An Excel sheet for inferring children's number-knower levels from give-*N* data

James Negen · Barbara W. Sarnecka · Michael D. Lee

Abstract *Number-knower levels* are a series of stages of number concept development in early childhood. A child's number-knower level is typically assessed using the give-*N* task. Although the task procedure has been highly refined, the standard ways of analyzing give-*N* data remain somewhat crude. Lee and Sarnecka (Cogn Sci 34:51–67, 2010, in press) have developed a Bayesian model of children's performance on the give-*N* task that allows knower level to be inferred in a more principled way. However, this model requires considerable expertise and computational effort to implement and apply to data. Here, we present an approximation to the model's inference that can be computed with Microsoft Excel. We demonstrate the accuracy of the approximation and provide instructions for its use. This makes the powerful inferential capabilities of the Bayesian model accessible to developmental researchers interested in estimating knower levels from give-*N* data.

Keywords Cognitive development · Cognitive modeling · Bayesian inference · Bayesian cognitive modeling · Number-word learning · Numerical cognition

Children learn very early (often as young as 2 years old) to recite the number-word list in order (Fuson, 1988). But at the beginning, the words are merely placeholders—children recite the list without knowing what the individual number words mean. Over time, children fill in the words with meaning, one at a time and in order (Carey, 2009; Sarnecka & Lee, 2009). The child's progress on this front is called

J. Negen (✉) · B. W. Sarnecka · M. D. Lee
Department of Cognitive Sciences, University of California, Irvine, CA, USA
e-mail: jnegen@uci.edu

their *number-knower level*, or just *knower level*. A child who does not yet know any number word meanings is called a *pre-number-knower* or *0-knower*. A child who only knows the meaning of "one" is a 1-knower; knowing "one" and "two" makes her a 2-knower; and so forth. After reaching the 3-knower or 4-knower level, children become cardinality-principle-knowers (or *CP-knowers*, for short). That transition happens when they figure out the cardinality principle of counting, which allows them to infer the meanings of all higher number words through counting.

Number-knower level has been a useful construct in several lines of research. For example, Ansari et al. (2003) used it to examine deficits caused by Williams Syndrome; Le Corre and Carey (2007) used it to investigate when and how the approximate-number system becomes linked to integer concepts; Sarnecka, Kamenskaya, Yamana, Ogura and Yudovina (2007) used it to examine cross-linguistic variation in number development; Carey (2009) used knower levels as a window into the general process of conceptual change.

The task most commonly used to assess knower level is the give-*N* task. In this task, children are given a large bowl of small items and told that they are going to play a game with a puppet. The experimenter asks the child to give a certain number of items to the puppet—for instance, "Can you give Mr. Bunny TWO bananas?" The numbers requested generally include "one" through "six" (a few studies include higher numbers), with about three trials per number word.

In the past, researchers typically inferred the knower levels of individual children either by ad hoc heuristics (e.g., Sarnecka & Gelman, 2004; Wynn, 1992) or by looking for convergence in a titrating method (e.g., Barner, Chow & Yang, 2009). In both of these paradigms, children are given credit for knowing a certain number word when

some cutoff in performance is reached (often 2/3 correct); however, the cutoffs are not motivated by any particular theoretical principles.

As an alternative to relying on heuristics to analyze give-N behavior, Lee and Sarnecka (2010, in press) developed a formal Bayesian model of how children with different levels of number knowledge behave on the task. Their model provides a principled basis for knower-level inference. However, implementing the model would require most developmental researchers to learn a host of new technical skills; since the model does not provide a simple formula for the estimate of a child's knower level, a computational approach is required to put the model to use. Here we present a reasonable approximation of the Lee and Sarnecka (2010, in press) model that only requires the user to interact with Microsoft Excel, thus taking away much of the technical burden.

In this article, we first describe the model itself in detail, and then describe how it is approximated in the Excel sheet. After that, we describe the two give-N data sets used to (1) create the point estimates that are needed for the approximation and (2) test the quality of the approximation. We then describe the results of applying the model to the calibration data set. After that, we compare the model's inference to the Excel sheet's inference, as well as to inferences by a popular ad hoc method. The results suggest that the Excel sheet provides estimation of knower levels comparable to that obtained with the full Bayesian model, and that it provides several advantages over ad hoc methods. Finally, we provide instructions for using the sheet itself.

