

A Model of Knower-Level Behavior in Number Concept Development

Michael D. Lee, Barbara W. Sarnecka

Department of Cognitive Sciences, University of California, Irvine

Received 4 September 2008; received in revised form 2 March 2009; accepted 9 June 2009

Abstract

We develop and evaluate a model of behavior on the Give-*N* task, a commonly used measure of young children's number knowledge. Our model uses the knower-level theory of how children represent numbers. To produce behavior on the Give-*N* task, the model assumes that children start out with a base rate that makes some answers more likely a priori than others but is updated on each experimental trial in a way that depends on the interaction between the experimenter's request and the child's knower level. We formalize this process as a generative graphical model, so that the parameters—including the base rate distribution and each child's knower level—can be inferred from data using Bayesian methods. Using this approach, we evaluate the model on previously published data from 82 children spanning the whole developmental range. The model provides an excellent fit to these data, and the inferences about the base rate and knower levels are interpretable and insightful. We discuss how our modeling approach can be extended to other developmental tasks and can be used to help evaluate alternative theories of number representation against the knower-level theory.

Keywords: Number concept development; Knower-level theory; Bayesian modeling; Give-*N* task

1. Introduction

A basic challenge in understanding human cognitive development is to understand how children acquire number concepts. Since the time of Piaget (1952), number has been one of the most active areas of research in the field. One prominent current theory about the origin of integer concepts is the “knower-level” theory (Carey, 2001; Carey & Sarnecka, 2006; Wynn, 1990, 1992; see also Le Corre & Carey, 2007; Le Corre, Van de Walle, Brannon, & Carey, 2006; Sarnecka & Gelman, 2004; Sarnecka, Kamenskaya, Yamana, Ogura, & Yudovina, 2007).¹

Correspondence should be sent to Michael D. Lee, Department of Cognitive Sciences, University of California, Irvine, CA 92697-5100. E-mail: mdlee@uci.edu

The knower-level theory asserts that children learn the exact cardinal meanings of the first three or four number words in order. That is, children begin by learning the meaning of “one” first, then “two,” then “three,” and then (for some children) “four,” at which point they make an inductive leap, and infer the meanings of the rest of the words in their counting list. In the terminology of the theory, children start as NN-knowers (for “no number”) or “Pre-number-knowers,” progress to one-knowers once they understand “one,” through two-knower, three-knower, and (for some children) four-knower levels, until they eventually become CP-knowers (for “cardinal principle”). Thus, the cardinal meanings of “one,” “two,” “three,” and sometimes “four” are learned in a completely different way than the meanings of “five” and higher number words. The former are learned gradually, one at a time; the latter are learned all at once, by induction (for reviews, see Carey, 2001, 2004; Carey & Sarnecka, 2006). Our concern is mainly with the early part of this process, involving the learning of small-number words.

An important task for the knower-level theory is a widely used one known as the “Give-*N*” task (e.g., Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; Fuson, 1988; Schaeffer, Eggleston, & Scott, 1974; Wynn, 1990, 1992). In this task, children are simply asked to give some number of objects (usually small toys) to the experimenter (or an experimenter substitute, such as a puppet). The behavioral data are just a set of question–answer pairs, recording how many toys the child was asked to give and how many they actually gave.

The knower-level theory makes a number of strong predictions about children’s performance on the Give-*N* task. For example, it predicts that children at a given knower level, when asked about a higher number whose exact meaning they do not know will avoid giving any set size they *can* name. In practice, this means that children’s guesses about unknown number words will be lower bounded by their knower level. This is because children learn the number words in sequence. For example, if a child understands only the number “one” (i.e., is a “one-knower”) he or she might mistakenly give three toys when asked for two, but he or she will not give one toy because he or she knows what “one” means, and they know that none of the other number words means one (Wynn, 1990, 1992).

Following this line of reasoning, the performance of a child on the Give-*N* task should be highly diagnostic in assessing his or her knower level, and so the task potentially provides an important developmental measure. It is not easy, however, to determine knower levels from raw Give-*N* data because there are task-specific influences on behavior that need to be accounted for in determining knower level. For example, it is empirically quite likely that a no-number-knower, whatever he or she is asked for, will give one toy, or two toys, or a small handful of toys, or the whole basketful of them. So if the basket of toys the child selects from has 15 toys in total, answers like 1, 2, 3, and 15 are more likely than numbers like 8, 9, or 10, but this is just a task-specific quirk of the Give-*N* procedure.

This behavior is a problem for diagnosis because, for example, it might lead a two-knower to give (apparently correctly) three toys when asked for three, but only because it is a default number to give when the instructions are not meaningful, not because he or she actually understands the concept “three.” The same two-knower is very unlikely to give eight toys when asked, though, because that is not a default response.

More generally, it is not straightforward to test directly the predictions of the knower-level theory against Give- N data because the theory itself does not provide a complete description of behavior. It provides a detailed theory of how children represent numbers but does not fully explain how these representations lead to behavior on developmental tasks. What is observed experimentally is a (potentially complicated) mixture of the representations of number concepts children have and the decision processes they use to transform their understanding to action. Accordingly, our ability to understand Give- N task behavior, measure children's knowledge of number concepts, and evaluate the knower-level theory itself in detail all depend on having a complete model of task performance.

This paper develops such a behavioral model and evaluates the model directly against previous data. We begin by describing the data, and then the model, at first intuitively and then formally. We present the results of applying the model to the data, and we finish with a discussion of possible extensions and applications of our approach.

