Copyright OUP

Chapter 11

Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism

Fei Xu, Kathryn Dewar, & Amy Perfors

11.1 Introduction

A tremendous amount of work has been conducted on infants' representations of physical objects during the last 25 years (see Baillargeon, 2004; Spelke, 1994, for reviews). One particular perceptual dimension, object shape, has received a great deal of attention in both the literature on object representations and generalization (e.g., Baldwin et al., 1993; Rosch et al., 1976; Shutts et al., Chapter 8; Waxman, 1999; Welder & Graham, 2001) and the literature on early word learning, since the initial vocabulary of most infants contains a large number of count nouns that refer to object categories (e.g., Nelson, 1973). This chapter focuses on a case study of how object representations in infants interact with early word learning, in particular the nature of the so-called "shape bias." We will begin with a short review of the controversies in this subfield, and use it to illustrate the two dominant views of cognitive development, which can be roughly classified as nativist or empiricist. We will then present both theoretical arguments and new empirical evidence for a rational constructivist view of cognitive development (see also Xu, 2007a). Our goal is to argue for a new approach to the study of cognitive development, one that is strongly committed to both innate concepts and representations, as well as powerful inductive learning mechanisms. In addition to discussing the shape bias and how it relates to object representations, we will also briefly discuss the generality of the approach.

11.2 Case study: The shape bias in word learning

The debate began with a seminal paper by Landau et al. (1988). They provided the first evidence that 3-year-old children, when taught a new count noun for

Copyright OUP

an object, generalized the word to other objects that shared the same overall shape as the original object, and ignored variations on other perceptual attributes of the objects such as color, texture, or size. Landau et al. (1988) argued that shape is, perhaps, the most important and salient perceptual dimension that underlies children's rapid learning of count nouns that map onto object categories. Furthermore, these researchers suggested that children could learn that words refer to distinct shapes, based on the correlational information in the input between shape and word usage.

A few years later, Soja et al. (1991) reported a series of experiments providing evidence that the ontological distinction between objects and substances, or individuated versus nonindividuated entities, is of great importance in early word meaning. They showed that 2-year-old children, who have not mastered the count and/or mass syntax of English, generalized novel words for objects based on shape similarity whereas they generalized novel words for substances based on material kind. These researchers argued that early words refer to kinds, not just distinct shapes. Although Soja et al. were not explicit about what they thought the status of the shape bias was, one reading of their work is that children assume that words for objects refer to kinds, and that shape is a particularly salient cue (perhaps even innately given) for kind membership.

Numerous follow-up studies have been conducted in the last two decades, and the debate continues at present—the journal *Developmental Science* devoted a special section on the topic recently (see Markson et al., 2008; Samuelson & Bloom, 2008). Both sides of the debate have conducted ingenious experiments to provide empirical evidence for their point of view. However, neither is convinced by the arguments marshaled by the other side.

Three controversies persist in this debate and they remain unresolved. First, is the shape bias innate or learned? Second, do early words (count nouns) refer to distinct shapes or distinct kinds? Third, what is the mechanism for acquiring the shape bias if it is learned?

Nativists and empiricists give different answers to the first question about innateness. If the shape bias is construed as a solution to the induction problem of word meaning (a la Quine, 1960; Soja et al., 1991), a natural conclusion is that it may have to be innate. That is, if the point of positing the shape bias were to eliminate other hypotheses (e.g., count nouns refer to objects with a shared color; count nouns refer to objects with a shared texture; etc.), then it would be a bit odd to say that the bias itself is learned (although see more discussion in later sections of this chapter). Thus, one version of the shape bias, perhaps the strongest version, is to posit that young children have a bias to pay attention to object shape and to generalize count nouns based on shape similarity. Alternatively, the shape bias may be the product of some

Copyright OUP

learning process. One candidate under consideration is that the shape bias is the result of associating object properties with word usage (e.g., Colunga & Smith, 2005; Samuelson, 2005; Smith et al., 2002).

A careful reading of the literature suggests that nativists, who are particularly concerned with positing biases to solve Quine's induction problem, have mostly claimed that young children assume that 'count nouns refer to object kinds' and have not discussed in great detail the status of the shape bias itself (e.g., Markman, 1989; Soja et al., 1991). However, the notion of kinds is quite abstract and it is not clear how a child is supposed to discover it: What are the perceptual dimensions that signal that objects belong to the same kind? We see two solutions: one is to say that some perceptual dimensions are privileged from the beginning, for instance, shape; the other is to say that the shape bias is learned from the input. Nativists seem to be more sympathetic to the latter view (e.g., Macnamara, 1982; Markson et al., 2008; Soja et al., 1991), *yet* there is no learning mechanism proposed in any of their papers!

Nativists and empiricists also give different answers to the second question: Do early words refer to kinds or distinct shapes? The nativist view suggests that children assume that early words refer to distinct kinds and shape as a cue to kind membership (e.g., Diesendruck & Bloom, 2003; Markson et al., 2008; Soja et al., 1991). This view grants the child some notion of kind that is more abstract than just clusters of perceptual properties such as shape, texture, and color. The reason why the shape bias is important is that shape turns out to be a reliable cue for identifying the kinds in a child's environment. In contrast, the empiricist view suggests that compared to other perceptual dimensions of objects such as texture or color, shape correlates better with word usage. Because the child is trying to figure out how to generalize words, he or she notices this correlation between shape and word usage, and concludes that words (count nouns) refer to objects that share a common shape. Most empiricists would get rid of the notion of kind entirely; for them, the notion of a kind may be reduced to clusters of perceptual properties of objects (e.g., Colunga & Smith, 2005).

Lastly, nativisits and empiricists (unsurprisingly) give different answers to the question of learning. If the shape bias is innate and it provides a solution to the problem of induction in word learning, then no learning mechanism is needed. Curiously, most nativists do not claim that the shape bias is innate (see the following for some possible reasons for this). However, this leaves the nativists in an uncomfortable position of not having a learning story at all. In contrast, the empiricists have provided detailed studies and models of how associative mechanisms may be responsible for the emergence of the

Copyright OUP

shape bias in early word leaning (e.g., Colunga & Smith, 2005; Smith et al., 2002).

