

Language acquisition and conceptual development

Edited by

Melissa Bowerman and Stephen C. Levinson

Max Planck Institute for Psycholinguistics

2001



CAMBRIDGE
UNIVERSITY PRESS

4 How domain-general processes may create domain-specific biases

Linda B. Smith

Indiana University

How do we come to understand the world so that we can act in it, so that we talk about it? Where do our rich understandings – of causality and intention, of number and objects, of space and time, of language – come from? Many have looked at the diversity of human knowledge and at the certainty with which children acquire it all and concluded that the diversity is there to begin with. They have suggested that knowledge acquisition is driven by special mechanisms dedicated to specific domains (e.g. Fodor 1975, 1983; Chomsky 1980; Gleitman 1986; Keil 1989; Gelman 1990; Spelke, Breinlinger, Macomber, & Jacobson 1992; Leslie 1994).

The argument for the domain-specificity of cognitive development is easy to make along four lines.

1. What needs to be acquired is unique to each domain. Thus, nothing beyond the most superficial metaphors are likely to exist across domains.
2. One cannot get something from nothing. In cognitive development, this idea is often presented in terms of the problem of induction and the need for constraints. Briefly, there are an infinite number of objectively correct generalizations from any data set. Therefore, learning cannot happen without prior content-specific constraints on the generalizations that are formed.
3. The empirical evidence on children's learning strongly implicates domain-specific constraints on what is learned. Children exhibit learning biases specific to the content being learned.
4. The smartness of young children's domain-specific learning stands in stark contrast to the inadequacies of their general problem-solving skills. The general cognitive processes of infants and very young children just do not seem smart enough on their own to yield the diverse competence evident in human cognition.

Despite the coherence of this line of reasoning, the present chapter offers an opposing view. The larger idea, borrowed from development as it is understood in modern biology and embryology, is that specificity is constructed out of processes of great generality, that one can get something much more from something much less through a *history* of activity.

This view is presented in the context of what is known about children's initial generalizations of newly learned nouns. Children are very smart learners of object names; so smart that they seem to learn a lexical category from hearing just one object named once. This chapter presents evidence that this smart and domain-specific learning is made in the history of activity of quite general and dumb processes. The central idea is this: domain-general processes when at work in particular learning contexts self-organize to form context-specific learning biases. They do so by creating contexts for new learning that feed on themselves, propelling development to a seemingly certain end and creating destiny.

1 Novel word generalizations

Consider a commonplace example: a 24-month-old child sees a tractor for the very first time. This particular tractor is big, red, snorting, and pulling machinery. The child is told "That's a tractor." The referent of *tractor* is clearly underdetermined from this one naming episode. Nonetheless, evidence from experiments and observations indicate that, after seeing this one tractor and hearing it named, the child will in the future correctly recognize and name other tractors, even tractors different from the original in size or color, in not snorting, in not pulling machinery (e.g. Clark 1973; Mervis 1987). How can a young child know before learning the lexical category that being tractor-shaped is what is critical to being called *tractor*?

In an effort to understand the smartness of children's early word learning, many researchers have used an artificial word learning task (e.g. Gentner 1978; Landau, Smith, & Jones 1988). In these experimental studies, children are presented with a novel object, it is named with a novel pseudoword, and then the children are asked what other objects have the same name. For example, in one study, Landau, Smith, and Jones (1988) showed two- and three-year-old children the novel wooden object shown at the top of figure 4.1. The object was named with a novel count noun, for example, "This is a *dax*." The children were then presented with the test objects also depicted in figure 4.1 and asked about each, "Is this a *dax*?"

The children in this study extended the novel name to test objects that were the same shape as the exemplar regardless of variations in texture or size. The degree of their selective attention to shape was remarkable. In one case, the original exemplar object was 2 cubic inches in area and made of wood. Children extended the name *dax* to same-shaped test objects even when those objects were 100 times the exemplar in size, and to objects made of sponge or chicken wire. This systematic selective attention to shape in the context of learning a novel count noun has now been demonstrated in many different studies and by many different experimenters (e.g.

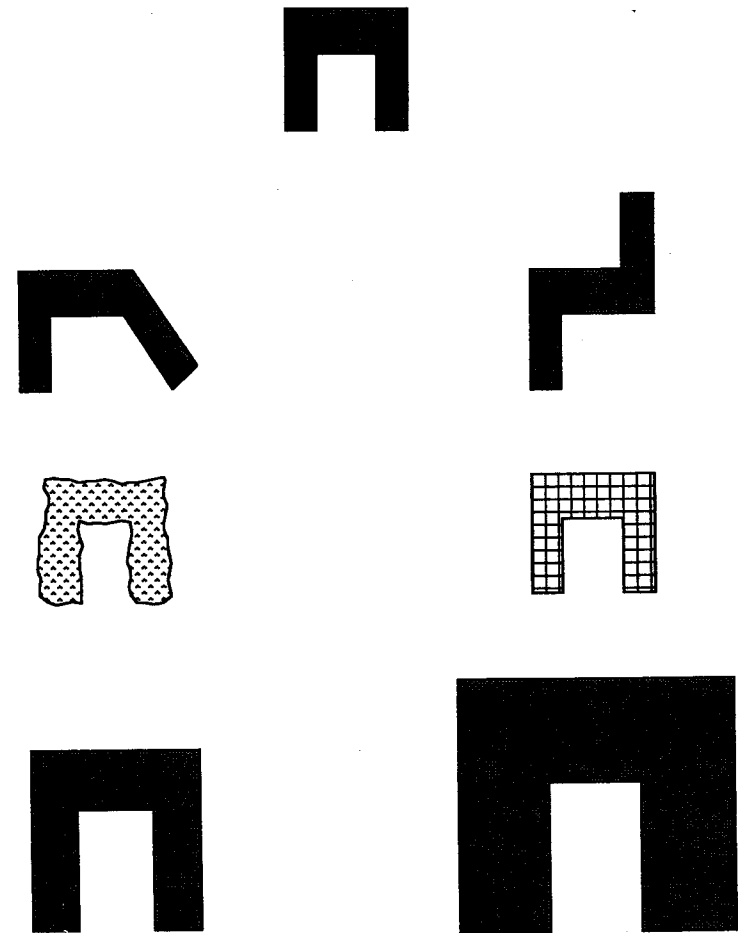


Fig. 4.1 Sample stimuli from Landau, Smith, & Jones (1988). All stimuli were three-dimensional objects made of wood, wire, or sponge.

Soja 1992; Imai, Gentner, & Uchida 1994; Keil 1994). It is a useful attentional bias: it promotes the learning of common nouns, lexical categories that typically refer to objects of similar shape (Rosch 1978; Biederman 1987).

Critically, in their study, Landau, Smith, & Jones contrasted children's attention to shape in the naming task with that in a nonnaming task. In the nonnaming task, they used the same stimuli depicted in figure 4.1 but asked two- and three-year old children to make similarity classifications, to pick out test objects that were "like" the exemplar. In this task, children did not

attend selectively to shape but based their judgments on the wholistic similarity of the test object to the exemplar. That is, they picked out as “like” the exemplar test objects that were not too different from it overall, regardless of the dimension of difference. Apparently, young children do not go around in their everyday lives always attending selectively to the shape of things. Rather, they attend to shape when learning object names.

The selectivity of children’s attention in the naming task and their nonselectivity in the nonnaming task illustrate well the idea of exceptional competence in special domains amidst the inadequacies of general cognitive skills. There is much evidence that young children consistently fare poorly in selective attention tasks, in tasks such as speeded and nonspeeded classification, discriminative learning, same/different judgments. In these tasks, young children seem unable to respond to only one attribute of an object while ignoring other attributes (see Aslin & Smith 1988, for a review). But the evidence from artificial word learning tasks indicates that same-aged children attend quite selectively to shape in the service of learning an object name. These facts are consistent with the idea of a domain-specific mechanism that is set apart from general attentional processes.

Other facts, however, suggest that if such a mechanism exists, it is complex and malleable. As research on children’s artificial word learning has continued, it has become clear that children do not simply attend to the shape of things in the context of naming. Rather, the properties to which children attend are exquisitely tailored to the specific properties of the named object and the linguistic context. Moreover, these attentional biases change with age. The evidence for each of these conclusions is reviewed in turn.

1.1 The properties of the named object

1.1.1 Eyes Jones, Smith, & Landau (1991) examined the effect of the property of eyes on 36-month-old children’s generalization of a novel noun. They reasoned that for many-eyed objects, texture (that is, being furry, or scaly, or feathered), as well as shape, is crucial for lexical categorization. A further impetus for the study was Massey & Gelman’s (1988) finding that young children pay particular attention to texture when making judgments about which kinds of eyed objects (statues vs. living things) can move on their own.

