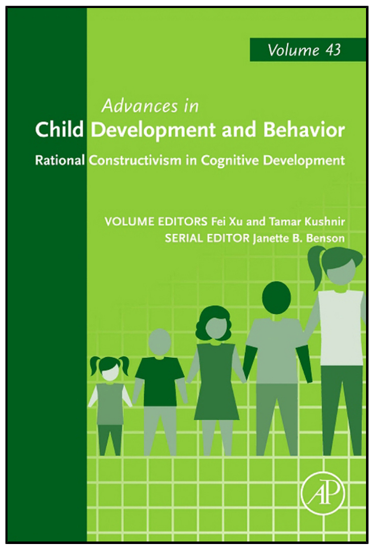
Author's personal copy

Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.

This chapter was originally published in the book *Rational Constructivism in Cognitive Development*. The copy attached is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research, and educational use. This includes without limitation use in instruction at your institution, distribution to specific colleagues, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:
http://www.elsevier.com/locate/permissionusematerial

From Sarnecka, B. W., & Negen, J., (2012). A Number of Options: Rationalist, Constructivist, and Bayesian Insights into the Development of Exact-Number Concepts. In J. B. Benson (Serial Ed.) & F. Xu & T. Kushnir (Vol. Eds.), *Rational Constructivism in Cognitive Development* (pp. 237–268). Elsevier Inc.: Academic Press.
ISBN: 9780123979193
Copyright © 2012 Elsevier Inc. All rights reserved.
Academic Press

Author's personal copy

CHAPTER NINE

A Number of Options: Rationalist, Constructivist, and Bayesian Insights into the Development of Exact-Number Concepts[1]

Barbara W. Sarnecka* and James Negen
Department of Cognitive Sciences, University of California, Irvine, CA 92697-5100
*E-mail: sarnecka@uci.edu

Contents

Abstract

The question of how human beings acquire exact-number concepts has interested cognitive developmentalists since the time of Piaget. The answer will owe something to both the rationalist and constructivist traditions. On the one hand, some aspects of

[1] This work was supported by NSF grant DRL-0953521 to the first author.

Advances in Child Development and Behavior, Volume 43
ISSN 0065-2407,
http://dx.doi.org/10.1016/B978-0-12-397919-3.00009-5
© 2012 Elsevier Inc.
All rights reserved.

Rational Constructivism in Cognitive Development, First Edition, 2012, 237–268

Author's personal copy

numerical cognition (e.g. approximate number estimation and the ability to track small sets of one to four individuals) are innate or early-developing and are shared widely among species. On the other hand, only humans create representations of exact, large numbers such as 42, as distinct from both 41 and 43. These representations seem to be constructed slowly, over a period of months or years during early childhood. The task for researchers is to distinguish the innate representational resources from those that are constructed, and to characterize the construction process. Bayesian approaches can be useful to this project in at least three ways: (1) As a way to analyze data, which may have distinct advantages over more traditional methods (e.g. making it possible to find support for a null hypothesis); (2) as a way of modeling children's performance on specific tasks: Peculiarities of the task are captured as a prior; the child's knowledge is captured in the way the prior is updated; and behavior is captured as a posterior distribution; and (3) as a way of modeling learning itself, by providing a formal account of how learners might choose among alternative hypotheses.

1. THE PROBLEM

1.1. Exact-Number Concepts

This chapter is concerned with how children learn concepts for exact numbers, especially numbers above four. Other writing on this topic has used the terms "natural numbers" or "positive integers," both of which are also correct. The natural numbers are the "counting numbers"—one, two, three, . . . and so on. They are a subset of the whole numbers (which are comprised of the natural numbers and zero), which in turn are a subset of the integers (the whole numbers plus negative numbers, excluding fractions and decimals), which are a subset of the rational numbers (i.e. anything that can be expressed as a ratio of two integers), which are a subset of the real numbers (i.e. anything that can be plotted a number line, including all rational numbers, plus nonterminating, nonrepeating decimals such as π and the square root of 2).

We use the term "exact numbers" for a few reasons. First, the term "natural number" is occasionally (and mistakenly) taken to mean that these concepts are "natural" in the sense of being innate, unlearned, or shared with other species. Not so. Exact numbers such as 42 (and even those as low as five and six) are not "natural" in that sense. They are constructed during childhood, based on cultural input. And as far as we know, they are unique to humans (really large number concepts, such as the concept 2014, certainly seem unique to humans).

The terms "natural number" and "positive integer" may also leave some readers confused about what, exactly, we think children know. Adults with

Author's personal copy

some mathematical training may have an explicit concept "natural number," which includes beliefs such as, "The natural numbers start at one and go on forever; there is no highest natural number" or, "Adding or multiplying any two natural numbers together produces another natural number." But we want to be clear in this chapter that we are *not* claiming that children have such explicit knowledge about "natural numbers" or, "positive integers" as mathematical objects.

Instead, we are interested in children's ability to represent exact numerical quantities of five or more. How can a child represent the information that there are, for example, 8 blocks in a tower, 12 friends on the playground, or 24 cookies in the oven? That is, how are the words "eight," "twelve" and "twenty-four" understood by the child?

Children do represent and reason about numbers long before they understand the formal properties of the natural numbers or the positive integers in an explicit, mathematical sense. Consider the following quotations, both from the same child (the first author's 6-year-old son, JS). The first quotation demonstrates that JS represents at least some natural numbers. The second demonstrates his confusion about countable infinity, which is a property of the natural numbers.

(1) JS (age 5 years, 11 months): "If you have a thousand dollars and you lose a hundred, that's the same as if you have ten dollars and you lose one."

(2) JS (age 6 years, 6 months): "Luca said that a googolplex is the highest number, but he was wrong. There is no highest number."

BWS: "Yes, because you can always add one to any number and get a higher one."

JS: "Until infinity."

BWS: "Yes."

JS: "And after infinity, it starts from the highest negative number, and counts back."

BWS: "It does what?"

JS: "It counts all the way back to zero. It's a big loop. And infinity and zero are the ends."

As these examples show, a child can represent natural-number concepts without explicitly representing the formal properties of the natural numbers as a set. To avoid giving the impression that we are talking about the latter, meta-numerical type of knowledge, this chapter uses the terms exact numbers and "exact-number concepts" for the mental representations of exact numerical quantities such as five, six, seven, eight, and higher natural numbers.

Author's personal copy

1.2. What Makes the Acquisition of Exact-Number Concepts Interesting?

Are numbers a cultural invention? It seems indisputable that at least some are. Take the number π, for example. No one knows anything about π until they hear about it from someone else, and there are plenty of people in the world who never acquire a concept of π at all. (Of course, whoever originally formulated the concept π was an exception to this statement, but it is true for all the rest of us.)

On the other hand, research over the past 40 years has shown that other types of numerical concepts are not cultural inventions, but are the outputs of cognitive systems that have evolved through natural selection. Most obviously, the approximate number system (often abbreviated *ANS*) allows humans and other animals to represent approximate numerical quantities of at least up to several hundred.

Separately from the ANS, humans and other animals also have the ability to create mental models for small sets of up to three or four individuals. This ability is sometimes called parallel individuation. As that name suggests, this is not a number system, but a system for identifying and tracking individuals (which may be objects, noises, actions, etc.) It is not a number system because it does not include any symbol for the *number* of objects in the set. Instead, it maintains a separate symbol for each individual being tracked. Number is represented only implicitly.

What makes exact-number concepts interesting is that a number like 42 cannot be represented by either of these innate systems. The ANS is only approximate, and parallel individuation only works for up to three or four items. So, how can numbers like "exactly 42" be represented? The answer is that the representational system supporting the concept "exactly 42" is constructed over a period of months or years during early childhood.

This is why any plausible account of the origins of exact, large number concepts must be both rationalist and constructivist. It must be rationalist in specifying the role played by those innate systems that represent some numerical content, and it must be constructivist in explaining how we go beyond those innate systems. Following Carey (2009), we will argue that exact-number concepts are a cultural invention, which must be rediscovered/reconstructed by each individual child during development, based on cultural input. The exact-number system, once acquired, has vastly more representational power than the innate systems, and forms the basis for all later-acquired number concepts (e.g. negative numbers, rational numbers, real numbers, etc.).

