J. Child Lang. **25** (1998), 267–291. Printed in the United Kingdom

Why is a pomegranate an *apple*? The role of shape, taxonomic relatedness, and prior lexical knowledge in children's overextensions of *apple* and *dog*[*]

SUSAN A. GELMAN
University of Michigan

WILLIAM CROFT
University of Manchester

PANFANG FU
University of Michigan

TIMOTHY CLAUSNER
University of Southern California

GAIL GOTTFRIED
Occidental College

(*Received* 16 *October* 1996. *Revised* 23 *August* 1997)

ABSTRACT

Children's overextensions (e.g. referring to a pomegranate as *apple*) raise intriguing questions regarding early word meanings. Specifically, how do object shape, taxonomic relatedness, and prior lexical knowledge influence children's overextensions? The present study sheds new light on this issue by presenting items that disentangle the three factors of shape, taxonomic category, and prior lexical knowledge, and by using a novel comprehension task (the screened-alternative task) in which children can indicate negative exemplars (e.g. which items are NOT *apples*). 49 subjects in three age groups participated (*M*s = 2;0, 2;6, and 4;5). Findings indicate: (1) Error patterns differed by task. In

---

[*] This work was supported by a University of Michigan Presidential Initiatives Fund grant to the first two authors, and NSF Grant 91-00348 and a J. S. Guggenheim Fellowship to the first author. A portion of the work was presented at the International Conference on Infant Studies, April, 1996 in Providence, RI. We are grateful to the parents and children who participated in the study, and to the University of Michigan Children's Center for their cooperation. Thanks to Cindy Andress, Dana Rosen, Allison Schloss, Jennifer Case and Lisa Haverty for their able research assistance. Address for correspondence: Susan A. Gelman, Department of Psychology, 525 E. University Ave., University of Michigan, Ann Arbor, MI 48109-1109, USA; e-mail: gelman@umich.edu.

production, errors were overwhelmingly due to selecting items that matched the target word in BOTH shape and taxonomic relatedness. In comprehension, more errors were based on either shape alone or taxonomic relatedness alone, and the nature and frequency of the overextensions interacted with prior lexical knowledge. (2) Error patterns also differed markedly based on the word being tested (*apple* vs. *dog*), in both comprehension and production. (3) As predicted, errors were more frequent in production than comprehension, though only for children in the two younger age groups. Altogether, the study indicates that overextensions are not simply production errors, and that both taxonomic relatedness and object shape play a powerful role in early naming errors.

INTRODUCTION

Toddlers have long charmed parents with their overextension errors: they say 'doggie' upon seeing a tiger in the zoo, or 'moon' upon seeing half a grapefruit. Just what these errors imply for acquisition, however, is a thorny issue. Do two-years-olds truly think that the word *moon* refers to a sliced grapefruit, or is this simply an error in children's productive use? The question gains special significance in light of recent suggestions that children may have a 'shape bias' when learning new words (Landau, Smith & Jones, 1988; Baldwin, 1992; Smith & Jones, 1993; Imai, Gentner & Uchida, 1994; Golinkoff, Shuff-Bailey, Olguin & Ruan, 1995). For example, in a word-learning experiment, a preschool child might learn that a high-heeled shoe is a *dax* and then extend the word *dax* to a sliding board. High-heeled shoes and sliding boards are roughly the same shape but are taxonomically unrelated. A finding that children tend to extend words based on shape alone would have broad implications for how language functions, suggesting either that children's word-learning biases undergo fundamental changes over development (from shape-based to taxonomically based) or that we need to reconceptualize what words mean even for adults.

Overextensions have the potential to provide particularly important evidence regarding the hypothesized shape bias. Unlike errors demonstrated in word-learning experiments, overextensions are not laboratory induced: they are unprompted, spontaneous, and may persist even after extensive input. Given that we do not yet know how word learning in the laboratory and word learning in real life may differ, overextensions are just the sort of converging evidence needed for more general conclusions regarding children's early linguistic knowledge.

The present paper examines the role of shape in children's overextensions, using a relatively new method that is more sensitive for revealing the status of early errors. In the remainder of the introduction below, we first discuss prior evidence regarding the role of shape in overextensions and early categorization, turn next to the methodological issue of distinguishing errors in production from errors in comprehension, and finally describe the design of the present study.

Role of shape in overextensions

Consistent with the 'shape-bias' claim, it is commonly reported that overextensions are based on perceptual features, especially shape. Clark's (1973) seminal review observed that most productive overextensions reported in published diary studies were based primarily on perceptual features (including movement, shape, size, sound, taste, and texture), although she noted that it is often difficult to determine which of two or more features is crucial (e.g. size and shape often covary). Similarly, features necessary to describe overextended word use often include perceptible parts that influence overall shape, such as absence of limbs (*fish*) or four legs (*dog*; Thomson & Chapman, 1977). Moreover, in a study with 13-month-old infants, Behrend (1988) found that some errors in comprehension appeared to be perceptually rather than taxonomically based (e.g. children looked longer at a wooden ring when hearing the word *cookie*). However, Behrend also acknowledged that further testing would be needed to examine the relative importance of conceptual and perceptual similarity (dimensions which, he noted, are not independent).

Other studies have found that overextensions are based on taxonomic relatedness. For example, in a study of children between 1;0 and 1;8, Rescorla (1980*a*) noted that overextensions tend to fall within appropriate taxonomic boundaries (e.g. *truck* was used for a variety of vehicles, *horse* was used for animals, and *apple* was used for fruits). Likewise, Huttenlocher & Smiley (1987) find that children's object names respect adult-like taxonomic boundaries.

In all of this work, it is acknowledged that perceptual and taxonomic relatedness are naturally confounded in the real world (cf. Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Rosch, 1978) and thus difficult to isolate. Indeed, we are unaware of any studies that systematically disentangle shape-based overextensions from other kinds of responses. For example, in one error characterized as 'related to shape', the child used *cat* to refer to dogs, cows, sheep, and horses – thus matching the referents on both shape and taxonomic kind (Clark, 1973). Likewise, Barrett (1978) characterized most overextensions in a sample reported by Leopold (1939) as 'perceptual', although they too tended to be marked by both perceptual and taxonomic similarities (e.g. *Papa* for adults). Similarly, although Thomson & Chapman (1977) listed PERCEPTUAL features necessary to describe children's overextensions (e.g. four legs, presence of eyes), they correctly noted that these may be perceptual analogues of abstract ontological features such as animacy. Thus, numerous examples of overextensions characterized as 'perceptual' are additionally bounded (implicitly) by taxonomic kind, and *vice versa* (i.e. taxonomic overextensions are also perceptually similar). More focused comparisons of carefully selected items are needed in order to determine if children favour shape over taxonomic kind (or *vice versa*) in their overextensions.

Role of shape in early categorization

In contrast to the overextension literature, studies of early categorization have attempted to disentangle shape-based from taxonomically based categories, often finding that quite young children are capable of forming categories based on dimensions that are orthogonal to shape. For example, by 9 months of age, infants sort together different basic-level animal categories (e.g. dogs and fish) and separate birds-with-outspread-wings from airplanes (Mandler & McDonough, 1993). Ten-month-olds classify together containers differing in shape and distinguish between same-shaped objects that differ in their capacity to contain (Kolstad & Baillargeon, 1996). By age 2 years, children weight substance more heavily than shape on a match-to-sample task on which the items are non-solid masses (Soja, Carey & Spelke, 1991). In addition, toddlers form inductive inferences based on taxonomic kind even when it conflicts with shape. Gelman & Coley (1990) found that by 2;8, children draw more inferences from a pterodactyl to a stegosaurus than from a pterodactyl to a bluebird. Language is crucial to this effect: without hearing a label, children draw inferences based on perceptual similarities (including shape). Altogether, these studies clearly demonstrate that words can convey information regarding properties other than shape at this age.

