Category Development in Early Language

Leslie Rescorla
Bryn Mawr College, lrescorl@brynmawr.edu

Let us know how access to this document benefits you.

Follow this and additional works at: http://repository.brynmawr.edu/psych_pubs

Part of the Psychology Commons

Custom Citation

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/psych_pubs/14

For more information, please contact repository@brynmawr.edu.
Category development in early language*

LESLIE A. RESCORLA
Yale University

(Received 1 March 1980)

ABSTRACT
Developing knowledge of the vehicle, animal, and fruit categories was traced in six children from 1;0 to 1;8. Data from mothers' language diaries and from bi-monthly sessions with the children were pooled to analyse the growth, content, and internal structure of the three categories over time. The children developed some grasp of most of the focal concepts in each category, but they made fewer differentiations than adults do. Overextension of a single concept term to encompass a cluster of related referents was common. The frequent discrepancies between comprehension and production of concept terms highlighted the importance of examining both modes. The data showed marked individual differences in style of category acquisition.

INTRODUCTION
It is widely accepted that pre-school children are unable to grasp the structural principles of hierarchically organized conceptual categories (Anglin 1977, Bruner, Oliver & Greenfield 1966, Inhelder & Piaget 1964, Vygotsky 1962). However, it is also clear that young children have a working grasp of categorical relations and a wide knowledge of basic categories: they know that dogs and cats and elephants are animals, they can generate a list of vehicles, and they can sort or cluster by category in learning tasks (Anglin 1977, Faulkender, Wright & Waldron 1974, Goldberg, Perlmutter & Myers 1974, Rossi & Rossi 1965). Even the toddler who is learning single words appears to have some rudimentary grasp of basic categories, some perception of kinship or similarity between different members of a common category. Among the clearest evidence the child from one to two years gives of this sense of categorical relation is his use of an overextended term such as car to label a collection of related referents such as trucks and buses (Bloom 1973, Clark 1973, Leopold 1939, Rescorla 1980).

* This research was conducted while the author was a Natural Science Foundation and National Institute of Mental Health pre-doctoral fellow. The author wishes to thank Katherine Nelson, William Kessen, and Gail Ross for their comments on earlier drafts of the manuscript. The author is also most grateful to the children and mothers who made the study possible. Address for correspondence: Yale Child Study Center, 333 Cedar Street, New Haven, Connecticut 06510.
The body of diary literature on early vocabularies (Chamberlain 1904, Guillaume 1927, Leopold 1939, Lewis 1951, Moore 1896), current theoretical accounts based on old diary data (Anglin 1977, Bloom 1973, Clark 1973) and recent data on early word use (Gruendel 1977, Nelson 1973a, Rescorla 1980) all indicate that young children frequently use an overextended term to encompass a category of related referents, followed by gradual differentiation within the category as more specific terms to label category members are acquired.

Existing accounts suggest other characteristics of this process of categorization by overextension followed by progressive differentiation (Bloom 1973, Clark 1973). Children are thought to vary in the degree to which they use overextended terms to label category members. Additionally, there is thought to be variation in the manner and rate with which the differentiation process occurs. Leopold's (1939) data suggest that some clustering or subcategorizing occurs as the child acquires more labels. Rosch (1973) argues that categories have complex internal structure, with some category members being more prototypical or central than others. Finally, it appears that discrepancies between comprehension and production for labels of category members commonly occur, such as the child calling several animals *doggie* but knowing their correct names in comprehension (Bowerman 1976, Huttenlocher 1974, Thomson & Chapman 1977).

While the existing literature suggests such aspects of the categorization process, few accounts of the process have been well documented in data. Most of the relevant data come from old diary studies. While these pioneering studies are rich data sources, they are highly variable in quality, often lacking in important information and limited to the vocabulary acquisition of single children. Clark (1973) has argued that the diary data on differentiation within semantic domains are sparse and sketchy. In recent years, new vocabulary data have been reported on groups of children (Benedict 1979, Greenfield & Smith 1976, Gruendel 1977, Nelson 1973a, Rescorla 1980). However, none of these more current studies has focused on the evolution of categories in early word use.