2. Give- N data

We consider previous data, presented as data set 1 by Sarnecka & Lee (2009), including 82 monolingual speakers of English, ages 2–4 years (mean 3 years, 7 months; range 2 years, 11 months to 4 years, 6 months), tested at preschools in Irvine, California, or at a university cognitive development laboratory in Cambridge, Massachusetts. As part of their participation in other studies, each child completed an intransitive counting task, where the experimenter simply asked them to “count to 10.” Our data include only those children who counted to 10 perfectly. Thus, we can be sure that every child was familiar with the number words “one” to “ten.”

Table 1 provides some examples of Give- N behavior, by detailing all the data for three of the 82 children in the full data set. (The full data set is presented in the Appendix.) Each row in Table 1 corresponds to a question, asking the children for “one,” “two,” “three,” “four,” “five,” “six,” “eight,” or “ten” toys. The entries in the columns for each child correspond

Table 1
Sample Give- N behavior for three children, showing the number of toys they gave, when asked for “one,” “two,” “three,” “four,” “five,” “six,” “eight,” or “ten” toys

Asked	Child A	Child B	Child C
“one”	1, 1, 1	1, 1, 1	1
“two”	2, 5, 5	2, 2, 2	2
“three”	5, 2, 6	3, 3, 3	3, 3
“four”	3, 4, 4	4, 6, 8	4, 4
“five”	10, 15, 2	6, 5, 8	15, 15
“six”	—	—	—
“eight”	14, 2, 3	3, 7, 7	—
“ten”	7, 2, 3	5, 5, 3	—

Note: Multiple entries show the child's response on multiple trials asking for the same number. Empty cells mean that the child was never asked for that number.

to how many toys each child actually gave when asked each question. Multiple entries are the same child's responses over multiple trials asking for the same number. So, for example, child A gave two, five, and five toys again on the three trials where "two" toys were requested.

There are several interesting features to note in the sample data in Table 1. One is that, if a child correctly gives a number, he or she also tends to avoid giving that number of toys when asked for a different number word. For example, child B responds correctly when asked for "one" and "two" and also does not give one or two toys when asked for larger numbers, even though the child makes many errors for these higher number questions. This pattern suggests that child B understands the meanings of the words "one" and "two." It is not clear, however, whether the child understands the meaning of "three" because, although he or she always gives three when asked, he or she also gives three in error when asked for "eight" and "ten." As these examples make clear, an application of the knower-level theory must account for both aspects of number knowing: giving the correct number when asked, and not giving that number when asked for something else.

A second observation about the data in Table 1 is that errors follow a nonobvious distribution. Some errors fall near the target number and could be attributed to miscounting, or to the use of estimation rather than counting (e.g., when child B is asked for "five" but gives six.) However, other errors fall far from the target, as when child A and child C give 10 and 15 toys for the word "five," respectively. In fact, giving all 15 toys is a common error.

Finally, Table 1 also shows how behavior can adhere to the knower-level theory assumption that children learn the number words in order. While child A seems only to understand "one," child B seems to understand "one" and "two," and child C seems to understand "one," "two," "three," and "four." All of these interesting features of the data set need to be explained by our model.

3. A knower-level model

One powerful way to build models of simple human inferences—such as how many toys to give—is to adopt a rational or computational-level perspective (Marr, 1982). The idea is to use rational principles, in this case the framework for inference provided by Bayesian statistics, as a working theoretical assumption about the goals of cognitive processes. We do not assume that children are doing formal Bayesian computations, but rather that the goal of human cognitive processes is to approximate these computations. In this sense, a computational model provides an account of *why* cognition behaves as it does, without making a commitment to *what* actual processes humans use, or *how* those processes are implemented in neural systems.

This approach has been successfully employed throughout psychological modeling, in areas including vision, causal learning, property induction, categorization, and decision making (for overviews, see Chater, Tenenbaum & Yuille, 2006; Griffiths, Kemp & Tenenbaum, 2008; McClelland, 2009). It is ideal for our purposes because we seek a way of relating the knower-level theory of number-concept representation to behavioral data, so that we can use experimental data to draw conclusions about developmental states.

Adopting a computational approach lets us build a quantitative model of behavior, with a rational justification, but without speculative theorizing about mental processes.

Using the computational Bayesian perspective, we assume that the child’s decision-making process on the Give-*N* task has four parts.

1. *Base rate.* Initially, a child has a “base-rate” distribution which expresses the probability of giving each possible number of toys. This base-rate distribution can be thought of as the child’s a priori bias toward or against each possible response, even before any particular number has been requested. Behaviorally, for a Give-*N* task with 15 toys, the base rate represents the probabilities that children would give 1,...,15 toys if they were asked to give objects in a completely nonnumerical way (e.g., if they were asked, “Can you give me fish?” and English did not make a singular–plural distinction).² In Bayesian terms, the base-rate distribution is the prior the child has over appropriate Give-*N* behavior, in the absence of any other information.
2. *Instructions.* When the experimenter gives an instruction, this is used to update the base-rate probabilities and create a new distribution of likely responses. In Bayesian terms, the instruction is the datum on which inference about appropriate Give-*N* behavior is based.
3. *Updated belief.* The updated distribution will depend critically on the child’s knower level. In Bayesian terms, the knower-level theory provides the likelihood function through which data updates prior to posterior beliefs. We describe this carefully below.
4. *Behavior.* Finally, the child will give some number of toys. This is the actual behavior observed in the task and recorded by the data. The probability of each possible response is expressed by the updated distribution. In Bayesian terms, observed behavior is sampled from the updated belief distribution.

Two concrete examples of this decision-making process are shown in Fig. 1. One example is shown in each row, as a sequence of the four stages connected by arrows. Both relate to a child we assume to be a three-knower, with a fixed base rate. This base rate is shown by the leftmost bar graphs, and, gives the initial probability that the child will give 1,...,15 toys.

