sqrt(1/\rho)$$
$$\beta \sim Ga(.1,.05) \qquad \lambda \sim Ga(.1,.05)$$

Results

Figure 3 shows a complete summary of the posterior distributions for the subjective prior types as well as the posteriors for the psychological variables that parameterized the subjective priors. We used people's predictions to infer the posterior probability that their subjective prior knowledge for each domain was best characterized by a Normal, Erlang or Pareto distribution. Although the inference allowed for uncertainty about the form of the

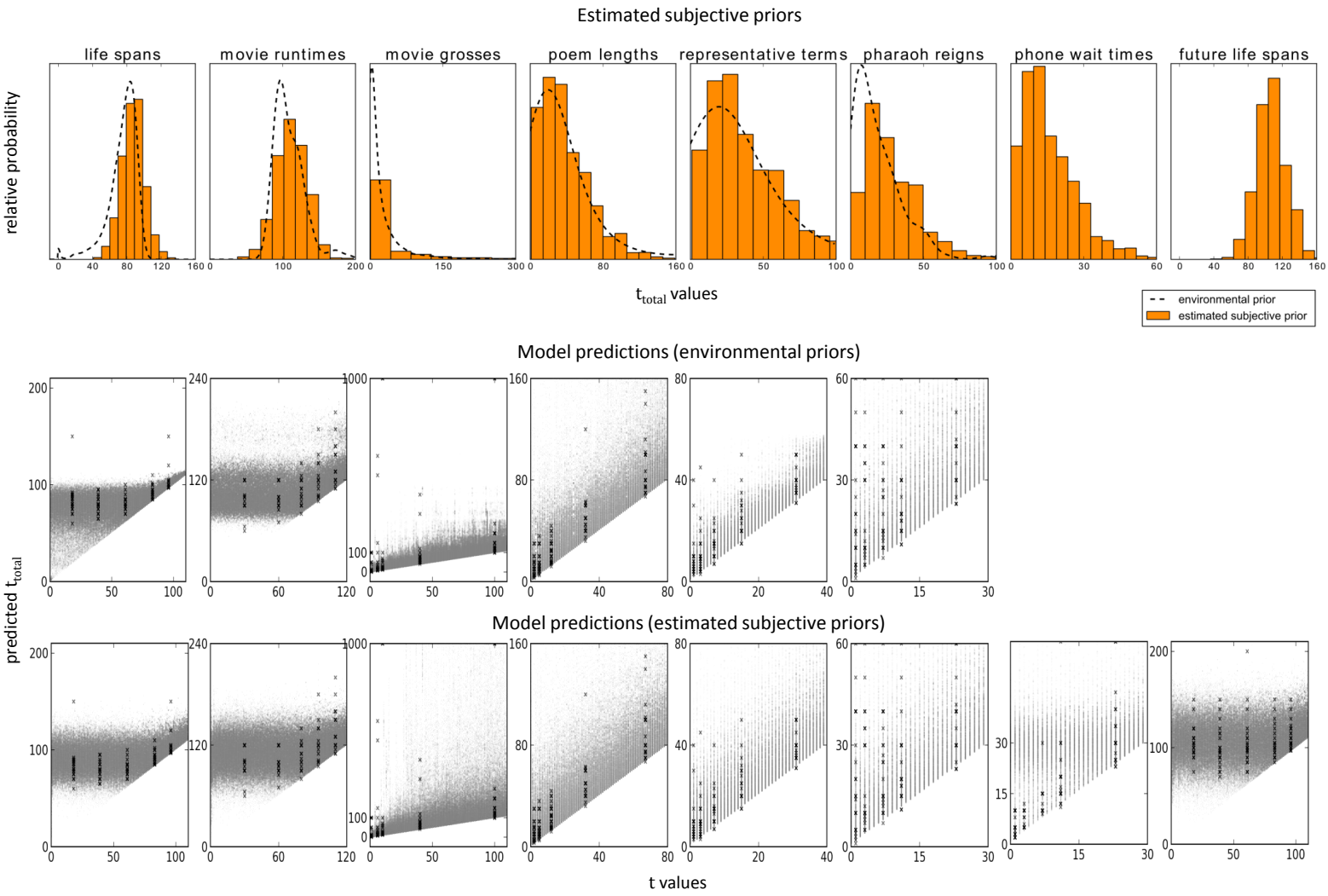

Figure 4. Estimated subjective priors and model predictions. The first row shows our estimates of people's subjective prior beliefs compared with the environmental distributions collected by Griffiths and Tenenbaum (2006). The bottom two rows overlay people's actual responses (black marks) with the posterior predictive distributions (gray shaded areas) of the Bayesian cognitive models for new (unobserved) responses. The posterior predictive probabilities of responses for the environmental prior model (second row) and the estimated subjective prior model (third row) are proportional to the darkness of the gray areas.

subjective prior—in which case some posterior probability would have been assigned to more than one of the possible forms—in every domain, all of the posterior mass was assigned to a single type of distribution.

The top row of Figure 4 shows the estimated subjective priors that people used to make predictions in comparison to the true environmental distributions that were collected by Griffiths and Tenenbaum (2006). The estimated subjective distributions were generated by sampling a prior type and parameter values from the posterior distributions and then using them to generate a sample.

