1998

Children's and adults' ability to generalize properties to novel objects :: implications for theories of categorization.

Kristen N. Asplin
University of Massachusetts Amherst

CHILDREN'S AND ADULTS' ABILITY TO GENERALIZE PROPERTIES TO NOVEL OBJECTS: IMPLICATIONS FOR THEORIES OF CATEGORIZATION

A Thesis Presented

by

KRISTEN N. ASPLIN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

September 1998

Psychology

CHILDREN'S AND ADULTS' ABILITY TO GENERALIZE PROPERTIES TO NOVEL OBJECTS: IMPLICATIONS FOR THEORIES OF CATEGORIZATION

A Thesis Presented

by

KRISTEN N. ASPLIN

Approved as to style and content by:

Charles E. Clifton, Chair

Gary F. Marcus, Member

Marvin W. Daehler, Member

Melinda Novak, Department Head

Department of Psychology

ACKNOWLEDGMENTS

I would like to thank Gary Marcus for his continued help despite his move to New York. I would especially like to thank my entire committee for their advice and comments on revising this manuscript. Finally, I would like to thank my late father, Raymond E. Hoffmann, Jr., M. D. for inspiring me to learn.

ABSTRACT

CHILDREN'S AND ADULTS' ABILITY TO GENERALIZE PROPERTIES TO NOVEL OBJECTS: IMPLICATIONS FOR THEORIES OF CATEGORIZATION

SEPTEMBER 1998

KRISTEN N. ASPLIN, B. A., GUSTAVUS ADOLPHUS COLLEGE
M. S., UNIVERSITY OF MASSACHUSETTS AMHERST
Directed by: Professor Charles E. Clifton

This study investigated people's ability to draw inferences from known objects to new kinds of objects. An illustrative example is the ability to generalize a rule, such as a member of species X will have offspring of species X, to a newly discovered species Y. This ability remains unexplained by certain connectionist models and models of categorization other than the Theory Theory.

This thesis utilized a method from Keil (1989), where objects were transformed to appear like other objects. Fifteen, seven- and ten-year-olds and adults each, were told stories of surgeons who transformed natural kinds objects to look like other natural objects (e.g. raccoon to skunk), artifacts to look like different artifacts (e.g. umbrella to flag) and novel natural objects to look like familiar natural objects (e.g. hoatzin - a bird that climbs trees with its wings - to goose).

Younger children usually stated that all objects had changed and that offspring would look like the transformed object. Ten-year-olds and adults, however, often responded that natural kinds had not changed and would have offspring like their original appearance. More importantly, no differences between responses to novel and familiar natural scenarios were found. All participants who judged that a plant or animal would have babies like the original object made this judgment for both novel and familiar scenarios.

To more fully examine the participants responses, their explanations of their answers were categorized into three types of reasoning: perceptual, functional/behavioral, and biological. As expected, perceptual reasoning decreased in frequency with age, and biological increased with age on the natural scenarios. Functional reasoning increased on the artifact scenarios. Examining the justifications explored an unexpected pattern of responses, where participants believed the object to have changed, but the offspring would appear like the original object.

In this experiment participants treated the novel and familiar scenarios equally. This result provides exceptionally clear evidence that children's and adults' theories of natural kinds must include a rule or variable-manipulating system, and they can generalize that rule to novel natural kinds. Therefore, a rule or variable-manipulating system is necessary in all models of categorization.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

The last decade has seen a new direction in the development of models of categorization. Previous models relied on features or properties of objects to categorize the world (Smith & Medin, 1981; Rosch, 1978). The newer Theory Theory builds on previous models by adding a second level of explanation. Proponents of the Theory Theory (e.g. Gopnik & Meltzoff, 1997) postulate that, besides using features to categorize an object, a person has a naïve theory, or concept, of why an object belongs in a category. While this new hypothesis can account for many previously unexplained facts, as will be described below, there are still more facts waiting to be explained. One phenomenon that remains unexplained is people's ability to generalize information from previous knowledge to new kinds of objects. People can generalize known properties to a new object when they assign it to a known category. For example, imagine you are seeing your first tiger, but you have already seen small cats, lions, and leopards before. You can infer that, as a feline, the tiger could be a carnivore and would therefore be dangerous. This thesis attempts to show that models of categorization must include a mechanism for this type of generalization or inference. Marcus (in press) proposes that a rule system or an algebraic, variable-manipulating system can provide that mechanism, and this mechanism can best account for the data that are presented here.