Methodological issues

As numerous scholars have observed, overextensions in productive speech may reflect errors of use rather than underlying word meanings. Bloom (1973) distinguishes word MEANING and word USE with the following example: 'It is almost as if the child were reasoning, "I know about dogs, that thing is not a dog, I don't know what to call it, but it is like a dog!"' (p. 79). In other words, with a limited vocabulary, children might overextend simply to cover gaps in their lexical knowledge. Thus, comprehension studies are needed to resolve the problem of what errors in production imply.

There are a variety of thoughtful comprehension studies in the literature. Nearly all conclude that overextensions are more frequent in production than comprehension (Kay & Anglin, 1982; Rescorla, 1980*b*; Mervis & Canada, 1983; Mervis & Mervis, 1988), although the degree of mismatch continues to be debated. Whereas most researchers conclude that some true overextensions in comprehension do exist (Anglin, 1977; Thomson & Chapman, 1977; Rescorla, 1980*a*; Kay & Anglin, 1982; Kuczaj, 1982; Mervis & Canada, 1983; Dromi, 1987), others (most notably Fremgen & Fay, 1980, and Chapman & Thomson, 1980) suggest that overextensions in comprehension reflect performance limitations entirely – i.e. there are no true overextensions in comprehension.

Despite the usefulness of examining comprehension in order to probe the source of overextension errors, by and large comprehension tasks themselves introduce additional interpretive problems. In a nutshell, standard forced-choice comprehension tasks can either underestimate or overestimate lexical knowledge. For example, suppose a child is asked for 'a dog', and is provided with two choices: a collie and a horse. A child who selects the collie may believe that both pictures are called *dogs*, but that the collie is a more typical instance (Kuczaj, 1982). In other words, seemingly correct use does not necessarily demonstrate correct extension of the word. Conversely, seemingly incorrect use may mask adult-like knowledge. For example, suppose a child is asked for 'a dog' while shown pictures of a horse and a book. The child, feeling pressure to make a selection, might select the horse despite knowing that it is not a dog. In evaluating examples of both overestimation and underestimation of lexical knowledge, the difficulty is that children have no opportunities to report category boundaries (e.g. which instances are NOT dogs).

Kuczaj (1982, 1986) used a method that reduces task demands and allows some opportunity to indicate category boundaries: on each of a series of trials, children saw multiple objects at once (e.g. in Kuczaj, 1982, children saw six objects: two appropriate exemplars, two objects to which the term had been overextended in production, and two obvious non-exemplars) and were asked to show/give the experimenter an X (e.g. 'a doggie'). After each choice, the question was repeated for that set until the child declined to pick any more objects. Because there were multiple instances per set, children presumably would feel more free to refrain from picking after the first one or two choices. Indeed, children rarely selected obviously incorrect exemplars (doing so on only 10% of trials), thus appropriately indicating word boundaries. However, children typically did select exemplars that had been overextended in production (62% of trials). Overall, the method has advantages over the standard two-alternative forced-choice task. Nonetheless, it does not entirely rule out the possibility of pressures on child subjects to continue selecting past the point they find appropriate. In other words, given that the experimenter directs the child to continue selecting objects even after the appropriate exemplars have been chosen, some children may decide to show the experimenter any related items (including overextended items), stopping only when they see no reasonable link between the experimenter's request and the objects remaining. Thus, although this method may reduce task demands, it would be beneficial to obtain converging evidence on the issue using a different procedure.

Other studies use preferential looking tasks, in which coders note how often the child gazes at each of two possible screens when a word is presented (Behrend, 1988; Naigles & Gelman, 1995). This method has the advantage of placing no explicit task demands on the child to choose one picture vs. another. Thus, in contrast to forced-choice tasks, there is no need to select just one picture, as the child is free to gaze at either or both pictures throughout the trial. However, this aspect of the task, though an advantage in some ways, also means that children's responses can be difficult to interpret. Because gaze does not require an explicit choice, it can reflect factors other than word-referent matching. For example, children who hear 'dog' may look at a cow because it REMINDS them of a dog, not because they think the cow is appropriately called a *dog*. Likewise, if the child is asked to 'Look at the dog' when the array does not include a dog, children may look at the exemplar they believe MOST CLOSELY RESEMBLES the referent named by the experimenter's word (Behrend, 1988). Casting a glance is not the same as matching word to referent, and instead may reflect children's awareness of item similarity or relatedness in semantic memory.

Not surprisingly, Hoek, Ingram & Gibson (1986) conclude: 'This area appears to be one where methodological advances are particularly necessary' (p. 492). One means of addressing these concerns was discussed by Dromi (1987, pp. 42–43):

> As Schlesinger (1982) argues, a procedure that could prove the hypothesis of only a production deficiency would be one that showed that children refused to choose a picture when asked to do so if no picture of a correct referent was included in the array presented to them. Such an experimental procedure cannot be attempted with children as young as those tested by Thomson and Chapman or Fremgen and Fay.

For example, it would be ideal to use a task in which children can inform us not only which instances are *dogs*, but also which instances are NOT *dogs*.

Although Dromi concluded that using such a procedure is not feasible with young two-year-olds, the task we have adopted allows us to do just that. The method is derived from studies recently conducted by Jean Hutchinson and her colleagues (Hutchinson & Herman, 1991; Hutchinson, Inn & Strapp, 1993), examining word-learning in children 1;2 to 2;2. Hutchinson's purpose was not to examine overextensions, but rather to distinguish between children's use of a mutual exclusivity principle and a strategy to fill lexical gaps when learning new words in an experimental context. However, the procedure is also suitable for examining overextensions. The gist of the task is as follows. For each trial, the child is presented with two pictures at a time. One of the two pictures is visible, and the other is hidden behind a cardboard screen. Children hear a word and are asked to select which picture is named by the word. In effect, subjects need to decide whether or not the visible picture matches the word. For example, the child is asked to find a 'dog' and is confronted with a visible picture of a cow and a screened picture. The correct answer, of course, is to choose the screened picture. A child who chose the cow instead would have displayed an overextension. We refer to this task as the screened-alternative task. Hutchinson *et al.* (1993) found above-chance responding on this task with children as young as 19 months of age.

Design of present study

The present study was designed to address two main questions: (1) Do children overextend in comprehension, when they are given a way of indicating semantic boundaries (i.e. which instances are NOT named by a word, using the screened-alternative task)? (2) If children do overextend on the screened-alternative task, then what are the bases of their overextensions? Specifically, we manipulated shape, taxonomic relatedness, and prior lexical knowledge. We included prior lexical knowledge because several theorists predict that children would be more likely to overextend a word to an item that does not already have a name than to an item that has a known name (Clark's (1988) principle of contrast; Markman's (1989) mutual exclusivity principle; Merriman & Bowman's (1989) lexical gap principle; Mervis & Bertrand's (1994) N3C principle).[1] For example, children should be more likely to overextend *moon* to an object without a known name (such as a pomegranate) than to an object whose name is already known (such as an apple).