Here we take issue with both sides. We will argue that neither view is descriptively or explanatorily adequate. We begin with pointing out a few problems with each of the two standard views.

From the nativist prespective, one thorny problem is that as soon as one looks carefully at the early lexicon, one discovers that the count nouns that refer to object categories vary in terms of the salient perceptual dimension along which they are generalized. For example, for the garden-variety artifact objects, shape is a good indicator for kind membership. However, for animal kinds (which are abundant in the early lexicon), shape is not quite as important and both texture and shape matter for generalization. For food kinds (which are also abundant in the early lexicon), color is perhaps the most important perceptual dimension and shape is not particularly important or salient (e.g., Jones & Smith, 2002). This may be one reason why most nativists do not make the strong claim that the shape bias is innate. However, this leaves the nativists with two problems: one is that we no longer have a solution for the problem of induction in word learning; the other is that nativists offer no mechanisms for how the shape bias is learned. This view now looks neither descriptively nor explanatorily adequate.

The empiricists run into problems as well. Several studies have shown that 2-year-old children take shape to be a heuristic in learning words, but only under some circumstances. Gelman and Ebeling (1998) showed that if a shape is created accidentally, children do not use it as the basis for their word generalization. It is only when a shape is created intentionally that it becomes the basis for word generalization. Similarly, Bloom and Markson (1998) and Diesendruck and Bloom (2003) showed that shape is a proxy for kind and the shape bias is not specific to word learning. In addition, the elegant and impressive training study conducted by Smith et al. (2002) seems to provide evidence *against* a pure associative learning view of the shape bias. In this study, 17-month-old infants were brought into the laboratory once each week for a total of 8 weeks and given a set of training objects with perfect shape regularity while being taught a small number of count nouns. The results showed that the training facilitated the emergence of the shape bias for the novel words they had been exposed to; moreover, the children who were in the experimental condition, and not the ones in the control condition, showed a faster rate of vocabulary acquisition outside of the laboratory! On the one hand, this is a beautiful demonstration that a bias may be trained with a relatively small amount of data; on the other hand, the rapid learning demonstrated by the children calls into question whether

Copyright OUP

associative learning mechanisms, which are often slow and laborious, provide an adequate account for these empirical findings.[1]

Smith et al. (2002) hypothesized that learning the shape bias amounts to making generalizations on two levels. On one level, the child learns that 'cups are cup-shaped, cats are cat-shaped, balls are ball-shaped', and so on. These are the first-order generalizations. On a higher level, the child posits a variable and forms a second-level generalization such as 'Xs are X-shaped'. This is a generalization over all count nouns in the lexicon, not just any particular count noun. By making the second-level generalization, the child is now in a position to apply the shape bias to all new count nouns that he or she might learn in the future. The second-level generalization provides a powerful tool for future learning. The ability to extract a variable, however, is not easily implemented in standard neural network models (see Marcus, 2001, for a critique). Submitting that young learners are capable of extracting variables of this kind is, itself, a concession from the more hardcore empiricist view of learning. Again, this view now looks neither descriptively nor explanatorily adequate.

For the rest of this chapter, we will attempt to characterize a new view—the rational constructivist view—of development (see also Xu, 2007a). This view differs from both of the standard views, and it provides a resolution to the longstanding debate on the nature of the shape bias. In the course of articulating this view, we borrow from two sources: one is the theory of induction explicated by Nelson Goodman (1955), and the other is the machinery of Bayesian inference. We will also report new empirical results that provide support for this view of development.

Most of the studies we see in the literature on the nature of the shape bias have been conducted with 2-, 3-, and 4-year-old children. Recently though, there have been a few studies investigating the importance of shape in property induction, and object categorization and individuation in infants (e.g., Graham et al., 2004; Waxman, 1999; Welder & Graham, 2001; Xu et al., 2004). These studies provide evidence for the privileged role of shape, but mostly they have not focused on word learning per se. We have recently completed several studies to address the three controversies in the shape bias debate systematically in 9- to 12-month-old infants (Dewar & Xu, 2007, 2008, in press; Perfors et al., 2007). These studies investigate the onset of the shape bias, the issue of innateness, and the issue of the learning mechanism.

[1] The reader may think that these training studies provide the best evidence for a learned shape bias. However, a possible alternative hypothesis for the effects of training is that perhaps children were more ready for being trained on the dimension of shape, so the laboratory exposure was particularly effective. That is, the training may have served the role of 'triggering' a small but preexisting bias and simply accelerated the rate of development.

Copyright OUP

11.3 When do we first see some sensitivity to shape in a word learning context?

Analyses of spontaneous language production by young children suggest that the shape bias may emerge as early as 17–24 months (Samuelson & Smith, 2005). What about infants who are at the beginning of word learning: that is, the beginning of comprehending words for object categories? We developed a new method to investigate this question with 9-month-old infants (Dewar & Xu, 2007). In these looking-time experiments, infants watched some events unfold on a puppet stage. We first familiarized the infant with some object pairs. A box was presented on the stage, with the experimenter sitting behind it. After drawing the infant's attention to the box, 'Look, [baby's name]!' the experimenter opened the front panel of the box and revealed a pair of objects sitting side by side inside the box. The object pair consisted of either two identical objects (e.g., two identical toy fish) or two different objects (e.g., a toy frog and a toy dog). The two different objects differed in shape and color (e.g., a toy frog-like animal and a toy dog-like animal). On alternate familiarization trials, the infant saw either an identical pair or a different pair. A total of eight familiarization trials were shown to each infant; the second set of four was a repetition of the first set of four. Thus, we have given each infant some idea as to what to expect inside the box: a pair of identical objects or a pair of different objects.