Author's personal copy

Finally, we take up the question of how Bayesian inference can be useful to this project, and we discuss three ways that it has already been used (Table 9.1). First, Bayesian methods can be used to analyze data. Depending on the data set and the question being addressed, these methods may have advantages over more traditional, frequentist methods. In the example we discuss, Bayesian inference makes it possible to find positive support for the null hypothesis, rather than simply rejecting or failing to reject the null. This can sometimes be a distinct advantage. Second, Bayesian methods can be used to model subjects' performance on specific tasks. In this case, peculiarities of the task are captured as a prior; the subject's knowledge is captured in the way the prior is updated; and the subject's observed behavior is captured in the posterior distribution. Third, Bayesian methods can be used to model learning itself. In this case, Bayes provides a formal account of how learners might choose among alternative hypotheses.

Table 9.1 Three ways of using Bayesian inference in this research

Application	Prior	Evidence	Posterior
Bayesian data analysis (to understand data)	Prior belief about population parameters (means, standard deviations, etc.)	Data collected in an experiment	Updated belief about what the population parameters are likely to be
Bayesian task modeling (to understand a task)	Contaminant influences on the child's behavior: task demands, pragmatics, order effects, etc.	The child's knowledge and/or perceptions	Probabilistic description of how a child will behave in the task with a given state of knowledge and/or perceptions
Bayesian concept-creation modeling (to understand how a concept could be acquired)	Prior preferences in a space of possible truths about the world	Typical input that a child would receive	Inferences about the world that the child is likely to make (i.e. knowledge that the child develops)

Author's personal copy

2. WHY ANY REASONABLE ACCOUNT OF THESE PHENOMENA MUST BE RATIONALIST

2.1. The Innate, Approximate Number System

Any effort to understand human numerical cognition must begin with the ANS. Readers who are already familiar with the ANS should feel free to skip the following section, which provides a brief description of the ANS in nonhuman animals, human infants, and adults.

The ANS is cognitive system that yields a mental representation of the approximate number of individuals in a set (e.g. Feigenson, Dehaene, & Spelke, 2004). Number is represented by a physical magnitude in the brain, and this magnitude is proportional to the actual number of individuals perceived. For example, if a person sees sets of 20 and 40 items, the neural magnitude for the set of 40 will be about twice as large as the neural magnitude for the set of 20 (e.g. Nieder & Miller, 2003). For this reason, representations of number in the ANS are often called "analog-magnitude" representations.

A key signature of the ANS (across development and across species) is that the discriminability of any two set sizes is a function of the ratio between them (for review, see Carey, 2009). It is equally as difficult to tell 8 from 16 items as it is to tell 16 from 32 items, or 50 from 100, or 80 from 160, because all of these cases compare sets with a ratio of 1:2.

Note that the discriminability of set sizes is not determined by their absolute difference. The comparison 8 versus 16 has a ratio of 1:2 and an absolute difference of 8. This is of the same difficulty as the comparison 80 versus 160, because the ratio in the latter case is still 1:2, even though the absolute difference (80) is ten times greater. On the other hand, discriminating 8 versus 16 is much easier than discriminating 152 versus 160. In this case, the absolute differences are both 8, but the ratio in the second comparison is much smaller (approximately 1:1.05).

This property of ratio dependence gives rise to two effects often mentioned in the literature. The first is the magnitude effect, which says that if the absolute difference between two set sizes is held constant, lower numbers are easier to discriminate than higher ones. For example, 5 and 10 are easier to tell apart than 105 and 110. The second is the distance effect, which says that if you are comparing two numbers to the same target, the one that is farther away from the target should be easier to discriminate from it. For example, it is easier to tell the difference between 10 and 15 than between 10 and 11. Both of these effects reflect the fact that discriminability in the ANS is a function of ratio.

Author's personal copy

2.1.1. Approximate Number Representation in Nonhuman Animals

The mental representation of approximate numbers is widespread among species. In a landmark study by Platt and Johnson (1971), rats were trained to press a bar some number of times in order to receive a food reward. After the rat had pressed the bar the required number of times, a food pellet would appear in the feeder. The rat had to leave the bar and run over to the feeder to find out whether the food was there. If the rat stopped pressing too soon, no reward would be in the feeder. If the rat pressed the bar more than the required number of times, the reward would be there, but the rat would have wasted some effort by pushing the bar more times than necessary. In this way, rats were motivated to press the bar just the number of times needed to get the food.

Different groups of rats were trained on different numbers. For example, one group was trained to press the bar 4 times, another group was trained to press it 8 times, and still other groups of rats were trained on the numbers 16 and 24. The results were clear. Each group of rats learned to press the bar as many times as needed. The mean number of presses in each group was actually one to two presses higher than the number trained, reflecting the fact that the rats were a little bit conservative. (Better to press the bar an extra time or two than to risk an empty feeder.)

In Platt and Johnson's (1971) study, number was correlated with other variables. Other studies deconfounded those variables and showed that rats do actually represent the *number* of presses and not just the total amount of energy expended in pressing the bar, or the total time spent pressing the bar. In a creative and early example, Mechner and Guevrekian (1962) showed that depriving rats of water makes them respond faster and with greater energy, but does not make them change the number of presses. To do this, the rat must represent number separately from duration and energy expenditure (see Meck & Church, 1983, for a related finding).

Note that rats do not perform perfectly. They do not press the bar *exactly* the right number of times on every trial. And the distribution of their errors is an important clue that the system they are using is the same ANS found in humans and other animals. The errors reveal scalar variability—a key signature of the ANS. Formally, scalar variability means that the ratio of the standard deviation of the subject's estimates to the mean of those estimates is a constant. In the case of the ANS, the mean estimate is equal to the target number (i.e., the number that the subject is trying to guess), so the spread of errors around each target number is a fixed proportion of the target number itself (see Fig. 9.1). This proportion differs for different subjects. The smaller it is, the more accurate the subject's estimation ability.

Author's personal copy

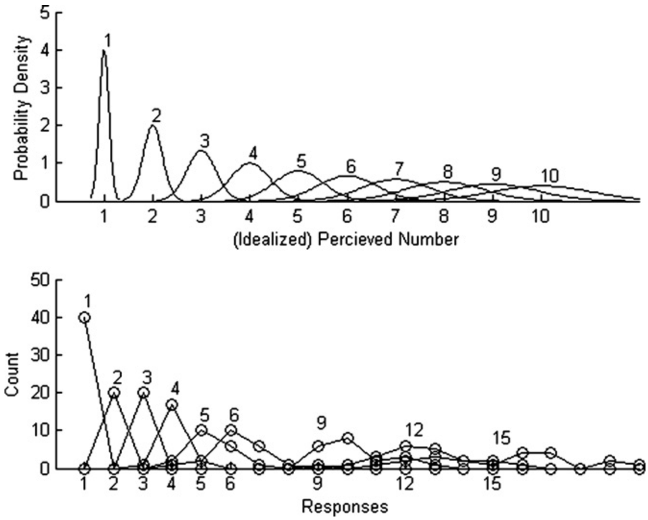


Figure 9.1 On top is an idealized set of distributions with scalar variability. The line with the "1" above it is the distribution of perceptions when a participant is shown 1 item, the line with a "2" is a response curve to 2 items, and so on. Note that the mean is equal to the correct number and the standard deviation increases linearly with the mean. On the bottom are some actual responses taken from adults (Negen & Sarnecka, under review). Participants were asked to tap a space bar 1, 2, 3, 4, 5, 6, 9, 12, and 15 times. Again, the lines are labeled with the correct responses. Though the real data are much noisier, one can still see the mean approximately matches the correct number and the standard deviation increases with the mean.

Looking at Fig. 9.1, we can see that errors on ANS estimation tasks follow a predictable pattern: The mean of guesses for every target number is the same as the target number itself because errors fall symmetrically above and below the target. In the case of Platt and Johnson's (1971), rats, the mean fell slightly above the target because task incentivized conservative behavior, as mentioned above. But in studies without such a reward structure, the mean of estimates is typically equal to the target. Also, errors close to the target are more frequent than errors far away from it. For example, if the target number is 32, then 31 will be a more common error than 21.