To examine these issues, we studied two-year-olds, two-and-one-half-year-olds, and four-year-olds. There are two reasons for examining these different time points. The first reason is that productive overextensions are reported to decrease rather rapidly between 2;0 and 2;6; thus, examining both comprehension and production over this age range should provide some information concerning the basis of the developmental change. Specifically, two-year-olds enable us to examine an age when overextensions in production are relatively frequent; two-and-one-half-year-olds allow us to examine an age when productive overextensions are waning. The second reason is that by including the oldest group (four-year-olds), we can compare the present data to prior studies of the shape bias. Such studies primarily include three- to

---

[1] There is currently much debate concerning the question of whether early word-learning can be accurately characterized in terms of principles such as these (see Nelson, 1988; Tomasello, 1992, for alternative theoretical perspectives). In the present paper we refer to these principles in a descriptive manner, and do not mean to imply that they necessarily serve as explanations of children's lexical use.

five-year-old children (see Landau *et al.*, 1988; Imai *et al.*, 1994). Although four-year-olds make few, if any, overextension errors, they are important to study so as to provide a bridge to the shape-bias work. By including the oldest group, we will be able to make some inferences concerning the source of any differences that might arise between the present study of overextensions and past work on the shape bias (i.e. whether they are task-based vs. age-based).

A final aspect of the design that should be noted is that the comprehension task included only two test words: *apple* and *dog*. Given the age of the subjects, it is difficult to obtain data from many trials. Overextension studies typically include either a large set of words with a small number of subjects (often six or fewer), or a larger set of subjects with a small set of words (cf. Naigles & Gelman, 1995). In this study the latter strategy seemed preferable because (as argued in the Procedure section), multiple trials per word were needed to disentangle the dimensions of task (comprehension vs. production), object shape, taxonomic relatedness, and prior lexical knowledge. The words *apple* and *dog* were selected because they are commonly reported to be overextended in the literature (Thomson & Chapman, 1977; Rescorla, 1980*a*; Kuczaj, 1982), and it was possible to obtain appropriate materials for these words.

METHOD

Subjects

There were 49 child subjects, divided into three groups: Group I ($N = 19$; ages 1;7 to 2;2; mean age 2;0), Group II ($N = 17$; ages 2;3 to 3;1; mean age 2;6), and Group III ($N = 13$; ages 4;0 to 4;9; mean age 4;5). For convenience, we refer to these as 'younger two-year-olds', 'older two-year-olds', and 'four-year-olds', respectively. All of the younger two-year-olds and 12 of the older two-year-olds were tested in their own homes; the remaining children (five of the older two-year-olds and all of the four-year-olds) were tested in their school, a university-based preschool. Parents of those children tested at home ($N = 31$) provided MacArthur Communicative Development Inventory (MCDI) ratings of their child's vocabulary. An additional 10 children were tested but either did not complete the task ($N = 7$) or had difficulty on the control items on the comprehension task ($N = 3$; see Results). In addition, 23 college students provided ratings of the experimental materials (see below).

Design

Each child received three tasks: a production task, a training task, and a comprehension task (the screened-alternative task). The production task and the comprehension task are the experimental tasks. The training task was

TABLE 1. *Pictures presented to children in the production task*

red apple	cocker spaniel	sedan	baseball
blue apple	poodle	race car	whiffle ball
orange	cow	school bus	balloon
pomegranate	hippo	tractor	top

TABLE 2. *Adult ratings of visible pictures presented on experimental trials of the comprehension task* (*items that were designed to yield high scores are indicated in bold italics*)

	overall sim.[a]	shape sim.[b]	category sim.[c]	*N*
Target word: *apple*				
red apple	***6·8***	***7·0***	***7·0***	13
blue apple	***5·7***	***7·0***	***5·3***	13
orange	4·3	***6·2***	***7·0***	13
pomegranate	4·7	***5·9***	***6·3***	13
banana	3·4	1·1	***7·0***	13
starfruit	3·2	1·4	***6·8***	10
baseball	3·5	***5·8***	1·0	13
pink candle	4·2	***5·8***	1·3	9
Target word: *dog*				
cocker spaniel	***7·0***	***6·9***	***7·0***	13
poodle	***6·5***	***6·5***	***6·9***	13
cow	3·0	***5·1***	***7·0***	13
hippo	3·1	***5·5***	***6·8***	13
chicken	2·3	1·9	***6·9***	13
whooping-crane	2·3	2·2	***6·9***	13
dog-chair	1·8	***4·1***	1·0	13
sawhorse	2·6	***4·8***	1·0	10

[a] Mean responses to the question, 'To what extent is this object similar OVERALL to an apple [for target word *apple*]/dog [for target word *dog*]?', from 1 = not at all similar to 7 = extremely similar.

[b] Mean responses to the question, 'To what extent is this object similar in SHAPE to an apple [for target word *apple*]/dog [for target word *dog*]?', from 1 = not at all similar to 7 = extremely similar.

[c] Mean responses to the question, 'To what extent is this a fruit [for target word *apple*]/animal [for target word *dog*]?', from 1 = not at all a member to 7 = an extremely good member.

included solely for the purpose of clarifying the comprehension task for the children; thus, responses during training were not analysed. For the PRODUCTION task, the factors were age (younger two-year-olds, older two-year-olds, four-year-olds) and prior lexical knowledge. Age was a between-subjects variable; prior lexical knowledge was a within-subjects variable, with words divided into HIGH-FAMILIAR vs. LOW-FAMILIAR. For the

COMPREHENSION task, the factors were age, prior lexical knowledge, taxonomic relatedness, and shape similarity. Age was a between-subjects factor; the other were within-subjects factors. Finally, parents filled out the MCDI vocabulary inventory (toddler version), with several additional words typed onto the form (*hippopotamus*, *lizard*, *pomegranate*, *starfruit*, *candle*, *sawhorse*).[2]

Materials

Materials were photographs of real objects and/or realistic coloured drawings, organized into three photo albums (one album each for the production task, training task, and comprehension task; see Procedure, below). Pictures used in the production task are listed in Table 1. Pictures used in the comprehension task are listed and described in Tables 2 and 4 (experimental items and control items, respectively). Tables 2 and 3 provide

TABLE 3. *Parental MCDI ratings for 13 words referring to pictures used in the experimental trials of the comprehension task*

	Apple set		Dog set	
	word[a]	say (%)[b]	word[a]	say (%)[b]
target items	apple	97	dog [or woof-woof]	100
high-familiar items	ball	97	chair	90
	banana	93	chicken	74
	orange	77	cow [or moo]	89
low-familiar items	candle	48	sawhorse	0
	starfruit	0	whooping-crane	*
	pomegranate	0	hippo	48

* *whooping-crane* was inadvertently omitted from the MCDI.

[a] Words that appeared on the MCDI, corresponding to pictures on experimental trials of the comprehension task.

[b] Percentage of children ($N = 31$) reported by their parents to say each word.