The stimuli used by Jones *et al.* are shown in figure 4.2. In both conditions, the exemplar objects were made of painted wood; the only difference between the two conditions was that in one the exemplar and test objects were eyeless and in the other, they had eyes. Thus, in both conditions, the experimenter named the exemplar, “This is a dax,” and asked about each

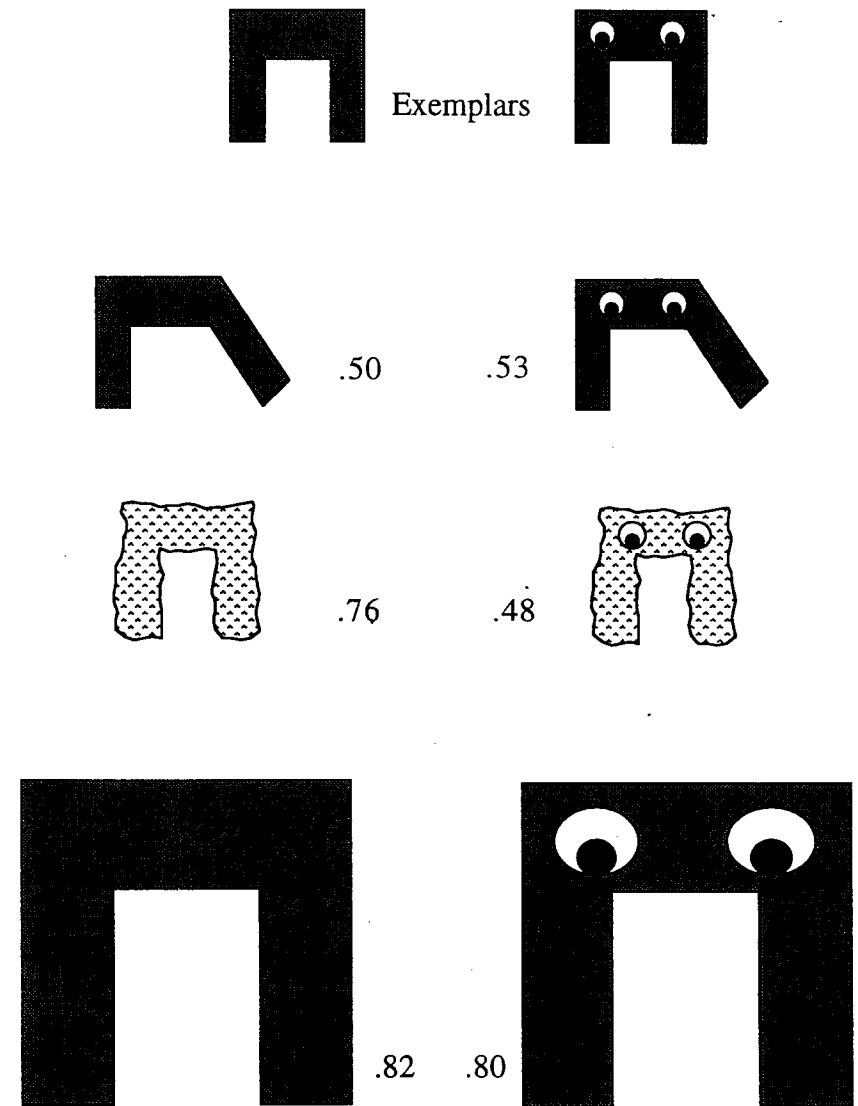


Fig. 4.2 Sample stimuli from Jones, Smith, & Landau (1991) and the proportion of times 3-year-old children extended the name of the exemplar to each of three kinds of test object.

test object, "Is this a dax?" The key results are summarized in the figure; next to each test object is the proportion of times children agreed that the novel noun named that test object. As is apparent, when the objects were eyeless, the children extended the name to same-shape objects. But when the objects had eyes, the children extended the name primarily to the same-shaped-and-same-textured test object. Thus, there is a shape bias for eyeless objects and shape-plus-texture bias for eyed objects.

Additional experiments showed that the presence of eyes affected children's attention only in a naming task. In a similarity judgment task, children attended holistically both when the objects had eyes and when they did not. Thus, while it is naming that triggers selective attention, the trigger is not just to one property. Indeed, children's smart word learning biases are *context sensitive*.

1.1.2 Rigidity Eyes are not the only contextual cue that causes the form of the learning bias to change. Soja and her colleagues (Soja, Carey, & Spelke 1991; Soja 1992) examined children's novel word generalizations when the named objects were rigid things and when they were nonrigid substances. In the Rigid-object condition, the exemplar and two test items were made of substances that held their shape, for example hardened clay or wood. In the Substance condition, the stimuli were formed from nonrigid materials such as shaving cream mixed with gravel. The rigid and nonrigid stimulus sets were formed to present the same shapes as illustrated in figure 4.3. Also shown in the figure are the two test objects – one which matched the exemplar in shape and one which matched the exemplar in material. The procedure in both conditions was the same. Children aged from 2 to 2½ years old heard the exemplar named: "This is my mell." They were then asked to pick "the mell" from the two test objects.

Soja *et al.* found that the children formed different lexical categories when the exemplar was rigid than when it was nonrigid stuff. Specifically, *mell* was generalized to the same-shape test object when the named exemplar was rigid but to the same-material test object when the exemplar was nonrigid. Thus, there is a bias to attend to shape when learning names for rigid things, but a bias to attend to material when learning names for non-rigid substances.

Soja *et al.* also found that this systematic attention to shape in the one case and to material in the other was specific to the task of generalizing a novel name. Children did not show systematic attentional biases with either the rigid objects nor the nonrigid substances in a nonnaming similarity-judgment task. Again, it is specifically the task of naming that triggers selective attention, but the precise properties attended to depend on the perceptible properties of the named object.

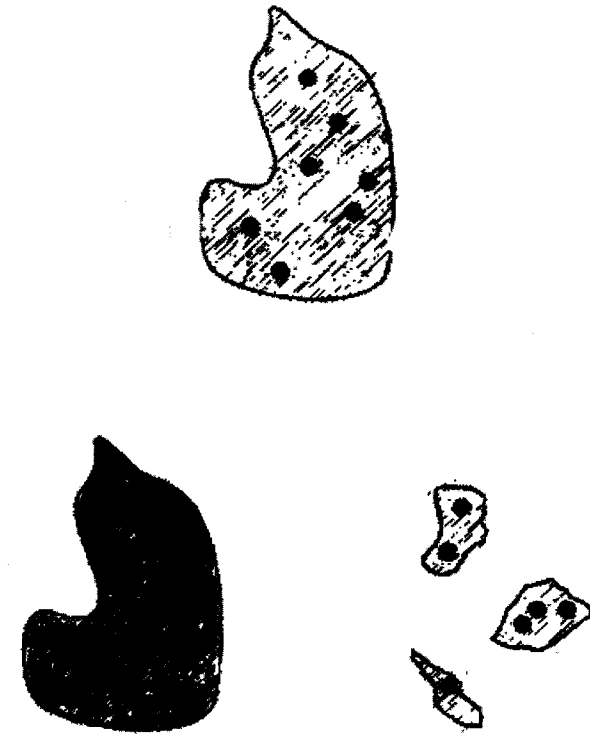


Fig. 4.3 Sample stimuli from Soja (1992). On object trials, the exemplar and test objects were made of rigid substances such as wood. On substance trials, the exemplar and test objects were made of nonrigid substances such as foam.

1.1.3 Sneakers Eyes and rigidity are properties that signal biologically important distinctions and thus children's special sensitivity to these properties in the context of naming could be interpreted in terms of an evolved mechanism that incorporates specific content. Jones & Smith (1998) reported evidence that suggests that the object properties that modulate naming may be an open rather than closed set. They replicated the original Eyes study, but instead of using eyes, they used sneakers. Examples of the sneakered stimuli are shown in figure 4.4. The mean proportion of times 36-month-olds extended the novel name to each test object or said that each test object was "like" the exemplar is also given in the figure. As is apparent, in the naming task only, children attended principally to texture when the stimuli had sneakers. It is unlikely that children are specially prepared to learn how sneakered things are named. Rather, it seems that children have

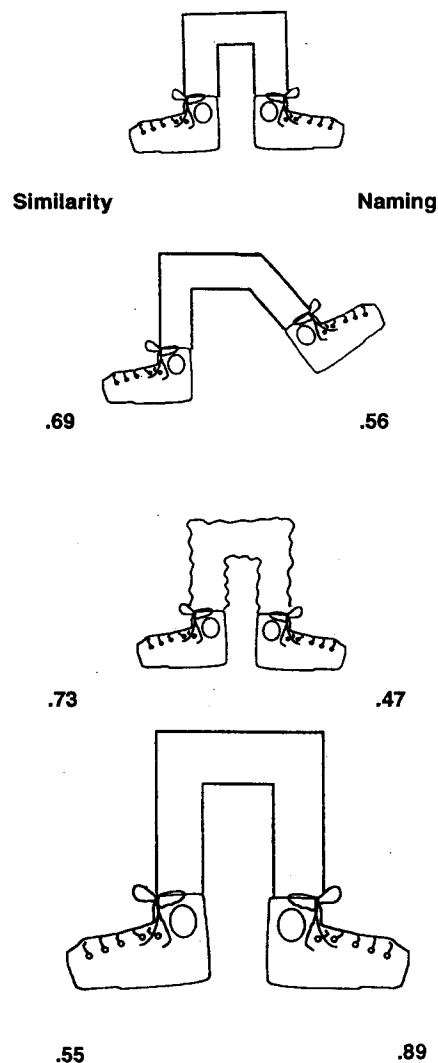


Fig. 4.4 Sample stimuli from a replication of Jones *et al.* that put sneakers rather than eyes on the stimuli and the mean proportion of times that 3-year-old children judged the test object to be "like" the exemplar in the Similarity judgment task and judged the test object to have the same name as the exemplar in the Naming task.