Early Models of Categorization

The classical view of categorization defines concepts in terms of a list of necessary and sufficient conditions (see Smith & Medin, 1981 for a review). Such a list of conditions would amount to a definition resembling those found in a dictionary. For example, we might define an odd number as the set of numbers that give a remainder of one when divided by two, or we might define a chair as something with four legs, an upright back support, and a flat space parallel to the ground that can support a human in a seated position.

This model is inadequate, however, since researchers have been unable to enumerate the list of necessary and sufficient conditions for most concepts. Instead, for most categories, there are exceptions for nearly every feature that has been proposed. To continue with the chair illustration from above, there are three-legged chairs, back-less chairs, and bean bag chairs which do not seem to have any features in common with other chairs. Even if we exclude bean bags as aberrant chairs, we might only retain the feature "space that can support a seated human." This reduction of features to only that which is necessary produces the opposite problem: the remaining feature is not sufficient to select only the objects in the world that are chairs. Not everything that can accommodate a seated person is a chair (e.g. table, desk, etc.). Caught between being too general and too specific, researchers have been unable to identify the necessary and sufficient conditions for all but a few concepts (Smith & Medin, 1981). These concepts include, notably, mathematical terms like "odd number" where the necessary and sufficient condition is "a number that, when divided by two, gives a remainder of one."

This inadequate model of categorization was replaced in the 1970's with a class of models that use features in more flexible ways to define a category. The family resemblance, prototype and similarity models, postulated by Rosch and other researchers, no longer use static lists of necessary and sufficient conditions. For these models, a category is a similarity space whose dimensions are features. This category space has a center which is sometimes called a prototype (e.g. Rosch, 1978). In the prototype theory, the prototype is the ideal instantiation of the concept. By using a comparison function, an individual exemplar is measured against the prototype or feature dimensions of the category. Unlike the all-or-none list of features in the classic view of categorization, this type of comparison allows for grading of exemplars. An exemplar can be a better or worse member of a category (e.g. a dining room chair vs. a bean-bag chair), based on its similarity to the ideal exemplar or category center. These models can also account for fuzzy boundaries for a concept, as they allow the similarity function to determine whether an object is a member of one of two similar categories. If an exemplar is somewhat similar to the prototypes for two categories, but not a good exemplar of one (e.g. is a tomato a fruit or a vegetable), the boundary between the two categories may be unclear.

Problems with Feature-Based Models

These similarity models still rely on features to categorize objects, and therefore cannot explain certain observations in categorization. Murphy and Medin (1985) and Keil (1994), among others, have argued that although feature-based models can explain much of the way our minds categorize the world, another level of explanation is also necessary.

Armstrong, Gleitman, and Gleitman (1983) tested whether all categories could fit into either the classical view of categorization or into the family resemblance view. According to the classical view, categories are strictly defined by necessary and sufficient features; they have no graded membership or internal structure. An exemplar is either a member of the category or it is not, and there is no mechanism to compare members within the category. In the family resemblance or similarity models, categories have graded structure and graded membership based on similarity to a prototype, or other feature dimensions. With these two types of models as options for the structure of categories, categories either have both graded structure within a category and graded membership between categories, or they do not.

To test if all categories had graded structure, the researchers first gave participants a list of items from the categories of odd numbers and fruit (among other category lists) and asked them to rate the "goodness of fit" of each of the examples to the category label on a seven point scale. A "very good" example of an odd number was rated a 1, and a "very poor" example of an odd number would be rated a 7. The researchers found that the number "3" received an average rating of 1.6, and the number "447" a rating of 3.7. Similar differences in the ratings for fruit were also obtained. A cherry was rated 1.7 and a coconut was rated 4.8. It seemed that both categories did have graded structure within the category. To test if both categories had graded membership participants were asked whether it made sense to ask someone about the degree of membership that 447 had as an odd number, and coconut had as a fruit. All of the participants said that it did not make sense to ask this question for the odd numbers, but 57 percent said it did make sense for the fruit. The participants seemed to agree that all odd numbers are odd numbers, whether or not they are "good" odd numbers. In contrast, judging the graded membership of the category of fruit seemed to make sense to the subjects.