---

[2] Because there are wide individual differences in early lexical development, the MCDI was also used to supplement the comparisons involving age. First, we conducted a correlation between age (in days) and lexical level (as measured by number of words checked off by the parent on the MCDI), for the 31 subjects for whom we had MCDI ratings, and discovered that in this sample, age was a reasonably good proxy for lexical level (correlation of 0·46, $p < 0·05$). Second, we reanalysed all the data using MCDI scores rather than age as the basis for grouping the youngest subjects (i.e. those below 4;0). Specifically, the number of vocabulary items checked off by each mother was counted to derive a single score, ranging from 53 to 635. The low-vocabulary group ($N = 15$) had MCDI scores ranging from 53–371 ($M = 266·33$); the medium-vocabulary group ($N = 16$) had scores ranging from 391–635 ($M = 540·19$). We considered the four-year-olds ($N = 13$) to constitute the high-vocabulary group. All results reported below that were significant using age as the grouping variable were also significant when MCDI scores were used as the grouping variable.

TABLE 4. *Control items used in the comprehension task*

word[a]	visible picture[b]
POSITIVE EXEMPLARS	
spoon	spoon
flowers	flowers
tree	tree
cup	cup
NEGATIVE EXEMPLARS	
door	bowl
crayon	lizard
car	rabbit
hat	scissors

[a] Words provided by the experimenter on the control comprehension trials.

[b] Pictures that were visible (unscreened) during the control comprehension trials. Each trial also included a picture that was covered with a cardboard screen.

descriptive information regarding the experimental items used in the comprehension task.

As shown in Table 2, four of the 16 experimental trials were actual instances (i.e. for the word *apple*, actual instances included a red apple and a blue apple; for the word *dog*, actual instances included a cocker spaniel and a poodle). On the remaining experimental trials, the pictures were not actual instances, and varied from actual instances in shape and/or category membership (verified by adult ratings, provided in Table 2). They also varied in familiarity (as verified by MCDI ratings; see Table 3). Thus, for example, orange and pomegranate were the same shape as an apple and of the same taxonomic category (fruit); baseball and round candle were the same shape as an apple but from a different taxonomic category; banana and starfruit were of the same taxonomic category but differed in shape. Aside from the actual instances, half the pictures had a name familiar to the child (e.g. *orange*, *baseball* and *banana*) and half did not (e.g. *pomegranate*, *round candle* and *starfruit*).

Adult ratings

A set of adults (*N* ranged from 9 to 13 per picture) rated each of the pictures used in the experiment. Each subject made three sets of ratings, in the following order (designed to minimize contamination from one task to the next): overall perceptual similarity (e.g. 'To what extent is this object similar OVERALL to an *apple*?', from 1 = not at all similar, to 7 = extremely similar), shape similarity (e.g. 'To what extent is this object similar in SHAPE to an *apple*?', from 1 = not at all similar, to 7 = extremely similar), and taxonomic relatedness (e.g. 'To what extent is this a *fruit*?', from 1 = not at all a member, to 7 = an extremely good member).

MacArthur ratings

The MacArthur Communicative Development Inventory (toddler version) (Fenson, Dale, Reznick & Bates, 1994) was administered to parents of those children tested in their own homes. Table 2 presents means ratings for names of the pictures used in the comprehension task.

PROCEDURE

Each subject received three tasks, in the following order: a productive naming task, a training task, and a comprehension task. For the PRODUCTIVE NAMING TASK, subjects were simply asked to name 16 pictures (see Table 1), one at a time. This provides a standard productive measure of overextensions. The pictures were presented in a small photo album, with one picture per page, in a separate random order for each child.

Next, there was a TRAINING TASK, to teach children to point to the screened picture when the visible picture did not match the word they heard. Subjects saw a photo album in which two pictures were presented at a time (on facing pages). For each picture-pair, one picture was fully visible and the other picture was fully screened. The screens were opaque pieces of cardboard. On each page, the child was asked, 'Where's the X?' or 'Can you find the X?' or 'Can you point to the X?' where X was the name of either the visible picture or the screened picture. On each page, the child was asked about both the screened picture and the visible picture, one at a time. For example, when the pair was a flower (visible) and a telephone (screened), children were asked 'Where's the telephone?' AND 'Where's the flower?' Children received feedback on every trial (including explicit corrections when errors were made), and following every screened trial they were also shown the picture that was underneath the screen. The training pictures were presented in a fixed order in the book; the questions were also presented in a fixed order. All trials in training were clear-cut (as in the flower/telephone example). The words tested (paired as they were presented) were: *leaf*, *cat*; *TV*, *butterfly*; *house*, *coat*; *flower*, *telephone*; *baby*, *bed*; *shoe*, *ice cream*; *socks*, *frog*.

Thirdly, children received a COMPREHENSION TASK. Children saw another photo album, structured similarly to that of the training task with two pictures presented at a time on facing pages. As in the training task, on every page one picture was visible and the other was screened with an opaque piece of cardboard. Subjects received 24 trials: 16 experimental, eight control. The experimental trials were those on which children could over- or under-extend. Eight of the experimental trials corresponded to the word *apple* and eight corresponded to *dog* (see Table 2). The control trials were unambiguous: half with a clear match visible (e.g. *spoon* when a spoon was visible; POSITIVE EXEMPLARS in Table 4), half with a clear mismatch visible (e.g. *hat* when a pair of scissors was visible; NEGATIVE EXEMPLARS in Table

4). Control trials were included to ensure that training had been effective, and that children understood and attended to the task. The experimenter simply asked, 'Where's the X?' or 'Show me/point to the X'. She asked about only one word per page. No feedback was given on this task, and children were never shown the screened pictures. Pages were arranged in modified random order (i.e. no more than two trials in a row using the same target word, with control trials spaced fairly equally across the session), with a separate order for each subject. However, the left–right order of pictures (i.e. whether the visible picture was on the right or on the left) was kept constant across subjects, and counterbalanced across all the factors of the task.

In any forced-choice procedure with young children, it is important to ensure that experimenters are not providing subtle cues to the children. We took several precautions to guard against inadvertent cueing on the comprehension task. First, the experimenters were not informed of the hypotheses underlying the research. Second, the experimenters were carefully trained (a) not to look directly at either picture choice but rather to look down at their answer sheet while awaiting the child's response, and (b) to leave their hands in their lap after asking each test question. These safeguards were employed so that researchers could not use gaze or gesture to influence children's response choices. Third, as noted above, the experimenters provided no feedback to children during the comprehension task proper.

The older subjects were generally tested in one session, whereas the children in the two younger groups were tested in two to four sessions. For the younger children, the first session was introductory: the experimenter played with the child, looked through a distractor picturebook with the child (none of the pictures overlapped in shape or taxonomic category with the items in the experimental book) and (for those children tested at home) gave the mother the MCDI to fill out. At the second session, the experimenter tested the child on the productive naming task and the training task. Occasionally the experimenter was also able to complete the comprehension task during the second session. More typically, the comprehension task was conducted during a third session. Occasionally a fourth session was required to complete the comprehension task.

RESULTS

As mentioned earlier, we excluded from further consideration children who had difficulty on the control items in the comprehension task. These items simply measured whether children understood and complied with the task, namely, whether they selected the visible picture when it clearly matched the word, and selected the screened picture when the visible picture did not match the word. We excluded children who erred on more than two out of eight control questions ($N = 3$ subjects). Focusing just on those children who met criterion, subjects were correct on the control trials an average of 7·16, 7·41 and 7·69 trials out of eight, at each of the three ages. There were no significant differences in responses on the control items across the three age groups, $F(2,47) = 2{\cdot}06$, $p > 0{\cdot}13$.