There are a variety of methods appropriate to the study of early category formation. Ross (1980) used a habituation-dishabituation paradigm with children of 1;0 to 2;0 and found evidence for categorization of such superordinate domains as animals, food, and furniture. Ross and colleagues (in preparation) have recently used action schemes as an index of concept learning and generalization in toddlers. Riccuiti (1965) and Nelson (1973b) have reported data on sequential exploration and spatial grouping by toddlers as an index of their perception of dimensional and categorical similarity. Finally, Anglin's (1977) recent book explores knowledge of hierarchical categories in children two years and older using pictorial stimuli and both comprehension and production of category names as dependent measures.

While these various experimental approaches are essential to delineating the variables which influence early categorization, there is also a strong need for better naturalistic data documenting the process as it unfolds spontaneously in
children from one to two years of age. Group data on category development need to be examined so that individual differences can be explored. Investigation is needed into how the categorization process may differ across various kinds of taxonomic categories. The time course of category evolution needs to be examined. Finally, discrepancies between comprehension and production of labels for category members need to be a major focus of investigation.

The data presented here address these needs. The data trace category development in three semantic domains: vehicles, animals, and fruits. These particular categories were chosen because each contains words which are almost universal in early vocabularies, which are acquired at a young age, and which are very prone to overextension (*car, truck, dog, cat, apple*). In addition, both the animal and vehicle domain have been widely discussed in both the diary and theoretical literature (Clark 1973, Leopold 1939). The three categories present interesting contrasts, however, in such characteristics as number of category members, extent of internal structure or subcategorization, and degree of discriminability possessed by concept exemplars.

**METHOD**

Six first-born children were studied, three boys and three girls drawn from middle-class families. These children were participants in a longitudinal study of early language development (Rescorla 1980). Each child was studied for at least six months, starting around 1;0 and continuing until the child had a productive vocabulary of at least 75 words (between 1;6 and 1;8). Children were seen in their homes for a 1- to 2-hour session every two weeks.

The main data source was a diary of productive language kept by each mother, detailing each new word the child acquired and all the objects and events to which it was applied. Extensive data on word comprehension were also collected, particularly in the early months when the children's comprehension vocabularies were relatively small. In addition to the diary material, supplementary data relevant to the three target categories of the study were collected in discussions with the mothers during each visit. These discussions covered the child's current comprehension and production of names of vehicles, animals, and fruits, as well as the particular objects the child associated with each word.

The three target domains were also explored by means of loosely structured probes with the children during each visit. The child's production and comprehension of the names of category members were assessed by presenting an array of toys or pictures of category exemplars. 'Foil' items were chosen from both target categories and irrelevant categories: that is, a target exemplar (car) could be presented with either related vehicle types (truck, plane) or with foils from other domains (cup, shoe). Because the procedures were geared to each child's knowledge, attention span, and level of co-operation, the probe data are properly considered as quasi-naturalistic rather than experimental.
Data from the diary records, from discussions with the mother, and from the probes with the children were combined to produce a developmental history of each target category for each of the six children. The same modes of analysis were used for all three domains, to look at both general trends and individual differences. The primary quantitative aspect of category development analysed was the number of distinct category members during each month of acquisition (e.g. dog, cat, horse are members of the category of animals). Presence of a category member in either comprehension or production was sufficient for inclusion. That is to say, if the child understood the word truck but had no label for it, or called all trucks car, truck would still be included as a distinct category member. To obtain a measure of the generality of each category member term for the child, an arbitrary dichotomous rating was used: application to many exemplars (more than 6) vs. few exemplars. The content and internal structure of each category over time were examined. Overextension of category member terms to related referents was investigated as indication of cluster formation or subcategorization within the category. Finally, discrepancies in category development as reflected in differences between comprehension and production were a focus of the analysis.

RESULTS

Vehicles

The quantitative data on development of the vehicle category appear in Fig. 1. Each child’s acquisition curve is plotted, showing the number of category members within the domain at monthly intervals. For the children who terminated their participation before 1:8, the number of category members obtained by termination was used as a conservative estimate of acquisitions in the following months, although these children probably continued to acquire more category members.