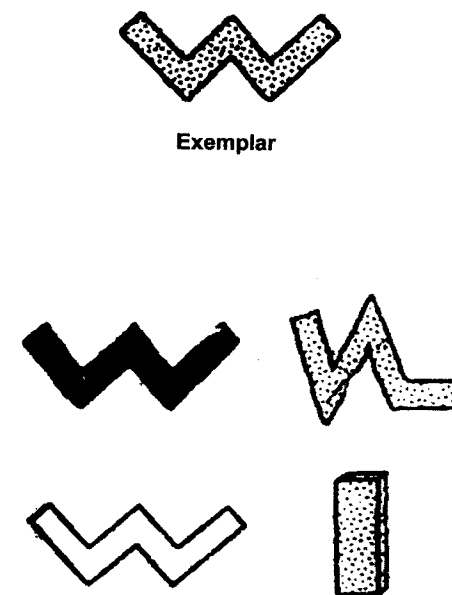


Fig. 4.5 Stimuli used in Smith, Jones, & Landau (1992). The colors of the exemplar and test objects were realized with glitter.

learned something about the correlation between sneakered things, eyes, and the relevance of texture to lexical categorization. These results suggest that children will learn and make use of any regularity when naming an object. Children's smart attention in the service of naming derives from mechanisms apparently open to arbitrary influence.

1.2 The role of syntax

The form of the word learning bias also depends on the syntactic frame in which the novel word is placed. One relevant contrast is that between the syntactic frame associated with count nouns, "this is a dax," and one associated with adjectives, "this is a dax one." A number of investigators have shown that children attend to different object properties in these two linguistic contexts (Smith, Jones, & Landau 1992; Landau, Smith, & Jones 1992; Waxman & Markow 1998; see also Au & Laframboise 1990). One relevant study (Smith, Jones, & Landau 1992) investigated 36-month-olds' shifting attention to color and shape when asked to interpret a novel word presented in the two syntactic contexts. The stimuli, shown in figure 4.5, varied in shape and color. The colors were realized by putting glitter in paint. In two separate experiments, these stimuli were presented either

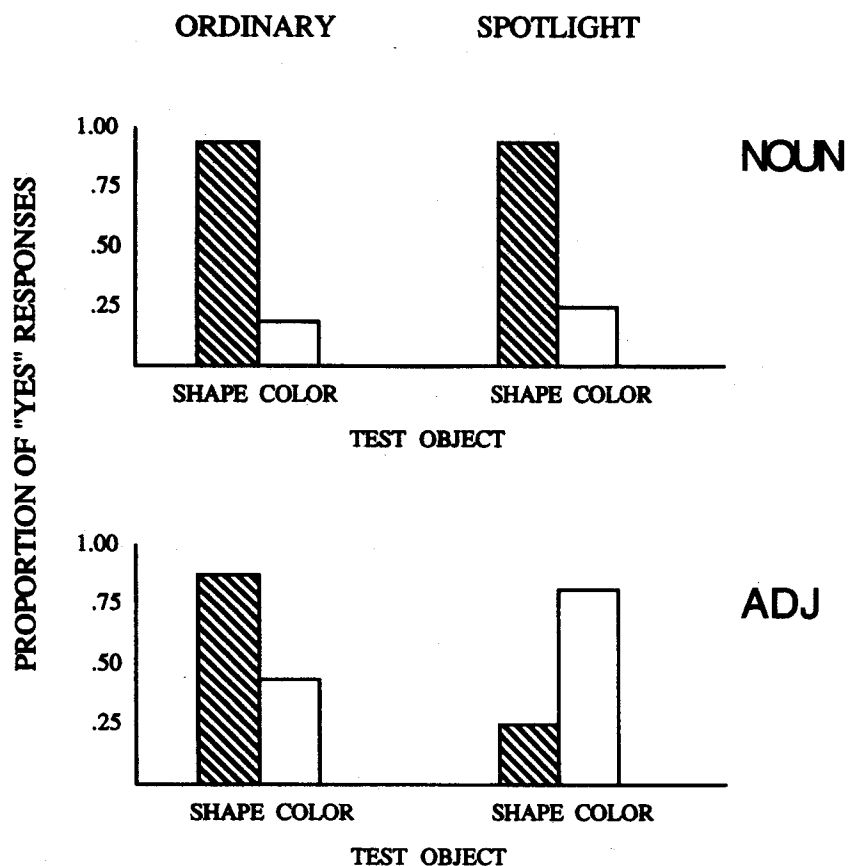


Fig. 4.6 Mean proportion of times the novel word in either a count noun or adjectival frame was extended to test objects that matched the exemplar in either color only or shape only under ordinary illumination or under a spotlight.

under ordinary illumination or in a darkened chamber with a spotlight. The effect of the spotlight was to make the glitter sparkle.

The results are shown in figure 4.6. The spotlight had little effect when the novel word was in a noun frame but controlled attention when the novel word was in an adjective frame. Children understood the novel noun as referring to objects the same shape as the exemplar under both ordinary and spotlight illumination. But children understood the novel adjective differently in the two lighting conditions: as referring to particularly shaped objects under ordinary illumination and as referring to the sparkling color under the spotlight. In other words, attention to shape in the context of a novel noun withstands the challenge of sparkling glitter but attention in the

context of a novel adjective is organized differently when there is a sparkling color vs. when there is not. These results show that children's attentional biases in the service of word learning depend on *both* the properties of the labelled object and the linguistic context.

Soja's (1992) findings about two-year-old children's interpretation of novel words in count- and mass-noun frames makes the same point. In these experiments, she used the rigid and nonrigid stimulus sets depicted in figure 4.3 but named the exemplar either with a count noun, "This is a mell," or with a mass noun, "This is some mell." She found that children's novel word generalizations were different in the two contexts. Children were more likely to attend to shape in the context of a count noun than a mass noun. Moreover, syntactic frame interacted with object properties. Shape choices dominated given rigid objects in both syntactic contexts but substance choices dominated only in the context of a mass-noun frame *and* a nonrigid substance. Soja concluded that children possessed knowledge about both count/mass syntax and the perceptual properties critical to distinguishing objects and substances.

These findings about object properties and syntactic frames constitute quite strong evidence that children know a great deal about how words map to categories. Their attentional biases in the context of word learning are intelligently dependent on the kind of object and the kind of word.

1.3 Attentional biases change with development

Where does this knowledge about how words map to different kinds of categories come from? Figure 4.7 summarizes the many developmental experiments that have been conducted on children's biased word learning (Landau *et al.* 1988; Soja *et al.* 1991; Landau *et al.* 1992; Smith *et al.* 1992; Soja 1992; Waxman 1994). Each curve summarizes the developmental pattern given specific stimulus properties and/or syntactic context. Briefly, the evidence suggests that by twenty-four months there is an early tendency in naming tasks to attend to shape that gets both stronger with development and more specific to specific contexts. Descriptively, other biases seem to grow out of, differentiate out of, this earlier less-articulated shape bias.

The data summarized in figure 4.7 all derive from children learning English. If I were to add the growing evidence from children learning other languages – Korean, Japanese, Spanish, or Yucatec Mayan – I would have to add other curves. The work of Bowerman & Choi, Gentner & Imai, Gaskins & Lucy summarized in this volume, as well as studies by Waxman (1994), strongly indicate that children learning different languages develop different attentional biases. In all these languages, children seem to have domain-specific attentional biases that are smartly specific to the language being learned. How can this be?

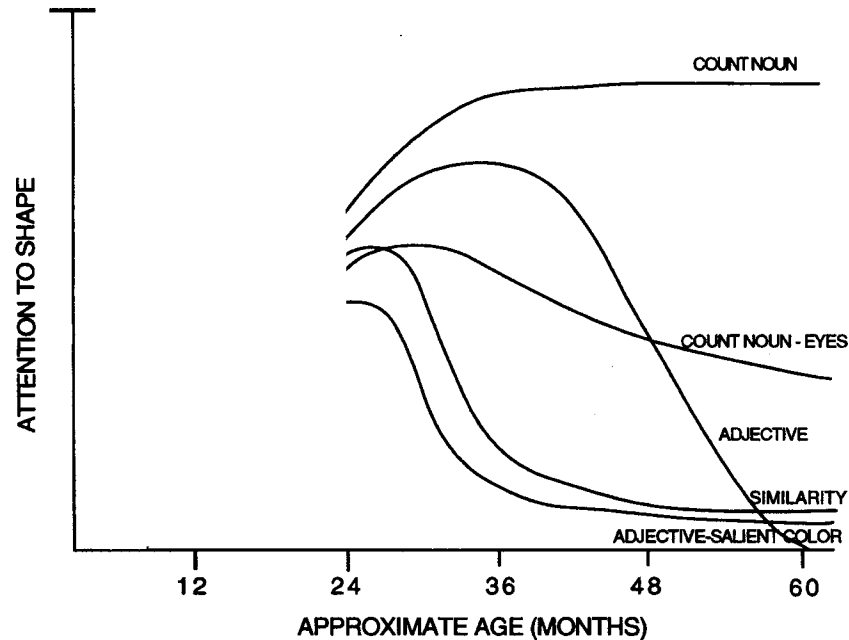


Fig. 4.7 Theoretical developmental trajectories summarizing findings across experiments.