Model description

Lee and Sarnecka (2010, in press) developed a probabilistic generative model of children's behavior on the give-N task. The model is based on the idea that children's answers can be understood as Bayesian reasoning based on (a) their prior knowledge of number concepts and (b) the data provided by experimenter instructions on each trial. Thus, one contribution of the model is that it specifies a base rate—a distribution of prior probabilities describing what "chance" responding on the give-N task looks like.[1] This base rate, which is inferred from give-N

data, corresponds roughly to the answers children give when they have no idea how many objects have been requested (because they don't know the meaning of the number word used in the prompt). This base rate is then modified, following Bayes's rule, when the child is asked for a particular number of items. To be specific, several things happen when the child is asked for N items:

If the child knows the meaning of the word N (e.g., "two"), the likelihood of them giving N (e.g., 2) items becomes higher than it was in the base rate. The evidence value, v, represents the size of this change.

All of the other number-word meanings that the child knows become *less* likely responses (i.e., less likely than they were in the base-rate distribution) by the same factor v. For example, when a 3-knower is asked for "two" items, the likelihood of giving 2 goes up and the likelihoods of giving 1 and 3 go down, all by the same factor v.

If the child does not know the meaning of N (e.g., when a 3-knower is asked for "six"), all known numbers (e.g., 1, 2, and 3) become less likely responses as described above, and all unknown numbers (e.g., 4, 5, 6, 7, 8, 9, . . .) retain the relative probabilities that they had in the base rate. (Their absolute probabilities go up, because the entire distribution must sum to 1.)

This model provides a principled way to predict patterns of data coming from the give-N task, sorted by knower level.[2] This makes it possible to infer a child's knower level from the data.

Following Lee and Sarnecka (2010, in press), we used a graphical model as our implementation, as shown in Fig. 1. Graphical models are a standard language in machine learning, statistics, and (more recently) cognitive science (e.g., Griffiths, Kemp & Tenenbaum, 2008; Jordan, 2004; Koller, Friedman, Getoor & Taskar, 2007; Lee, 2008, 2011; Shiffrin, Lee, Kim & Wagenmakers, 2008). They specify how unobserved psychological variables generate the observed data. The model takes the form of a graph, whose nodes represent the variables and the data. The links indicate the dependencies between them. Discrete variables are indicated by square nodes; continuous variables are indicated by circular nodes. Stochastic variables are indicated by single-bordered nodes; deterministic

---

[1] It is intuitively easy to understand that, all else being equal, children are more likely to give some set sizes than others. For example, children often give all of the objects in the bowl to the puppet; they rarely give all but one. The base rates are very high for giving 1 item, 2 items, or all 15 items. The curve in the middle (from 3 to 14) is high for small numbers and decays steadily as it moves to larger ones. For a visualization, look ahead to Fig. 2.

[2] For readers who are curious what the patterns are like: When the model is applied to data, it turns out that pre-number-knowers just give the base-rate number of items. 1-knowers almost always give 1 when asked for "one," but rarely give 1 when asked for anything else. Of course, 1-knowers sometimes give 2 for "two," because the overall probability of giving 2 objects (specified by the base rate) is high. Likewise, 2-knowers give 2 when asked for "two" and rarely give 2 for anything else, and so forth. CP-knowers have a very good chance of getting any number correct, because they can use counting to solve the task. (For children who do not yet understand the cardinality principle, counting is not available as a strategy.)

variables (included for conceptual clarity) are indicated by double-bordered nodes. Shaded nodes are the observed data; unshaded nodes are latent variables. Finally, encompassing plates are used to denote independent replications of the graph structure within the model.