Fig. 1. Vehicles: number of category members per child across months of acquisition.
Development of the vehicle category was concentrated between 1;3 and 1;8. Mean number of category members rose from 1.67 at 1;3 to 9.00 types at 1;8, or at a rate of about two category members added per month. As will be shown later, the vehicle category occupied the middle position with regard to number of category members, with the animal domain being considerably more differentiated and the fruit category somewhat less so.

Inspection of Fig. 1 shows that the children varied in their degree of category differentiation. The range in terminal number of category members was from 13 types for Daniel to 4 types for Erica. With regard to generality of category members, five of the six children had between 3 and 5 vehicle words with which they identified many instances or exemplars, and one child (Daniel) had 9 such vehicle words. All six children appeared to have developed a differentiated concept of at least 3 classes of vehicles by the end of the study, labelled by the words car, truck, and plane. Five of the six children also had boat and bike as vehicle concepts. This suggests that the children covered roughly the full range of vehicles according to adult taxonomy, although within this range they made fewer differentiations than the adult category permits.

Overextension of vehicle words to closely related vehicle types indicated the presence of internal structure or cluster formation which appeared to parallel the adult taxonomy. For instance, four of the six children subsumed helicopters under the label plane; two of these children used plane or copter to cover blimps, rockets, and gliders. Another common cluster was single-person open vehicles: four children applied the word bike to tricycles and/or motorcycles as well as to bicycles at some point in their development. A third distinct type might be called large commercial vehicles. Five of the six children used the word truck (or bus) to cover such vehicles as buses, trucks, trains, bulldozers, cement mixers, and fire engines.

In addition to this cluster formation, all six children showed a clear focus on the concept car as a primary organizing principle for much of the vehicle domain. They each had a period of about a month in which car had a normal extension; each then overextended the word to a range of other vehicles. All the children overextended far enough to include trucks, five extended the term to buses and trains, and four children used the term in the sufficiently broad extension to include vehicles such as bikes, a toy plane, and strollers.

Following this period of overextension of the term car, all six children showed some narrowing of the concept, as distinct category members began to emerge from the overextended conglomerate. Erica, Andrea and Evan acquired other vehicle words such as bike, truck, and bus but they still occasionally called such vehicles car at the end of the study. Daniel, Donald and Rachel narrowed car to normal extension, and Daniel even began to make finer distinctions such as taxi car and beach car by the end of the study.

The final point to be made about development of the vehicle category is that
several children showed some discrepancy between comprehension and production of vehicle words. Often a child would respond to a word such as truck or bus for a given toy and then later refer to it as car. More striking was the case of Rachel. During the period when she overextended car to a wide range of vehicles, she was able to pick out all the same objects in response to their correct name; these included motorcycle, bike, truck, plane, and helicopter. Once she acquired productive labels for these concepts, they began to emerge from the car cluster. Indicative of how the process operated, her first label for airplane was sky car.

Animals

The quantitative data on development of the animal domain appear in Fig. 2. The animal category experienced most of its development from the ages of 1;0 to 1;5, after which point few new category members were added. Mean number of category members rose from 3.50 types at 1;0 to 21.83 types at 1;8 (with 20 types being attained at 1;5).

Comparison with the vehicle category shows that animals were an earlier interest for the children, but that development in this domain peaked at about 1;5. The animal category developed much more differentiation than did the vehicle domain, with more than double the number of different category members by the
terminal point of development (21·83 vs. 9·00 types). This difference seems largely attributable to the fact that the animal domain has many more potential members. It did not seem to reflect a significant differential in the children's interest in the two domains.

There was a very wide range in number of animal category members across children, as shown in Fig. 2. Three children were clustered at the bottom of the range; the other three children all had more than double the number of category members. It is interesting to note that the three children with significantly more animal types were the same three children who were at the top of the distribution for the vehicle domain. The data on generality of category members showed a similar pattern. Evan, Erica, and Andrea each had 4 or fewer animal types with many applications; Daniel and Rachel each had 10 types with many exemplars, and Donald had 22 different animal types with numerous instances.

Because the children were so diverse in their development, the performance of the children with fewer category members will be used as the minimal range of extension of the domain in discussion of category content and structure. All six children had dog and cat in their category in some form. In addition, all children identified at least one large mammal (horse or cow) and one small mammal (rabbit). Five out of six children manifested the animal concepts bird and fish and four children identified some types of insects with the word bug or ant. It seems, therefore, that most of the children identified at least one animal from the major classes of the animal domain.