Armstrong, Gleitman and Gleitman had found that while the category of odd number had graded structure within the category (as in a family resemblance model of a category) it was still rigidly defined by a set of features, as in the classical view of categorization. The similarity class of models, like those of Rosch and her colleagues, could not predict that graded membership and graded structure were separable. Therefore, some other mechanism or level of explanation must exist, which can decide between all-or-none vs. graded membership types of categories.

Another phenomenon for which the early models of categorization have yet to account is the differences existing between broad classes of knowledge. Previous models categories (e.g. Rosch, 1978) have shown that there are hierarchies for where more specific categories are contained within larger, more general categories. The large, general categories are called superordinate categories, and include things like "furniture" or "animals." One of the most significant superordinate distinctions that has been investigated is the separation of natural kinds from artifacts. Natural kinds are a class of objects (e.g. animals, vegetables and minerals) that occur in the environment without human intervention (Putnam, 1975).

Artifacts, which are produced by humans, make up another large class. How the two classes differ can be seen in how they are treated by science[1] (Gopnik & Meltzoff, 1997). We observe, discover, and constantly redefine the division of the natural kind objects. Large branches of science, such as botany, are devoted to discovering more about natural kinds. In contrast, there is no division of science that is devoted to discovering more about trash cans and how they differ from mailboxes. We humans, as creators of these artifacts, have defined the categories by their function.

Medin and Shoben (1988) have shown that transforming objects can point out one of the fundamental differences between artifacts and natural kinds. When participants are presented with a banana and a boomerang that have been straightened, the banana is still perceived as a banana but the boomerang is seen as a stick. This example illustrates that natural kinds (e.g. animals, plants, minerals) comprise a domain where only deep, structural changes (genetic engineering, nuclear fission, etc.) can change their identity. Artifacts form another domain where appearance and features match intended function. When the features are

---

[1] Occasionally there are intriguing exceptions within the class of artifacts, for example: art, computers, cars and other complex inventions (Keil 1991). These man-made objects seem to defy the simple, functional descriptions that define most artifacts, and can have fields of study associated with them. Despite this, all of the artifacts in this study are simple (like the mailbox), so the simple division of artifacts and natural kinds will suffice.

changed so that the function is modified or lost, the artifact has changed.

Keil (1989) used the procedure of transforming objects to test if children and adults do indeed make the distinction between these two categories. In his study, children from five to ten years-old were given pictures of objects, both natural (animal, vegetable and mineral) and artifacts, and were then told a story about how those objects were changed with plastic surgery. After seeing a picture of the transformed object, the children were then asked, with a somewhat informal interview, questions about the object.

Keil found that children of all ages believed the artifacts to had changed. The natural kinds evoked a different response pattern. Children in Kindergarten tended to believe that the objects were changed, but fourth-grade children stated that the animal remained the same, despite the transformation. Keil interpreted his findings by concluding that as children grow older, their understandings of these transformations are less reliant on the characteristic perceptual features and more reliant on their more biological understandings of natural kinds.

The Theory Theory

A new model of categorization was introduced to provide a second of explanation that can help explain some of the limitations of the early, feature-based models of categorization. The Theory Theory, which began in the philosophy of science literature (Morton, 1980), postulates that people have a framework of intuitive theories for a domain of knowledge. We can partially define these theories as a "host of mental 'explanations' ... a complex set of relations between concepts, usually with a causal basis" (Murphy & Medin, 1985). In other words, a framework of theories provides us with a broad basis for understanding a domain of knowledge (e.g. biology), and helps us relate one concept to another within that domain (Wellman, 1990). A theory underlying a particular concept goes beyond the exemplars, or features of the exemplars in that concept, and provides the reason the concept contains those exemplars and features. (Gopnik & Meltzoff, 1997, Murphy & Medin, 1985). A theory does not, however, need to be explicit or even available to the conscious mind. A theory most likely operates on a sub-conscious level. The Theory Theory does not refute the usefulness of models like the prototype or family resemblance models but provides a second, more abstract level to the explanation of our ability to categorize the world we live in. According to the Theory Theory view, both features (and/or exemplars) and the theoretical explanations for the category are needed at the same time. Keil (1994) succinctly explains this need for two levels of information. "Explanations do not amount to much if they do not have anything to explain, and raw tabulations quickly overwhelm any information gathering system if it does not partially order that information in terms of explanatory usefulness." Gopnik and Meltzoff (1997) agree, comparing categorization to scientific inquiry by saying we need both the hypothesis and the data to fully understand a phenomenon.