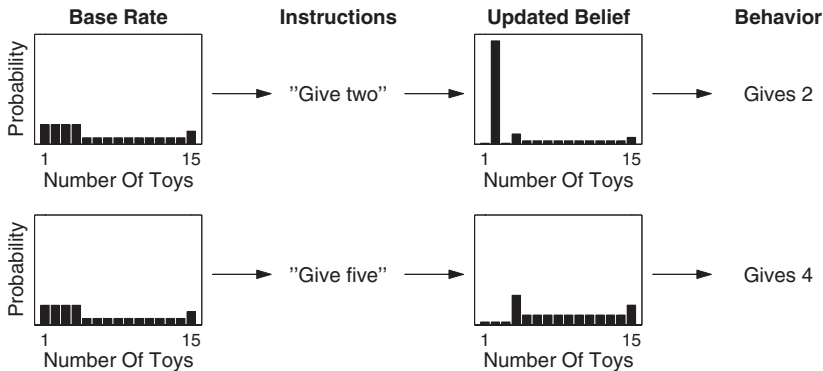


Fig. 1. Two examples of the decision-making process modeled for a three-knower. Note that the base-rate probabilities start out the same. The updated probabilities combine the base rate with the instruction, in a way that depends on the numbers (1, 2, and 3) that the child knows. Observed behavior is sampled from the updated probabilities.

Giving a “handful” of toys (i.e., 1,...,4) is most likely; giving all of the toys (i.e., 15) is also more likely; the other possibilities (i.e., 5,...,14) are less likely but still possible. To begin with, we are just assuming a plausible base rate to help us explain the model using concrete examples. Later we will use the model to infer the base rate from the actual data.

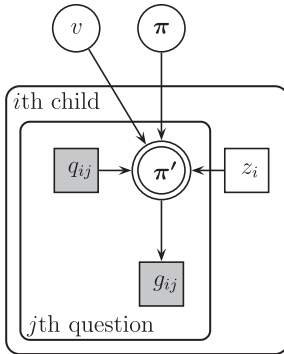
In the first row of Fig. 1, the child is asked to “give two” toys. This child, being a three-knower, knows what “one,” “two,” and “three” mean. So the child is very likely to give two toys and very unlikely to give one or three toys. This is reflected in the updated belief distribution. The other possible responses (4,...,15 toys) do not change in their relative probabilities, although they will change their absolute probabilities because the probabilities for one, two, and three have changed. That is, as 4 and 15 were more likely than 5–14 in the base rate, they will still be more likely after updating, although all the numbers 4–15 are less likely in absolute terms. All of these changes can be seen in the rightmost bar graph in the first row. Giving two is very likely, giving one or three is not at all likely, and giving 4,...,15 is unlikely, with 4 and 15 slightly more likely than the other responses. In this example, the most likely response is two. Thus, a three-knower who is asked for “two” will probably respond correctly.

The second row of Fig. 1 shows an example of another trial with the same hypothetical three-knower. This time, the child is asked to “give five” toys, but “five” is a number the child does not know. All the child knows that “five” does not mean one, two, or three. So the responses one, two, and three will become much less probable, but all of the other numbers (i.e., 4,...,15) will retain the same relative probabilities to each other. All of these changes can be seen in the rightmost bar graph in the second row. Giving four becomes the most likely response, followed by 15, followed by the other numbers the child does not know (i.e., 5,...,14). The numbers they do know (i.e., one, two, and three) are very unlikely responses. The actual behavior produced by the child is again just a sample from the updated belief distribution. In this example, the most likely response is four. Thus, a three-knower who is asked for “five” will probably respond incorrectly.

3.1. Graphical model implementation

Fig. 2 presents the graphical model we used to implement our model. Graphical models are a standard approach to implementing probabilistic models in machine learning and statistics, and they have more recently been used as a framework for implementing and analyzing models of cognition (for overviews, see Lee, 2008; Lee & Wagenmakers, 2008; Shiffrin, Lee, Kim, & Wagenmakers, 2008). In graphical models, variables are represented by nodes in a graph, and their connections show how they relate to each other. In Fig. 2, observed variables (i.e., data) are shown as shaded nodes, and unobserved variables (i.e., model parameters to be inferred) are shown as unshaded. Discrete variables are indicated by square nodes, and continuous variables are indicated by circular nodes. Stochastic variables are indicated by single-bordered nodes, and deterministic variables (included for conceptual clarity) are indicated by double-bordered nodes. Finally, encompassing plates are used to denote independent replications of the graph structure within the model.

In our implementation of the knower-level model in Fig. 2, the data are the observed q_{ij} and g_{ij} variables, which give the number asked for (the “question”) and the answer (the



$$\pi \sim \text{Dirichlet}(1, \dots, 1)$$

$$v \sim \text{Uniform}(1, 1000)$$

$$z_i \sim \text{Categorical}(\frac{1}{6}, \dots, \frac{1}{6})$$

$$\pi'_{ijk} \propto \begin{cases} \pi_k & \text{if } q_{ij} > z_i \\ v \times \pi_{ijk} & \text{if } q_{ij} \leq z_i < \text{CP and } k = q_{ij} \\ \frac{1}{v} \times \pi_{ijk} & \text{if } q_{ij} \leq z_i < \text{CP and } k \neq q_{ij} \\ v \times \pi_{ijk} & \text{if } z_i = \text{CP and } k = q_{ij} \\ \frac{1}{v} \times \pi_{ijk} & \text{if } z_i = \text{CP and } k \neq q_{ij} \end{cases}$$

$$g_{ij} \sim \text{Categorical}(\pi')$$

Fig. 2. Graphical model for behavior on the Give-*N* task using the knower-level theory of number representation.

number “given”), respectively, for the *i*th child on his or her *j*th question. The base-rate probabilities are represented by the vector π , which is updated to π' , from which the number given is sampled. The update occurs using the number asked for, the knower-level z_i of the child, and an evidence value v that measures the strength of the updating. The base rate and evidence parameters, which are assumed to be the same for all children, are given vague priors (i.e., ones that allow for a very large range of possible inferences).