2 An attentional learning account

One hundred years of research provides an answer in the most basic and universal processes of attentional learning (James 1890; Rescorla & Wagner 1972). When one perceptual cue is regularly associated with another, attention to the first mandatorily recruits attention to the second. The automatic control of selective attention by associative learning is one of the most widespread and well-documented phenomena in all of psychology. It is a fundamental process, evident in infants, children, adults, and nonhuman animals (e.g. MacIntosh 1965; Medin & Wattenmaker 1987; Lewicki, Hill, & Sasak 1989; Younger, 1990; Kruschke 1992). It is a process that may also be sufficient to explain young children's selective generalizations of a novel word to new instances.

The viability of this idea is suggested by the strong resemblance between the contextual control of children's attention in novel word learning tasks and the contextual control of selective attention that emerges in well-controlled studies of attentional learning. In such studies, some cue is regularly associated with attending to some property; and the presence

of that cue comes to recruit attention to the associated property (e.g. MacIntosh 1965; Rescorla & Wagner 1972; Lewicki *et al.* 1989; Younger 1990). In learning language, children repeatedly experience specific linguistic contexts (e.g. "This is a ____" or "This is some ____") with attention to specific object properties and clusters of properties (e.g. shape or color plus texture). By hypothesis, then, these linguistic contexts come to serve as cues that automatically control attention. In so doing, they make attention in the context of language learning different from attention in other contexts and they push word learning in certain directions that will depend on the language being learned. This account affirms that children know a lot about how words map to categories. But it implements this knowledge in very general learning and attentional processes, associative connections, and attention weights, which themselves have no domain-specific content.

There are three points in favor of this account that merit recognition.

(1) The posited processes are nondeliberative, nonthoughtful, nearly reflexive. This is advantageous because young children, two- and three-year-olds, are not typically deliberative and reflective. They are particularly poor testers of explicit hypotheses (e.g. Kemler 1978) and particularly poor at strategically controlling their attention (see Aslin & Smith 1988). Thus, it is to the advantage of this account that the proposed mechanisms are also dumb and take no conscious thought.

(2) The posited processes are known to exist. Although some might question whether associative processes are sufficient to account for much in human cognition (e.g. Keil 1994), their existence is not in question. Associative processes of attentional learning *are* part of children's biology and thus could be the mechanism behind children's smart interpretations of novel nouns.

(3) These processes can be demonstrated to create attentional biases. The hypothesized mechanisms can be shown to work. For example, simulating learned attentional biases specific to specific contexts is easily achieved by connectionist networks of the most generic sort (e.g. Gasser & Smith 1991, 1998; Smith 1993, 1995).

These three points indicate the plausibility of the idea that children's domain-specific word learning biases originate in domain-general processes. In the next section, I present data from an ongoing series of experiments that provide further support for this account.

3 Tests of an attentional learning account

The proposal is that learned cues that automatically shift attention underlie the entire pattern of results summarized by the changing curves in figure

4.7. The idea is that early in word learning, the linguistic context “This is a _____” becomes associated with attention to the shape of rigid things *because* many of the words that children learn refer to categories of rigid objects similar in shape. A learned shape bias may be first among learned attentional biases because the correlation between shape similarity and lexical category is the broadest and most general in language to children (see Smith 1995). Other more local attentional biases will develop from more local statistical regulations as more words are learned. Attention will become modulated by the presence of eyes, or nonrigidity, or a mass-noun frame. The long-range theoretical goal is to explain all of this and especially the emerging complexity of attentional biases. However, our initial tests concentrate on the origins of children’s attention to shape in the context of naming a rigid thing.

3.1 A longitudinal study

Many of the first nouns that children learn, nouns like *table*, *cat*, *bottle*, and *truck*, name objects similar in shape. The hypothesis is that learning enough nouns like this causes the act of naming to become a contextual cue that automatically recruits attention to shape. The key prediction that follows from this hypothesis is that a lexically specific shape bias should not be evident prior to language but should emerge only after some number of nouns has been acquired.

Susan Jones, Barbara Landau, and I tested this prediction in a longitudinal study of eight children from fifteen to twenty months of age. We tracked the children’s vocabulary growth by having parents keep diaries of all new words spoken by their child. We measured the emergence of a shape bias by having the children come into the laboratory every three weeks to participate in an artificial word generalization task. At the beginning of the study, the children had very few words in their productive vocabulary (less than 15). At the end of the study, each child had over 150 words and for each child more than half of these were nouns that named concrete objects. Thus, if the shape bias is learned from learning words, children should not show a shape bias at the beginning of the study but should show one at the end – after they had learned a number of object names.

The stimuli used in the laboratory task are shown in figure 4.8. They include an exemplar object made of wood and three test objects that matched the exemplar in either color, shape, or texture. Because our subjects were so young at the start of the study, we modified the novel word generalization tasks used in our previous experiments as follows: the task began with the experimenter putting the exemplar and three test objects on the table in front of the child. The experimenter picked up the exemplar and said “This is a dax. Look, this is a dax.” Then while still holding the exem-

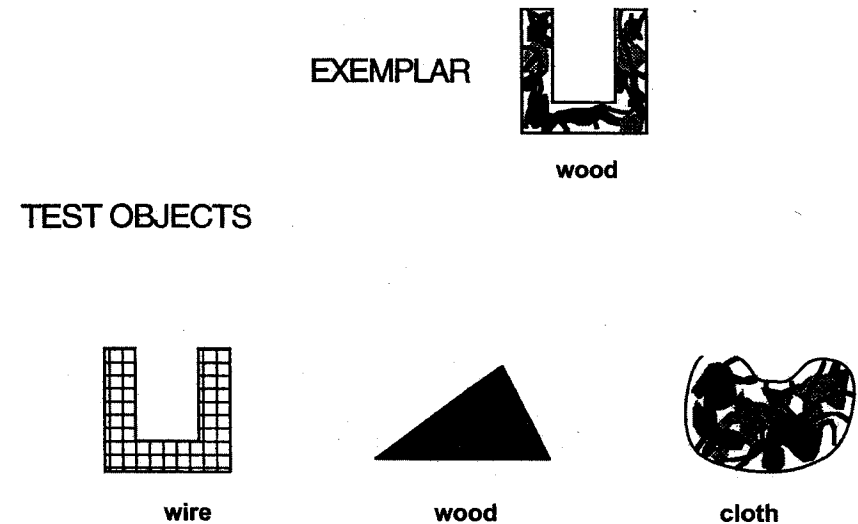


Fig. 4.8 Stimuli used in the longitudinal study of the development of the shape bias.

plar, the experimenter held out her other hand, palm up, and said, “Give me a dax. Give me another dax.”

The results are shown in figure 4.9. As is apparent, shape choices did not predominate early in the study but did by the end. More specifically, all eight children began to systematically extend the novel word *dax* to the same-shape test object *after* they had fifty object names (and about eighty total words) in their productive vocabulary. This point (fifty object names) in the word learning trajectory is *after* the spurt in noun acquisitions commonly known as the “naming explosion” as defined by Gopnik & Meltzoff (1987; see also Dromi 1987; Gershkoff-Stowe & Smith 1997). The developmental timing of biased attention to shape thus suggests that it may be the consequence of word learning – the consequence of learning some number of names for shape-based categories.

3.2 A cross-sectional study

We conducted a cross-sectional replication of the longitudinal study for two reasons. The first purpose was to ensure that the repeated laboratory visits had not somehow created a shape bias. The second purpose of the cross-sectional study was to determine whether the shape bias was lexically specific when it first emerged. It should be if it is the product of an associative link between naming and attending to shape.

Sixty-four children between the ages of eighteen and twenty-four

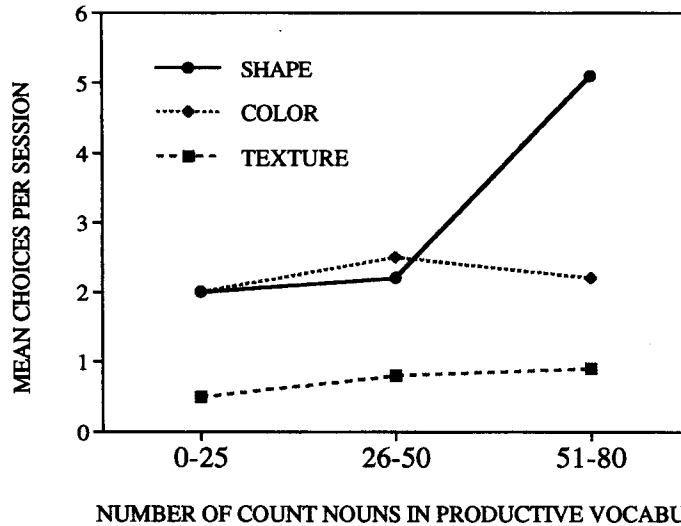


Fig. 4.9 Mean number of selections of test objects that matched the exemplar in shape, color, or texture as a function of the mean number of object names in the child's productive vocabulary.

months participated in the cross-sectional study. They came to the laboratory just once. Productive vocabulary was measured by having parents complete the MacArthur Toddler Communicative Development Inventory. This is a parent report measure which catalogues 625 of the most common early words and for which there is extensive normative and reliability data (Fenson, Dale, Reznick, *et al.* 1993). From this measure, children were assigned to four groups according to the number of nouns in their productive vocabulary: 0–25 nouns, 25–50 nouns, 50–75 nouns, and 75 and more nouns. Half the children at each level of productive vocabulary participated in a novel word generalization task. This Naming task was structured similarly to that used in the longitudinal study. Half the children participated in a Nonnaming task. All aspects of this task were the same as in the Naming task except the exemplar was not named. The experimenter merely held up the exemplar and said “Look, look at this” and then held out her hand and said “Get me one.”