Production

As a measure of overall accuracy, we first coded each response on the production task as correct or incorrect. All 16 trials were included (fruits, animals, toys and vehicles). Correct responses included standard basic-, subordinate-, or superordinate-level labels (e.g. *cow* for cow; *toy* for top, *truck* for tractor) as well as diminutive versions of these labels (e.g. *doggie* for dog). All other labels or responses were coded as incorrect. As expected, the number of correct names increased with age (Ms = 8·95, 10·94 and 12·38 at the three ages, out of 16 possible), $F(2,46) = 8{\cdot}69$, $p < 0{\cdot}001$. Newman-Keuls analyses indicated that the youngest children produced fewer correct labels than children in the middle and oldest groups, who did not differ significantly from one another, $p < 0{\cdot}05$.

We then looked specifically at children's incorrect responses. We subdivided overextensions into those based on similar shape (e.g. *moon* for whiffle ball), those based on taxonomic relatedness (e.g. *bird* for poodle), and those based on both similar shape and taxonomic relatedness (e.g. *apple* for orange; *kitty* for cow). Other incorrect responses included anomalous responses (*horse-boony* for apple) and non-naming responses (e.g. 'I don't know'; 'has a wheel' for a race car). Analyses of overextensions confirm that the developmental increase with age in the proportion of correct labels was due primarily to a decrease with age in overextensions (Ms = 5·00, 3·71, 2·00), $F(2,47) = 7{\cdot}88$, $p < 0{\cdot}002$, with the oldest children producing fewer overextensions than children at either of the two younger ages, $p < 0{\cdot}05$. There were no significant changes with age in the number of anomalous or non-naming responses.

Furthermore, as Table 5 shows, most overextensions were based on a combination of both shape and taxonomic relatedness (e.g. *apple* for pomegranate; *kitty* for cow). A(3)age × (3)error type ANOVA revealed a main effect for age, $F(2,46) = 5{\cdot}53$, $p < 0{\cdot}01$, with better performance among the four-year-olds than among either of the younger groups, $p < 0{\cdot}05$, Newman-Keuls. There was also a significant main effect of error type, $F(2, 92) = 67{\cdot}36$, $p < 0{\cdot}0001$. Errors based on both factors combined (e.g. *orange* for apple) were more frequent than errors based on either taxonomic relatedness alone (e.g. *airplane* for car; *bird* for poodle) or those based on shape alone (e.g. *ball* for apple), $p < 0{\cdot}01$, Newman-Keuls; the latter two did not differ significantly from one another. Simple-effects tests revealed that this effect of error type held up at each age group examined separately, all ps $< 0{\cdot}001$. Although the age × error type interaction was not significant ($F(4,92) = 2{\cdot}14$, $p = 0{\cdot}08$), follow-up simple-effects tests revealed that the

TABLE 5. *Mean number of overextensions on the production task (out of 16 trials) as a function of age group and error type*

	Error type		
	shape[a]	taxonomic[b]	both[c]
Age group			
Group I (younger two-year-olds)	0·58	0·32	3·10
Group II (older two-year-olds)	0·59	0·23	2·71
Group III (four-year-olds)	0·31	0·00	1·54

[a] Overextension based on shape only (e.g. *moon* for whiffle ball).
[b] Overextension based on taxonomic relatedness only (e.g. *bird* for poodle).
[c] Overextension based on both shape and taxonomic relatedness (e.g. *apple* for orange).

decrease in overextensions with age was significant only for the shape/taxonomic combined responses, $F(2, 137) = 9{\cdot}38$, $p < 0{\cdot}001$.

Finally, we examined variability in responses across individual items. Most of the pictures elicited relatively good performance: on 12 of the 16 pictures, at least half the children labelled the picture correctly, with performance on a given picture ranging from 55% correct (race car) to 98% correct (red apple). The majority of the errors were elicited by the remaining four pictures: hippo (41% correct), poodle (31% correct), top (14% correct), and pomegranate (2% correct). Similarly, overextensions were relatively low for the 12 high-performance pictures (overextensions ranging from 0–24% of subjects) but high for hippo (39%), poodle (41%), top (61%), and pomegranate (67%). From the MCDI, it is clear that *hippo* and *pomegranate* were relatively unfamiliar words for this sample of children (see Table 2), thus lending some support to the notion that overextensions are more frequent when children do not yet have a name for the object being labelled. However, the errors on *poodle* and *top* cannot be explained in this way, because the words *dog* and *top* (or *toy*) were familiar to most subjects. Nonetheless, the particular tokens displayed in the pictures (a manicured poodle and a small, light-bulb-shaped gyroscope) were atypical instances. Thus, children's difficulties on the production task may be explained largely in terms of prior lexical knowledge and typicality. As noted by other researchers, prototypical instances of a category are learned earlier and with relatively more ease than atypical instances of a category (Mervis & Pani, 1980; Mervis & Rosch, 1981).

Comparison of comprehension and production

The next question is how children perform in comprehension, relative to production. To examine this question, we look just at those pictures which were identical in the two tasks. These included four instances of apples and dogs (i.e. *red apple*, *blue apple*, *cocker spaniel*, *poodle*) and four distractors (i.e. *orange*, *pomegranate*, *cow*, *hippo*). We refer to the apple/dog instances as TARGETS and the non-apple/non-dog instances as NON-TARGETS. Thus, in production a child would see each picture and be asked for its name; in comprehension, a child would see each picture plus a screened alternative and be asked to point to 'an apple' (for the red apple, blue apple, orange and pomegranate sets) or 'a dog' (for the cocker spaniel, poodle, cow, and hippo sets). Thus, each child received four scores: (1) number of correct responses to the target items in comprehension, (2) number of correct responses to the non-target items in comprehension, (3) number correct of correct responses to the target items in production and (4) number of correct responses to the non-target items in production. Each score could range from zero correct to four correct, although the probability of responding correctly differed by task.[3] Mean scores for these categories are presented in Table 6.

TABLE 6. *Mean comprehension and production scores on the pictures that appeared on both tasks (in each cell, mean number of correct responses are out of a possible 4)*

	Age group		
	younger two-year-olds	older two-year-olds	four-year-olds
Target items (*red apple*, *blue apple*, *cocker spaniel*, *poodle*)			
Production task	2·53	3·00	3·54
Comprehension task	3·37	3·29	3·69
Non-target items (*orange*, *pomegranate*, *cow*, *hippo*)			
Production task	1·10	1·71	2·61
Comprehension task	2·16	2·47	3·08

We than compared comprehension and production using a (3) age: younger two-year-olds, older two-year-olds, four-year-olds × (2)item type; target, non-target × (2)task: comprehension, production repeated measures ANOVA. As expected, comprehension yielded more accurate responses than

---

[3] Chance performance was 50% for the comprehension task, in which children were asked to choose one of two pictures. It was considerably below that for the production task, in which children were asked to generate their own label for each picture. Errors on the two tasks were also asymmetrical, based on the structure of each task: for the comprehension task, the only errors detectable were either underextensions or overextensions of *apple* or *dog*. (For example, an underextension of dog would entail FAILING to apply the word *dog* to an actual exemplar of a dog.) For the production task, errors included a wider range of responses: underextensions of *apple* or *dog*, overextensions of *apple* or *dog*, underextensions of other words (e.g. 'don't know' in response to a pomegranate), and overextensions of other words (e.g. '*moo-moo*' [cow], used to refer to a hippo).

production, $F(1, 46) = 29.40$, $p < 0.0001$. Children were more likely to select the appropriate choice in the comprehension task than to label the pictures accurately in the production task, Ms = 6.02 and 4.83, respectively (each out of eight, derived by summing over target and non-target items and averaging over ages; see Table 6). Also as expected, accuracy improved with age, $F(2, 46) = 16.15$, $p < 0.0001$. Interestingly, there was also a trend toward an interaction between age and task, $F(2, 46) = 2.92$, $p = 0.06$. Follow-up simple effects tests revealed that the comprehension–production difference was significant for younger two-year-olds ($F(1, 46) = 14.83$, $p < 0.001$) and older two-year-olds, $F(1, 46) = 4.14$, $p < 0.05$; however, the oldest group did not show a significant difference in accuracy as a function of task, $p > 0.30$. Finally, there was a significant effect for item type: children were more accurate (in both comprehension and production) on target than non-target pictures (Ms = 3.24 vs. 2.19, each out of four; these were derived by averaging over scores for the two tasks and three ages), $F(1, 46) = 78.22$, $p < 0.0001$.