These signatures—the symmetrical distributions of estimates, the mean of estimates for each target being the same as the target itself, and the scalar variability in the estimates—are characteristic of the ANS. These same signatures, have been found in the numerical cognition of a variety of other species including crows, pigeons, monkeys, apes, and dolphins (Dehaene, 1997; Gallistel, 1990).

Author's personal copy

2.1.2. Approximate Number Representation in Human Infants

Given that this very useful cognitive system is widely shared among vertebrate species, it is not surprising that humans also have it. Researchers from several different laboratories have shown that preverbal human infants represent approximate numerical quantities using the ANS system (e.g., Brannon, Abbot, & Lutz, 2004; McCrink & Wynn, 2004; Xu & Spelke, 2000).

Rather than making infants press bars for food rewards, these studies rely on infants' tendency to get bored, or "habituate" when they see the same thing over and over again. In these studies, infants are shown different sets of a particular number, over and over until they get bored. The infants' boredom is measured by how long they look at the display. For example, many studies use a criterion of half the initial looking time. This means that if the infant initially looks at the display for 6 seconds, the researchers keep showing the same display until the infant only looks at it for 3 seconds (or less). At that point, the test trials are begun. If the infant looks at the new (test) display significantly longer than 3 s, the researchers conclude that the infant noticed some difference between the old, habituation displays and the new, test display.

In one important study of infant number representation, Xu and Spelke (2000) habituated one group of 6-month-old infants to displays containing 8 dots and another group of infants to displays containing 16 dots. Xu and Spelke made sure to test infants' representation of number rather than other correlated variables (e.g. the sizes of individual dots, the total area covered by each display, the total summed perimeter length of the dots, etc.).

Results showed that infants do represent number. Those who were habituated to the 8-dot displays recovered interest when they were shown a 16-dot display; those who were habituated to the 16-dot displays recovered interest when they were shown an 8-dot display. Later studies showed that 6-month-old infants also discriminate 16 from 32 dots and 4 from 8 dots (Xu, 2003; Xu, Spelke, & Goddard, 2005; see also Lipton & Spelke, 2004).

Supporting the idea that infants were using the ANS, infants' success was a function of the ratio between the two set sizes. Six-month-old infants succeed at a 1:2 ratio (4 vs. 8, 8 vs. 16, or 16 vs. 32 dots), but they fail at a ratio of 2:3 (4 vs. 6, 8 vs. 12, or 16 vs. 24 dots). By 9 months of age, infants succeed at the 2:3 ratios but fail at 3:4 (e.g. they fail to discriminate 6 vs. 8, 12 vs. 16, or 24

Author's personal copy

vs. 32 dots). Thus, human infants form analog representations of approximate numbers long before they learn anything about counting or number words.

2.1.3. Exact-Number Concepts are Connected to the Approximate Number System

The fact that approximate number representation is innate in humans does not necessarily mean that it underlies the acquisition of exact-number concepts. But there is evidence from both adults and children to suggest that exact-number concepts, once they are acquired, are mapped onto analog representations in the ANS.

For example, many studies (e.g. Moyer & Landauer, 1967) have shown adults pairs of written Arabic numerals (e.g. 7 and 9) and asked them to indicate which was numerically greater. Participants' responses show both distance and magnitude effects in terms of reaction times (and sometimes in terms of error rates, though performance is often at ceiling in terms of accuracy). In other words, it takes people longer to judge that $8 > 7$ than that $7 > 6$ (magnitude effect). It also takes people longer to judge that $7 > 6$ than that $8 > 6$ (distance effect).

Recent studies have tried to specify the kinds of mappings that exist between ANS representations and exact-number words in adults (Izard & Dehaene, 2008; Sullivan & Barner, 2010). In one such study, Jessica Sullivan and David Barner asked adult participants to estimate (by saying a number word) the number of dots in an array. The arrays were too large for parallel individuation and were shown too fast for verbal counting, forcing participants to rely on the ANS. Results suggested that for relatively low number words (up to 30 or so), adults seemed to have a direct, individual ANS mapping for each number word. (That is, people have an ANS-based estimate of how many "twelve" is, how many "twenty-one" is, and so on, up to about 30.) This was indicated by the fact that estimates for numbers below 30 were not biased when participants were given misleading information about the range of set sizes used in the experiment. On the other hand, estimates for larger numbers were biased by this type of information. For example, if participants were told to expect arrays of up to 750 dots, when in fact the most numerous array shown had only 350, participants' estimates of numerosity were systematically biased upward, but only for arrays of more than about 30 dots. This suggests that words for numbers higher than about 30 are mapped to the ANS, but they are mapped in terms of an overall structure. That is, people know the order of the number words, and they expect later numbers to be mapped onto larger ANS magnitudes, but they

Author's personal copy

don't maintain a separate representation of the magnitude of each individual number.

Recent evidence suggests that an ANS-to-number-word mapping emerges soon after children have learned the first few exact numbers. We (Negen & Sarnecka, under review) showed 3- and 4-year-old children cards with pictures of one to four items. The children's task was to say how many pictures were on the card. (These children did not yet understand how to use counting to solve the problem.) Using Bayesian methods of data analysis (discussed in more detail in Section 5), we found evidence that children's answers were drawn from an underlying distribution in which variability was scalar. In other words, we found evidence for a mapping between ANS representations and the number words "one" through "four". Later (about 6 months after acquiring all of the counting principles), children learn to extend this mapping out to all of the number words they know (Le Corre & Carey, 2007). By around 5 to 6 years old, children map number words to ANS magnitudes much as adults do.

Thus, in both adults and children, exact-number concepts are mapped to magnitude representations in the ANS. For this reason, any plausible theory of how exact-number concepts are acquired must be somewhat rationalist. It must at least recognize that exact numbers are mapped to the ANS ideally, it would offer an account of how and when ANS representations (which are, by definition, approximate) become connected to representations of exact numbers.

2.2. Another Innate System Relevant to Exact-Number Concepts

Infants also create and maintain working-memory models of small sets of individuals (objects, sounds, or events). Up to three individuals at a time can be represented this way. For example, when infants are habituated to displays of two objects, they recover attention when shown one or three (Antell & Keating, 1983; Bijeljac-Babic, Bertoncini, & Mehler, 1991; Feigenson, 2005; Starkey & Cooper, 1980; Wood & Spelke, 2005; Wynn, 1992a, 1996). Unlike the ANS, this is not a system that represents number *per se*. It is a system that represents individuals. The system does not include any symbols for "two" or "three," but instead maintains a separate symbol for each individual. Other information about the individual (such as its type and properties) can also be bound to these symbols. Thus, whereas the ANS system could represent the content "approximately 10," the parallel

Author's personal copy

individuation system could represent the content, "big duck, little car, little doll."

Because it maintains a separate symbol for each individual, only a limited number of individuals can be represented at any one time. For infants, constraints on attention and working memory seem to place this limit at three (e.g. Feigenson, Carey, & Hauser, 2002). That is, up to three individuals can be tracked at once. If four or more individuals are presented, the infant is not able to track them. When this constraint is exceeded, performance falls to chance in many tasks. For example, Feigenson et al. (2002) found that children could correctly discriminate that $2 < 3$ but not $2 < 4$. This is especially odd because if they simply failed to represent the fourth item in the set of 4, the $2 < 4$ problem would reduce to $2 < 3$, which they pass. Presenting more than three items seems to make the system simply "shut off" and fail to produce any useful output.

Although it does not include any symbols for numbers, this parallel individuation system is relevant to exact-number concepts for several reasons. First, exact-number concepts require some notion of *individual* or *one*, and this notion comes from the attention and memory mechanisms that identify and track individuals. (No representation of *exactly one* exists in the ANS, where number is always approximate.)

Second, the parallel individuation system supports judgments of numerical identity: Is this object or sound or event *the same one*, or *a different one* than the object/sound/event that came before it? Number concepts require that different individuals can be identified as such because number is a property of sets, which are comprised of separate individuals. Without criteria for individuation and some representation of the separateness of different individuals (i.e. some criteria for determining numerical identity), no numerical content could be represented at all.