The key questions of this experiment are whether children systematically select test objects the same shape as the exemplar, whether they do so only in the Naming task and not in the Nonnaming task, and whether their selective attention to shape emerges only after they have acquired some number of nouns. The data are shown in figure 4.10 and, as can be seen, the

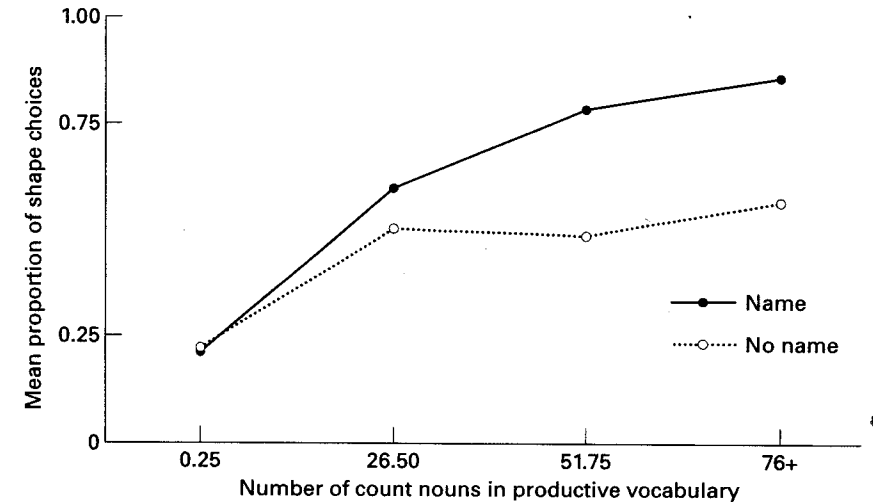


Fig. 4.10 Mean number of selections of test objects that matched the exemplar in shape in the Nonnaming and Naming tasks for children grouped according to the number of nouns in their productive vocabulary.

answer to each of these questions is “yes.” There is a rise in shape choices as a function of vocabulary growth and it is specific to the Naming task. After children have acquired more than 50 nouns, it is the linguistic context of naming a novel object that recruits attention to shape. These results along with those of the longitudinal study fit the idea that learning words *creates* a shape bias by creating a contextual cue so regularly associated with attention to shape that the presence of that cue automatically shifts attention to shape.

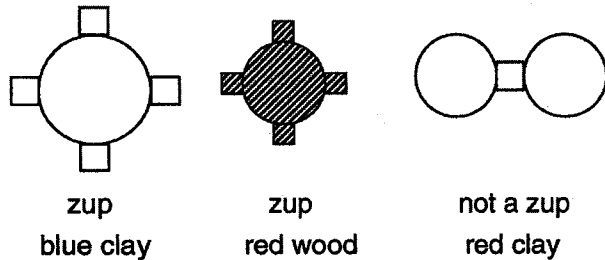
3.3 A training study

Correlation, of course, is not causation. In order to provide stronger evidence that a lexically specific shape bias is the *consequence* of learning names for rigid things, we attempted a training experiment. The goal was to create biased attention to shape by teaching lexical categories well-organized by shape to children who did not yet know many words. The subjects were seventeen months of age and produced an average of 12 nouns at the start of the study. These children came to the laboratory for seven weeks and were given extensive training on four different novel categories – all well-organized by shape. The top of figure 4.11 illustrates the training stimuli for one lexical category.

Each lexical category was trained as follows: the two exemplars for a

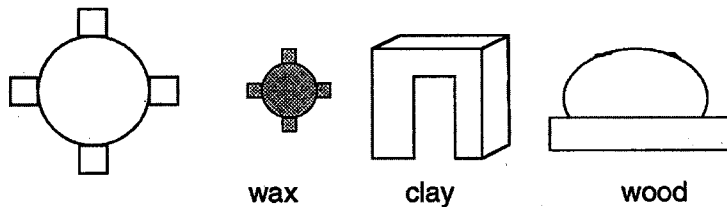
Training set

The zup set



Test set

Week 8: trained lexical categories



Week 9: novel lexical categories

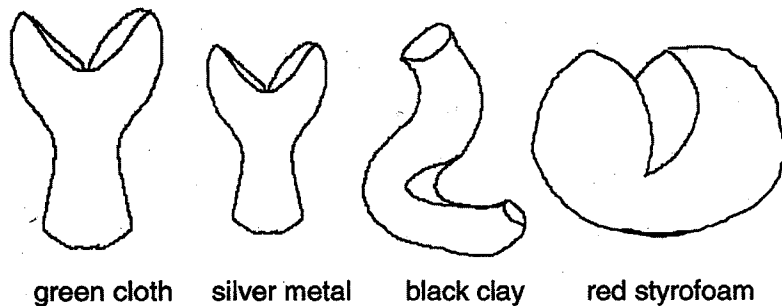


Fig. 4.11 Sample stimuli from the Training experiment: top – training stimuli; middle – stimuli for the test of generalization of the trained object name (week 8); bottom – stimuli for the test of generalization of novel names (week 9).

category were placed on the table and named (e.g. “This is a *zup*. Here is another *zup*.”) As illustrated in figure 4.11, these two exemplars differed in many ways but were identical in shape. The experimenter and child played with these two objects for five minutes during which time the experimenter repeatedly named the objects (e.g. “Put the *zup* in the box. Can you put the *zup* in the wagon?”). Midway in a play session with one pair of exemplars, a nonexemplar for that category (see figure 4.11) was briefly placed on the table. The experimenter announced that this just-introduced object was not a member of the category (e.g. “That’s not a *zup*!”) and then removed it. This nonexemplar matched each exemplar in one nonshape attribute but differed from both exemplars in shape. This nonexemplar thus provides the child with negative evidence as to the kinds of things that are *not* in the lexical category.

For the first seven weeks of the experiment, the children were trained as described above on each of four lexical categories. In weeks 8 and 9 of the experiment, the children participated in two test sessions that asked them to generalize what had been learned over the first seven weeks. The first test session, in week 8, measured children’s generalizations of the trained lexical categories to new instances. The structure of this task was identical to the novel word generalization task used in the longitudinal and cross-sectional studies described earlier. The stimuli used to test generalization of the *zup* category are illustrated in the middle section of figure 4.11. The test began with the experimenter placing one of the trained exemplars on the table along with three novel objects: one which matched the exemplar in material, one in color, and one in shape. The experimenter picked up the exemplar and said “This is a *zup*,” then asked, “Get me a *zup*.” For these children after their seven weeks of training, the exemplar is not a novel object and the label provided is not a novel name. Thus, if the children have learned that the specifically trained names refer to objects of a particular shape, they should generalize these already-learned names to the novel object that is the same as the exemplar in shape.

At week 9, the children were tested in a novel word generalization task structured in the same way as the generalization task at week 8. However, as illustrated by the sample stimulus set at the bottom of figure 4.11, all the objects and names were new. This generalization task thus tests the critical prediction that learning specific categories well-organized by shape transforms the task of naming into a contextual cue that automatically shifts attention to shape. If the seven weeks of intensive training on shape-based categories has caused the linguistic context of naming to cue attention to shape automatically, then these children should form and generalize *novel* names on the basis of shape.

This experiment also included ten Control children. These children did

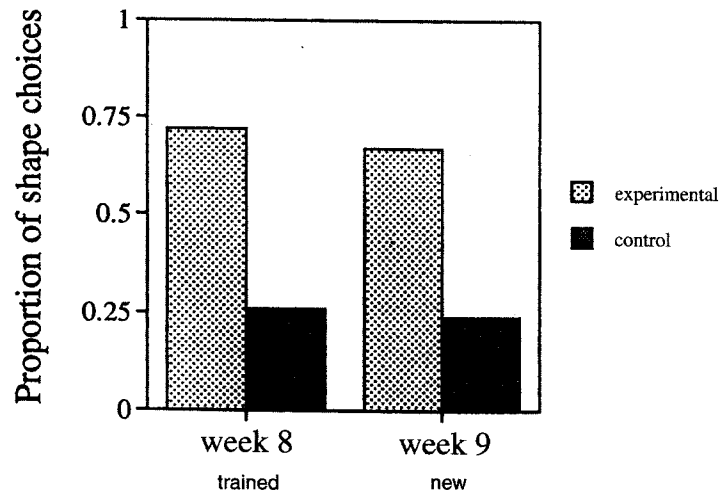


Fig. 4.12 Proportion of choices of test object the same shape as the named exemplar on the two generalization tests: week 8 – trained name and exemplar; week 9 – novel name and exemplar.

not participate in the seven weeks of training but were tested in the generalization tasks of weeks 8 and 9. The Control children were matched to the Experimental children in the number of nouns in their productive vocabulary at week 8 (mean = 26 for both groups). Since the Control children have learned few nouns and have not received intensive training on shape-based lexical categories, the expectation is that these children will not selectively attend to shape in the generalization tasks.