As an additional means of comparing comprehension and production, we conducted three correlations among: (a) the number of words produced correctly on the production task (out of 16 possible), (b) the number of words comprehended correctly on the comprehension task (out of 16 possible, excluding control items), and (c) the total number of words reported to be used (out of 686 possible; tallied from the MCDI). These correlations included only those children whose parents had completed the MCDI ($N = 31$, all from the two younger age groups). As expected, there was a significant positive correlation between scores on the production task and scores on the MCDI (0.57), $p < 0.01$, indicating that children with a higher productive vocabulary (as reported by their parents) are more accurate on our production measure. In contrast, the comprehension task did not correlate significantly with the MCDI (0.22). Moreover, comprehension and production showed no correlation (0.00), suggesting that they are independent measures.

Comprehension errors

To investigate the source of children's comprehension errors, we conducted a four-way ANOVA that focused on the 12 experimental trials, excluding the control trials. The factors were: age (younger two-year-olds, older two-year-olds, four-year-olds), word (apple, dog), item type (taxonomic, shape, both), and familiarity (familiar, unfamiliar). For example, a familiar taxonomic trial was one in which the visible picture was a banana and the child was asked for an 'apple'; an unfamiliar shape trial was one in which the visible picture was a sawhorse and the child was asked for a 'dog'. The dependent variable was the number of correct selections when asked for the target word.

There was a main effect of age, $F(2, 46) = 5.47$, $p < 0.01$, as a result of the four-year-olds performing better than each of the younger two groups

($Ms = 62\%$, 66% and 82%, for younger two-year-olds, older two-year-olds, and four-year-olds, respectively), $ps < 0.05$, Newman-Keuls. There were no significant interactions involving age. The remaining significant effects involved word, item type, and familiarity. Each main effect and each two-way interaction involving these three factors was significant, ps ranging from 0·05 to 0·0001. However, all of these effects were subsumed under the significant three-way interaction involving word, item type, and familiarity, $F(2, 92) = 4.25$, $p < 0.02$ (see Table 7).

TABLE 7. *Comprehension task, percent correct (i.e. percent of trials on which subjects avoided the visible foil and selected the screened alternative instead)*

	high familiar	low familiar
Visible foil matches *dog* on:		
Shape only (chair, sawhorse)*	93	81
Taxonomic category only (chicken, whooping crane)*	89	85
Both taxonomic category and shape (cow, hippo)*	86	81
Visible foil matches *apple* on:	familiar	unfamiliar
Shape only (ball, candle)*	87	16
Taxonomic category only (banana, starfruit)*	79	55
Both taxonomic category and shape (orange, pomegranate)*	62	29

* Descriptions of visible foils, with familiar picture listed first, followed by unfamiliar picture.

To interpret this complex interaction, we consider *dog* and *apple* separately. For the word '*dog*', there were no significant effects based on item type or familiarity. In other words, children generally selected correct pictures of dogs and ruled out pictures of things that were shaped like dogs or were taxonomically related to dogs, even when the items were unfamiliar. Performance was consistently high in all cells, at each of the three ages (overall $M = 86\%$). Even in the lowest cell (i.e. younger two-year-olds on an unfamiliar item for which shape and taxonomic kind were combined) children chose correctly on 68% of trials.

For the word *apple*, however, performance was lower overall ($M = 55\%$ correct), varying as a function of both familiarity and item type. Children were more accurate on familiar vs. unfamiliar *apple* items ($Ms = 76\%$ vs. 33%), correctly excluding an orange more often than correctly excluding a pomegranate, for example. Moreover, the KINDS of errors on *apple* varied by familiarity. Specifically, when items were FAMILIAR, errors were more often based on a combination of both shape and taxonomic kind (i.e. pointing to an orange as an *apple*) than on either dimension alone (i.e. pointing to a banana or a baseball as an *apple*, $ps < 0.05$, Newman-Keuls. For UNFAMILIAR trials, errors were least often based on taxonomic kind (i.e. pointing to a starfruit as an *apple*) than on either shape or the combination of shape-and-taxonomic kind (i.e. pointing to a round candle or a pomegranate as an *apple*), ps $< 0{\cdot}05$, Newman-Keuls.

DISCUSSION

This study was concerned with two primary questions: (1) Do children overextend in comprehension, when using a novel task that allows children to indicate which instances are NOT named by a given word? This can be considered the primary METHODOLOGICAL concern of the present research, and is discussed in the section below entitled 'Methodological issues: comprehension vs. production' (2) What are the bases of children's overextensions? Specifically, what is the role of shape, taxonomic relatedness, and prior lexical knowledge in guiding children's overextension errors? This more theoretical concern is discussed in the section below entitled 'Bases of overextension: shape, taxonomical relatedness, prior lexical knowledge, and particular lexical items'.

Methodological issues: comprehension vs. production

As predicted, the screened-alternative (comprehension) task yielded greater accuracy than the standard production task, especially in the youngest age group. Whereas in PRODUCTION younger two-year-olds lagged considerably behind older two-year-olds in accuracy, in COMPREHENSION the younger two-year-olds performed nearly as well as the older two-year-olds. Nonetheless, children in all age groups (even four-year-olds) did at times overextend in comprehension, on a task for which demands to overextend were minimized. Recall that children were given an opportunity to indicate directly that an item should not be labelled by a given word, and successfully did so on the control trials (e.g. indicating that a pair of scissors was not a *hat*). Despite this, subjects consistently pointed to inappropriate referents on certain experimental items. These data strongly suggest that some overextensions are not merely production errors.

How distinct are comprehension and production? Overall better performance on comprehension vs. production does not provide evidence for the two being separate processes. Indeed, such a result would occur if the two tasks simply pose different levels of difficulty. However, two additional results suggest that comprehension and production may be more fundamentally distinct (see also Clark & Hecht, 1983). First, comprehension and production did not correlate significantly with one another, even though performance on the production task correlated significantly with maternal reports on the MCDI. Second, the PATTERNS of performance differed on the two tasks, with comprehension yielding more errors based on EITHER shape or taxonomic relatedness, and production yielding more errors based on

BOTH shape and taxonomic relatedness (combined). More research is needed to examine in greater detail whether and to what extent comprehension and production can be considered distinct processes.