In our implementation of the knower-level model (Fig. 1), the data are the observed q_{ij} and g_{ij} variables. They give the number asked for (the "question") and the answer (the number "given"), respectively, for the ith child on his or her jth trial.[3] The base-rate probabilities are represented by the vector π, which is updated to π' according to the Bayesian updating process described above. This probability space is visualized in Fig. 2. This process depends on the knower level z_i of the child and an evidence value v that measures the strength of the updating. The child's actual behavior (i.e., the number of items given) represents a sample from the distribution expressed by π'. The base-rate and evidence parameters, which are assumed to be the same for all children, are given uninformative priors (i.e., initial distributions that assume very little, allowing for a very large range of possible inferences). There is a separate knower-level parameter for each child, with no prior preference for any of the six knower levels.

Lee and Sarnecka (2010, in press) used this graphical model implementation (and associated computational sampling methods) to demonstrate that the model can (1) describe children's behavioral data, (2) make useful inferences about the knower level of specific children, and (3) describe the base-rate and evidence-strength parameters. One of the most obvious and useful applications of the model is to infer a child's knower level from their give-N behavior. The goal of this article is to provide an approximation of the full model that is nonetheless accurate in making these estimations and that can be implemented in Microsoft Excel.

Approximation in the Excel sheet

Next we describe an approximation that provides estimation of knower levels comparable to that obtained with the full Bayesian model. To create the approximation, we eliminated all of the continuous distributions in the full model, thereby removing the need for Markov chain Monte Carlo techniques. The node π'—the final probability of every possible response, given the knower level and the request—was filled with point estimates; the parameters π and v were removed entirely. In the full model, π' is subject to implied distributions (through π, v, and the

updating logic) that need to be integrated out. In the approximation, π' is just a set of summary point probabilities—for example, the chance that a 1-knower will give 1 when asked for "one" is 83.33%. These probabilities were found by applying the full model to a calibration data set (described in the next section) and calculating the posterior predictive.[4]

The remaining computational machinery is simple enough to be captured in a single equation. The probability that the ith child is at a given knower level Z can be written as

$$p(Z = z_i | \pi', q_{ij}, g_{ij}) = \frac{p(Z) \prod_j \pi'_{Z,qij,gij}}{\sum_{z=1}^{6} \left[p(z) \prod_j \pi'_{z,qij,gij} \right]}$$

Since π' is a set of point probabilities in the approximation, nothing further is needed for it to work. Following the usual Bayesian form, the top part of the equation is just the prior times the (simplified) likelihood function. Since each trial is assumed to be independent, calculating the likelihood just requires multiplying over the probabilities of all the observed responses. The bottom part is the normalization constant. Since there are a total of six discrete knower levels, the constant is just the sum of the top part over all knower levels. This approximation can work in Excel because no integration is required—with set values for π' and Z, the right-hand side of the above equation can be worked just by multiplying, adding, and dividing. Then, by simply repeating over the six possible values of Z, the full distribution is calculated.
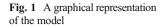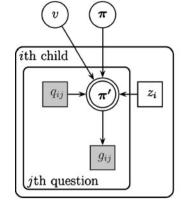
Data

We based our approximation on empirical data from experiments reported by Negen and Sarnecka (2010). For this data set, children were asked to give one, two, three, four, six, and eight items. Each request was repeated three times, for a total of 18 trials, which were presented in one of two pseudorandom orders. The present analysis includes only sessions in which the child completed at least 15 of the 18 trials, a total of 423 sessions. (The original data set included an additional 31 sessions in which the child failed to complete at least 15 trials. These sessions were excluded.)

The independent data used to compare the model inferences and Excel sheet inferences came from Lee and Sarnecka (in press). This data set includes data from 56

---

[3] Note that the jth trial does not necessarily involve the child being asked for j items. The value of q_{ij} is the number of items the ith child was asked for on the jth experimental trial completed.

[4] For readers who are unfamiliar with this term: the posterior predictive is the distribution of predicted future data, given the current observed data and the assumptions in the model.