The updating rule that defines π' decomposes into three basic cases, as explained in discussing the examples in Fig. 1. If a number k is greater than the knower-level z_i , then regardless of whatever number q they are being asked for, the updated probability remains proportional to the base-rate probability π_k for that number. If a number k is within the child’s knower-level range z_i , it either increases in probability by a factor of v if it is the number q being asked for, or decreases in probability by a factor v if it is not. For a child who is a CP-knower, his or her range encompasses all of the numbers. The final part of the graphical model relates to the behavior step, with the number of toys given being a draw from the probability distribution π' representing the updated beliefs.

The graphical model in Fig. 2 provides a *generative* probabilistic model of behavior on the Give-*N* task. This means that it provides a formal account of how data from the task are produced or generated. The model starts the generating process from the unknown psychological variables—the base-rate distribution and the evidence value parameters, which are the same for all children, and the child’s knower-level parameter, which varies from one child to the other—and then says how these variables interact with the task instruction (i.e., the question asked) to produce the observed behavior (i.e., the number of toys given).

The great strength of generative models is that by formalizing the process that produced data, inference can automatically be done using Bayesian methods.³ Intuitively, Bayesian inference works out what the base rate, knower level, and evidence value must have been to have produced the data that were actually observed. It does this simultaneously for all of the psychological parameters and for all of the children. As it knows what behavior any fixed set of parameters would produce, it can take actual observed behavior and infer what the parameters must have been. We think that generative models have an advantage in our context because the inferences they make come from a clearly articulated formal account of

how observed behavior was produced. This puts the modeling emphasis on psychological theorizing, rather than data analysis.

4. Modeling results

The graphical model implementation in Fig. 2 is an especially convenient formalization of our generative model because it makes it easy to do fully Bayesian inferences. We achieve this using standard WinBUGS software (Spiegelhalter, Thomas, & Best, 2004), which applies Markov Chain Monte Carlo computational methods (see, for example Chen, Shao, & Ibrahim, 2000; Gilks, Richardson, & Spiegelhalter, 1996; MacKay, 2003) to make inferences about model parameters and data. In particular, we applied our model to the data by collecting five independent chains of 5,000 samples, each with 1,000 samples of burn-in. The standard \hat{R} measure of convergence—which basically measures between- to within-chain sample variability—was between 0.99 and 1.01 for π , ν , and all 82 z_i variables, indicating good convergence (e.g., Gelman, Carlin, Stern, & Rubin, 2004, pp. 296–297).

We report the results in four parts. First we report the base-rate distribution inferred by the model. Second, we report the degree to which evidence (in the form of the experimenter's request) changes the base-rate distribution. Third, we report on the model's ability to assign a knower level for each child. Each of these analyses comes immediately from the posterior distribution over the π , ν , and z_i variables provided by the graphical model. Fourth, we examine the posterior prediction our model makes about data, which is a standard Bayesian way to examine goodness of fit between model and data.

4.1. Base rate

Fig. 3 shows the inferred base rate,⁴ which represents children's predisposition to give each possible number of toys before they are asked for any particular number. The distribution accords surprisingly well with what we might intuitively expect. That is, children seem predisposed to give either a small number of toys (1,...,5) or all 15 toys rather than something in between.

We want to emphasize that this base rate is entirely inferred from the data under the generative model of behavior we developed using the knower-level theory. Any possible combination of probabilities summing to one was given equal prior probability in our modeling; we did not "insert" this, or any other, base rate into the model in any way. The fact that a highly reasonable and interpretable base rate was inferred is one piece of suggestive evidence that the model is a useful one for Give- N data.

4.2. Evidence

The posterior distribution for the evidence ν was approximately Gaussian distributed, with a mean of 29.2 and SD of 7.4. This is a sensible result that is straightforward to

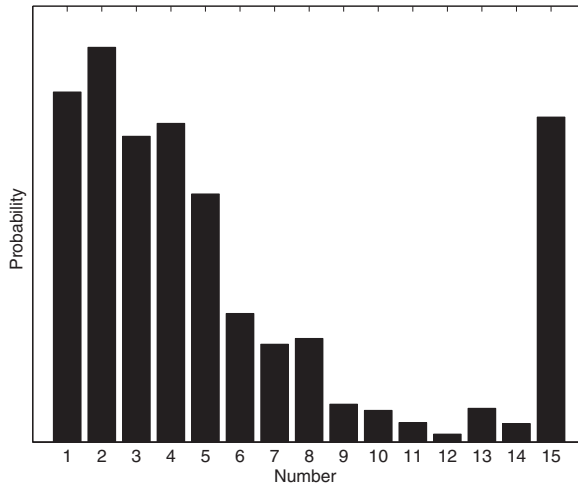


Fig. 3. Children's base-rate probabilities for giving 1,...,15 toys. This base rate was inferred by the knower-level model from the Give- N data.

interpret. It means that the instructions provided to children in the Give- N task had the effect, under our model, of increasing or decreasing the probability of any given response by a factor of about 30.

4.3. Posterior inference for knower levels

Fig. 4 shows the posterior distribution over the six knower levels (NN-, one-, two-, three-, four-, and CP-knowers) for each child, ordered from the smallest expected value to the largest. The noteworthy feature of this result is that most of the children are classified with high certainty into a single knower level. There are exceptions (e.g., child 3 in the first row; child 78 in the second row), but, for the most part, there is confidence in a single classification (indicated by a single, high peak for each child). In fact, over 89% of the children have a posterior mass for a single knower level that is at least twice as likely as any other alternative, and more than 68% of children have a classification that is fully 10 times more likely than any other.