Finally, parallel individuation supports at least some rudimentary "chunking" of individuals into sets. As described above, infants generally fail to track sets of more than three items. For example, if an infant sees one, two, or three toy cars placed inside a box and is allowed to reach inside the box to retrieve the toys, the infant's search behavior shows that the infant remembers whether there are one, two, or three items in the box. However, if four items (e.g. four cars) are placed in the box, the infant searches no longer than if only one car had been placed there, indicating that the infant can represent the information "car, car, car" but fails to represent "car, car, car, car." This is the set-size limit of three individuals, described above.

Author's personal copy

However, recent work by Feigenson and Halberda (2008) shows that if infants are shown two shoes and two cars going into the box, they will search for all four items—something about having two different types of item helps the infants to create two representations of two individuals each, rather than directly representing all four items. These abilities—to identify and track individuals, and to create higher-order, "chunked" representations of sets of individuals—are required for the representation of exact, large numbers.

3. CAREY'S RATIONALIST, CONSTRUCTIVIST ACCOUNT OF HOW EXACT-NUMBER CONCEPTS ARE ACQUIRED

3.1. Why Any Reasonable Account of these Phenomena must be Constructivist

Given these innate capacities, the reader might wonder whether it makes sense to talk about the "construction" of exact-number concepts at all. Is it likely that exact-number concepts themselves are innate? The answer is no, it is not likely, for the simple reason that none of the innate capacities are up to the job of representing exact, large numbers.

The system of exact numbers has enormous representational power. Using exact numbers, we can represent very large numbers, very precisely. On the day of this writing, J.S. (the 6-year-old mentioned in the anecdotes above) was heard complaining that, "The American flag is super hard to draw," because it has 50 stars, 13 stripes *"and the blue!"* He added in an irate tone that this totaled "*64 things to draw!*" and went on to express his sincere admiration for the Japanese flag.

The concepts of exactly 50, exactly 13, and exactly 64 simply cannot be formulated over ANS representations, which have approximate, real-number values, rather than exact, natural-number values. In other words, there is no way to represent "64 things" as distinct from 63.9 things, or 64.5 things, in the ANS. On the other hand, the parallel individuation system has no explicit representation of number at all. Number is represented only implicitly because there is a symbol maintained for each individual in the set. Because more attentional resources are required to represent each additional individual, the number of individuals that can be represented is strictly limited to three or four. Thus, parallel individuation cannot, on its own, support the representation of exact large numbers. This is why a theory of the construction of exact-number concepts is needed.

Author's personal copy

3.2. Conceptual-Role Bootstrapping

Carey has put forward an in-depth proposal for how exact-number concepts could be constructed, through a uniquely human kind of learning called conceptual-role bootstrapping (Carey, 2009; see also Block, 1986; Quine, 1960). Conceptual-role bootstrapping is not to be confused with semantic bootstrapping or syntactic bootstrapping, both of which are ways of solving mapping problems in the domain of word learning.

Conceptual-role bootstrapping is a way of solving a different problem—the problem of how to construct a representational system (for a given domain of knowledge) that is discontinuous with the representational system that the learner had before. "Discontinuous" means that the content of the new conceptual system cannot be formulated over the vocabulary of the old conceptual system. In practice, this may happen for either of two reasons: (1) The new system is incommensurable with the old one, in the sense that most or all of the old conceptual framework must be discarded to make way for the new one, or (2) The new system has massively greater representational power than the old one. Exact-number concepts fall into this second category; they are not incommensurable with the antecedent representations of the ANS and parallel individuation, but they have massively greater (not just incrementally greater) representational power.

3.3. Bootstrapping Exact-Number Concepts

Episodes of conceptual-role bootstrapping happen as follows. The learner first acquires a placeholder structure—a set of symbols that are structured (i.e. they have some fixed relation to each other) but are not initially defined in terms of the learner's existing vocabulary of concepts. In this case, the set of placeholder symbols is the list of counting words and the order of the list is its structure. Importantly, the words are not (and cannot be) initially defined in terms of the learner's existing vocabulary of number concepts, which include only approximate representations of number from the ANS. (Recall that the parallel individuation system contains no explicit representation of number at all.) Thus, the first step in bootstrapping exact-number concepts is to learn the placeholder structure—the list of number words and the pointing gestures that are deployed along with it. But these words and gestures are initially just placeholders, devoid of exact-number content.

Over a period of many months (often more than a year), the child gradually fills in these placeholder symbols (the counting words and gestures) with meaning. For example, children must learn that number words are

Author's personal copy

about *quantity* (Sarnecka & Gelman, 2004); that they are specifically about *discontinuous* quantity (i.e. discrete individuals such as blocks, rather than continuous substances such as water, Slusser, Ditta, & Sarnecka, under review; Slusser & Sarnecka, 2011a); and that numerosity (as opposed to, e.g. total spatial extent) is the relevant quantitative dimension (Slusser & Sarnecka, 2011b).

Of course, children must also learn the exact meaning of each number word, and how they do this is very interesting. Recall that parallel individuation supports mental models of one to three individuals and that children can under some circumstances "chunk" individuals into nested representations. In order to learn the meaning of the word "one," the child must create a summary symbol for states of the nervous system when exactly one individual is being tracked (Le Corre & Carey, 2007).

The role of the ANS in this process is a matter of some debate. On the one hand, the ANS contains no representation (not even an implicit representation) of exactly 1. On the other hand, the ANS does contain summary symbols for numerosities, and 1 is discriminable from 2 in this system, even if the representations of 1 and 2 are real-number approximations rather than natural numbers. Furthermore, recent evidence suggests that even those children who know only a few number words (e.g. "one," and "two") do have ANS representations defined for those number words (Negen & Sarnecka, under review). All of which suggests that ANS representations are somehow recruited even in the early stages of exact-number-concept construction.

Children take a long time to learn the meanings of number words. Their progression is most clearly illustrated by the changes in their performance on the Give-N task (Wynn, 1992b). In this task, the child is given a set of objects (e.g. a bowl of 15 small plastic bananas) and is asked to give a certain number of them to a puppet. For example, the child might be asked to "Give *five* bananas to the lion." The somewhat surprising finding is that many young children who count perfectly well (i.e., they recite the counting list correctly while pointing to one object at a time) are unable to give the right number of bananas to the lion. Instead of counting to determine the right set size, they just grab one banana, or a handful, or they give the lion all the bananas. Even when children are explicitly told to count the items, they do not use their counting to create a set of the requested size (Le Corre, Van de Walle, Brannon, & Carey, 2006).

Author's personal copy

Studies using the Give-N task have shown that children move through a predictable series of performance levels, often called number-knower levels (e.g. Condry & Spelke, 2008; Le Corre & Carey, 2007; Le Corre et al., 2006; Lee & Sarnecka 2010, 2011; Negen & Sarnecka, in press; Sarnecka & Gelman, 2004; Sarnecka & Lee, 2009; Slusser & Sarnecka, 2011a, 2011b; Wynn, 1990). These number-knower levels are found not only in child speakers of English but also in Japanese and Russian (Sarnecka et al., 2007).

The number-knower levels are as follows. At the earliest (i.e. the "pre-number-knower") level, the child makes no distinctions among the meanings of different number words. On the Give-N task, pre-number knowers might always give one object or might always give a handful, but the number given is unrelated to the number requested. At the next level (called the "one-knower" level), the child knows that "one" means 1. On the Give-N task, this child gives exactly 1 object when asked for one and gives 2 or more objects when asked for any other number. After this comes the "two-knower" level, when the child knows that "two" means 2. Two knowers give 1 object when asked for "one" and 2 objects when asked for "two," but they do not reliably produce the right answers for any higher number words. The two-knower level is followed by a "three-knower" and then a "four-knower" level.

After the "four-knower" level, however, it is no longer possible to learn the meanings of larger number words (five, six, seven, etc.) in the same way as the small numbers have been learned. This is because the innate systems of number representation do not support the mental representation of 5 in the way that they support the representation of 1 through 4. Specifically, parallel individuation does not allow for the tracking of five individuals at a time, and the difference between 5 and 6 is not easily discriminable by young children via the ANS.