Bases of overextension: shape, taxonomical relatedness, prior lexical knowledge, and particular lexical items

The present study was designed to enable us to disentangle shape, taxonomic relatedness, and prior lexical knowledge as bases for children's overextension errors. Of central interest is whether children's overextensions can be characterized as SHAPE-BASED, as suggested by classic studies of overextensions (e.g. Clark, 1973), which would be consistent with more recent claims regarding early categorization (e.g. Imai *et al.*, 1994). The present data argue against a strong shape bias in overextensions, in three ways. First, children were usually correct in comprehension. Even when presented with objects of the same shape AND same taxonomic kind, children typically refrained from extending a word erroneously. Second, both in production and in comprehension of familiar items, when children DID overextend, it was typically to items that matched the target word in both shape and taxonomic relatedness, rather than to shape preferentially. Third, in comprehension of familiar items, children were as likely to overextend based on taxonomic relatedness alone as on shape alone. For example, when asked for an *apple*, children picked a banana as often as they picked a baseball. All of these findings suggest that shape, though salient, has no special priority in two- to four-year-olds' semantic representations.

Indeed, the present study in some respects underestimates the importance of taxonomic kind in children's lexical extensions. Note that taxonomic relatedness as measured here was at the superordinate level, which provides a particularly stringent test (see Golinkoff *et al.*, 1995). If we had included basic-level (rather than superordinate-level) taxonomic matches, presumably children would have used taxonomic relatedness even more often in their naming and comprehension. Moreover, all items presented to children in this study were two-dimensional pictures. (This aspect of the design was obviated by logistical concerns, including the difficulty of using live animals as stimuli!) Two-dimensional pictures presumably highlight the salience of shape (because a standard, unvarying perspective is presented) and lower the salience of taxonomic kind (which is often conveyed by subtle texture, size, and movement cues). It would be informative to vary these cures to determine the conditions under which shape vs. taxonomic kind are made salient.

Two further factors seemed to be especially predictive of children's word use: prior lexical knowledge and the particular word studied. Children were much more likely to overextend in comprehension when they did not know the name of the item (e.g. *pomegranate*) than when they did (e.g. *orange*).

The importance of prior lexical knowledge is wholly consistent with a variety of current theories (Clark, 1988; Markman, 1989; Merriman & Bowman, 1989; Mervis & Bertrand, 1994), all of which predict that children should avoid overextending a word to an item that already has a known name (but see Nelson, 1988; Tomasello, 1992, for alternative theoretical accounts). Perhaps more surprisingly, prior lexical knowledge also influenced the nature and frequency of children's overextension errors. Thus, when children were shown FAMILIAR items and asked to select the *dog*/*apple*, they were more likely to select incorrectly a picture that shared both shape and taxonomic kind with the *dog*/*apple* than a picture that shared either shape alone or taxonomic kind alone. When shown UNFAMILIAR items, children's errors were more often based on either shape or shape-and-taxonomic-kind than on taxonomic kind alone. It appears, then, that when children see unfamiliar items, they make use of shape in determining label extension (perhaps because shape is the most accessible information one can obtain about an unfamiliar thing), but when they see familiar items they identify them on the basis of both shape and taxonomic kind.

Finally, the particular word studied also exerted tremendous effects. As did Naigles & Gelman (1995), we explored overextensions by conducting a 'case study' of a small set of words (in this case, *apple* and *dog*). Unexpectedly, results indicated that *apple* was overextended much more often than *dog*. This difference was found despite comparable adult ratings of similarity and taxonomic relatedness across *apple* and *dog* items, and despite comparable MCDI ratings of the two words.[4] It also is not the case that children made more errors OVERALL with the word *apple*: in a *post hoc* analysis, we discovered that UNDEREXTENSIONS (i.e. failing to apply the word

---

[4] Across all items, the average shape similarity score for apples is very close to that for dogs (5·02 vs. 4·62). Perhaps more telling is the fact that, even when one considers just those items for which shape similarity is equivalent across the two sets (taxonomic matches and taxonomic + shape matches, with average similarity scores of 3·65 for *apple* and 3·67 for *dog*), there is again a large difference in performance on the comprehension task (85% correct on *dog*, 56% correct on *apple*). Thus, differences in performance between the two words cannot be attributed to shape similarity.

Similarly, with regard to familiarity, MCDI scores are nearly identical across the animal and dog sets, if we assume a score of 0 on *whooping-crane* (which was inadvertently omitted from the inventory): 59% of the time mothers reported their children knew the words corresponding to pictures in the *apple* set; 57% of the time mothers reported their children knew the words corresponding to pictures in the *dog* set. Even if one excludes the low-familiar items matching on taxonomic category only (*whooping-crane* and *starfruit*), in order to make no assumptions regarding the familiarity of *whooping-crane*, one is again left with equivalent familiarity scores on the MCDI (69% on the remaining words in the *apple* set; 67% on the remaining words in the *dog* set) – but a large difference in children's accuracy on the comprehension task (55% correct on the remaining *apple* items; 86% correct on the remaining *dog* items). Thus, differences in performance between the two words cannot be attributed to item familiarity.

dog to an actual dog instance) were more frequent for *dog* than *apple* in both comprehension (37% vs. 6%) and production (37% vs. 15%).[5] Having tested only two words, we do not have sufficient information to account for why they differed. Indeed, one limitation of the present research is that it was able to focus on only two words in comprehension.[6] However, it is intriguing to note that *dog* and *apple* are from different ontological domains (animal vs. inanimate object; Keil, 1979; Gelman, 1990). In future work, it would be interesting to explore whether the frequency or nature of overextensions varies by domain, more generally.

We return to the role of shape in children's overextensions. Although we have suggested that shape is not paramount (i.e. shape does not overpower taxonomic kind), it is still the case that shape was often a prominent factor, at times in conjunction with taxonomic kind and at times alone. Why did children overextend based on shape, when they did so? Here we speculate beyond the present data to suggest the importance of multiple factors. First, shape-based errors were more frequent when the items were unfamiliar to children, suggesting that shape may be relatively more accessible than other factors and used as a default (see also Imai *et al.*, 1994, Medin, 1989, for similar arguments). In other words, when one knows little or nothing about an object, shape can still readily be derived from a glance, whereas taxonomic information may be less accessible (especially from pictures). Relatedly, shape may be the basis of classification errors. As Mervis & Canada (1983) note, young children attend to different features than older children and adults, and may overextend because they are misclassifying an item. The salient features they attend to may often include shape (e.g. the roundness of a round candle may lead children to think erroneously that it is an apple; the wick, though more diagnostic, may not be noticed).

Second, children may at times overextend based on shape because the same-shaped object is treated as a REPRESENTATION of the item being named (see also Soja *et al.*, 1992, for a related argument). For example, children at all ages often selected the pink, round candle as an exemplar of the word

---

[5] This analysis was conducted by considering comprehension and production errors to the four target items (*red apple*, *blue apple*, *cocker spaniel*, *poodle*; see Table 6). A child was considered to UNDEREXTEND *apple* if he or she either failed to produce the word *apple* (production task) or failed to point to the apple (comprehension task) on the 2 apple target items. Likewise, a child was considered to underextend dog if he or she either failed to produce the word *dog* (production task) or failed to point to the dog (comprehension task) on the two dog target items.