When inferring a discrete latent variable like a knower level, highly peaked posteriors are a suggestive indication that the model is a useful one. When models are badly misspecified, Bayesian inference tends to mix over a wide range of possibilities to try and fit the data, making interpretation difficult. What the peaked distributions in Fig. 4 show is that the model leads to confident predictions about the knower level of most children.

We also note that, in those cases where the posterior distribution shows uncertainty about a child's knower level, uncertainty is invariably distributed over neighboring knower levels. For example, the model shows uncertainty about whether child 3 is a two- or three-knower. There is no case where the posterior distribution covers two levels that are not adjacent. For example, there is no case where the distribution is split between two- and CP-knowers. This is not an assumption built into the model, which treats the knower levels as a set of nominally

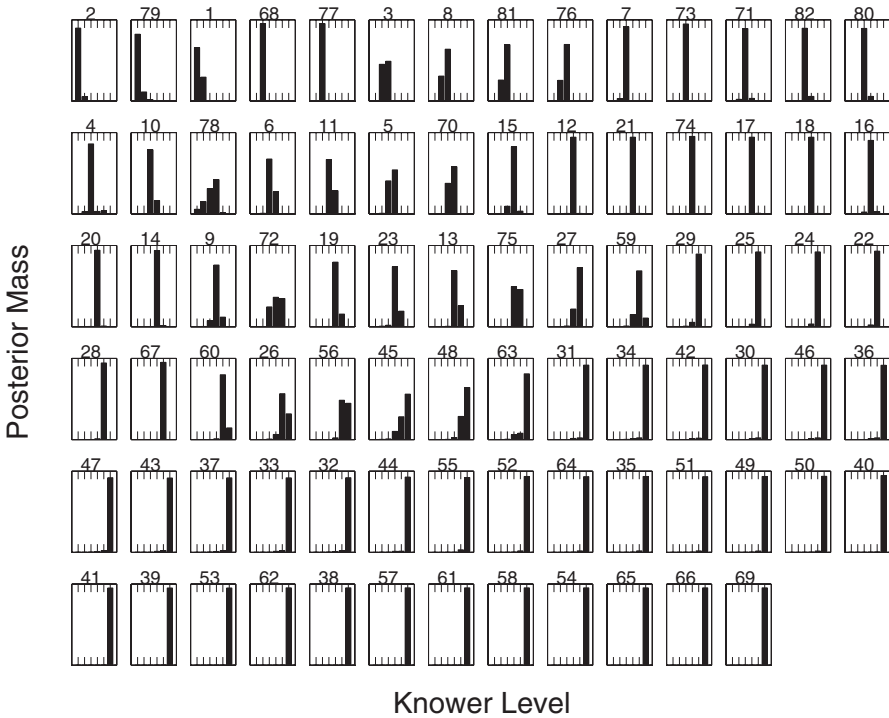


Fig. 4. The posterior distribution over knower levels for all 82 children, ordered from those most likely to be NN-knowers to those most likely to be CP-knowers. Each panel corresponds to a child, with the x-label corresponding to (from left to right) NN-, one-, two-, three-, four-, and CP-knower levels. The y-axis corresponds to posterior probability for each knower level.

scaled possibilities. Accordingly, the patterns of uncertainty seen in Fig. 4 provide suggestive support to the claims of knower-level theory that children learn number concepts in order.

4.4. Posterior predictions for knower levels

Finally, we assess the model more directly, using posterior prediction. This is a standard Bayesian approach, comparing the probability of data according to the model with the data actually observed. In a sense, posterior predictive analysis is a way of assessing goodness of fit, but it is important to understand that it automatically accounts for model complexity in ways that approaches like maximum-likelihood fitting do not. Each prediction is the average across the entire parameter space, as weighted by the posterior distribution for parameters, not the prediction that gives the maximum agreement at a specific set of parameter values. In this way, the posterior predictive guards against overfitting and constitutes a principled and useful way to evaluate whether a model provides an adequate account of data.

Fig. 5 shows the posterior predictions of the model for the NN-, one-, two-, three-, four-, and CP-knower levels. Each level corresponds to a panel, and each panel is organized with

each possible question (i.e., how many toys are asked for) along the *x*-axis and each possible answer (i.e., how many toys are given) along the *y*-axis. This organization means that each cell corresponds to a possible question-and-answer combination. The shading of each cell corresponds to the posterior probability of that number of toys being given when that question is asked, with darker shading indicating greater probability. The overlaid circles represent the behavioral data for those children classified into each knower level by the posterior inferences presented earlier, with their size showing how many toys were actually given when questions were asked.⁵

It is clear from this analysis that the model provides an excellent account of the data because the larger circles representing data almost always fall on darkly shaded regions, showing that the model expects this behavior. Note that Fig. 5 shows the posterior prediction of the model for all possible question-and-answer pairs, including for questions that were not asked as part of the current data set. As a consequence, there are many dark squares without circles in Fig. 5, corresponding to predictions the model makes for questions where data are not available. Obviously, these cases correspond to gaps in the available data, not failures in the prediction of the model.