The main results are shown in figure 4.12. At week 8, when the Experimental children were asked to generalize the trained names to new objects, they did so on the basis of shape. These children have clearly learned that the words we taught them refer to objects of a particular shape. The Control children, for whom this is a *novel* word interpretation task, did not systematically attend to shape.

At week 9, both the Experimental and Control children heard novel objects named by novel nouns. However, the Experimental children, but not the Control children, systematically generalized these newly learned names to other novel objects by shape. In brief, we taught the Experimental children four categories organized by shape but they learned more than just these categories. They learned to attend to shape when novel rigid objects are named. This generalized attentional shift *is* a learning bias. These children are now biased to learn about shape when a novel object is named.

3.4 Summary

I proposed that, in the course of early word learning, children learn contextual cues (linguistic context, object properties) that automatically shift attention to specific object properties in that context. Three preliminary experiments in an ongoing research project were presented. The results support the attentional learning account by showing (1) that the shape bias in early naming does not precede language learning, (2) that it is lexically specific when it first emerges, and (3) that its emergence is accelerated by training on lexical categories well-organized by shape.

Clearly, more evidence is needed. Accelerating a bias by training is not as powerful evidence for learning as, for example, teaching a new and arbitrary bias. We are attempting such a stronger test of our attentional learning account. However, the evidence to date on the origins of the shape bias are encouraging. They suggest that at least in this case of word learning, young children's domain-specific smartness may originate in domain-general processes – in ordinary associative learning and the general mechanisms of selective attention.

4 Self-organizing biases

Several theorists in cognitive development have expressed doubt that the smartness of children's early word learning could emerge from general processes such as associative connections and the automatic control of attention weights (see e.g. Markman 1989; S. A. Gelman & Medin 1993; Keil 1994). In these views, the directedness of early word learning is too smart, too constrained, to be based on unconstrained associative processes. There is, however, much power in a single learned attentional bias, the power to start learning in directions that will build on themselves, the power to *create* a destiny. This point can be made by considering the cascading effects of learning just one context cue that automatically shifts attention.

Consider a two-year-old – say one of the children in the Experimental condition of the training study – who has learned enough for “That’s a _____” in the context of a rigidly shaped object to cue attention to shape automatically. Imagine further that this is the *only* linguistically relevant attentional cue that the child has learned. From all that is known about associative learning, one would expect, as illustrated in figure 4.13, a generalization gradient around the original learning: the child’s attention to shape should be maximal in contexts most similar to the original learning and should decrease in word learning situations as those situations become increasingly dissimilar to the context of original learning.

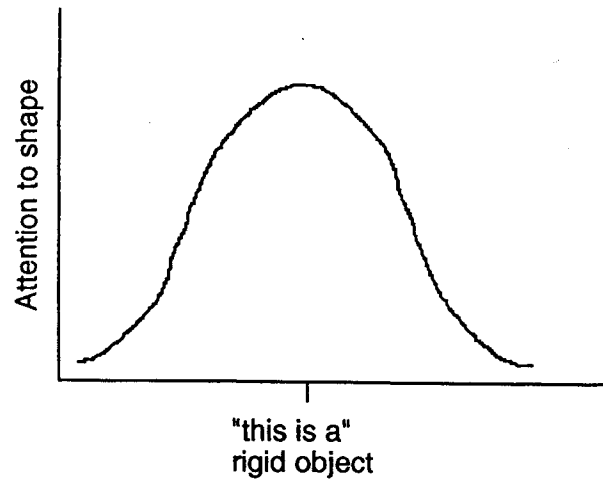


Fig. 4.13 A hypothetical generalization gradient.

By this account, given only a learned association between the count-noun frame, object rigidity, and attention to shape, one would expect the pull of shape to be less if the novel word were embedded in an adjectival rather than a count-noun frame. Because the pull to shape of the utterance "This is a dax one" is less than "This is a dax," the child will be likely to learn something different about a novel word in the two different frames. The child might, for example, learn that the word refers to sparkling colors in one frame but shape in the other. In this way, "knowing" only about count nouns means learning differently about nouns and adjectives.

The purpose of this account is not to "explain away" children's knowledge about adjectives. Rather, the proposed explanation is interesting because it shows what might be the mechanism through which an understanding of the semantic forces of nouns and adjectives develops. If children are less likely to attend to shape in the context of an adjective – whatever the reason for this decreased likelihood – then they are more likely to learn differently about adjectives and nouns. Put another way, *because* the generalized association to shape is weaker and more easily challenged in the context of a novel adjective than a novel noun, children with just a learned association between count nouns and shape will be already *biased* to learn differently about adjectives and nouns. In this sense, they do "know" something about adjectives but that knowledge is implemented through a single association that is not, strictly speaking, "about" adjectives at all. Nonetheless, such a "dumb" mechanism may start the development process in the right direction just as well as specific knowledge about

adjectives. A simple generalization gradient creates an opportunity for learning differently in different contexts – learning that may culminate in the sophisticated use of lexical contrast (e.g. Au & Markman 1987; Au & Laframboise 1990) and in crosslinguistic differences in learning about nouns and adjectives (see Waxman 1994).

A similar account may be the right one for young children's differential interpretations of mass and count nouns. Recall that Soja (1992; Soja, Carey, & Spelke 1991) found that children's generalizations of a novel noun were modulated by both count/mass syntax and the rigidity of the labeled object: children attended to shape more in the context of a count noun and a rigid object but attended to material more in the context of a mass noun and a nonsolid substance. Soja concluded that children possessed knowledge both about count/mass syntax and the perceptual properties critical to distinguishing objects and substances.

This knowledge, however, could be implemented only by a strong association between a count-noun frame, rigidity, and attention to shape. Figure 4.14 shows a redrawing of Soja's data. The proportion of shape choices are shown as a function of the hypothesized dissimilarity of each kind of test trial from the context of hypothesized original learning: a count-noun naming a shape-based category of rigid objects. Children's shape choices in Soja's experiments clearly suggest a generalization gradient; shape choices decrease as similarity from the prototypical naming context decreases. But this generalization gradient means that these children are effectively biased to learn differently about how rigid and non-rigid objects are named and to learn differently about novel words in mass- and count-noun frames. Again, knowledge about one kind of word creates – by its very existence – biases for learning about other kinds of words.

Notice also what is happening with development. The differences between younger and older children in Soja's experiments suggest discriminative learning; the generalization gradient is steeper for older than for younger children just as would be expected if older children now have two contrasting associations – the original count-noun-rigid-shape association and the newer mass-noun-nonrigid-material association.

Such changes in the generalization gradients with development are predicted by the attentional learning account. Learning creates context cues that shape future learning; this future learning in turn creates new contexts for new learning. Each new word the child learns will change what the child knows about learning words – adding to, strengthening, weakening associations among linguistic contexts and attention to object properties. The history of these events will be laid on top of one another, making development. And thus, when we look back from adulthood at the developmental

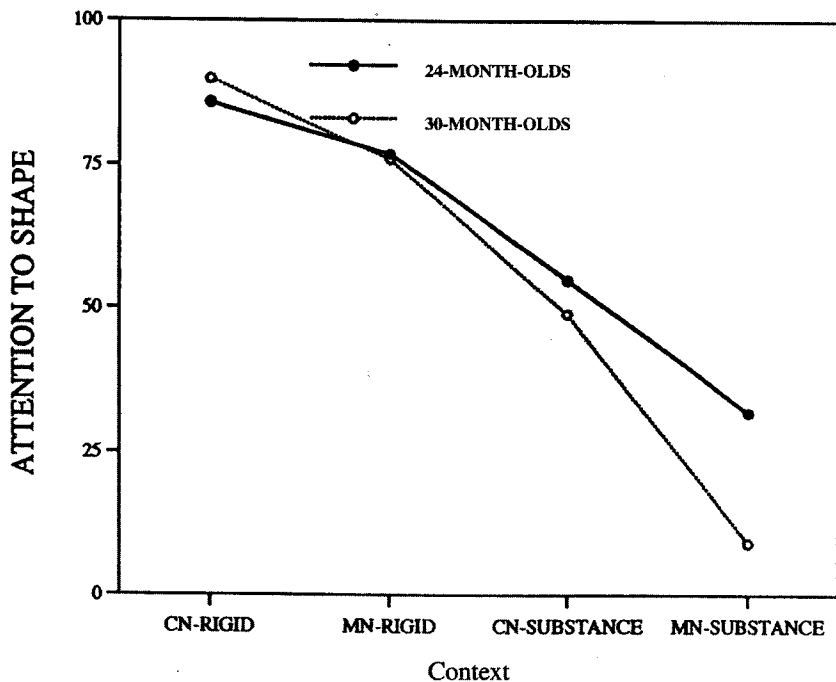


Fig. 4.14 Mean percentage choice of the test object that matched the exemplar in shape in Soja (1992) as a function of the hypothesized dissimilarity of the testing condition from the context of original learning. Chance equals 50 percent. Redrawn from the data reported in Soja (1992).

trajectory that brought us to maturity, it will look predestined. In fact, however, it will have been made in a history of possibilities.