Thus, the meaning of "five" must be learned differently from how the meanings of "one" through "four" were learned. Carey's proposal is that children learn the meanings "five" and all higher numbers when they induce the cardinal principle of counting (Gelman & Gallistel, 1978; Schaeffer, Eggleston, & Scott, 1974). The cardinal principle of counting makes the cardinal meaning of every number word dependent on its ordinal position in the counting list. (In other words, the cardinal principle guarantees that for every list of counting symbols, the fifth symbol must mean 5, 13th symbol must mean 13, the 64th symbol must mean 64, etc.) At this point the meaning of the ordered list of placeholder symbols (i.e. the counting words) becomes clear.

Author's personal copy

To understand the cardinal principle, is to understand the logic of exact numbers. This requires an implicit understanding of succession (the idea that each number is formed by adding one to the number before it) and of equinumerosity (the idea that every set of numerosity N can be put into one-to-one correspondence with any other set of numerosity N; Izard, Pica, Spelke, & Dehaene, 2008). Supporting this idea, recent empirical studies find that children who understand the cardinal principle of counting (as measured by the Give-N task) do indeed show an implicit understanding of succession and equinumerosity as well. Only cardinal-principle knowers know that adding one item to a set means moving one word forward in the counting list; whereas adding two items to a set means moving two words forward in the list (Sarnecka & Carey, 2008). Similarly, only cardinal-principle knowers show a robust understanding that two sets with perfect 1-to-1 correspondence *must* be labeled by the same number word, whereas two sets *without* 1-to-1 correspondence must be labeled by different number words (Sarnecka & Wright, in press).

4. THREE WAYS THAT BAYES CAN HELP WITH THIS PROJECT

Research exploring the development of exact-number concepts can make use of Bayesian inference in different ways. Here, we discuss three of them: Bayesian data analysis, Bayesian task modeling, and Bayesian concept-creation modeling. What all of these Bayesian approaches have in common is that they involve some sort of prior information, which is weighted against some form of evidence to form a posterior distribution. The approaches differ in the kinds of information captured by the prior, the evidence, and the posterior.

4.1. Using Bayes to Analyze Data (Agnostic Bayesianism)

The first approach is *Bayesian data analysis*. Of the three, this is the one supported by the largest statistical literature (e.g. Gelman, Carlin, Stern, & Rubin, 1995; 2003). It requires no theoretical commitments about the developing mind, because Bayesian methods are used only to analyze data. This is sometimes called *Agnostic Bayes* (e.g. Jones & Love, 2011) because it does not require any commitment to the idea that the mind itself makes inferences in a Bayesian way.

Author's personal copy

In Bayesian data analysis, priors are formulated over things that researchers want to estimate in a data set, such as the rate of correct responses to a question or the effect size of a given between-group difference. The evidence is the set of sample observations. The posterior is an updated belief about the population.

For example, imagine that we have a sample of 24 children, and we ask each of them the same question about whales: Is a whale a mammal, or a fish? We will accept only two possible responses—"mammal" (the correct answer) or "fish" (the incorrect answer). In this group of children, 22 answer "mammal" and the other 2 answer "fish". We want to estimate how many children in the population of interest will say that whales are mammals. To do this, we first set a prior, saying that the set of responses can equally be anything from 0% correct to 100% correct. This distribution is known as a flat prior or Beta(1,1).

Our data are the 22 "mammal" and 2 "fish" responses that we collected from the children. These data combine with the prior to form the posterior, Beta(23,3). This posterior is a probability distribution for the rate of "whales-are-mammals" responses in the population of interest, given our data and the prior, and assuming that our sample was randomly drawn from that population. This posterior distribution is shown in Fig. 9.2.

This posterior allows us to interpret the data without calculating a p-value. The probability density around a 50% correct-response rate in the population (i.e. the probability that the children in the population have no knowledge of what whales are, and that they children in our sample

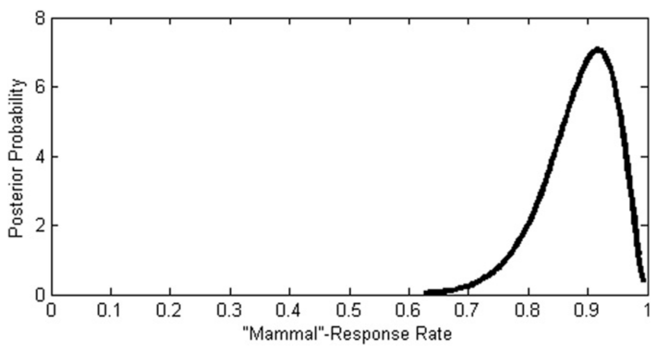


Figure 9.2 The posterior distribution over rates of "whales-are-mammals" responses in the population, given a flat prior and 22/24 "yes" answers in our data set. Intuitively, it should make sense that the highest probability is at 22/24 and very little probability exists below about 70% correct.

Author's personal copy

answered randomly) is extremely low (less than 0.001) compared to the peak. From this, it is already reasonable to infer that the data do not reflect chance responding. Furthermore, a 95% credible interval stretches from about 74% to 97% on this posterior distribution, meaning that we can conclude with 95% certainty that somewhere between 74% and 97% of children in the population of interest actually will say that whales are mammals. Further reinforcing this, we can calculate a Bayes factor, which is an expression of preference for one hypothesis over another. In this case, the alternative hypothesis (the rate could be anything from 0% to 100%) is preferred over the null (a correct-response rate of 50%, indicating chance responding) by a factor of 2431. This is considered extremely strong evidence. Space precludes a review of all of the ways that posteriors can be formed and interpreted (for review, see A. Gelman et al., 1995; 2003). But this basic approach—forming priors, calculating posteriors, looking at confidence intervals, and calculating Bayes factors—can be applied in virtually any case where researchers would otherwise use t-tests, analyses of variance (ANOVAs), regressions, and so on.

One clear advantage of this type of data analysis over more traditional methods is that the Bayesian methods make it possible to find positive support for the null hypothesis, rather than simply rejecting or failing to reject it. In classical hypothesis testing (using t-tests, ANOVAs, etc.), a null hypothesis can be rejected if enough evidence is found against it, but evidence can never be found *for* the null.

With a Bayesian approach, it is actually possible to use a prior as the alternative hypothesis in a way that allows either hypothesis (the null or the alternative) to be preferred after the data are taken into account. This is particularly well understood in the case of t-tests, for which there even exists a simple online calculator (Rouder, Speckman, Sun, Morey, & Iverson, 2009). There is also an excel sheet available for approximating this kind of approach for ANOVAs, though it requires researchers to separately calculate sums of squares used in the usual F-tests (Masson, 2011; for technical details, see Dickey and Lientz, 1970).

We know of only one number-concept development study using this type of agnostic Bayesian method (Negen & Sarnecka, under review). The paper asks whether children who know the meanings of only a few number words (e.g. "one," "two," and "three") have already mapped those words to representations in the ANS. Operationally, the question is whether children's responses on a number-word task are drawn from an underlying distribution with scalar variability. (As mentioned above, scalar variability is a key signature of the distribution of ANS representations in the brain.)

Author's personal copy

Previous studies (e.g. Cordes, Gelman, Gallistel, & Whalen, 2001) have only been able to test for the absence of this signature. To test whether the signature could actually be inferred from the data, we calculated a Bayes factor. The null hypothesis (i.e. that variability was scalar) was preferred by a factor of about 14. In other words, it was 14 times more likely that the data came from an underlying distribution with scalar variability than that the data came from an underlying distribution where variability was (linearly) non-scalar. This is very strong evidence for the null hypothesis. In general, this type of analysis is useful in situations where researchers want to present evidence *for* the null hypothesis—for example, to argue that subjects *are* guessing at random, that two means *are* the same, that variables *are* unrelated, and so on.

Because Bayesian data analysis tends to result in the calculation of Bayes factors, it is also useful when several models are being compared frequently. A current example is the debate over logarithmic and linear performance in bounded number-line tasks (e.g. Siegler, Thompson, & Opfer, 2009). Most studies to date have compared linear and logarithmic models by (1) calculating the median response for each child, (2) finding the best fit for the linear and logarithmic models, and (3) counting the number of children fit better by each model.