The benefit of showing the full range of model predictions in Fig. 5 is that it makes graphically clear how the model formalizes the key assumptions of the knower-level theory

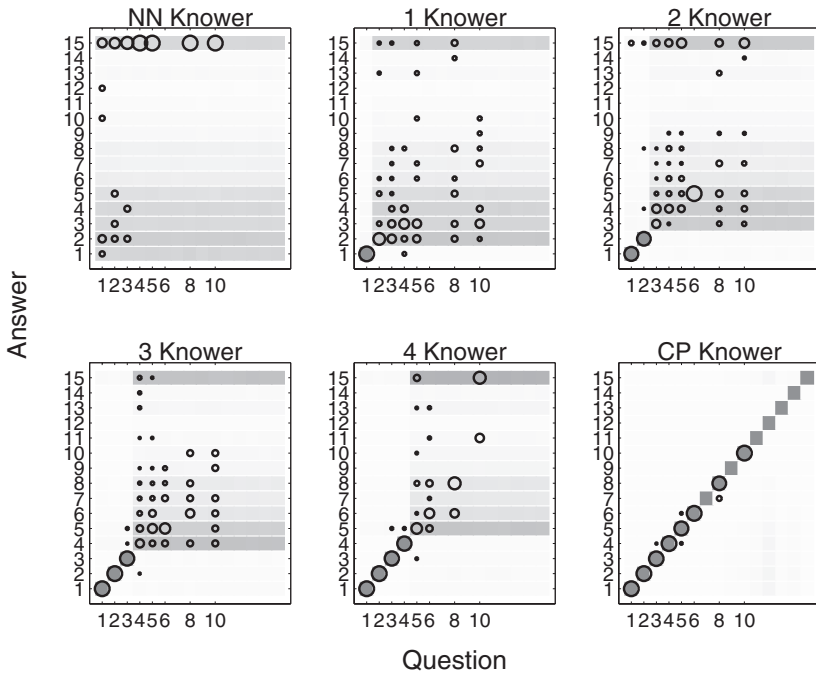


Fig. 5. Posterior prediction for each of the knower levels, for each possible question-and-answer combination. The shading represents the posterior prediction of the model, with darker combinations of questions-and-answers being more likely. The overlaid circles represent the behavioral data, showing the frequency of responses for all children classified by the model as having each knower level.

and how those assumptions are borne out by the experimental data. For the NN-knowers, the model is able to capture the nonobvious pattern of errors we noted earlier, giving the highest probability to the numbers 1–5 and 15, as observed in the data. The base rate is responsible for these good predictions because, for a NN-knower, the experimenter's instructions provide no additional information, and the base rate is the sole guide for behavior.

For one- to four-knowers, the model predicts that all of the numbers that are understood will be used correctly. That is, they will tend to be given when asked for, and they will not be given in error when asked for a different number. Those numbers larger than the knower level, however, continue to follow base-rate probabilities. In the posterior predictive display in Fig. 5, this leads to a distinctive pattern whereby predictions for small numbers are largely on the diagonal (i.e., correct responses), but numbers above the knower level have predicted errors consistent with the base rate. The superimposed data show that this pattern of predictions reflects actual behavior very well. There are only a few data points that violate the expected pattern, and those are explained by the probabilistic nature of our account of decision making, as captured by the evidence parameter.

Finally, a similar story holds for CP-knowers, who are inferred to understand all of the numbers. The model predicts correct behavior for all of the numbers, and the data again show very few exceptions.

5. Discussion

A basic goal of empirical science is to relate formal models to experimental data. Among the many benefits of this endeavor are the ability to make inferences about unobserved but substantively meaningful parameters, and the ability to make direct predictions about empirical observations. Our development of a model using the knower-level theory of number development was motivated in this way, and we think it is successful.

Our results show how the formal model allows us find the base-rate distribution for the Give- N task, a measure of how much task instructions influence behavior, and the knower level of each child. The base rate quantifies the “chance” distribution for the Give- N task. From the outset, it seemed unlikely that this distribution was uniform because some responses seem more likely than others due to the nature of the Give- N task. However, as chance responding is never directly observed, it would be extremely difficult to quantify the appropriate distribution without a formal model. Thus, the ability of the model to infer the base rate shown in Fig. 3 provides an insight into the nature of the Give- N task that otherwise would not be available. Similarly, knowing how much instructions in the Give- N task influence behavior is useful task-specific information.

Perhaps the greatest benefit of being able to infer the base rate and evidence, however, is that it enables knower-level theory to be applied cleanly to the problem of measuring children's understanding of number words. This is seen in the ability of the model to make inferences about knower levels, as shown in Fig. 5. Assessing knower levels has previously been done by applying ad hoc heuristics to behavioral data and has failed to account for the

nonobvious chance distribution captured by our base rate. For this reason, applying our model provides a sharper inference about an important developmental variable.

Our posterior predictive assessment of model fit shows how we are able to assess the knower-level theory directly in terms of observed raw data. This is possible because our model provides a complete generative account of how behavior on the Give-*N* task is produced. The knower-level theory is the cornerstone of this account but is supplemented with simple rational assumptions that specify how children transform their understanding of number concepts into actual behavior. Without these additional mechanisms, empirical evaluation of the knower-level theory would have to rely on less direct statistical tests of properties of the data and would not be amenable to making quantitative predictions about Give-*N* behavior. For these reasons, we think our model is a good example of the benefits of adopting a generative approach to psychological modeling.

It would be straightforward to apply our model to data from alternatives to the Give-*N* task, such as the “What’s-On This-Card?” task, in which children produce number words for sets presented visually (Gelman, 1993; Le Corre & Carey, 2007; Le Corre et al., 2006). We would expect the base rate inferred from these data to be quite different, but the model itself, including the key knower-level theoretical commitments, to be unchanged. Indeed, one way to understand the benefits of our model is that it separates, in a formal way, the task-specific base-rate effects on behavior from the effects coming from a child’s understanding of number concepts. This separation serves to “factor out” the task specifics and focus on the fundamentally important psychological concept of knower levels.