5 Developmental process

When we look at development, it seems directed – so directed that it is hard to imagine how it is not predetermined and prescribed from within. Since this directedness is obvious in so many content-specific domains, it would also seem that there must be domain-specific mechanisms and processes that direct that development. The evidence presented in this chapter suggests how specialized developmental mechanisms could themselves be a product of development, the outgrowth of the history of activity of more fundamental and universal processes.

This idea, that specialization emerges from general processes, is the very idea of biological development and one whose truth can be seen at multiple

levels of analysis. Consider, as one example, how animals get their basic body parts – the parts and organs that emerge in orderly fashion in the first weeks of life after conception (for more detail, see Wolpert 1971; Gierer 1981; Marx 1984a, b; Cooke 1988). The initial state from which these parts emerge is a seemingly homogenous and formless single cell. It makes copies of itself that are all the same. Formation of body parts occurs when there are about 10,000 of these cells amassed in an undifferentiated heap. But at this point, cells are already marked by their position in the mass to become distinct body parts. They are so marked by gradients of substances. These gradients are the consequence of pervasive processes not specific to these cells or this kind of organism; they arise, for example, from the “mundane” effects of gravity and the mechanical effects of molecular structure in the cell and at its surface. These gradients emerge out of the geography of this amorphous mass of cells. Remarkably, the effect of these gradients is to switch on and off the regulating genes in the nucleus of individual cells. These genes have their effects by making proteins; in so doing, they change the activity of once-identical cells into different kinds. They also change the environment of those cells, thus creating the context for more changes. In one probabilistic event after another, arms and legs and stomachs and lungs are created with seeming certainty. It happens because each event creates the context for, and thus constrains, what can happen next.

There are two points to note from this example: (1) it is the specific *history* of events and the contexts they create that makes specificity, that makes some cells destined to be of one kind and others to be of another kind; and (2) the processes that make development, that make the special parts of the body – protein synthesis, gravity, and the mechanical and bio-mechanical actions of cells – are very general processes. They are what all life is made of.

A similar story at the psychological level has been offered by O'Reilly & Johnson (1994) in their recurrent network model of imprinting in precocial birds. Their account presents a parallel case to embryology in that the specialized process of imprinting is a consequence of a pervasive external reality and the most general processes known to make up intelligence. The external regularity is the fact that objects tend to persist across space and time. The general processes are Hebbian learning, excitatory connections, and lateral inhibition. None of these is a prescription for imprinting or its sensitive period. But together in a particular history of events, O'Reilly & Johnson have shown they are enough to create a bias to fixate on the first conspicuous object seen after hatching.

The architecture of O'Reilly's & Johnson's model, illustrated in figure 4.15, is explicitly based on what is known about the patterns of connectivity in the chick brain. The network consists of three layers of units; each unit is

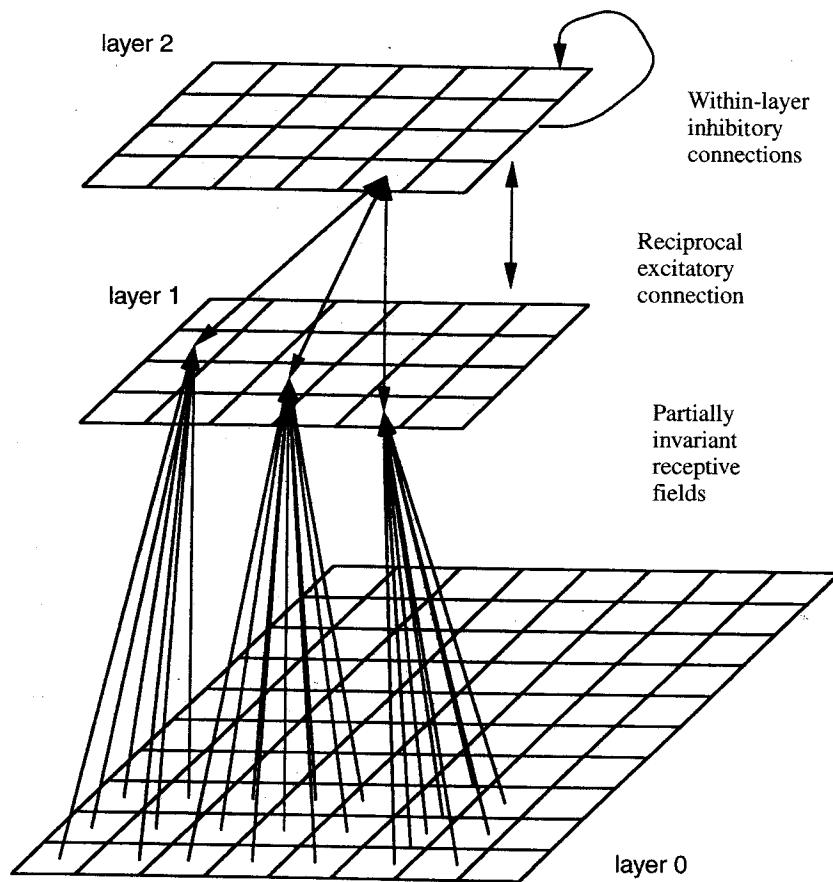


Fig. 4.15 The architecture of O'Reilly & Johnson's network.

represented by an individual rectangle in the grid. Spatially close units in the input layer (Layer 0) connect to the same units in Layer 1, creating spatially organized receptive fields, a common pattern of connection in many regions of the brains of many different kinds of animals. In the network, the connection weights between these two layers can change through Hebbian learning; thus the receptive fields on Layer 1 can vary within limits. The units on Layer 1, in contrast, project in an invariant fashion to those on Layer 2, but again in a manner that is common throughout sensory systems. Lateral inhibition is present in each layer of the model in the form of relatively large negative weights between all units in layer. Reciprocal (bidirectional) excitatory connections exist between Layers 1 and 2. Finally, there are recurrent connections. The activity in a unit

depends not just on its immediate input from below but also on its own past activity and the past activity of the network as a whole. These building blocks are all the common stuff of brains and networks.

The experience O'Reilly & Johnson gave the network is also common stuff – looking at individual objects presented one at a time for varying durations, experiences that might correspond to ten minutes or one hour of viewing the same object. That is all there is to the model. But the outcome of these experiences is preferential recognition of the first object seen and the emergence of a self-terminating sensitive period. Quite simply, if early experience consists of an object that persists for sufficient duration, the strength of that bias cannot ever be overcome; through lateral inhibition and recurrent connections, it maintains itself. If, in contrast, early experience consists of sufficiently many different nonpersisting objects, no preference emerges and the possibility of developing such a preference is lost. The point of this example, like the one from embryology, like the attentional learning account of children's word learning biases, is that something special can *develop* out of the history of activity of general, indeed mundane, processes that subserves many other specializations.

Some readers may be questioning how this approach can explain what seems to be truly special about human cognition and truly smart – systematic belief systems and conscious, self-reflective thought. Could these possibly develop out of mundane, generic cognitive processes such as those considered here? The answer seems inescapable to me. If human cognition is a biological process, if it has a material cause, then it is made of processes, of bits of matter and connections, that are themselves *much less than what they yield*.

It is commonplace in psychological theorizing to impute to internal workings a "copy" of externally observed behavior (Smith & Thelen 1993). Thus, sucking in infants is explained by a sucking reflex, statements of beliefs by internally represented beliefs, categorization by represented categories, and syntax by an innate grammar. In each case, an abstract and sometimes truly elegant icon of the behavior to be explained is proposed as the mechanism that produces the behavior. There is a failure of imagination here that evokes the preformationism of eighteenth-century embryology (see Ausabel 1957).

It was once seriously believed that a miniature but fully formed little human being existed in the sperm and when implanted in the uterus, this little being fed off the egg and grew in bulk until it was big enough to be born. This preformationist theory eschews the complex developmental processes that we now know drive the formation of new structures in embryology. The eighteenth-century proponents of a preformationist embryology did not, of course, explicitly deny developmental process; the truth is, they

could not imagine it. They could not imagine that the organism at conception is undifferentiated and *develops* its structure through a series of complex and cascading interactions, through a series of small steps that are each mundane and ordinary.

As students of cognitive development, we are like the early preformationists. It is hard for us to imagine how beliefs, conscious thought, theories about space and time could emerge from anything *less* than beliefs, conscious thought, or theories. This is perhaps a failure of imagination to be replaced by scientific fact.