If one considers only the relatively simple log and linear models, this method seems adequate. However, it would be more formally rigorous to use a Bayes factor. This would also allow for the strength-of-preference to be calculated for each individual child, which may be useful. Finally, a Bayes factor naturally punishes models that make over-broad predictions, so it would allow for rigorous comparisons between the simple log and linear models and also models that have more parameters in them (e.g. Barth, Slusser, Cohen, & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Slusser, Santiago & Barth, under review).

At the moment, the use of Bayesian data analysis is not widespread among developmental scientists. We see at least three reasons why this is so. First, there is very little training available in how to use these methods, although some progress has been made on this front with a few authors posting free, online training books (e.g. Wagenmaker & Lee, in preparation).

Second, because the methods are relatively unfamiliar to reviewers, authors are required to explain the analysis at much greater length than would be needed for traditional, frequentist methods; they must explain both how the analysis was done and why they used Bayesian methods

Author's personal copy

instead of frequentist ones. This turns every paper into something of a statistics tutorial—even when the authors are not interested in convincing anyone else to use Bayesian methods but simply want to present their empirical work. This problem would presumably decrease over time, if the methods were used more widely.

Third, virtually nothing exists in the way of "friendly" software (GUI-based, standardized, professionally supported) to help with the analysis in any but the simplest of cases. (E.g., there is no Bayesian equivalent of SPSS.) This problem does not even seem to be recognized as a problem; most statistical software is still being developed and released in R, which is text based and largely decentralized.

4.2. Using Bayes to Model Subjects' Behavior on a Task

A second way to adopt a Bayesian approach to studying number-concept development is to use *Bayesian task modeling*. This method allows us to separately model the demands of a task and the knowledge state of the subject and to think about how those combine to create the observed behavior.

This method has been used to model children's behavior on the Give-N task (Lee & Sarnecka, 2010, 2011). In this task, the child is asked to produce sets of a certain number (e.g. "Please put *three* bananas on the lion's plate."). The prior captures the base rate of responses for each task. This is roughly how children would respond in the absence of any numerical information. For example, if you could somehow ask for "banana(s)" in a way that did not provide any singular/plural or other cues about how many bananas were wanted.

Lee and Sarnecka (2011) inferred their prior from a large set of Give-N data, by aggregating across all the wrong guesses that children made. The resulting prior is shown in Fig. 9.3 (left panel). Children have a high probability of giving just 1 item, a somewhat lower probability of giving 2, still lower and approximately equal probabilities of giving 3–5, and significantly lower probabilities of giving any number larger than that. However, there is a bump up at 15.

This distribution is intuitively sensible. If children understand that they should give something from the bowl but have no information about how many things they should give, it seems reasonable that they should give one item or a handful of items (each object is about 2 cm in diameter, so children can typically grab two to five objects at once). Nor does it seem surprising (to

Author's personal copy

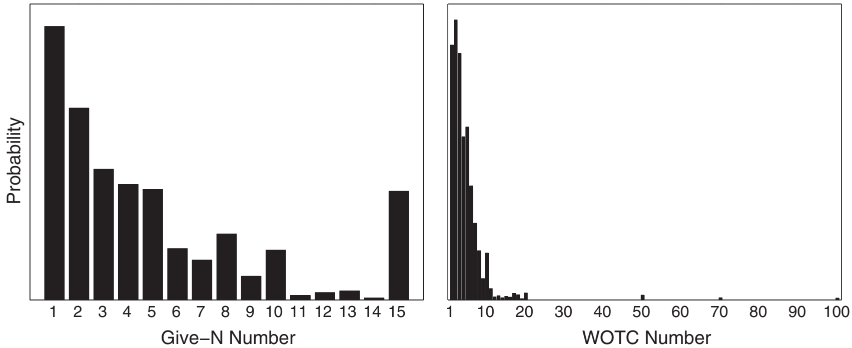


Figure 9.3 Inferred base rates (i.e. priors) for the Give-N task and What's-On-This-Card tasks from Lee & Sarnecka, 2011

anyone who has spent time with preschoolers) that it is relatively common for children to give the entire set of 15 items, either by dumping them all onto the lion's plate at once or by placing one item at a time on the plate until there are none left.

If the child knows the exact meanings of any exact number words (e.g. *one*, *two*, and *three*), this information changes the base rate for that child. A three-knower will usually give the correct number of items when asked for "one," "two," or "three" and will very rarely produce those set sizes when asked for any other number. The intuitive operation of the model is illustrated in Fig. 9.4. The child depicted is a three-knower, reflected by the fact that the numbers 1, 2 and 3 are underlined in the first thought bubble, representing the prior. The prior probability of any given set size being produced (as shown in Fig. 9.3) is represented here by the size of each numeral, with numerals for higher-probability set sizes appearing in larger type.

If the child hears the request, "Give me *two*," the posterior probability (illustrated in the thought bubble on the upper right) is very high for 2 and very low (too small to be pictured) for any other number. In other words, this simplified model predicts that children who are three-knowers will *always* give 2 objects when asked for two.

If the request is, "Give me *five*," then the probabilities for 1, 2, and 3 immediately drop to very near zero. (In the figure, these numerals do not appear in the lower-right thought bubble.) This reflects the fact that the child is a three-knower, and three-knowers know that whatever "five" means, it cannot mean 1, 2, or 3 (Wynn, 1992b). What remains are all the other numbers of objects the child could give, each of which has the same probability (relative to all the alternatives) as it did in the prior. Chances are

Author's personal copy

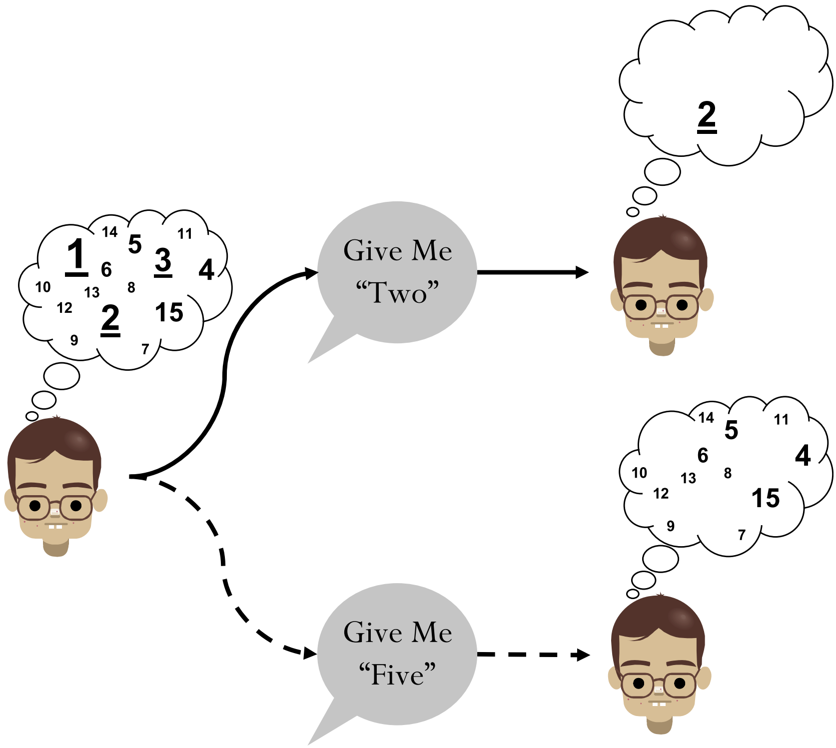


Figure 9.4 Intuitive operation of Lee and Sarnecka's (2011) model, showing a child who is a three-knower responding to instructions to "give two" or to "give five" (Lee & Sarnecka, 2011). For color version of this figure, the reader is referred to the online version of this book.

that the child will produce a set of 4, 5, or 15 in response to this request. Other set sizes are less likely to be produced, as reflected by their smaller numerals.

The key point here is that the base rate has a large impact on the observed performance. It answers the question, for example, why might a child give four items instead of six for a given request, if that child does not know what either "four" or "six" mean? The answer is, because the prior (base-rate) probability of giving four items is higher.