As with the domain of vehicles, some clusters of animals were evident in the data. The most prevalent cluster was large quadruped, shown by four of the six children at some point in their development: horse or heehaw for horses, cows, goats, donkeys, giraffes, and camels. The other clusters which emerged tended to be found in only a few children: bzz for any flying bugs and duck for ducks, geese, swans, pigeons.

With regard to the overall structure of the animal category, the sample divided into two groups. The three children with many animal concepts (Donald, Daniel, and Rachel) did not show any single dominating animal concept overextended to cover most of the domain. Each of these three children was surprisingly accurate in his application of animal words, making fine differentiations with very few errors. Although they would occasionally say dog or cat inappropriately, such as calling a dog with pointy ears cat, these instances were extremely rare and of minor significance compared to the multitude of times such words were applied correctly.

The other three children showed a strikingly different pattern. All three, at some point in their development, subsumed a wide range of animals under one animal concept. Evan, who showed the least extreme pattern, overextended the word dog from about 1;2 to 1;5. He consistently called cats dog for about two months, and then acquired the word cat; in addition, Evan used dog occasionally
for an assortment of different mammals (squirrel, giraffe, lamb, teddy bear). Andrea also overextended the word *dog*. For a period of three to four months she consistently labelled cats *dog*; she also used the label for a variety of mammals (giraffe, camel, horses, bears), and a few times for a turtle and a frog. By the end of the study, Andrea used *dog* only for dogs and one wolf picture. Lastly, Erica showed an extreme degree of overextension of the word *cat*. From 1;1 to 1;8 she used the word *cat* consistently to refer to dogs, and frequently to name a diverse collection of other animals including non-mammals (bear, coyote, lamb, giraffe, rabbit, horse, camel, and chicken). While Erica showed only marginal comprehension of the word *dog* through most of the study, in the last month she acquired the word *goggie* which she used for a variety of animals including dogs, although *cat* continued to be her main label for dogs.

There were numerous discrepancies between comprehension and production for the animal domain. For instance, Andrea showed some comprehension of the word *cat* during the months she consistently called cats *dog*. Similarly, Erica comprehended the names of some of the animals which she called *cat*. On the other hand Evan showed no comprehension of *cat* during the period he called cats *dog*; but when he acquired the word *cat* he no longer applied *dog* to cats.

**Fruits**

The quantitative data on development of the category of fruits appear in Fig. 3. The main period of category expansion occurred between 1;2 (mean number of types 1.67) and 1;6 (mean number of types 5.33). The terminal mean for the sample was 6.33, making fruits the least differentiated of the three target domains. This is hardly surprising, given the rather limited size of the fruit category in the
typical adult taxonomy. The children were rather similar in their degree of differentiation of the category. Each had between 5 and 8 category members, with the exception of Erica, who had only 2. The data on generality of types showed a similar pattern.

With respect to category content, the most common fruit type was banana, present in all six children. Five children identified apples and grapes, four identified oranges, and three children recognized peaches, cherries, plums, and melons as distinct types. It would seem that most of the major fruit types were included in the fruit category for many of the children, with the exception of the class of berries which was represented in only two children. There was very little evidence of subgroups or clusters within the fruit domain. One child showed some confusion between peach and plum for a time, using both names interchangeably for both fruits. The only other notable case was Rachel's use of the term *lardì* for a variety of berries (strawberries, blueberries, and raspberries); this word derived from the name of her mother's greengrocer's delivery man, Mr Lamberti.