Fig. 1 A graphical representation
of the model

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

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

$$z_i \sim \text{Categorical}\left(\tfrac{1}{6},\dots,\tfrac{1}{6}\right)$$

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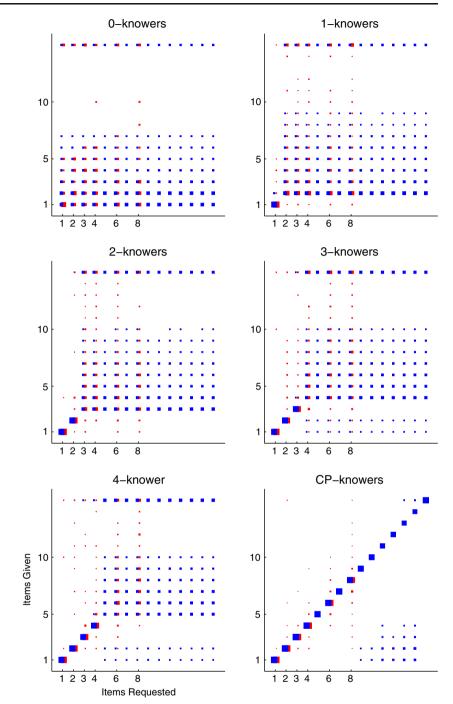
$$g_{ij} \sim \text{Categorical}(\pi')$$

children who were asked for one, two, three, four, five, eight, and ten items. Each number was requested three times, for a total of 21 trials per child.

Results

We did fully Bayesian inference on the calibration data, using the same computational sampling method as Lee and Sarnecka (2010, in press). More specifically, two chains were run, each with 2,000 burn-in samples and 25,000 data collection samples, for a total of 50,000 samples. Chain convergence was good, with the standard R-hat statistic being very close to 1 for all of the variables sampled.

The model inferred that among the children tested, 9 were 0-knowers, 48 were 1-knowers, 50 were 2-knowers, 53 were 3-knowers, 67 were 4-knowers, and 196 were CP-knowers. The inferred posterior predictions are shown in Fig. 3, broken down by knower level. The full numeric breakdown is Sheet 2 of the Excel sheet itself. The model inferred an evidence value v of 16.94 ($SD=0.69$), which is slightly lower than values estimated from other data sets by Lee and Sarnecka (2010, in press)—a drop from about 29 and 23, respectively.[5] In other words, if the experimenter makes a request with a number word that the child knows, the correct response becomes about 17 times more likely than it was in the base-rate distribution, and other number-word meanings the child knows become 17 times less likely. For 0-knowers, the inferred base rate is the same as the posterior predictions.

Comparison with heuristic and fully Bayesian methods To see how the inferences made by the Excel sheet compare with a typical heuristic inference or a fully Bayesian inference, we looked at a different data set of 56 children, first reported in Lee and Sarnecka (in press). According to the heuristic, a child gets credit for knowing a number if her correct answers outnumber her errors by at least 2:1. The child's inferred knower level is the highest number that she knows. The data were run though (a) this heuristic, (b) the Excel sheet, and (c) a fully Bayesian method, after being appended to the calibration data set.

Figure 4 shows all of the posterior distributions for every child from each inference method. For the purposes of comparing the Excel sheet with the heuristic, we used the most likely posterior knower level as a point estimate. There are 38 cases in which this estimate matches the knower level inferred by the heuristic. Of the remaining 18 cases, in only 1 case is the estimate from the Excel sheet a lower knower level than the estimate from the heuristic. This primarily occurred because the heuristic requires the same proportion of correct answers for every knower level, whereas the model is more "lenient" on larger sets (except for CP-knowers, who are expected to count with good accuracy for any set size). Intuitively, it makes sense that larger sets should be more difficult to generate; for example, 3-knowers should not be required to maintain the same level of accuracy as 1-knowers. Children at higher knower levels also have more opportunities to demonstrate knowledge, and thus more opportunities to make performance errors.