Note that this kind of modeling does not commit the user to the idea that children make any of the calculations involved, either explicitly or implicitly. Formally, this is a computational model (a model of how cognitive parameters and task demands lead to observed behavior) rather than an algorithmic model (a model of exactly how the various cognitive processes

Author's personal copy

are implemented). A Bayesian operation takes a base rate of responding (the prior) and updates it using the child's knowledge (the data) to produce an actual, observed rate of responding (the posterior). This does not imply that the child represents any of these concepts. It is simply a description of how task demands and different states of knowledge combine to form different patterns of behavior in the task.

One way that Bayesian task modeling can be very useful is in allowing researchers to investigate the "psychological reality" of theoretical constructs across tasks. Using this type of modeling, we can assess children on multiple tasks that are believed to tap the same underlying knowledge and then compare performance across tasks, even if the task demands and resulting performance data are very different.

To continue with the earlier example, Lee and Sarnecka (2011) tested children on two tasks, both of which are supposed to reveal the child's number-knower level. One was the Give-N task mentioned above; the other was the What's-On-This-Card task (Gelman, 1993; Le Corre & Carey, 2007; Le Corre et al., 2006). The Give-N task asks children to produce the set corresponding to a given a number word; the What's-On-This-Card task asks children to produce the number word for a given set.

The priors for each task were different because the kinds of behavior possible on each task were different. For example, the What's-On-This-Card prior (Figure 9.3, right panel) did not have a bump up at 15 because the bump up at 15 was an artifact of the Give-N task. It reflected the fact that there were 15 items in the bowl set before the child and that children often dumped out and handed over all the items. On the other hand, any number word a child could think of was a possible response on the What's-On-This-Card task, whereas on the Give-N task, the only possible responses were the numbers 1–15.

Number-knower levels are most often assessed using the Give-N task. But if they are a psychologically "real" phenomenon (rather than an artifact of Give-N task demands), then the number-word knowledge inferred for children on the Give-N and What's-On-This-Card tasks should be the same, despite the different task demands. And indeed, this is what Lee and Sarnecka found for most children. Furthermore, by combining information from the two different tasks, Lee and Sarnecka's (2011) model was able to diagnose the knower levels of many children with a much higher degree of certainty than was possible using the data from either task alone.

In practice, it is often the case that researchers who use Bayesian task modeling will also want to use Bayesian data analysis. Some authors have

Author's personal copy

argued that this is necessary to realize the full potential of the approach (e.g. Kruschke, 2010; Lee, 2010, 2011a, 2011b). Estimating the parameters of a model like the one described above is actually a very hard problem. A widely accepted alternative is to sample from the posterior (rather than attempting to fully describe it) and then examine the samples. This process, known as Markov-Chain Monte-Carlo, is well studied and is implemented in several free software packages (e.g. WinBUGS; Thomas, 1994).

4.3. Using Bayes to Model Learning Itself

The third way to adopt a Bayesian approach in this research is to use *Bayesian concept-creation modeling*, where Bayesian methods are used to model the creation of a new concept. Here, the prior is some set of beliefs in a model learner's virtual mind—some set of bets or preferences about what is probably true in the world. If the model is to be cognitively plausible, these prior beliefs should be ones that could plausibly be attributed to human learners at the outset of the learning episode. In the case of exact-number concept creation, for example, priors should reflect the known limits of the ANS and/or parallel individuation system.

The evidence is the input received by the learner. Here, the requirement for a cognitively plausible model is that the input must match real-world experiences that children actually have. For example, researchers trying to model word learning might base the input on transcripts of natural, child-directed speech.

The posterior is a distribution over various inferences a child could make. The explanatory value of the model depends on this posterior distribution giving most of its mass to inferences that children actually do make. In other words, at the end of the learning episode, the model learner must represent the knowledge in question. For example, a model learner acquiring English count/mass syntax should agree that number words cannot quantify over mass nouns (e.g. **the three furniture*). That is, any rule that accepts **the three furniture* should have low posterior probability.

This approach is particularly helpful for addressing arguments over learnability. Philosophers have famously argued that any set of data can be fit equally well by an infinite space of hypotheses. For instance, the English language could be a subject-verb-object language up until January 1, 2025, and then suddenly switch to being a verb-subject-object language. All the data available at the time of this writing are equally consistent with this 2025-change hypothesis and the alternative, no-change hypothesis. The

Author's personal copy

intuition that this 2025-change hypothesis is silly, counterproductive, needlessly complex, and/or confusing is met with the counterargument that (1) these objections just reflect a bias toward what we already somehow know and (2) there is no formal way to measure complexity.

Rips, Asmuth, and Bloomfield (2006, 2008) have put forward such an argument about the development of exact-number concepts. Specifically, researchers are challenged to explain how young children (who know only some exact numbers) could infer that numbers keep going in a linear progression, rather than following a modular principle, such that at some arbitrary number (e.g. 10), the numbers stop counting up and start over again at 1. (Rips and colleagues point out that some notational systems, such as days of the week, months of the year, and hours of the day, do have a modular structure—so modular systems must in principle be learnable by children.) In other words, the numbers the child knows are like all the years in which English has been a subject-verb-object language. No matter how many there have been, the next one could be different.

Recent work by Piantadosi, Tenenbaum, and Goodman (in press) has shown that a Bayesian model learner can overcome this hurdle and construct exact-number concepts with a linear progression, even when the evidence is theoretically consistent with either a modular or a linear system. The model works by describing various systems for matching number words with meanings as lambda calculi. An example of a one-knower might be

$$\lambda\, S\, .\ (\text{if (singleton? S) “one” “two”}),$$

which outputs "one" if given a set S with 1 item and otherwise outputs "two". The prior favors calculi that (a) are short, (b) use fewer elements of recursion, and (c) re-use primitives. In many ways, this formally captures the intuition of certain systems being "simpler".

Piantadosi and colleagues address the question of how children could infer a linear number system, given the available evidence, rather than a modular number system. The answer is that the model's prior prefers systems that can be described in a shorter calculus, with fewer primitives. (Note that this is a formal definition of "simpler," undermining the claim that there is no way to measure complexity.)

By this definition, linear systems are less complex than modular systems, which require all the machinery of a linear system in order to get from 1 to 10 (or whatever the highest number of the module is), and then additional machinery to tell the user to stop and start again from 1. When fed true-to-life number-word input, Piantadosi's model learner selects the

Author's personal copy

correct (linear) hypothesis because it has the greatest posterior likelihood given the data and the prior, even against the infinite space of other hypotheses.

It is not that Piantadosi's model learner *cannot* learn a modular system—it is just that positive evidence for a modular system must be provided in order to overcome the simpler, linear hypothesis. When the learner is fed modular information (real-world number-word input that has been altered to reflect a modular system), a modular system is indeed what it learns. This is intuitively appealing when one considers the learning trajectories of real children: Children represent at least some exact numbers by about age 4; at that age, most of them have not yet learned the cycles of hours in a day, days in a week, or months in a year. However, they are not fundamentally incapable of learning the cyclical systems and neither is the model learner. It just takes the model longer to learn modular systems ("longer" meaning that it requires more data), which mimics the learning trajectories of real children.

One objection to Piantadosi's model might be that it simply does not consider a very rich space of hypotheses. But the same general method can be used to address a space of hypotheses of any size. For example, the model does not consider that the meaning of number words might change at some specified future date, because the model learner does not have access to date information. However, even if the model were modified to include this information, the date-change hypothesis would be doomed from the outset because it would require an expression such as "if the date is before X" and then two full models of number-word meanings (one for all dates before the change and another one for all later dates). Such a model would be much less likely under the prior (which prefers shorter descriptions with fewer primitives) and thus would not be selected. This argument holds equally true for any other alternative that would encumber the correct system with conditional variance, for which there is no negative evidence.

In very general terms, Bayesian concept-creation modeling provides a way of separately modeling prior biases and observations and for both of these to be interesting, well-specified, research-supported, necessary components of the concept-creation process. This is exciting because a similar approach has been useful in explaining human induction in other areas (e.g. Perfors, Tenenbaum, Griffiths, & Xu, 2011). Indeed, the approach seems so flexible across domains that in some cases, domain-general priors may eventually replace domain-specific constraints.