6 Conclusion

The phenomena considered in this chapter, attentional biases that are specific to and promote word learning, are themselves much simpler than the phenomena usually interpreted in terms of domain-specific belief systems (see, for example, ch. 7 by Carey, ch. 3 by Spelke & Tsivkin, and ch. 2 by Gopnik in this volume). Still, the evidence suggests that these word learning biases may emerge out of processes both more general and simpler than the conceptual and linguistic knowledge embodied in these biases. These simple attentional biases, once created, may then help to create more thoughtful beliefs and belief systems about kinds of categories and about language (see Smith & Heise 1992; Thelen & Smith 1994, for further discussion).

Evidence from several other chapters in this volume also points to a created rather than predetermined developmental destiny: Slobin's arguments against predetermined biases in favor of associations among semantic concepts and closed-class words; Bowerman's & Choi's evidence on the diversity of spatial concepts; and the considerable evidence in many chapters on crosslinguistic variation. All of these point to domain-specific knowledge first as a product of development and then as a cause that shapes development further.

In conclusion, development and cognition may be more wonderful than a diversity of mechanisms that is there to begin with. Children's initial generalizations of novel words present an intriguing case. These generalizations are smart; they are domain-specific, embodying knowledge about how kinds of words are mapped to kinds of objects. But the processes that make this domain-specific smartness may themselves be general and dumb.

NOTE

The research program reported in this chapter was supported by NIH grant HD28675.

REFERENCES

- Aslin, R. N., & L. B. Smith. 1988. Perceptual development. In M. R. Rosenweig & L. W. Porter (eds.), *Annual Review of Psychology* 39: 435–474.
- Au, T. K., & D. E. Laframboise. 1990. Acquiring color names via linguistic contrast: the influence of contrasting terms. *Child Development* 61: 1808–1823.
- Au, T. K., & E. M. Markman. 1987. Acquiring word meaning via linguistic contrast. *Cognitive Development* 2: 217–236.
- Ausabel, D. 1957. *Theories and problems of child development*. New York: Grune and Stratton.
- Biederman, I. 1987. Recognition-by-components: a theory of human image understanding. *Psychological Review* 94: 115–147.
- Chomsky, N. 1980. *Rules and representations*. New York: Columbia University Press.
- Clark, E. V. 1973. What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 65–110.
- Cooke, J. 1988. The early embryo and the formation of body pattern. *American Scientist* 76: 35–41.
- Dromi, E. 1987. *Early lexical development*. New York: Cambridge University Press.
- Fenson, L., P. Dale, J. S. Reznick, E. Bates, J. Hartung, S. Pethick, & J. Reilly. 1993. *MacArthur Communicative Development Inventories*. San Diego: Singular.
- Fodor, J. A. 1975. *The language of thought*. New York: Thomas Y. Crowell.
1983. *The modularity of mind: an essay on faculty psychology*. Cambridge, MA: MIT Press.
- Gasser, M., & L. B. Smith. 1991. The development of a notion of sameness: a connectionist model. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum, 719–723.
1998. Learning nouns and adjectives: a connectionist account. *Language and Cognitive Processes* 13: 269–306.
- Gelman, R. 1990. First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples. *Cognitive Science* 14: 79–106.
- Gelman, S. A., & D. L. Medin. 1993. What's so essential about essentialism? A different perspective on the interaction of perception, language, and concrete knowledge. *Cognitive Development* 8: 113–139.
- Gentner, D. 1978. What looks like a jiggy but acts like a zimbo?: a study of early word meaning using artificial objects. *Papers and Reports on Child Language Development* 15: 1–6.
- Gershkoff-Stowe, L., & L. B. Smith. 1997. A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology* 34: 37–71.
- Gierer, A. 1981. Generation of biological patterns and form: some physical, mathematical, and logical aspects. *Progress in Biophysics and Molecular Biology* 37: 1–47.
- Gleitman, L. 1986. Biological disposition to learn language. In W. Demopoulos & A. Marras (eds.), *Language learning and concept acquisition: foundational issues*. Norwood, NJ: Ablex, 3–28.
- Gopnik, A., & A. N. Meltzoff. 1987. The development of categorization in the

- second year and its relation to other cognitive and linguistic developments. *Child Development* 58: 1523–1531.
- Imai, M., D. Gentner, & N. Uchida. 1994. Children's theories of word meaning: the role of shape similarity in early acquisition. *Cognitive Development* 9: 45–76.
- James, W. 1890. *The principles of psychology*, vol. 1. New York: Dover Publications, Inc.
- Jones, S. S., & L. B. Smith. 1998. How children name objects with shoes. *Cognitive Development* 13: 323–334.
- Jones, S. S., L. B. Smith, & B. Landau. 1991. Object properties and knowledge in early lexical learning. *Child Development* 62: 499–516.
- Keil, F. C. 1989. *Concepts, kinds, and cognitive development*. Cambridge, MA: Bradford/MIT Press.
1994. Explanation, association, and the acquisition of word meaning. In L. R. Gleitman & B. Landau (eds.), *The acquisition of the lexicon*. Cambridge, MA: MIT Press, 169–196.
- Kemler, D. G. 1978. Patterns of hypothesis testing in children's discriminative learning: a study of the development of problem-solving strategies. *Developmental Psychology* 14: 653–673.
- Kruschke, J. K. 1992. ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review* 99: 22–44.
- Landau, B., S. S. Jones, & L. B. Smith. 1992. Syntactic context and the shape bias in children's and adult's lexical learning. *Journal of Memory and Language* 31: 807–825.
- Landau, B., L. B. Smith, & S. S. Jones. 1988. The importance of shape in early lexical learning. *Cognitive Development* 3: 299–321.
- Leslie, A. M. 1994. ToMM, ToBY and Agency: core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (eds.), *Mapping the mind: domain specificity in cognition and culture*. Cambridge: Cambridge University Press, 119–148.
- Lewicki, P., T. Hill, & I. Sasak. 1989. Self-perpetuating development of encoding biases. *Journal of Experimental Psychology: General* 118: 323–338.
- MacIntosh, N. J. 1965. Selective attention in animal discrimination learning. *Psychological Bulletin* 64: 125–150.
- Markman, E. M. 1989. *Categorization and naming in children: problems of induction*. Cambridge, MA: Bradford/MIT Press.
- Marx, J. L. 1984a. New clues to developmental timing. *Science* 226: 425–426.
- 1984b. The riddle of development. *Science* 226: 1406–1408.
- Massey, C., & R. Gelman. 1988. Preschoolers' ability to decide whether pictured unfamiliar objects can move themselves. *Developmental Psychology* 24: 307–317.
- Medin, D. L., & W. D. Wattenmaker. 1987. Category cohesiveness, theories, and cognitive archeology. In U. Neisser (ed.), *Concepts and conceptual development*. Cambridge: Cambridge University Press, 25–62.
- Mervis, C. B. 1987. Child-basic object categories and early lexical development. In U. Neisser (ed.), *Concepts and conceptual development: ecological and intellectual factors in categorization*. New York: Cambridge University Press, 201–233.
- O'Reilly, R. C., & M. H. Johnson. 1994. Object recognition and sensitive periods: a computational analysis of visual imprinting. *Neural Computation* 6: 357–389.

- Rescorla, R. A., & A. R. Wagner. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (eds.), *Classical conditioning*, vol. 2. New York: Appleton-Century-Crofts.
- Rosch, E. 1978. Principles of categorization. In E. Rosch & B. B. Lloyd (eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum, 28–46.
- Smith, L. B. 1993. The concept of same. *Advances in Child Development and Behavior* 24: 216–253.
1995. Self-organizing processes in learning to learn words: development is not induction. *The Minnesota Symposia on Child Psychology*, vol. 28: *Basic and applied perspectives on learning, cognition, and development*. Mahwah, NJ: Lawrence Erlbaum, 1–32.
- Smith, L. B., & D. Heise. 1992. Perceptual similarity and conceptual structure. In B. Burns (ed.), *Percepts, concepts, and categories*. Amsterdam: Elsevier, 233–272.
- Smith, L. B., S. S. Jones, & B. Landau. 1992. Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology* 28: 273–289.
- Smith, L. B., & E. Thelen. 1993. *A dynamics system approach to development: applications*. Cambridge, MA: MIT Press.
- Soja, N. N. 1992. Inferences about the meanings of nouns: the relationship between perception and syntax. *Cognitive Development* 7: 29–46.
- Soja, N. N., S. Carey, & E. S. Spelke. 1991. Ontological categories guide young children's inductions of word meanings: object terms and substance terms. *Cognition* 38: 179–211.
- Spelke, E. S., K. Breinlinger, J. Macomber, & K. Jacobson. 1992. Origins of knowledge. *Psychological Review* 99: 605–632.
- Thelen, E., & L. B. Smith. 1994. *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Waxman, S. R. 1994. The development of an appreciation of specific linkages between linguistic and conceptual organization. In L. R. Gleitman & B. Landau (eds.), *The acquisition of the lexicon*. Cambridge, MA: MIT Press, 229–250.
- Waxman, S. R., & D. Markow. 1998. Object properties and object kind: twenty-one-month-old infants' extension of novel adjectives. *Child Development* 69: 1313–1329.
- Wolpert, L. 1971. Positional information and pattern formation. *Current Topics in Developmental Biology* 6: 183–223.
- Younger, B. 1990. Infants' detection of correlations among feature categories. *Child Development* 61: 614–621.