All three inference methods will lower the chances that the child is a 2-knower if the child gives 2 items in response to a different number word. However, the Excel sheet and the fully Bayesian method also consider how many trials have passed without the child erroneously giving a certain number. So, for instance, every time a child makes an error when asked for "four" and that error is _not_ giving 2, the posterior odds
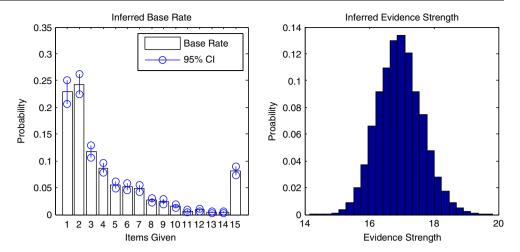
---

[5] This means that the Excel sheet expects behavior to be slightly less differentiated by knower level, which results in the inference being conservative—that is, the resulting posteriors will be more diffuse (appear less certain) than would happen if a higher evidence value were used. It is interesting that this difference in evidence value does not lead to large problems for the approximation when generalizing between these two sets (as shown in the next section). This may suggest that the method is robust to small changes in the evidence parameter.

Fig. 2 The probability of responses, organized by items requested and knower level. The six graphs separate out each knower level. In each graph, the x-axis indicates how many items are being requested. The y-axis indicates how many items the child gives. Large blue squares indicate that the particular request–response pair is very likely for that knower level in the model; smaller blue squares indicate lower probability. Red squares show how often those pairs actually occurred in the data, separated into the six graphs by the modal posterior knower level of each child (only available for requests of one, two, three, four, six, and eight items). There are large overlapping blue and red squares where children know the number word in the request (e.g., at the bottom left corner of the 1-knowers graph); the child is very likely to correctly answer when she knows the relevant number word, both in the model and the data. Following in both vertical and horizontal "stripes" from each large square, there are many very small squares—there is little chance of either responding incorrectly to that word or giving a set of that size for any other word

of being a 2-knower receive a small upward push relative to the odds of being a 0-knower or 1-knower. This is because the chance of a 0-knower or 1-knower erroneously giving 2 items is much higher than the chance of a 2-knower committing the same error. This allows indirect counterevidence to accumulate against a child's errors when asked for a number word.

The Excel sheet's inference is very close to the inference generated by fully Bayesian inference. In terms of maxima, there are no discrepancies. The mean absolute difference between the model's posterior over knower levels and the Excel sheet's approximate posterior is 0.2%, with a standard deviation of 0.68%. The largest absolute difference is 4.8% (for Child 50 being a 4-knower), where both inferences have the same basic shape, but the Excel sheet is slightly more peaked at the mode. More diffuse posteriors tend to be approximated less accurately; the user can be especially confident of the approximation when the posterior has a very strong mode.

Fig. 3 Inferred parameters of the model: the base rate (what the child might give if no number word is used) and the evidence strength (the *v* in the model description; a parameter controlling how much the probability of different responses gets modified by the child's knowledge of number words)

How to use the Excel sheet

The Excel sheet is available at www.cogsci.uci.edu/cogdev/Negen/Knower-LevelEstimater.xls. Figure 5 is a screen shot of the Excel sheet. The user enters data in the rows near the top labeled "question" and "response." Both of these must be in the range of 1 to 15. The Excel sheet is designed to handle all of the data from a single child at a time. In the question row, the user can enter the numbers requested from left to right, with the child's responses beneath. Trials do not have to be in order, as long as each response is entered below the corresponding question. The sheet will automatically calculate the likelihood of each question/response pair conditional on each knower level, in rows 6 to 11. Questions without responses do not affect the posterior distribution, so they can be omitted.

Prior probability is the place to enter prior weights for different knower levels. If used properly, priors are the best way to draw on all of the information available about a child. Imagine, for example, that the user knows something about the population from which a child was recruited: At least 50% of them are CP-knowers. Then, the prior probability of being a CP-knower should be set to 50% in the sheet. If the data don't strongly adjudicate between two knower levels, then the prior can help push the inference in the more likely direction. If the user doesn't have this kind of prior information, all of the prior probabilities can just be set to the same value.