After the familiarization trials, the test trials began. On each test trial, the same box was presented on the stage, with its front panel closed. The experimenter looked into a top opening of the box and labeled the objects inside. On a two-word trial, she said in infant-directed speech, 'I see a zav! I see a fep!' thrice; on a one-word trial, she said, 'I see a wug! I see a wug!' thrice (Fig. 11.1). The question was whether the infant expected the contents of the box to be predicted by the labeling information. For an adult, hearing two distinct labels will lead to the expectation that the box contains two different objects whereas hearing one label repeated will lead to the expectation that the box contains two identical objects. What about 9-month-old infants? After the labeling information was presented, the experimenter opened the front panel of the box to reveal either two identical objects or two different objects. The experimenter lowered her head and gaze, and the infant's looking times were recorded. If the infant had the same expectations as an adult, she should look longer at the identical objects upon hearing two labels because that was the unexpected outcome. Conversely, the infant should look longer at the different objects upon hearing one label repeated because that was the unexpected outcome. The results were consistent with these predictions—infants' looking times were predicted by the linguistic information provided.

Copyright OUP

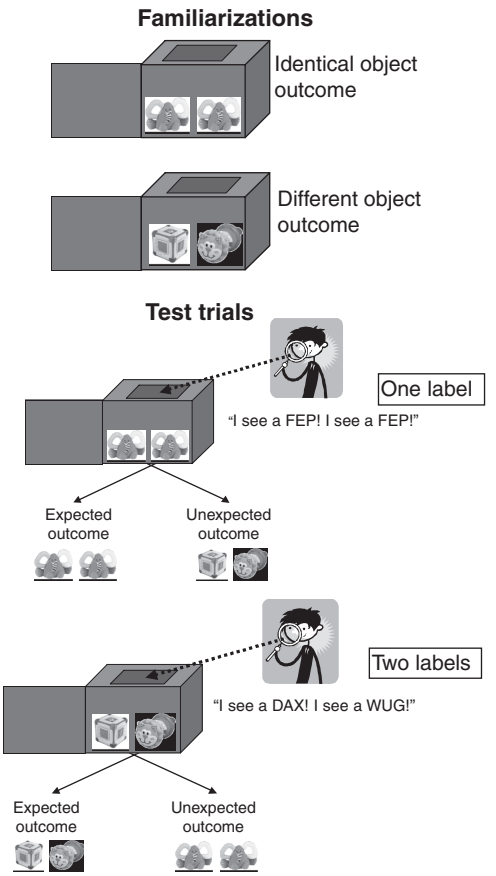

Fig. 11.1 Familiarization and test events from Dewar and Xu (2007).

So far we have only shown that 9-month-old infants expect distinct labels to refer to different-looking objects. In the next two experiments, we used the same methodology but used objects that differed only in shape or only in color. The rationale was that if shape was privileged early on, we might find that infants expect distinct labels to refer to different-shaped objects but not different-colored objects.

In the second experiment, infants were familiarized with pairs of objects that were either identical or differed only in shape. On the test trials, either two distinct labels or one repeated label were provided. The box was then opened to reveal its contents and looking times were recorded. We found that infants looked longer at the identical object pair upon hearing two distinct labels whereas they looked longer at the different-shaped object pair upon hearing one label repeated.

Ch11.indd 269 2/23/2009 6:24:17 PM

Copyright OUP

In the third experiment, infants were familiarized with pairs of objects that were either identical or differed only in color. Again, either two distinct labels or one repeated label were provided on the test trials. We found that infants' looking times were not predicted by the number of labels; they looked longer at the different-colored object pair on all trials.

Taken together, the results of these three experiments suggest that at as young as 9 months of age, infants expect distinct labels to refer to different-shaped, but not different-colored, objects. The sensitivity to the perceptual dimension of shape appears to emerge at the very beginning of word learning (Dewar & Xu, 2007).[2]

Is the sensitivity to shape innately given or learned? The studies we have described so far do not address this question. With older children, both corpus analyses and word-learning experiments suggest that names for different domains of objects—animals, artifacts, food objects—have different perceptual correlates. Shape is a very salient and reliable cue for artifact categories whereas both shape and texture matter for animal categories. For food objects, color and texture are more reliable cues for category membership than shape (e.g., Jones & Smith, 2002). What about for infants? It is conceivable that infants start off considering shape as the most important perceptual dimension for word meaning, and later on they start to use different perceptual dimensions for different domains given the input.

To investigate the origin of the shape bias, we conducted another series of experiments with food objects. The general methodology was the same as Dewar and Xu (2007), but we used objects that looked like food objects. The objects were made out of Fema (a material similar to play dough that, once baked, becomes solid) and they were made to look like a slice of watermelon or a bunch of grapes, and so on. Before the experiment started, the experimenter picked up each of the food object and pretended to eat it. The rest of the experiment used the same procedure as Dewar and Xu (2007). In each of two experiments, the object pairs differed only in shape or only in color. With 9-month-old infants, we found no sensitivity to shape: their looking times were not predicted by the labeling information. With 9-month-old girls, but not boys, we found some sensitivity to color: the girls' looking times were predicted by the labeling information when they were presented with different-colored food objects (Dewar & Xu, 2008). Preliminary results from

[2] The reader may think that these experiments contradict the results of the Smith et al. (2002) training studies. However, in Dewar and Xu (2007), no specific mapping between a word and an object was required. All the infants had to do was the use the number of labels to predict which two objects should be revealed inside the box. Thus, the information-processing demand was lower in these experiments, and that might account for the early sensitivity to shape we have found.

Copyright OUP

12-month-old infants suggest color sensitivity for the whole age group when given labeling information. We are currently investigating if shape also emerges as a reliable cue for category membership with food objects by 12 months.