How can the Theory Theory explain the facts found by Armstrong, Gleitman and Gleitman? According to this view, a person's intuitive theory about a category may contain an explanation for why, within a domain such as mathematics, categories can have rigid, non-graded membership. This explanation might include the information that constructs, such as odd number, are mathematical definitions that absolutely constrain the set of possible exemplars. While the theory is constraining the membership of the category, our feature-calculations may still provide a graded "goodness" judgment when forced.

The difference between natural kinds and artifacts, illistrated through the transformations performed on them by Keil (1989) can also be explained by the Theory Theory. People's naïve theories of natural kinds may contain an idea of an "essence" in natural objects that remains the same despite superficial transformations (Gelman and Hirschfeld, 1997). This essence could tell us that the banana remains a banana, even when straightened. Since the theory behind the identity of a boomerang, as an artifact, would have no such essence associated with it, would allow for identity changes by superficial transformations.

To briefly summarize, the Theory Theory has been able to offer explanations for differences between categories. This model, as opposed to the earlier models of categorization, can explain why a category can either have fixed or graded membership. The Theory Theory can also explain why objects from artifact categories can be transformed by simply changing the perceptual features, whereas natural kinds remain unchanged by these transformations.

Conceptual Change

In his 1989 study, Keil not only found that transformed artifacts and natural kinds are treated differently, but also that people's intuitive theories about these transformations are not the same at different ages. The present study did not measure participants' responses longitudinally, so it cannot directly measure how a theory behind a category or concept might change over time. However, I will discuss here how adults' theories might have changed with maturity and experience and therefore differ from those of children.

Concepts, and the theories behind them, must change as we learn and age. Carey's (1985) work with the concept of alive is an experimental example of a conceptual change. Carey (1985) introduced participants to a novel property; children were taught that people have spleens, and adults were taught that people possessed an omentum (a membrane in the abdomen). She then asked them if other things (dogs, bees, worms, flowers, clouds, garlic presses) had spleens or omentums. Children seemed to use the general properties of animacy and similarity to humans to create their inferences. That is, they inferred that all animate objects had a spleen, but were less likely to make this inference as the object became less similar to humans. The adults attributed an omentum to an object much more often when the object was a vertebrate than an invertebrate. They also seemed to distinguish between plants and other inanimate objects. Carey concluded that younger children understand life as "animacy" and adults understand it as "living" with sub-categories for plants, invertebrates, vertebrates and mammals. These two frameworks (animacy and living) differ, but both are used to make inference judgments of this new property to known objects.

Keil (1989) and Carey (1989) both found that a theory behind a concept was different at different ages. Carey's studies found that while younger children's theory of alive might have only included animate objects, the theory seems to develop with age and/or experience to include the plant kingdom. The oldest children's theory of alive also subdivides alive into sub-classes such as vertebrates and invertebrates. It seems that there are several stages in understanding the concept "alive." Keil's studies on transformations showed that the understanding of essence goes through several stages as well. He ran his transformations experiment in two ways. With younger children (two to three year-olds) the transformation story only said that the first object was wearing a costume, and therefore looked like the second object. Keil found that three year-olds could understand that the essence of the object remained unchanged in these simple costume changes, but that two year-olds could not. With older children (five to nine year-olds) he used the deeper transformations that a plastic surgeon can perform in his stories of change. As stated above, the five year olds could not understand that the essence of the object remained unchanged after the surgery, but the nine year-olds could. So costume changes seem to be understood very early as not truly representing identity changes, but the changes provided by sugeons take more age and experience to understand. These stages in children's theories of animals and plants, as shown in the studies of the concepts of alive and essence reflect a ever-growing understanding of natural kinds.