Finally, the model-based approach we have adopted has the potential to contribute to the most basic questions of theory evaluation and comparison. An alternative theory of how children initially represent exact numbers involves an analog magnitude scale (e.g., Dehaene, 1997; Gallistel, 1990). There are various possibilities, including mechanisms based on scalar estimation and counting processes (e.g., Cordes, Gallistel, & Gelman, 2001; Whalen, Gallistel, & Gelman, 1999) for using this theory to develop a model of Give-*N* task behavior. With a rival to the current model in place, it would be possible to evaluate both directly against experimental data, using standard quantitative criteria measuring their descriptive adequacy and predictive ability (see Myung, Forster, & Browne, 2000; Shiffrin et al., 2008). While formal model-based evaluations are certainly not the only criteria for choosing between competing theories, they can provide important evidence that is difficult to obtain by other means. Accordingly, we believe that models, like the one we have presented, constitute an important, but currently underdeveloped, line of research needed to evaluate and improve our theories of how children represent numbers.

Notes

1. Of course, the knower-level theory is not the only well-developed account of how children represent numbers. We discuss how the modeling approach we adopt in this

paper can address more general questions of theory evaluation and comparison in the Discussion section.

2. We thank a reviewer for this psychological interpretation of the base rate.
3. It is important to distinguish between the two distinct ways Bayesian inference is being used in our study. One is as a theoretical assumption about how the child uses instructions as data to update his or her base rates. The other is as a statistical framework for relating a cognitive model to behavioral data, for the purposes of inferring parameter values and producing model predictions. These two uses are quite independent. It would be possible to develop a cognitive model of Give-*N* behavior that did not involve Bayesian assumptions about the mind, and it would be possible—although technically challenging—to do statistical analysis on our model of Give-*N* behavior using standard frequentist approaches.
4. Technically, Fig. 3 shows the expected posterior predictive distribution for the base rate. This is a convenient way to summarize visually the most important properties of the 15-dimensional joint posterior distribution of parameters using the one-dimensional data space of numbers 1, ..., 15.
5. To classify each child using the posterior distributions in Fig. 4, we took a conservative approach and assigned the first knower level with posterior mass greater than the prior mass (i.e., the first knower level for which the data provided positive evidence). Generally, of course, this procedure just gives each child his or her obvious classification based on Fig. 4 (e.g., child 2 is classified as a NN-knower), but in the rarer ambiguous cases our approach is conservative (e.g., child 79 and child 1 are also classified as NN-knowers, despite their being some possibility that they are one-knowers).

Acknowledgments

This research was supported by NICHD grant 00234342 to the second author. Massachusetts data collection was supported by NSF REC grant 0337055 to Elizabeth Spelke and Susan Carey. We thank Josh Tenenbaum and two reviewers for their very helpful comments. We also thank the children and families who participated in the original studies, the preschools hosting that research, and UCI Cognitive Development Lab Manager Emily Carrigan and Research Assistants John Cabiles, Alexandra Cerutti, Jyothi Ramakrishnan, Sarah Song, Dat Thai, and Gowa Wu for their help with data collection.

References

- Carey, S. (2001). Evolutionary and ontogenetic foundations of arithmetic. *Mind and Language*, 16(1), 37–55.
- Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus*, 133(1), 59–68.
- Carey, S., & Sarnecka, B. W. (2006). The development of human conceptual representations. In M. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 473–496). New York: Academic Press.

- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7), 287–291.
- Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. New York: Springer-Verlag.
- Cordes, S., Gallistel, C. R., & Gelman, R. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, 8, 698–707.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development*, 60, 1158–1171.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. In D. L. Medin (Ed.), *The psychology of learning and motivation. Advances in research and theory* (pp. 61–96). London: Academic Press.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.) (1996). *Markov Chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall/CRC.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling* (pp. 59–100). Cambridge, MA: Cambridge University Press.
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105, 395–438.
- Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of counting principles. *Cognitive Psychology*, 52(2), 130–169.
- Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15(1), 1–15.
- Lee, M. D., & Wagenmakers, E.-J. (2008). *A course in Bayesian graphical modeling for cognitive science*. Unpublished course notes, University of California, Irvine. Available at: <http://www.socsci.uci.edu/~mdlee/bgm>. Accessed on August 7, 2009.
- MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, England: Cambridge University Press.
- Marr, D. C. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.
- McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, 1(1), 11–38.
- Myung, I. J., Forster, M., & Browne, M. W. (2000). A special issue on model selection. *Journal of Mathematical Psychology*, 44, 1–2.
- Piaget, J. (1952). *The child's conception of number*. New York: Routledge.
- Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, 92, 329–352.
- Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, J. B. (2007). From grammatical number to exact numbers: Early meanings of “one,” “two,” and “three” in English, Russian and Japanese. *Cognitive Psychology*, 55, 136–168.
- Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge in early childhood. *Journal of Experimental Child Psychology*, 103(3), 325–337.
- Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, 6, 357–379.
- Shiffrin, R. M., Lee, M. D., Kim, W.-J., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32(8), 1248–1284.

- Spiegelhalter, D. J., Thomas, A., & Best, N. G. (2004). *WinBUGS version 1.4 user manual*. Cambridge, England: Medical Research Council Biostatistics Unit.
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193.
- Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology*, *24*, 220–251.