The results of this second series of experiments show that shape is not a privileged perceptual dimension for all object categories. As early as we can test, infants show differential sensitivity to shape or color depending on the domain of objects under consideration. It was remarkable and surprising to find sensitivity to color as early as 9 months, albeit only in girls. These two sets of studies strongly indicate a *learned* shape bias that is driven by environmental input.

11.4 **Do early words refer to kinds or shapes?**

Have we settled the disagreements between the nativists and the empiricists vis a vis the origin and the nature of the shape bias? Not yet. The empirical results we just described indicate a *learned* bias, but we still do not know if early words and/or count nouns refer to distinct shapes or distinct kinds. On the latter view, shape is a reliable cue to kind membership and infants observe this regularity from the input quite early on.

Several studies with 2- to 4-year-old children show that shape is a proxy for kind membership. For example, Gelman & Ebeling (1998) showed 2- and 3-year-old children line drawings roughly shaped like various namable objects. For half the participants, each line drawing was described as depicting a shape that was created intentionally (e.g., someone painted a picture). For the remaining participants, each drawing was described as depicting a shape that was created accidentally (e.g., someone spilled some paint). Participants were asked to name each picture. The findings suggest that children used shape as the basis of naming primarily when the shapes were intentional. Although shape does play an important role in children's early naming, other factors are also important, including the mental state of the picture's creator.

Similarly, Bloom and Markson (1998) found that when a picture was ambiguous (e.g., resembling both a balloon and a lollipop), 3- and 4-year-olds named the picture based on the creative intent of the artist. These results demonstrate that the sameness of shape is not sufficient in determining children's naming preferences: something can be shaped like a lollipop, but not called 'a lollipop'. These studies suggest that for young children, labels for objects (and pictorial representations of those objects) are not wholly determined by shape similarity; shape may be a proxy for kind membership (see also Cimpian & Markman, 2005). A recent study provides evidence that infants as young as 18 months take into account both perceptual and conceptual knowledge when extending a novel word. Booth et al. (2005) found that when targets were described as

Copyright OUP

artifacts, infants extended on the basis of shape; however, when targets were described as animates, infants extended on the basis of both shape and texture. These findings suggest that shape is being used as a cue for kind membership in children's naming (see also Shutts et al., Chapter 8).

By 18 months of age, children expect words for object categories to refer to kinds. However, it remains an open question whether infants who are at the beginning of word learning hold the same expectations. It may be the case that infants expect words to refer to distinct shapes and, later in development, they realize that shape is correlated with kind membership.

How can we get at the question of kind versus shape in infants? In a recent study, we borrowed an idea from previous studies with preschoolers, and developed a methodology that could be used to test much younger infants: a group of 10-month-olds. One way to distinguish between shape versus kind is by investigating infants' expectations about nonobvious properties (Baldwin et al., 1993; Graham et al., 2004; Welder & Graham, 2001). The notion of kind implies that members of the same kind share deep, internal, nonobvious properties, and their shared surface perceptual properties are caused by the shared deeper causal properties, a la psychological essentialism (Gelman, 2003; Medin & Ortony, 1989). If early count nouns refer to distinct shapes, we should see no effects of nonobvious properties. If, on the other hand, early count nouns refer to kinds, we may expect to see labeling predict internal, nonobvious properties of objects.

We therefore designed a study that pitted perceptual similarity against labeling, and we asked if 10-month-old infants would use the labeling information to predict internal, nonobvious properties, in this case, sound properties (Dewar & Xu, in press). We used a between-subject design: half of the infants were familiarized with identical object pairs, and the other half of the infants were familiarized with different-looking object pairs (see Fig. 11.2). On the familiarization trials, infants were shown one pair of objects (identical or different) that made the same sound when shaken and one pair of objects (identical or different) that made different sounds when shaken. On the test trials, pairs of objects were placed on the stage. The experimenter picked up each of the two objects and labeled it with either the same label or two distinct labels. The objects were then shaken to demonstrate their sound properties. The objects then sat stationery and silent on the stage and the infant's looking times were recorded.

We analyzed the effects of object appearance (identical or different), the number of labels (1 vs. 2), and the sound properties (identical or different). The analysis of variance revealed a two-way interaction between the number of labels and the sound properties, and no other main effects or interactions.

Copyright OUP

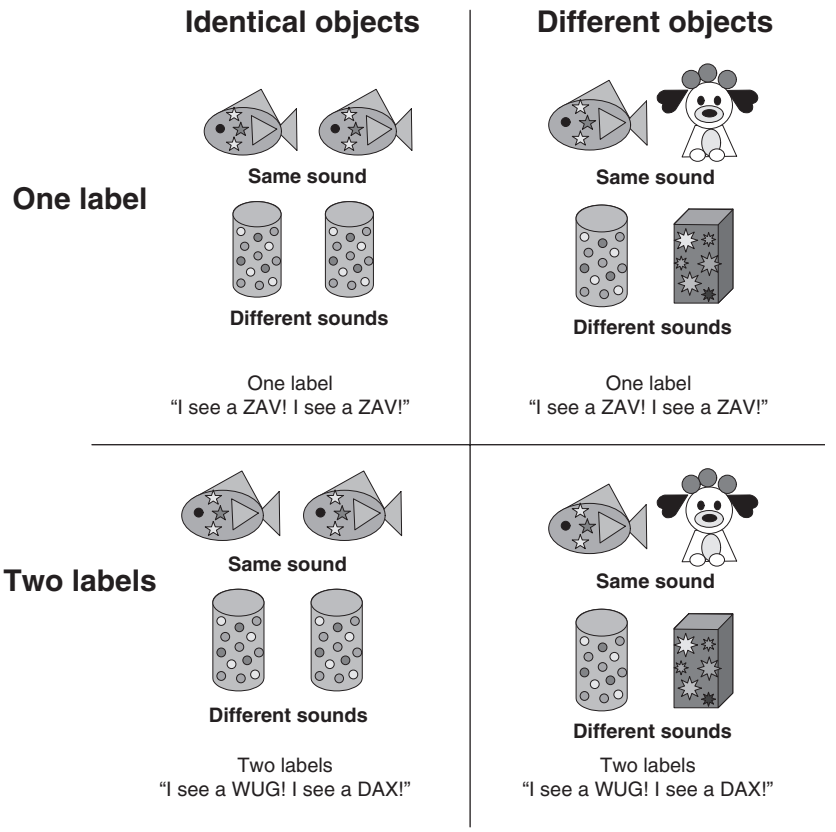

Fig. 11.2 Test trials for all four conditions (identical vs. different objects; one vs. two labels) from Dewar and Xu (in press). See colour plates.