Gopnik and Meltzoff (1997) provide an explanation for how such a conceptual change might occur. According to their view, we come into the world with very simple theories about very broad categories, such as animate vs. non-animate motion (Spelke, Phillips & Woodward, 1995). Gopnik and Meltzoff also propose that we are endowed with a drive to explain the world by using the simplest theory that will be able to predict our environment. According to this model, a cycle of conceptual change is started when counterevidence is encountered. At first, counterevidence is ignored as noisy data. Later, an auxiliary hypothesis may be added to cover some of these examples. Finally, as the preponderance of evidence against the theory builds, the theory can be changed to replace the old one.

The changes in peoples' intuitive theories might also be similar to scientific theory development throughout history. From the vantage point of history, we can see that the development of a new theory to replace the old might not be automatic. Even when the auxiliary hypotheses are becoming more numerous and cumbersome, no person may have the insight to develop a new theory. Also, not everyone who is introduced to a new theory may accept it. Scientists disagreed with Copernicus for many years after he developed the theory of a heliocentric universe to replace the unwieldy notion of crystal spheres. This old hypothesis had many exceptions and an increasingly complex set of auxiliary hypotheses (in the form of spheres around spheres...) (Kuhn, 1962). Fortunately, we do not have to replicate the genius of insight in every new generation. While historical theory change might be slow in acceptance, conceptual theory change can occur rapidly as a child (or adult) sees that their current intuitive theory is inadequate and finds a replacement theory by learning from various other sources provided by society and experience (Gopnik & Meltzoff, 1997). Conceptual change would probably most often occur at a subconscious level; just as theories are not often conscious, their change in status would probably not be as well.

How can we measure changes in a theory, when theories are constructs, possibly even subconscious, within the mind? One indirect way was used effectively by Keil (1989) in his transformations study. Keil simply measured whether the participants believed the object changed, or believed that it had stayed the same. The older children responded differently to the transformations of natural kinds than to the transformations of artifacts. According to the view of the Theory Theory, this could be due to changes in their theory of biology. The younger children treated all the objects as equally transformed. This confirmed that children's understanding of the transformations had changed, but did not illuminate how they changed.

Despite the fact that theories are usually operating at an intuitive level, it may be possible to question the participants directly about their theories. While they might not be able to place the entire theory into words, they may be able to offer a richer source of information about their theories than their responses to the transformations would show. In the course of Keil's experimental interview, the children gave numerous justifications for, and elaborations on, their answers. Keil briefly mentions that, when confronted with artifact transformations, older children tended to refer to function more than appearance. Unfortunately, no further analysis was done with this rich source of information. Therefore, one of the main goals of the current experiment is to do an a analysis of the justifications that participants provide for their answers.

Asking the participants to justify their answers should offer more insight into their theories of natural kinds and artifacts. Categorizing their justifications into separate types of reasoning may also show that perticipants who mention biological concepts in their justifications might show a more mature pattern of responding to Keil's transformation scenarios. The indirect (transformation answers) and more direct (justification) measures of the children's theories should validate each other.

The Nature of "theories"

As stated above, the Theory Theory postulates that intuitive theories are the second level of explanation behind a concept. Until this point, I have been relatively vague about the specific information or mechanisms that a theory might include. In fact, the study of the structure and mechanism of theories has not progressed very far, but from the writings of Murphy and Medin (1985), Carey (1985) and those of others (Keil, 1989; Gopnik & Meltzoff, 1997; Marcus, 1997, in press) we can begin to examine the structure and function of theories. Some of the proposed features of a theory include abstractness, causality, coherence and rules.

Abstractness implies that the theory is at a more removed, abstract level than the features or exemplars in the categories. Although both features and theories are necessary for complete categorization, the theory is separate from the features (Keil 1994). An illustration of the abstract nature of theories can be seen in categories without graded membership, like odd number (Armstrong, et. al., 1983). The Family Resemblance models did not separate graded membership between categories from graded structure within a category. Since, as illustrated above, the category "odd number" has fixed membership and graded structure, membership and structure must be separate functions, at least for some categories. A theory operating at a removed, more abstract level, can remain separate from the concrete feature calculation and override these calculations in the cases of categories with fixed membership.