Thus, Bayesian concept-creation modeling represents a convenient way of formally describing what we know to be true about development: That

Author's personal copy

the rationalists and the constructivists have always *both* been right (at least in part) because both priors and evidence matter. Even if a human infant and a puppy are raised by the same, loving human family, the baby will grow up to speak a human language and the dog will not, because of prior constraints. On the other hand, if a Japanese baby and a French baby are switched at birth, it is the baby raised in Japan who will learn to speak Japanese, and the one raised in France who will learn to speak French, because of the evidence in the environment. By creating models that take both these aspects of development seriously, Bayesian concept-creation modeling allow us to move beyond tiresome debates where each side emphasizes either prior constraints or learning, but no theory seems able to accommodate both.

Finally, it is worth mentioning that for all three types of Bayesian approaches discussed here, it would be possible to take a similarly probabilistic approach that retains much of the power of these models without using Bayes' rule. For instance, the model by Piantadosi and colleagues (in press) could rank hypotheses by some criterion other than posterior probability. One could design a scoring system wherein hypotheses earned points for good fit and desirable calculi, and the hypothesis earning the most points would be the winner. Such a model would likely lead to very similar conclusions but would not technically be Bayesian. The appeal of Bayesian formalisms is that they are already very well studied and well described and are therefore most convenient for researchers to use.

5. SUMMARY

The study of early number concepts is a thriving field that provides many insights into the developing mind. As the search for the origins of numerical thought continues, the future researcher has many options. A complete theory must be somewhat rationalist, because children are genetically endowed with at least some abstract numerical concepts. A complete theory must also be somewhat constructivist, because children clearly move beyond the innate building blocks of number, eventually acquiring much more complicated mathematical constructs such as integers, rational numbers, and so on. Bayesian approaches hold great promise in this area, whether as a way of analyzing data, of modeling subjects' performance on individual tasks or of modeling the creation of number concepts themselves.

Author's personal copy

AUTHOR NOTE

This material is based upon work supported by the National Science Foundation under DRL-0953521 to the first author. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development, 54*, 695–701.

Barth, H., Slusser, E., Cohen, D., & Paladino, A. (2011). A sense of proportion: Commentary on Opfer, Siegler and Young. *Developmental Science, 14*(5), 1205–1206.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*, 711–721. doi:10.1037/0012-1649.29.4.711.

Block, N. (1987). Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy, 10*, 615–678. doi:10.1111/j.1475-4975.1987.tb00558.x.

Brannon, E. M., Abbot, S., & Lutz, D. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition, 93*(2), B59–B68.

Carey, S. (2009). *The origin of concepts*. Boston: MIT Press.

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18*, 331–338. doi:10.3758/s13423-011-0059-z.

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*, 22–38. doi:10.1037/0096-3445.137.1.22.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal and nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review, 8*(4), 698–707.

Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics, 42*, 204–223.

Feigenson, L. (2005). A double dissociation in infants' representation of object arrays. *Cognition, 95*, B37–B48.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*(2), 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Science of the United States of America, 105*, 9926–9930.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: Bradford Books/MIT Press.

Gelman, R. (1993). .A rational-constructivist account of early learning about numbers and objects. In D. Medin (Ed.), *Learning and motivation, Vol. 30* (pp. 61–96). New York: Academic Press.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995; 2003). Bayesian data analysis.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Oxford: Harvard University Press.

Izard, V., & Deheane, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221–1247.

Author's personal copy

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2008). Exact equality and successor function: Two key concepts on the path towards understanding exact numbers. *Philosophical Psychology, 21*, 491–505.

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, 34*, 169–188.

Kruschke, J. K. (2010). Bridging levels of analysis: Comment on McClelland et al. and Griffiths, et al. *Trends in Cognitive Sciences, 14*, 344–345.

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 the counting principles. *Cognitive Psychology, 52*(2), 130–169.

Lee, M. D. (2010). Emergent and structured cognition in Bayesian models: Comment on Griffiths et al. and McClelland, et al. *Trends in Cognitive Sciences, 14*, 345–346.

Lee, M. D. (2011a). How cognitive modeling can benefit from hierarchical bayesian models. *Journal of Mathematical Psychology, 55*, 1–7.

Lee, M. D. (2011b). In praise of ecumenical bayes. *Behavioral and Brain Sciences, 34*(4), 206–207.

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

Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modeling. *Cognition, 120*, 391–402.

Lipton, J. S., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy, 5*(3), 271–290.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*(3), 679–690.

McCrink, K., & Wynn, K. (2004). Large number addition and subtraction by 9-month-old infants. *Psychological Science, 23*(3), 776–781.

Mechner, F., & Guevrekian, K. (1962). Effects of deprivation upon counting and timing in rats. *Journal of the Experimental Analysis of Behavior, 5*, 463–466.

Meck, W. H., & Church, R. M. (1983). *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 320–334. doi:10.1037/0097-7403.9.3.320.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*, 1519–1520.

Negen, J. & Sarnecka, B.W. (in press). Number-concept acquisition and general vocabulary development. *Child Development*

Negen, J., & Sarnecka, B. W. (under review). Early number learning and the approximate number system: Re-visiting the variability argument.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*, 149–157.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L.,, & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*, 302–321.

Piantadosi, S. P., Tenenbaum, J. B., & Goodman, N. D. (in press). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, *123*, 199–217. http://dx.doi.org/10.1016/j.cognition.2011.11.005

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogenous behavior chain: Effects of error contingencies. *Learning and Motivation, 2*(4), 386–414.

Quine, W. V. O. (1960). *Word and Object*. Cambridge: MIT Press.

Rips, L. J., Asmuth, J., & Bloomfield, A. (2006). Giving the boot to the bootstrap: How not to learn the natural numbers. *Cognition, 101*, B51–B60.

Author's personal copy

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences, 31*, 623–642.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16*(2), 225.

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., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*, 662–674. doi:10.1016/j.cognition.2008.05.007.

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. http://dx.doi.org/10.1016/j.cogpsych.2006.09.001.

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

Sarnecka, B. W. & Wright, C. E. (in press). The exact-numbers idea: Children's understanding of cardinality and equinumerosity. *Cognitive Science*.

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*, 357–379.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3*(3), 143–150.

Slusser, E., Ditta, A., & Sarnecka, B. W. (under review). Connecting numbers to discrete quantification: A step in the child's construction of integer concepts.

Slusser, E. B., & Sarnecka, B. W. (2011a, March). Extending representations of discrete and continuous quantities to early number-word learning. Paper given as part of symposium "The development of discrete and continuous quantification from infancy to childhood" (M. Le Corre & P. Cheung, Chairs) at the biennial meeting of the Society for Research in Child Development, Montreal, Canada.

Slusser, E., & Sarnecka, B. W. (2011b). Find the picture of eight turtles: A link between children's counting and their knowledge of number-word semantics. *Journal of Experimental Child Psychology, 110*, 38–51. doi:10.1016/j.jecp.2011.03.006, PMC3105118 [Available 2012/9/1].

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science, 210*(4473), 1033–1035.

Sullivan, J., & Barner, D. (2010). Mapping number words to approximate magnitudes: Associative learning or structure mapping?. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* Austin, TX: Cognitive Science Society.

Thomas, A. (1994). BUGS: A statistical modelling package. *RTA/BCS Modular Languages Newsletter, 2*, 36–38.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. Psychological Science*, 10*, 130–137.

Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: numerical discrimination and its signature limits. *Developmental Science, 8*, 173–181. doi:10.1111/j.1467-7687.2005.00404.x.

Wynn, K. (1996). Infants' individuation and enumeration of actions. Developmental Science*,* 7, 164–169. doi:10.1111/j.1467-9280.1996.tb00350.x.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*(2), 155–193.

Wynn, K. (1992a). Addition and subtraction in human infants. *Nature, 358*(6389), 749–750.

Wynn, K. (1992b). Children's acquisition of number words and the counting system. *Cognitive Psychology, 24*(2), 220–251.

Author's personal copy

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7*, 164–169. doi:10.1111/j.1467-9280.1996.tb00350.x.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89*, B15–B25.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month old infants. *Cognition, 74*(1), B1–B11.

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*(1), 88–101.