The end result, after entering the data and setting the prior, is a set of probabilities for the six knower levels (along with a graph to visualize them). This allows the user to see what knower level is preferred and how strongly. Provided also is the log-likelihood and the scaled log-likelihood of the data, since they may be more familiar to some researchers. In the example in Fig. 4, the child was asked for "one," "two," "three," "four," and "five." She gave 1, 2, 3, 3, and 6, respectively. These are entered in rows 3 and 4. The prior probability of each knower
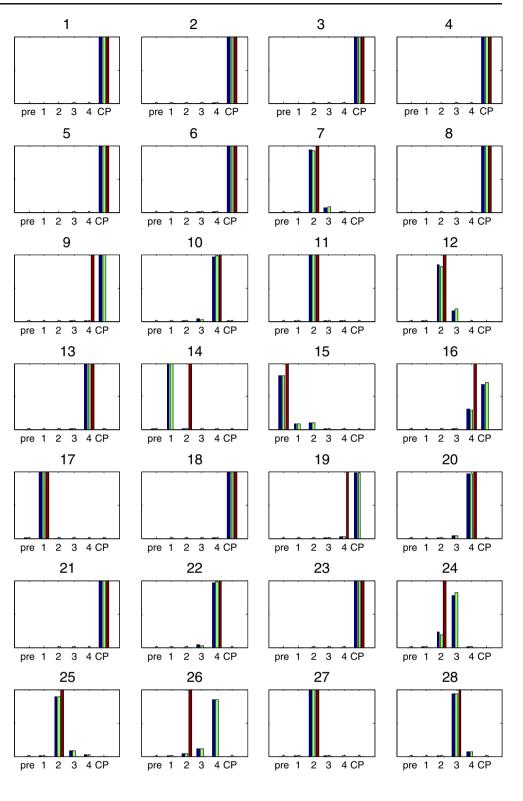
level is the same (cells B36 to B41). This leads to a confidence of about 60% that this child is a 2-knower (seen in the graph and under "Posterior Probability" in cells J36 to J41).

Model limitations The formal expression of the approximation assumes that the child (1) was never asked for more than 15 items and (2) had a total of 15 items available to give. This has two implications: (1) If the child was asked for more than 15 items, some other method of analysis must be used. However, it is rare for given-*N* tasks to involve asking for more than 10 items. (2) If the child had more or fewer than 15 items to give, there might be some problems in how the estimation works—specifically, the base rate might change. For example, children probably like to give 15 items because it is a maximal response, not because of anything special about the number 15. It is plausible that some mapping could be created to line up the base rates (maximal requests/responses mapped to 15, 15+ mapped to 10 or so), but this is an empirical question that we do not currently have the data to answer. As it stands, the sheet is only appropriate for experiments in which the child had exactly 15 items to give—though the method of the approximation could, of course, be applied to other numbers of items.
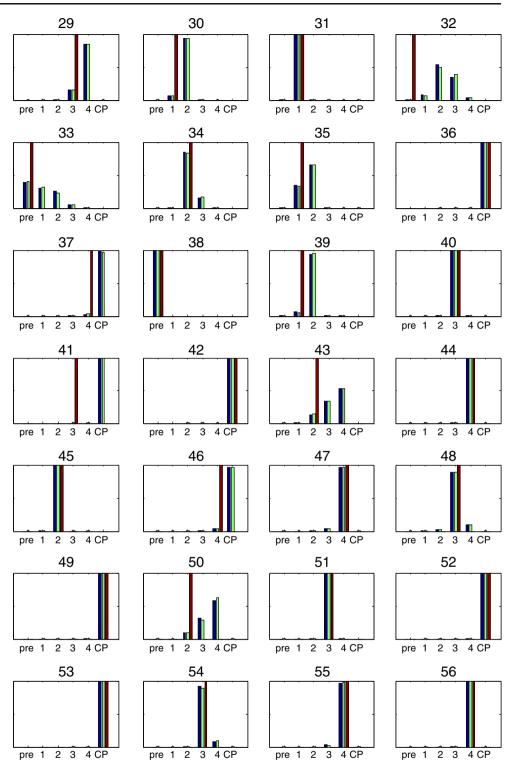
There are two other properties of the model that are worth discussing.[6] First, the model predicts very high-accuracy performance from CP-knowers. Children who respond accurately when asked for 1–4 items, but frequently miscount larger requests—even just by an item or two—are often judged to be 4-knowers by the model (e.g., try entering the following question–response pairs: 1–1, 2–2, 3–3, 4–4, 5–5, 6–5, 7–7, 8–7). This classification comes from a theoretical question, beyond the scope of this study: Are such children CP-knowers, since they are attempting to use counting to construct the sets, or are they 4-knowers,