Thus, regardless of object appearance, infants expected the object pair to make the same sound if one label was provided for both objects, and regardless of object appearance, infants expected the object pair to make different sounds if two distinct labels were provided. This was a surprising and striking result. For infants who live in contemporary society (at least in the United States or Canada), they are constantly exposed to toys that make different sounds; our impression is that there may not be a particular good correlation between object appearance and sound properties. Nevertheless, 10-month-old infants expected labeling to be predictive of an internal, nonobvious property, and they were willing to override perceptual similarity in favor of labeling information.

Copyright OUP

11.5 What is the mechanism of learning?

The previous three studies provide evidence that (1) the shape bias is learned and (2) shape is a proxy for kind membership. If the shape bias is learned, how is it learned? What is the mechanism of learning such that even 9-month-old infants have acquired the bias? One prominent proposal in the literature is that the shape bias is learned through associative mechanisms, often instantiated in neural network and/or connectionist models (e.g., Colunga & Smith, 2005; Samuelson, 2005). An alternative account, which we will argue for, is a hierarchical Bayesian model that captures the idea of overhypothesis formation in general.

Kemp et al. (2007) presented the basics of a Bayesian model of overhypothesis formation, and one of their chosen case studies was the shape bias. They (and we) borrowed the idea of an overhypothesis from the philosopher Nelson Goodman, whose well-known book *Fact, Fiction, and Forecast* (1955) is considered a classic on induction. One widely cited example that Goodman uses to illustrate the idea is the following:

> Suppose I show you a bunch of bags that are all identical. From the first bag I pull out a few marbles and they are all white. From the second bag I pull out a handful of marbles and they are all red. From the third bag I pull out some marbles and they are all green. If I ask you what you think the color of the next marble would be from the first, second, or the third bag, presumably you will answer with a high degree of confidence that the next marble will be white, red, and green, respectively. Now suppose I show you a new bag, and I pull out one marble that is blue. What do you think the color of the next marble will be from this new bag? Again, I think you will answer with a high degree of confidence that it is going to be blue (Fig. 11.3).

Goodman suggests, and we concur, that the learner is making both a first-order and a second-order generalization in this case, and perhaps with equal levels of confidence. In other words, the learner has formed an overhypothesis— 'Bagfuls of marbles are uniform in color'—and this second-order generalization allows the learner to make predictions about a brand new bag with new marble colors she has never seen before.

Several researchers have applied this general idea to address the problem of how children acquire the shape bias (e.g., Smith et al., 2002; Samuelson, 2005; Kemp et al. 2007). Smith et al. (2002) hypothesized that first, children acquire count nouns that refer to object categories, e.g., 'cat' refers to cats, 'horse' refers to horses, and 'chair' refers to chairs, etc. For each word and each category, children may form a first-order generalization, e.g., the word 'cat' refers to objects that are cat-shaped, the word 'horse' refers to objects that are horse-shaped, and the word 'chair' refers to objects that are chair-shaped, etc. Once a number of individual words and categories are learned, the child

Copyright OUP

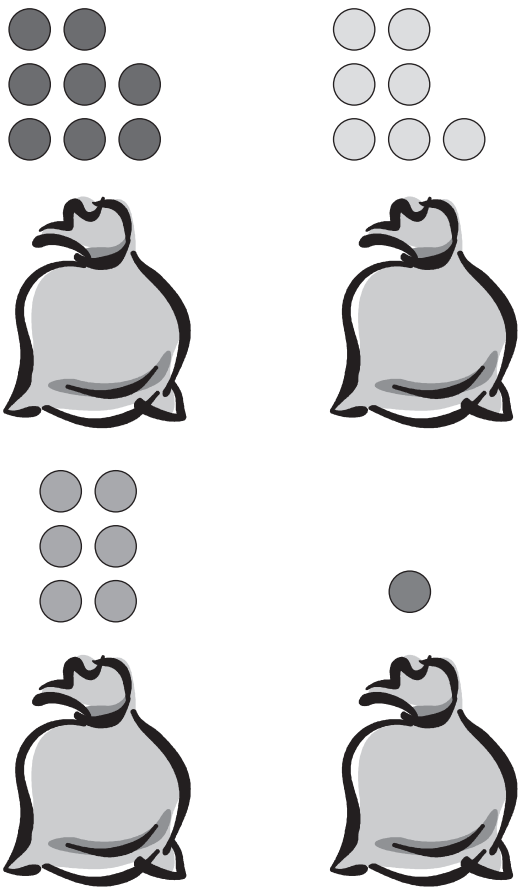

Fig. 11.3 A schematic of the thought experiment on overhypothesis formation by Goodman (1955). See colour plates.

may form a second-order generalization—that is, an overhypothesis—that 'Noun X refers to objects that are X-shaped'. This mechanism allows the child to go beyond the specific words and categories they have learned and to make a principled generalization about all count nouns that refer to object categories. In a set of elegant training studies, Smith et al. (2002) found that laboratory exposure to shape regularity and word usage with 17-month-old infants induced the shape bias and subsequently allowed these children to learn words faster outside of the laboratory.