Appendix: Full data set

Child	“one”	“two”	“three”	“four”	“five”	“six”	“eight”	“ten”
1	1, 2, 10	3, 5	4, 15					
2	12, 2	2	2					
3	1, 1	3, 2, 2, 2	2, 4, 3					
4	1, 1	2, 2	8, 5					
5	1, 1	2, 2	3, 4	5	4			
6	1	2, 2	3, 4, 3	5, 3	5, 4	5		
7	1, 1	2, 2, 4, 2	4, 7, 5					
8	1, 1, 1, 1	2, 2, 2, 2	3, 2, 2, 3, 4	3				
9	1	2	3, 4, 3	4, 6, 4	7, 6			
10	1	2, 2, 2	15, 3, 4, 15, 3	4	4			
11	1	2, 2, 2	3, 4, 4, 5, 3	4, 4, 4	6, 4, 4			
12	1, 1	2, 2	3, 3, 5, 3, 3	2, 15, 15	6			
13	1		3, 3	5, 4, 13	5, 5, 6, 11	5, 7		
14	1		3, 3, 3	4, 5, 6	5, 4, 5, 6	5, 5		
15	1	2, 2	3, 3, 5	4, 7	7	4		
16	1	2	3, 3	9, 14	15			
17	1		3, 3, 3	4, 5, 5	8, 4			
18	1	2	3, 3, 3	4, 5, 5	6, 4			
19	1	2	3, 3, 3	7, 4, 8	5	5, 8		
20	1		3, 3, 3	5, 7, 4	5, 8	4		
21	1		3, 3, 3	15, 6, 13	6			
22	1		3	4, 4	6, 5	5, 5		
23	1		3, 3	14	6	9		
24	1		3, 3	4, 4	8	8		
25	1	2	3, 3	4, 4	15, 15			
26	1		3	4	5	8, 6		
27	1		3	4, 4	5, 5, 4, 5	5, 5, 7		
28	1		3, 3	4, 4, 4	15, 13			
29	1	2, 2	3, 3, 3, 3, 5	4, 4, 5, 4	3, 8			
30	1		3		5	6		
31	1		3		5	6		
32	1		3		5	6		
33	1		3		5	6		
34	1		3		5	6		
35	1	2	3	4	5	6		
36	1		3		5	6		
37	1		3		5	6		

Appendix
(Continued)

Child	“one”	“two”	“three”	“four”	“five”	“six”	“eight”	“ten”
38	1		3		5, 5	6, 6		
39	1		3	4	5	6, 6		
40	1		3		5, 5	6		
41	1		3	4	5	6, 6		
42	1		3		5	6		
43	1		3		5	6		
44	1		3	4, 4	4, 5	6, 6		
45	1		3		5, 5	5, 6		
46	1		3		5	6		
47	1		3		5	6		
48	1	2	3	4	5, 5	5, 6		
49	1	2	3	4	5	6		
50	1	2, 2	3	4	5	6		
51	1	2	3	4	5	6		
52	1	2	3	4	5	6		
53	1		3		5, 5	6, 6		
54	1		3		5, 5, 5	6, 6, 6		
55	1		3	4	5, 6, 5, 5	6, 6		
56	1		3	4	5, 5, 5	7, 8, 6, 6		
57	1		3		5, 5	6, 6		
58	1		3		5, 5	6, 6		
59	1		3		5, 5, 5	6, 11, 13		
60	1		3	4, 4	5, 8, 5	8, 6, 6		
61	1		3		5, 5	6, 6		
62	1		3		5, 5	6, 6		
63	1, 1	2, 2	3	4, 4	4, 5	6		
64	1	2	3	4	5	6		
65	1, 1, 1	2, 2, 2	3, 3, 3	4, 4, 4	5, 5, 5		8, 8, 8	10, 10, 10
66	1, 1, 1	2, 2, 2	3, 4, 3	4, 4, 4	5, 5, 6		8, 8, 8	10, 10, 10
67	1, 1, 1	2, 2, 2	3, 3, 3	4, 4, 4	10, 15, 15		8, 6, 8	11, 15, 15
68	1, 1, 1	2, 5, 5	5, 2, 6	3, 4, 4	10, 15, 2		14, 2, 3	7, 2, 3
69	1, 1, 1	2, 2, 2	3, 3, 3	4, 4, 4	5, 5, 5		8, 8, 7	10, 10, 10
70	1, 1, 1	2, 2, 2	4, 15, 3	5, 15, 6	5, 15, 4		13, 13, 15	15, 4, 15
71	1, 1, 15	2, 2, 8	15, 15, 15	15, 15	15, 15		15, 9	15, 7, 15
72	1, 1, 1	2, 2, 2	3, 3, 3	4, 6, 8	6, 5, 8		3, 7, 7	5, 5, 3
73	1, 1, 1	2, 2, 2	4, 6, 4	6, 4, 4	15, 15, 15		3, 7, 5	9, 4, 7
74	1, 1, 1	2, 2, 2	3, 3, 3	4, 6, 5	4, 7, 5		6, 4, 8	5, 4, 7
75	1, 1, 1	2, 2, 2	3, 3, 3	4, 11, 8	5, 5, 9		10, 7, 6	10, 9, 6
76	1, 1, 1	2, 13, 15	3, 7, 15	3, 8, 3	3, 6, 7		5, 15, 15	10, 3, 8
77	1, 1, 1	2, 3, 6	2, 3, 2	3, 4, 2	2, 3, 2		2, 3, 5	3, 4, 3
78	1, 15, 15	2, 15, 15	4, 3, 15	4, 15, 15	15, 15, 15		5, 15	15, 15, 15
79	15, 15, 15	15, 15, 15	15, 15, 15	15, 15, 15	15, 15, 15		15, 15, 15	15, 15, 15
80	1, 1, 1	2, 2, 2	4, 4, 15	8, 15, 15	15, 6, 15		5, 15, 15	15, 14, 15
81	1, 1, 1	2, 2, 2	2, 8, 4	1, 3, 2	3, 13, 3		8, 6, 8	7, 9, 4
82	1, 1, 1	2, 2, 2	4, 3, 4	7, 9, 8	9, 8, 7		5, 4, 4	3, 4, 5