[6] On the production task, many more words than *apple* and *dog*, were studied, as children were present with a variety of different items (animals, fruit, vehicles, toys) and produced a variety of different words from a variety of semantic categories (e.g. *balloon*, *pig*, *car*, *moon*, *tiger*, *airplane*, *goat*, *train*, *turtle*, *doughnut*, *mower*, *blueberries*, *mousie*, *wheel*, *truck*, *camel*, *sheep*, *grape*, *giraffe*, *kitty*, *tato*). Thus, the primary limitation is on the comprehension task.

apple, even though adults had no difficulty identifying that it was not a fruit. In retrospect, we suspect that, not knowing that the item was a candle, children may have assumed it was a representation of an apple (e.g. a toy apple or a fake apple). To examine whether representational status (rather than shape *per se*) is driving such naming patterns, it would be important to examine children's naming patterns when the representational status of a set of items is varied independent of shape (e.g. including items that are accidentally the same shape as an apple, such as clouds or paint spills).

A third motivation for overextending based on shape may be that children treat certain words as if they are classifiers. For example, in Rescorla's (1980*b*) longitudinal study of overextensions, the most common shape-based overextension was with the word *ball*, which children extended to other round objects. We suspect, however, that *ball* in English functions as a shape classifier even for adults: any spherical mass is appropriately labelled a 'ball' (e.g. cheese ball, ball of wax, ball of yarn, ball of lint). If this analysis is correct, then use of the word *ball* for round objects may reflect children's sensitivity to this usage in the adult language. One might argue that it should not even be considered an overextension, if children are simply omitting the portion of the phrase that specifies the substance.

In sum, this study made use of a novel task and obtained results supportive of two claims in the literature: overextensions are in part usage errors, but overextensions in comprehension that indicate more enduring category or labelling errors ('true overextensions') do seem to exist. Although overextensions of both types decreased with age, developmental changes were more pronounced on the production task than the comprehension task. Moreover, by using items carefully selected to examine the influence of shape and taxonomic kind independently, we found that both superordinate-level taxonomic relatedness and shape are salient in children's early word meanings, as early as two years of age.

REFERENCES

Anglin, J. M. (1977). *Word, object, and conceptual development*. New York: Norton.

Baldwin, D. A. (1992). Clarifying the role of shape in children's taxonomic assumption. *Journal of Experimental Child Psychology* **54**, 392–416.

Barrett, M. D. (1978). Lexical development and overextension in child language. *Journal of Child Language* **5**, 205–19.

Behrend, D. A. (1988). Overextensions in early language comprehension: evidence from a signal detection approach. *Journal of Child Language* **15**, 63–75.

Bloom, L. (1973). *One word at time*. The Hague: Mouton.

Chapman, R. S. & Thomson, J. (1980). What is the source of overextension errors in comprehension testing of two-years-olds? A response to Fremgen and Fay. *Journal of Child Language* **7**, 575–8.

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.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language* **15**, 317–35.

Clark, E. V. & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology* **34**, 325–49.

Dromi, E. (1987). *Early lexical development*. C.U.P.

Fenson, L., Dale, P. S., Reznick, J. S. & Bates, E. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59** (5), v. 173.

Fremgen, A. & Fay, D. (1980). Overextensions in production and comprehension: a methodological clarification. *Journal of Child Language* **7**, 205–11.

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. & Coley, J. D. (1990). The importance of knowing a dodo is a bird: categories and inferences in two-year-olds. *Developmental Psychology* **26**, 796–804.

Golinkoff, R. M., Shuff-Bailey, M., Olguin, R. & Ruan, W. (1995). Young children extend novel words at the basic level: evidence for the principle of categorical scope. *Developmental Psychology* **31**, 494–507.

Hoek, D., Ingram, D. & Gibson, D. (1986). Some possible causes of children's early word overextensions. *Journal of Child Language* **13**, 477–94.

Hutchinson, J. E. & Herman, J. P. (1991). *The development of word-learning strategies in delayed children*. Paper presented at the Boston University Conference on Language Development.

Hutchinson, J., Inn, D. & Strapp, C. (1993). *A longitudinal study of one year-olds' acquisition of the mutual exclusivity and lexical gap assumptions*. Paper presented at the Stanford Child Language Research Forum.

Huttenlocher, J. & Smiley, P. (1987). Early word meanings: the case of object names. *Cognitive Psychology* **19**, 63–89.

Imai, M., Gentner, D. & Uchida, N. (1994). Children's theories of word meaning: the role of shape similarity in early acquisition. *Cognitive Development* **9**, 45–75.

Kay, D. A. & Anglin, J. M. (1982). Overextension and underextension in the child's expressive and receptive speech. *Journal of Child Language* **9**, 83–98.

Keil, F. (1979). *Semantic and conceptual development: an ontological perspective*. Cambridge, MA: Harvard University Press.

Kolstad, V. & Baillargeon, R. (1996). *Appearance- and knowledge-based responses of 10·5-month-old infants to containers*. Manuscript under revision.

Kuczaj, S. A. (1982). Young children's overextensions of object words in comprehension and/or production: support for a prototype theory of early object word meaning. *First Language* **3**, 93–105.

—— (1986). Thoughts on the intentional basis of early object word extension: evidence from comprehension and production. In S. A. Kuczaj & M. D. Barrett (eds), *The development of word meaning* (pp. 99–120). New York: Springer-Verlag.

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

Leopold, W. F. (1939). *Speech development of a bilingual child: a linguist's record. Vol. I: Vocabulary growth in the first two years*. Evanston, IL: Northwestern University Press.

—— (1940). *Speech development of a bilingual child: a linguist's record. Vol. III: Grammar and general problems in the first two years*. Evanston IL: Northwestern University Press.

Mandler, J. M. & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development* **8**, 291–318.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction*. Cambridge, MA: MIT Press.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist* **44**, 1469–81.

Merriman, W. E. & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development* **54**, 1–130.

Mervis, C. B. & Bertrand, J. (1994). Acquisition of the novel name – nameless category (N3C) principle. *Child Development* **65**, 1646–62.

Mervis, C. B. & Canada, K. (1983). On the existence of competence errors in early comprehension: a reply to Fremgen & Fray and Chapman & Thomson. *Journal of Child Language* **10**, 431–40.

Mervis, C. B. & Mervis, C. A. (1988). Role of adult input in young children's category evolution. I. An observational study. *Journal of Child Language* **15**, 257–72.

Mervis, C. B. & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology* **12**, 496–522.

Mervis, C. B. & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology* **32**, 89–115.

Naigles, L. & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: preferential-looking in a study of 'dog', 'cat', and 'cow'. *Journal of Child Language* **22**, 19–46.

Nelson, K. (1988). Constraints on word learning? *Cognitive Development* **3**, 221–46.

Rescorla, L. A. (1980*a*). Category development in early language. *Journal of Child Language* **8**, 225–38.

—— (1980*b*). Overextension in early language development. *Journal of Child Language* **7**, 321–35.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (eds), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* **3**, 382–439.

Schlesinger, I. M. (1982). *Steps to language*. Hillsdale, NJ: Erlbaum.

Smith, L. B. & Jones, S. S. (1993). Cognition without concepts. *Cognitive Development* **8**, 181–8.

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.

Soja, N. N., Carey, S. & Spelke, E. S. (1992). Perception, ontology, and word meaning. *Cognition* **45**, 101–7.

Thomson, J. R. & Chapman, R. S. (1977). Who is 'Daddy' revisited: the status of two-year-olds' over-extended words in use and comprehension. *Journal of Child Language* **4**, 359–75.

Tomasello, M. (1992). The social bases of language acquisition. *Social Development* **1**, 67–87.