Intra-concept coherence is the "glue" that holds a particular concept together (Murphy & Medin, 1985). It is how the individual parts of a theory behind a category are related to each other. For example, it can explain how the theory is related to the data, in the case of the graded structure of the category "odd number", and how the theory remains separate from the data in the graded membership of the same category. Inter-concept coherence tells us how a particular category is related to other categories within a larger domain of knowledge (Gopnik & Meltzoff, 1997). "Animal" is a superordinate category, and the category "monkey" is part of "animal". This coherence also tells us that toy monkeys and clouds are not related in the same fashion.

Finally, because we cannot observe all the relevant properties of an object during every individual observation of that object, we need some kind of inference mechanism is needed to make predictions about new items that we place in our category (Gopnik & Meltzoff, 1997). Just because a tiger is not currently eating, does not mean that it is not a acting as a carnivore at other times. Inferring that this tiger is a carnivore, like the other tigers we have witnessed eating, gives us much more information to help us act wisely around tigers. Inference can also bring information from the superordinate level of categories to the basic level. For example, knowing that members of the category of animals (e.g. tigers, geese) bear offspring, we can attribute the ability to produce offspring to a new animal species called a hoatzin, without having to observe a hoatzin giving birth.

These examples present a type of inference consistent with previous models of categorization. This type of inference involves generalizing a feature from known examples of a category to new ones. The only difference between the two examples is the size of the category within which the generalization is being made. In the first instance, we are generalizing the feature "is carnivorous" to a specific example of a tiger within the category of tigers. In the second, we are generalizing the feature "bears offspring" to a new species of animal, hoatzin, within the larger category of animals. Models of categorization before the Theory Theory are able to make these inferences, since they involve features already familiar to the models. The models only need to pull up the features associated with a category, and then generalize the features to the specific example. Since these models are entirely feature-based, they can easily apply the known features to the new instances.

There is, however, another type of inference that these models cannot account for (Marcus, in press). This inference involves new features (e.g. bears *hoatzin* offspring), not just generic, known features (e.g. bears offspring). While it may seem obvious that people can infer that a hoatzin has hoatzin babies, two important types of models cannot predict this ability. The first type of models that cannot account for generalization of new features includes feature-based models of categorization. To illustrate, we turn again to the case of transformations.

If we know that horses remain horses and have horse offspring despite being painted to look like zebras, we should also know that hoatzins remain hoatzins and have hoatzin offspring despite being altered to look like geese. Based on the fact that feature-based models are built entirely out of stored memory representations, Marcus (1997) has argued that these models cannot infer a new feature. "... you can't *remember* that geesabs [a novel animal] have geesab offspring: you have to *infer* it." (Marcus 1997, p. 6, emphasis original) Feature-based models are very efficient at weighing new examples against the other members in a category, and making a judgment on whether, and how well, the new example fits the category. They can also, after including the new example as a member of the category, generalize the *known* properties from the other instances of the category to the new instance. They cannot, however, infer *new* features (e.g. a new type of offspring) to new examples (e.g. a new species) based on their category membership.

A second class of models that cannot predict that a new instance of a category will have an appropriate new feature, includes feed-forward neural network models (e.g. Rumelhart & Todd, 1993). These neural networks consist of several layers of nodes: Input nodes, output nodes, and one or more layers of hidden nodes. These networks start with all the connections between the nodes randomly assigned a weight, or strength of connection between two nodes. The network then receives an input, which is usually an exemplar that contains several features which are represented as a distributed pattern of activation of the input nodes. For example, an hoatzin bird might be input to the system by activating nodes that represent "has wings," "has claws on its wings," "can climb trees," and so on, but not activating other nodes that represent "has gills," or "is made of cloth." The network then produces output by activating one or more of the output nodes, and is informed if the output was correct. Based on the feedback that the output was correct or incorrect, the system adjusts the weights of the connections between the nodes so that the network will be more or less likely to produce that output on the next trial. After numerous experiences with examples which contain features that represent every input and output node, the network is able to correctly weigh the connections between the nodes and produce the correct output for a particular input.