---

[6] We thank an anonymous reviewer for drawing our attention to these issues.
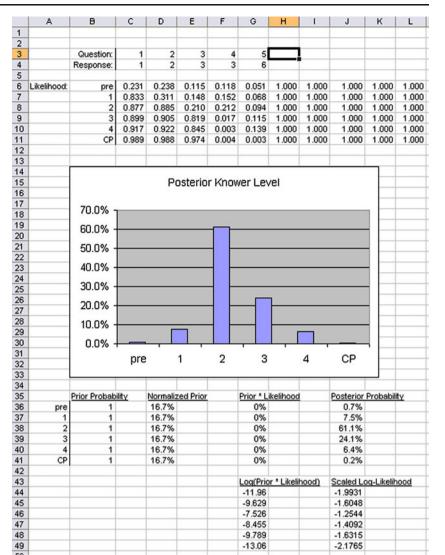
since they do not yet understand the counting principles well enough to apply them consistently? Future iterations of the model may have a separate way of treating responses that are likely to be counting errors. Examining Fig. 4, it appears that such a model could make reasonable gains in fitting the data better.

Some readers may also find something counterintuitive in how little or how much certain errors inform the end estimation. For example, consider the following question–response pairs: 1–1, 2–2, 3–3, 4–4, 5–5, 5–5, 6–5. The knower-level posterior now shows nearly equal probabilities for both 4-knower and CP-knower, with all other knower

Fig. 4 (continued)

levels being far less likely. Now try adding in 8–1. It is unusual for a CP-knower to make this error, estimated to happen only 2.7% of the time. This might cause the reader to expect the bar for CP-knowers to shrink. However, it is also an unlikely error for most other knower levels, too, so the impact is actually very small. (Remember, the sheet is always comparing knower levels against each other; evidence for/against all of the knower levels just washes out.) On the other hand, an error like 8–5 leads to different changes. This behavior does not obviously indicate knowledge of the words "one" through "four," but it pushes up the 4-knower bar dramatically. This happens because this is an error that a 4-knower—but not any

Fig. 5 A screen shot of the actual Excel sheet, with some example data filled in. There is more room for data entry off to the right

	A	B	C	D	E	F	G	H	I	J	K	L
3		Question:	1	2	3	4	5					
4		Response:	1	2	3	3	6					
6	Likelihood:	pre	0.231	0.238	0.115	0.118	0.051	1.000	1.000	1.000	1.000	1.000
7		1	0.833	0.311	0.148	0.152	0.068	1.000	1.000	1.000	1.000	1.000
8		2	0.877	0.885	0.210	0.212	0.094	1.000	1.000	1.000	1.000	1.000
9		3	0.899	0.905	0.819	0.017	0.115	1.000	1.000	1.000	1.000	1.000
10		4	0.917	0.922	0.845	0.003	0.139	1.000	1.000	1.000	1.000	1.000
11		CP	0.989	0.988	0.974	0.004	0.003	1.000	1.000	1.000	1.000	1.000

Posterior Knower Level (bar chart)

Y-axis: 70.0%, 60.0%, 50.0%, 40.0%, 30.0%, 20.0%, 10.0%, 0.0%
X-axis categories: pre, 1, 2, 3, 4, CP

	Prior Probability	Normalized Prior	Prior * Likelihood	Posterior Probability
pre	1	16.7%	0%	0.7%
1	1	16.7%	0%	7.5%
2	1	16.7%	0%	61.1%
3	1	16.7%	0%	24.1%
4	1	16.7%	0%	6.4%
CP	1	16.7%	0%	0.2%

Log(Prior * Likelihood)	Scaled Log-Likelihood
-11.96	-1.9931
-9.629	-1.6048
-7.526	-1.2544
-8.455	-1.4092
-9.789	-1.6315
-13.06	-2.1765

other knower level—is likely to make; lower knower levels are much more likely to give fewer items, and CP-knowers are much more likely to count out eight items correctly.