The new studies we described earlier in this chapter suggest that the shape bias emerges earlier than 17 months, at least when the infants are not required to map specific words to specific objects (Dewar & Xu, 2007, 2008). If the

Copyright OUP

shape bias is indeed learned through a mechanism of overhypothesis formation, we will need evidence from much younger infants. To date, two studies provide preliminary evidence that 8- to 12-month-old infants are able to form overhypotheses based on either shape or color after fairly limited exposure (Dewar & Xu, in preparation; Perfors et al. 2007).

In one experiment, we taught 8- to 12-month-old infants a visual rule: if two objects that collide have the same shape, an explosion effect occurs (i.e., interesting sounds playing and sparkles flying). If two objects that collide have the same color and texture, no effect occurs (see Fig. 11.4). During the training phase, the infants were shown combinations of five different shapes, five different colors, five different textures, and five different effects. Each habituation trial consisted of a continuous string of collision events, randomly chosen from the whole set. The habituation phase ended when the infants' looking times dropped to 50% or less of the initial level.

On the test trials, infants were shown two test events alternately: an expected event when two objects that shared a novel shape collided and a new effect occurred, and an unexpected event when two objects that shared a novel color and texture collided and a new effect occurred. If the infants had formed an

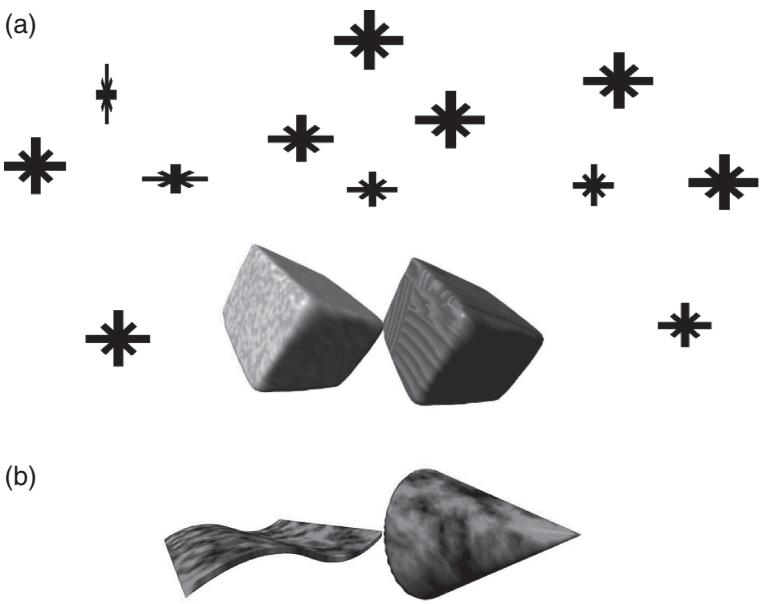

Fig. 11.4 Sample event from habituation trial. (a) When the objects match in shape, there is a sparkly effect when they collide. (b) When the objects do not match in shape, there is no effect.

Copyright OUP

overhypothesis about how a shape match predicts an effect, they should look longer at the unexpected event than the expected event, even though both involved new objects and new effects. These were the results we found. When we divided up our sample into younger and older infants (8:15 to 10:15 and 10:16 to 12:15), we found no differences in looking times between the two age groups. Thus, at both 9 and 11 months, infants were readily able to form an overhypothesis based on shape. (For various technical reasons, these results are preliminary and they need to be replicated.)

Given these findings, however, one might still argue that perhaps shape is special—it may be more salient and therefore it may be easier to form an overhypothesis over this dimension. In the second study, we used a different experimental paradigm and conducted a version of the bags-of-marbles experiment a la Goodman (Dewar & Xu, in preparation). In the experimental condition, the experimenter began by bringing out a rectangular box. She pulled out four ping-pong balls from the box and placed them into a small transparent container in front of the box; all four balls were white. Next, the experimenter brought out another box and set it beside the first box. She pulled out four yellow ping-pong balls and placed them in front of the second box. The experimenter repeated the sequence with a third box and four pink ping-pong balls. Then, she brought out a fourth box and pulled out three black ping-pong balls plus a white ping-pong ball—this was the unexpected outcome. The ping-pong balls were placed into the small container and looking time was recorded. Adults watching this event found the last draw surprising, because they have formed an overhypothesis about how each box contained uniformly colored ping-pong balls. For the expected outcome, the experimenter did exactly the same thing up until the last draw from the fourth box: instead of pulling the white ping-pong ball from the fourth box, she pulled it out of the first box, which in fact contained all white ping-pong balls. Thus, the outcomes of the expected and unexpected events were identical so as to rule out any intrinsic preferences for one outcome versus the other. We found that 9-month-old infants looked longer at the unexpected outcome than the expected outcome in the experimental condition.

In the control condition, the ping-pong balls placed into the first three boxes were drawn from all boxes (e.g., the first and the fourth white balls were drawn from the first box, the second white ball from the third box, and the third white ball from the second box), so the final display of three containers with four ping-pong balls of one color was the same as before but there was no overhypothesis to be formed about how 'boxes contain uniformly colored ping-pong balls'. The fourth box was brought out and the four ping-pong balls were drawn as in the expected and unexpected outcomes of the

Copyright OUP

experimental condition. Results showed that infants looked about equally at the two outcomes, suggesting that it was not the perceptual grouping or simply forming a first-order generalization with the last box that drove the effect in the experimental condition.

With very minimal exposure, infants seem to be willing to form an overhypothesis quickly, and this overhypothesis guides their inductive inference immediately. Which overhypothesis has been formed in these studies? At least two overhypotheses are consistent with the data we have collected so far: one is Goodman's 'bagfuls of marbles are uniform in color' (or our equivalent 'boxes of ping-pong balls are uniform in color', but they may vary on other dimensions); the other is 'bagfuls of marbles are identical within each bag'. Studies are underway to tease these apart. Are shape, color, and texture equally good candidate perceptual dimensions for forming overhypotheses? These are still open questions and studies are currently being conducted in our laboratory to address these issues.