As with the other two domains, it was relatively common for children to organize almost the entire domain around a single focal point. Four children showed this pattern with *apple* for part of their development, with much of the overextension involving pictorial referents. Andrea used *apple* to refer to a range of fruits, including bananas, oranges, peaches, lemons, pears, grapes and orange juice from 1;4 to 1;6. By the end of the study she had acquired the word *banana* and knew several other fruit words in comprehension, but she occasionally still used *apple* to refer to oranges and peaches. Evan showed a similar pattern, using *apple* from 1;5 to 1;6 for peaches, oranges, tomatoes, plums, and one onion. Donald's first dominant fruit word was *lalala* for banana. He used this term to label apples, oranges, pears, and cherries off and on for about two months, at which time he developed the word *apple*. He then proceeded to restrict *lalala* to bananas but adopted *apple* as the all-purpose fruit word, applying it to oranges, strawberries, pears, and tomatoes. Some overextensions of *apple* continued to the end of the study. Finally, Daniel presented the interesting picture of having *apple* as the dominant fruit type without a great deal of overextended use. This was manifested by his use of the word *apple* as an initial response to a variety of fruits for which the names were not firmly established (plum and melon especially). Daniel would see one of these fruits, call it *apple* and then correct himself and label it appropriately. He never did this with fruits he had been naming for a long time, such as orange or banana.

One more striking case deserves mention. Erica had almost no development of the fruit domain, beyond knowing the word *banana*. However, in the last month of the study she acquired the general term *fruit*, first in comprehension and then in production. She used this word to label apple-sauce, nectarines, and pears; *pear* was the only one of these fruits for which she understood the name. This was the only example in the data of the use of a true superordinate category term to
cover a variety of subordinate category members rather than subsuming them under an overextended application of a subordinate term.

Three children showed some examples of discrepancies between comprehension and production in the fruit domain. Andrea understood the words *orange* and *peach* during the same time period she called them both *apple*. Similarly, Donald had comprehension of *apple* and *orange* while calling them *banana*. He could identify pictures of strawberries in books but would call the same pictures *apple*. Finally, Rachel used the word *nana* (*banana*) consistently to refer to raisins, though she clearly comprehended the word *raisin*.

**DISCUSSION**

The study reported here traces the development of three taxonomic categories in six children. Over the time period from 1;0 to 1;8, the children acquired an increasing number of words denoting members of the taxonomic categories of vehicles, animals, and fruits. The animal category attained the greatest differentiation in terms of number of members identified, with the fruit category having the least differentiation. Individual differences were most marked in the animal category, with three of the children having twice as many animal concepts as the other three. The three children advanced in animal concepts also had more vehicle and fruit concepts.

As in other studies of early language (Benedict 1979, Huttenlocher 1974, Thomson & Chapman 1977), the children’s comprehension of words was frequently more advanced than their production. Thus, conclusions about a child’s grasp of category structure and content drawn from production performance alone tend to underestimate the child’s underlying competence. As recent studies have made clear, young children know more about the world around them than they are able to express in productive speech. Similarly, they probably also know more about commonalities and relationships between objects in their environment than can be assessed even by their comprehension of language. Investigation of this hypothesis requires methodologies other than the one used in this study, such as Ross’s habituation paradigm (Ross 1980) or procedures utilizing the child’s action–function schemes (Ross, Nelson, Wetstone & Tanouye, in preparation).

The data indicate that the children in this study perceived certain vehicle, animal, and fruit types as closely related clusters. Aircraft, large commercial vehicles, large quadrupeds, and round fruits were clearly treated as clusters by most of the children, as indicated by overextension of a key word to subsume the exemplars. This cluster formation is consistent with Rosch’s (1973) view that certain category members constitute core, focal concepts which serve as prototypical exemplars providing internal structure for categories.

That some of this cluster formation arose from failures of discrimination is
likely. There were surely cases in which the child did not notice or register the
difference between a garbage truck and a fire engine and thus called them both
truck. However, there were many cases in which the child clearly discriminated the
exemplars from one another, could identify them correctly in comprehension, and
yet labelled them with a single overextended term. Such behaviour seems a clear
indication of some sense of relationship or commonality between the over-
extended referents and thus the presence of a cluster structure within the
category.

The acquisition of subordinate clusters within larger superordinate categories
has been alluded to in previous literature (Clark 1973, Leopold 1939), but has not
been the focus of systematic research. The data reported here suggest that such
cluster formation is typical in early language development, in fact more typical
than the overextension of a single term to encompass an entire category which has
been so widely discussed in the literature. Because the present study was natur-
alistic rather than experimental, it leaves important questions about this cluster
formation process unanswered. One crucial issue is the relative contribution of
perceptual similarity, functional equivalence, and contextual contiguity in deter-
mining cluster formation. Existing research (Rescorla 1980) has indicated that
young children can use each of these three types of information as the basis
for word applications, both separately and in combination with one another. In
the clusters described in this study, these three factors tended to be completely'
confounded with one another, as they are in the real world. That is, commercial
vehicles such as trucks, buses and trains share common perceptual features
(large size, movement, wheels), have similar functions from the child's point of
view (carry people and/or goods, go places, can be pushed on the floor), and often
appear in the same context of time or place (can be seen from the window or on a
walk, appear together in book pictures). Experimental research to tease apart
these three determinants of cluster formation is clearly indicated.