Our estimates of people's subjective priors for life spans, movie runtimes, movie grosses, poem lengths, U.S. representatives' terms and pharaohs' reigns are remarkably similar in form to the true environmental distributions. The subjective priors for life spans, movie runtimes and pharaohs' reigns are shifted slightly to the right compared to the environmental distributions, suggesting that people's prior knowledge for these domains has the same form as the environmental statistics but may not be tuned perfectly to the environment.

People's subjective prior for waiting times was estimated in the same manner as the other priors even though the environmental data was not available. The estimated subjective prior for waiting times was consistent with an Erlang form. Griffiths & Tenenbaum (2006) were unable to provide estimates of these posteriors using the standard qualitative analysis, but did use non-Bayesian methods to fit people's responses and found that a prediction function based on a Power-Law (Pareto) prior provided the best fit. It is not immediately clear if our disagreement about the form of the subjective prior for phone waiting times is due to differences in our methodology or to differences in the predictions of our respective participants.

A subjective prior for future life spans was estimated even though it is based on a counterfactual scenario and therefore has no true environmental distribution. This subjective prior appears to have a similar form to the prior for actual life spans, but is shifted to the right with an average life span of 105.

The bottom two rows of Figure 4 overlay people's actual responses (black marks) with posterior predictive

distributions from the Bayesian cognitive model for new (unobserved) responses using the environmental prior (second row) and the estimated subjective prior (third row).

The posterior predicitve distributions are generally similar for both the environmental prior model and the estimated prior model. There are some differences in the predictions of the models which are consistent with differences between the estimated and environmental priors. For example, the estimated prior for life spans did not capture an increased risk of death for infants and therfore the estimated model predicts less deaths at a young age than the environmental model does. This can likely be attributed to the limited range of ages (18 to 96 years) presented to participants. The estimated models for movie grosses and representatives' terms tend to predict higher values than the environmental model, which is consitent with the tendancy of some participants to overestimate these values.

Discussion

We demonstrated that an integrative Bayesian approach—combining Bayesian data analysis with Bayesian models of cognition—allowed us to estimate people's subjective prior knowledge based on their responses in a simple prediction task. This approach allowed us to relax the assumption that representations of people's prior knowledge in a rational model should be veridical with environmental statistics.

Although we did not require environmental data to apply an integrative Bayesian approach, having this data allowed us to compare our estimates of people's subjective beliefs to real-world environmental data. We found that people's beliefs about the phenomena in our study were similar in form to the environmental statistics, but that they showed some deviations. At least one of these deviations—related to infant mortality in the life spans question—likely resulted from the limited range of response data that the model used to estimate subjective priors. Other differences between the estimated and environmental priors seem more likely to be the result of deviations between people's subjective beliefs and the environmental statistics. For example, some people tended to overestimate the total gross of movies and the lengths of representatives' terms and pharaohs' reigns. The integrative Bayesian approach is able to provide explanations and predictions that account for these human responses in a way that traditional rational analysis cannot.

Furthermore, in situations where a Bayesian model of cognition requires representations of people's prior beliefs and environmental data is unavailable or non-existent—like it was for telephone waiting times and future life spans in our study—an integrative Bayesian framework can still be used to infer subjective priors and make model predictions.

Taking an integrative Bayesian approach opens the door for researchers to take advantage of all of the methods that have been developed for Bayesian analysis of cognitive process models (Lee, 2008) and apply these methods to Bayesian cognitive models. In addition to the estimation of subjective priors and psychological parameters, this method also allows for individual differences in subjective prior beliefs (Hemmer, et al., in prep). This is important because if people's subjective priors are not tuned to the environment for a particular domain, then it is reasonable to assume that different people have different subjective priors.

Acknowledgements

Thank you to Michael Lee and James Pooley for helpful discussions; and to Tom Griffiths and Josh Tenenbaum for providing us with the environmental data that they collected.

References

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17*, 767–773.

Hemmer, P. & Steyvers, M. (2009a). A Bayesian account of reconstructive memory. *Topics in Cognitive Science, 1*(1), 189–202.

Hemmer, P. & Steyvers, M. (2009b). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic bulletin & review, 16*(1), 80–7.

Hemmer, P., Tauber, S., & Steyvers, M. (in preparation). Bayesian estimation in rational models.

Huszár, F., Noppeney, U. & Lengyel, M. (2010). Mind reading by machine learning: A doubly Bayesian method for inferring mental representations. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Jones, M. & Love, B.C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15, 1–15.

Lee, M.D. (2011). In praise of ecumenical Bayes. *Behavioral and Brain Sciences*, 34, 206–207.

Lee, M. D. & Sarnecka, B.W. (2010). A model of knower-level behavior in number-concept development. *Cognitive Science, 34*, 51–67.

Mozer, M., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science*, 32, 1133–1147.

Oaksford, M., & Chater, N. (Eds.). (1998). *Rational models of cognition.* Oxford: Oxford University Press.

Sanborn, A. N., & Griffiths, T. L. (2008). Markov chain Monte Carlo with people. *Advances in Neural Information Processing Systems 20*.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology, 60*, 63–106.

Vul, E., Goodman, N.D., Griffiths, T.L., & Tenenbaum, J.B.(2009). One and done? Optimal decisions from very few samples. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.