In the hierarchical Bayesian model for overhypothesis learning presented by Kemp et al. (2007), a small number of perceptual dimensions (e.g., shape, texture, and color) are assigned a uniform prior probability. The inferential mechanism (i.e., Bayesian belief updating or estimation of the posterior probability given the data) updates all these hypotheses simultaneously given the data. In addition, it estimates two second-level parameters, one corresponding to how uniform a particular feature is within a category and the other corresponding to how that feature is distributed across categories. The uniform prior assumption is an important one: it states that no perceptual dimension is privileged at the beginning. The input statistics will push the model to consider one or another perceptual dimension as a reliable cue for word meaning, for example, for artifact objects, shape will receive a greater posterior; whereas for food objects, color will receive a greater posterior given the data. This model ensures rapid inductive learning in part because every item contributes to the second-order inference, whereas a more bottom-up or associative learner would require many items before forming even first-order generalizations. This sort of rapid learning is one of the characteristics that distinguish a Bayesian model from a standard neural network model (e.g., Colunga & Smith, 2005).

How do we know which hypotheses children assign initial high prior probability to? We can think of two possible empirical tests. One is to show that given roughly the same amount of data, overhypotheses over *shape* or *color* (or a few other perceptual dimensions) can be formed fairly quickly and easily; the other is to investigate whether certain features, such as *being placed on the left side of the table*, are also subject to overhypothesis formation.

Copyright OUP

	Learned or innate?	Kinds vs. shapes?	Mechanism of learning
Standard nativist	innate	kinds	N/A
Standard empiricist	learned	shapes	associative learning
Rational constructivist	**learned**	**kinds**	**overhypothesis formation**

Fig. 11.5 Three views of the development of the shape bias.

The rational constructivist view differs from both the nativist and the empiricist view of cognitive development. Using the shape bias in word learning as an example, the rational constructivist view is committed to this bias being learned, being subject to object domain effects, and to be learned via domain-general inferential mechanisms such as overhypothesis formation (this is the non-nativist part). At the same time, this view accepts that early count nouns for objects map onto kinds, and that shape is a proxy for kind membership. Furthermore, the mechanisms we suggest are not 'dumb' correlation detectors but rather complex rational learning mechanisms (this is the non-empiricist part) (see Fig. 11.5 of the various views).

11.6 Generality of the approach: other learned biases in language and cognition

Other word-learning biases may be acquired in a similar way. We briefly describe three cases in this section. Some researchers have argued that the basic-level bias is learned (e.g., Callanan, 1989; Callanan et al., 1994; Xu & Tenenbaum, 2007). One possible learning mechanism for this bias is overhypothesis formation. The child learns early on that words for object categories tend to refer to objects that have high within-category perceptual similarity and low between-category similarity, i.e., the basic level, then she expects other count nouns for objects to work the same way. For example, all dogs share a basic shape, but we also have subordinate-level count nouns such as *poodle*, *terrier*, *Chihuahua*, and *German Shepard*. This learning story suggests that if most of the count nouns a child was exposed to referred to subordinate-level categories, the child might instead arrive at the generalization that count nouns tend to refer to subordinate-level categories. Thus, an overhypothesis about the level of category specificity to which count nouns tend to refer may be formed to guide future word learning.

In another line of research, we have charted the development of sortal concepts in infants, and suggested that word learning may be an important causal

Copyright OUP

factor in acquiring sortal concepts, which are a subset of our concepts that provide criteria for individuation and identity (see Xu, 2007b, for a review). Some of our empirical work suggests that at 9 months infants can use the presence of two distinct count noun labels to help them make the inference that two objects are behind an occluder in an object individuation task. By 12 months, online verbal labeling is not needed in the same task, at least for different-shaped objects, suggesting that by the end of the first year, infants may have formed a general rule of thumb about whether particular perceptual differences provide sufficient evidence for positing a second object in the event. One possible learning mechanism is that the younger infants learn piecemeal that certain perceptual features matter for object identity and hearing linguistic labels provide strong evidence for that. By tracking word usage and object features over time, the infants may form an overhypothesis about which perceptual features may be important for object identity. Once that overhypothesis is formed, infants no longer need explicit labeling for object individuation.

Lastly, biases for other aspects of language can be acquired via overhypothesis as well. Wonnacott et al. (2007) present evidence from adults learning the word order of an artificial language that adults not only learn the specific frequency distribution of individual verbs (i.e., whether they appear in a Verb-Subject-Object order or a VOS-particle order) but also track the overall distribution of verb types (e.g., if majority of the verbs allow the alternation, then they generate novel alternations for other verbs in the language.) Again, the general idea of an overhypothesis applies: what the adults may have learned by tracking verb-type information is that they entertain the hypothesis 'in this language, word order is flexible' if enough verb types support it (see Perfors, 2008, for a Bayesian model that captures these phenomena).

Overhypothesis formation is presumably a domain-general mechanism, not specific to word learning, or to language. Shipley (1993) was probably the first psychologist to draw our attention to Goodman's work (1955) and to apply it to developmental psychology. She points out that a child learning about animal species may acquire knowledge about specific animal kinds, and then form an overhypothesis about all animal kinds. For example, children may learn that cows eat grass, rabbits eat carrots, and monkeys eat bananas. They may also learn that fish live in water, goats live on mountains, and dogs often live in people's homes. From these data, children may form the overhypothesis that animal kinds have characteristic diets and they have characteristic habitats. Once this second-order generalization is formed, children would expect new animal kinds they encounter to have these properties, and this may give them pointers when seeking new knowledge. We have followed up on these

Copyright OUP

studies in the laboratory to address the issue of how domain-general these learning mechanisms are.