As previous literature has suggested, overextension followed by gradual dif-
ferentiation was a common process in these children's early category develop-
ment. As the children grew older, they began to make increasingly fine distinc-
tions within categories. Overextended terms became more constricted in their
denotation and new words were acquired in production to label category mem-
bers, as Leopold (1939), Clark (1973), and others have described.

The data indicate that the children acquired words to refer to most of the
classes composing the adult categories of vehicles, animals, and fruits. For
example, almost all the children had words denoting the major vehicle classes,
except for trains: car, truck, plane, boat, and bike. A similar pattern was evident
in the animal domain, where domestic pets, large and small mammals, birds, fish,
and insects were all identified as types by many of the children. Thus the chil-
dren's categories came to approximate the adult semantic system in range and
internal structure although they were less exhaustively differentiated.
The degree to which these six children had a grasp of vehicles, animals, or fruits as superordinate categories composed of subordinate category members is the most complex question posed by this research. As reported before, only one superordinate term was acquired by these children (fruit). However, many of the children used an overextended concept label to refer to a wide range of category members: four children used car and apple in this way, two used dog and one used cat. As was the case with the cluster data, comprehension evidence indicated that the children were often able to discriminate and identify distinct category members within this domain of overextension.

On the basis of these data, it thus seems reasonable to argue that the children had a rudimentary grasp of some of these categories as superordinate groupings; that is to say, they had some awareness that dogs, cats, and horses were distinct types with their own names but that in some general sense they formed a category of related entities which they could denote by dog. Rosch (1973) makes a similar argument with reference to older children having a practical grasp of class in cluster relations while lacking awareness of the logical relationships in such categorical structures.

A good example in the data of the intuitive grasp of superordination being suggested here occurred when Daniel at 1;5 was presented with a new assortment of Matchbox vehicles. As he watched them being dumped out of the bag, he began to say car, car! several times in great excitement; he then proceeded to inspect each carefully and to classify it as either car or truck, wavering back and forth between the two terms for ambiguous cases such as an ambulance. Of course, this particular example does not answer the question of whether car for Daniel meant wheeled road vehicle or vehicle in a more general sense.

This study raises some interesting questions about individual differences in style of category development in early language. Certainly, the six children in this study differed widely both in the degree to which they overextended a term to encompass an entire domain and in the amount of differentiation they achieved within each category. For example, only two of the children overextended a single animal term beyond the domain of mammals and only one child was making distinctions in the vehicle category such as taxi car vs. beach car at the end of the study. Looking at such patterns of overextension in a larger sample of children might reveal distinct typologies or strategies of category acquisition, somewhat in the manner of the strategies Nelson (1973a) found in vocabulary development. One strategy might involve a global sense of a category, conveyed by use of a single overextended term, followed by gradual differentiation within the domain. A second strategy might consist of gradually building up subordinate clusters from core organizing concepts within the category; children using this strategy might have a grasp of the entire domain which would be detectable by some non-verbal methodology, but they might not label the domain as such by a broadly overextended term.
In summary, this study suggests that children manifest considerable knowledge about basic categories by the time they are two years old. While it can be assumed from other literature (Anglin 1977) that toddlers do not yet have an understanding of the structural properties of hierarchical systems, they seem to have a working grasp of categorical relations. Their use of words suggests that they perceive a kinship or relationship between entities which form part of a categorical structure in the adult taxonomy. Further research of a more experimental nature is required to elucidate the relative contributions of perceptual, functional, and contextual factors in eliciting this perception of kinship. Thus, this research supports the notion that as children progress through the single-word period their mastery of basic categories becomes more highly differentiated and more internally structured, while their verbal labels become more circumscribed in application with increases in vocabulary.

REFERENCES