Point estimates and interpreting the posterior The end result of the Excel sheet is a posterior distribution over knower levels. We conclude this article by providing some brief guidance on what this distribution means and how it can be interpreted. In the example in Fig. 5, the hypothetical child has roughly a 61% posterior probability of being a 2-knower. This posterior probability has a very direct interpretation: It is the probability that a child is at a specific knower level, given the data and the assumptions in the model. This is actually somewhat simpler than the results of many other forms of statistical inference; it is not the probability of an observed statistic under a null, but is actually the straightforward probability that the child is a 2-knower.

However, for many researchers, posterior probabilities will not be familiar; point estimates are the outcome of most classical estimation methods. Point estimates are also needed for methods like regression analyses or ANOVAs. (For most cases, this is not a problem, because for many children, the posterior is so peaked that any reasonable summary will suggest the same knower level.)

If the user does not feel comfortable interpreting the posterior directly, then the choice of the point estimate should depend on the purpose of the analysis. For most cases, the mode is a sensible choice, but there could be exceptions. For example, suppose a researcher wants to argue something in the following form: "Even before they learn the cardinality principle, children already know X" (e.g., Sarnecka & Gelman, 2004). In this case, mistakenly classifying a CP-knower as being at some lower knower level would be problematic; it could undermine the validity of the whole argument. On the other hand,

misclassifying a less knowledgeable child as a CP-knower would not undermine the argument; it would only remove a small amount of data. In such a case, a sensible estimate would be the highest knower level with posterior probability that exceeds the prior probability. This method would tend to sort children into higher knower levels where the data were ambiguous, and thus would be unlikely to misclassify a CP-knower. Researchers with very particular needs could even explicitly define a utility function—though a full discussion of that topic is outside the scope of this article.

Summary

Knower levels are one of the most useful landmarks of number-concept development in young children. Data from the standard give-*N* task can be analyzed using a generative model of the task, thereby drawing inferences in a more principled way than heuristic methods. The model's estimates also provide information about how certain we can be of the knower-level classification for each child, and they empower the researcher to decide ambiguous cases in the way that is most appropriate for the problem at hand. The software presented in this article computes a close approximation to the model's inference while remaining quick and easy to use. It is available at www.cogsci.uci.edu/cogdev/Negen/Knower-LevelEstimater.xls.

References

Ansari, D., Donlan, C., Thomas, M. S. C., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology, 85,* 50–62. doi:10.1016/S0022-0965(03)00026-2

Barner, D., Chow, K., & Yang, S.-J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology, 58,* 195–219. doi:10.1016/j.cogpsych.2008.07.001

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer.

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: Cambridge University Press.

Jordan, M. I. (2004). Graphical models. *Statistical Science, 19,* 140–155.

Koller, D., Friedman, N., Getoor, L., & Taskar, B. (2007). Graphical models in a nutshell. In L. Getoor & B. Taskar (Eds.), *Introduction to statistical relational learning*. Cambridge: MIT.

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. doi:10.1016/j.cognition.2006.10.005

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review, 15,* 1–15. doi:10.3758/PBR.15.1.1

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55,* 1–7. doi:10.1016/j.jmp.2010.08.013

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science, 34,* 51–67. doi:10.1111/j.1551-6709.2009.01063.x

Lee, M. D., & Sarnecka, B. W. (in press). Number knower-levels in young children: Insights from a Bayesian model. *Cognition*.

Negen, J., & Sarnecka, B. W. (2010). Analogue magnitudes and knower-levels: Re-visiting the variability argument. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1252–1257). Austin: Cognitive Science Society.

Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition, 92,* 329–352. doi:10.1016/j.cognition.2003.10.001

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103,* 325–337. doi:10.1016/j.jecp.2009.02.007

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. 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. doi:10.1016/j.cogpsych.2006.09.001

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32,* 1248–1284. doi:10.1080/03640210802414826

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24,* 220–251. doi:10.1016/0010-0285(92)90008-P