11.7 Conclusions: the basic commitments of rational constructivism

The case study on the shape bias is meant to illustrate the basic ideas behind a rational constructivist approach to studying development. On the one hand, we are committed to a representational view of the mind, and we submit that certain conceptual primitives such as *OBJECT*, *AGENT*, *NUMBER*, and *KIND* may be innately given (or learned very early), and they provide the foundation for later development. At the same time, we suggest that specific biases such as the shape bias or the basic-level bias are learned, and we provide some suggestions and evidence for the existence of powerful, domain-general inductive learning mechanisms such as overhypothesis formation and probabilistic reasoning (e.g., Girotto & Gonzalez, 2007; Gopnik et al., 2004; Kushnir et al., in press; Schlottman, 2001; Teglas et al., 2007; Xu & Garcia, 2008; Xu & Tenenbaum, 2007). Our hope is that the existence of these inductive learning mechanisms, perhaps best instantiated as computational models within a Bayesian framework, gives us the tools we need to avoid ascribing very specific, innately given biases to the young human learner. Furthermore, these mechanisms themselves need to be discovered and studied in a great deal more detail. In each case, a detailed learning story may be in sight, one that meets the challenges of human inductive learning—making meaningful and principled generalizations based on limited amount of data.

Acknowledgment

We thank Susan Carey, Stephanie Denison, Charles Kemp, Mijke Rhemtulla, Joshua Tenenbaum, and Henny Yeung for helpful discussions. This research was supported by grants from NSERC and SSHRC Canada to F. Xu.

References

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science*, *3*, 89–94.

Baldwin, D., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: evidence from exploratory play. *Child Development*, *64*, 711–728.

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science, 9*, 200–204.

Booth, A. E., Waxman, S. R., & Huang, Y. T. (2005). Conceptual information permeates word learning in infancy. *Developmental psychology*, *41*, 491–505.

Copyright OUP

Callanan, M. A. (1989). Development of object categories and inclusion relations: preschoolers' hypotheses about word meanings. *Developmental Psychology*, *25*, 207–216.

Callanan, M. A., Repp, A. M., McCarthy, M. G., & Latzke, M .A. (1994). Children's hypotheses about word meanings: Is there a basic level constraint? *Journal of Experimental Child Psychology, 57*, 108–138.

Cimpian, A., & Markman, E. M. (2005). The absence of a shape bias in children's word learning. *Developmental Psychology*, *41*, 1003–1019.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: a role for associative learning. *Psychological Review*, *112*, 347–382.

Dewar, K. & Xu, F. (in preparation) Overhypothesis formation in 9-month-old infants. Manuscript in preparation.

Dewar, K. & Xu, F. (in press) Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science.*

Dewar, K., & Xu, F. (2007) Do 9-month-old infants expect distinct words to refer to kinds? *Developmental Psychology*, *43*, 1227–1238.

Dewar, K. & Xu, F. (2008) Do 9-month-old infants expect distinct labels to refer to kinds: the effect of domain. Poster presented at the 16th Biennial International Conference on Infant Studies, Vancouver, Canada.

Diesendruck, G., & Bloom, P. (2003) How specific is the shape bias? *Child Development*, *74*, 168–178.

Gelman, S. A. (2003). The *Essential Child*. Oxford: Oxford University Press.

Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, *66*, B35–B47.

Girotto, V., & Gonzalez, M. (2007). Children's understanding of posterior probability. *Cognition*, *106*, 325–344.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Gopnik, A., Glymore, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, *111*, 3–32.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). 13-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, *5*, 219–232.

Kemp, C., Perfors, A., & Tenenbaum, J.B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307–321.

Kushnir, T., Xu, F. & Wellman, H. (2008) Preschoolers use statistical sampling information to infer the preferences of others. In V. Sloutsky, B. Love, & K. MacRae (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Macnamara, J. (1982). *Names for Things: A Study of Human Learning*. Cambridge, MA: MIT Press.

Marcus, G. (2001). *The Algebraic Mind*. Cambridge, MA: MIT Press.

Markman, E. M. (1989). *Categorization and Naming in Children*. Cambridge, MA: MIT Press.

Copyright OUP

Markson, L., Diesendruck, G., & Bloom, P. (2008). The shape of thought. *Developmental Science*, *11*, 204–208.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 179–195). New York: Cambridge University Press.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*.

Perfors, A. (2008). *Learnability, Representation, and Language: A Bayesian Approach*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Perfors, A., Kemp, C., Tenenbaum, J. B., & Xu, F. (2007) Learning inductive constraints. Talk presented at the Biennial Meeting of the Society for Research in Child Development, March 29–April 1, Boston, MA.

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

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Samuelson, L. K. (2005). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15–20 month olds. *Developmental Psychology*, *38*, 1016–1037.

Samuelson, L. K., & Bloom, P. (2008). The shape controversy: What counts as an explanation of development? Introduction to the special section. *Developmental Science*, *11*, 183–184.

Samuelson, L. K., & Smith, L. B. (2005). They call it like they see it: spontaneous naming and attention to shape. *Developmental Science*, *8*, 182–198.

Schlottman, A. (2001). Children's probability intuitions: understanding the expected value of complex gambles. *Child Development*, *72*, 103–122.

Shipley, E. (1993). Categories, hierarchies, and induction. *The Psychology of Learning and Motivation*, *30*, 265–301.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–19.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. *Cognition*, *38*, 179–211.

Spelke, E. S. (1994). Initial knowledge: six suggestions. *Cognition*, *50*, 431–445.

Teglas, E., Girotto, V., Gonzalez, M., & Bonatti, L. (2007) Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 19156–19159.

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*, B35–B50.

Welder, A. N., & Graham, A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, *72*, 1653–1673.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2007) Acquiring and processing verb argument structure: distributional learning in a miniature language. *Cognitive Psychology*, *56*, 165–209.

Copyright OUP

Xu, F. (2007a). Rational statistical inference and cognitive development. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind: Foundations and the Future*, vol. 3 (pp. 199–215). Oxford University Press.

Xu, F. (2007b). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, *11*, 400–406.

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology*, *49*, 155–190.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 5012–5015.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.

Ch11.indd 284 2/23/2009 6:24:18 PM