Such a network cannot capture the hoatzin example. Imagine that a network has been assigned the task of determining which kind of offspring a particular animal would have. The input nodes might be the features of the parent animal, and the output nodes could be types of offspring. This network would have received many examples of animals and their offspring, which then would become familiar to the system. One example might be to input the features of a cat (claws, fur, etc.) and then the system would select an output node that represented the offspring. If the system was incorrect, the weights would be readjusted, and the next time the system received the cat features as input, it would be more likely to activate the "kitten" node. It would not, however, have experienced instances that involve a particular output node, "hoatzin baby." Since the system was told that activating the "hoatzin baby" node was incorrect in all of the thousands of previous examples, the connection leading to this node would have a acquired a set of connection weights such that it tends not to be activated. Before receiving the novel hoatzin input, no given input would activate this unfamiliar output node; the network would need many examples of the hoatzin to be familiarized with the novel input and output to then readjust the connections enough to produce the novel output. Therefore, these models are incapable of producing an output that involves an unfamiliar node (Marcus, in press).

What type of mechanism is able to capture the hoatzin example, then? The ability to generalize new features to new instances of a category requires a variable-manipulating mechanism (Marcus, in press). This mechanism might use a rule, such as "For all X, Y, such that X is a member of species Y, X has offspring of type Y." In this case, X would be a specific instance (Mary) of the unfamiliar species Y (hoatzin). So since Mary is a member of the hoatzin species, Mary will have offspring of the type hoatzin. This rule will not only work when Mary looks like a hoatzin, but also in the more complex example where Mary has been altered to look like a goose. As long as Mary is still included in the category of hoatzins, the above rule will predict that Mary will have hoatzin babies. The use of variables in the rule allows for the generation of novel features based on the pattern established by the other examples of the category.

While the categorization and neural network models described above may be unable to solve a problem that would involve producing novel output, models that can represent variables will not have this limitation (Marcus, in press). A model such as the LISA model of analogy (Hummel & Holyoak, 1997) can represent variables and therefore could solve the problem posed in the current experiment. Any model of categorization that does not include a variable representation system will be unable to make inferences requiring novel responses.

Therefore, I will test if people use variable representations within an inference system, by testing whether or not they infer new properties to unfamiliar objects, i. e. specifically inferring that hoatzins have hoatzin babies (Marcus, 1997). I expect that participants can treat both novel and familiar items equally, and can predict that hoatzins, altered to look like geese, can have hoatzin babies. If these predictions are borne out, the results found here would provide exceptionally clear evidence that a rule or variable-manipulating system is being used. This thesis will show that children and adults can indeed make these inferences, and therefore argues that models of categorization must include some kind of inference mechanism to be able to accurately portray our ability to categorize the world. Unlike other models of categorization, the Theory Theory could easily incorporate this mechanism in the structure of theories, thereby offering a more complete model of categorization than has previously been shown.

The Experiment

This experiment utilizes transformation experiments on natural objects and artifacts (after Keil, 1989) as a valuable tool for examining the theory of natural kinds. The first goal of this thesis, therefore, is a simple replication of Keil's (1989) experiment. In this experiment, children and adults were tested using a procedure very similar to Keil's (1989) plastic surgery transformations of natural kinds and artifacts, as has been described above. I hope to again show that as children grow older, they have a better understanding of the stability of a natural kind through a superficial transformation

The second goal of this experiment is to categorize people's types of reasoning during the process of giving an answer. The justifications given by younger children should reflect reasoning based on surface features, or as Keil described them, "characteristic features." As children grow older, their justifications are expected to reflect their increasing understanding of the difference between natural kinds and artifacts by including biological explanations.

The final goal of the experiment is to begin to examine the structure of theories. Previous research has shown that children can attribute novel properties to known animals (Carey, 1985), and known properties to novel categories based on the category label (Davidson & Gelman, 1990). These studies cannot tell us, however, how inference within a category takes place. We will ask children to generalize two known properties to new natural kind objects: natural kinds do not change when only perceptually altered; and a natural kind object will produce offspring of the same species as itself. Testing participants on novel objects will assess their ability to generalize variable properties from a familiar instance to a new instance within a domain. If participants can make these inferences, especially in regard to the heredity questions, these results would be best explained by a rule or variable system, similar to the one stated above, operating as a part of the participants' inference mechanism.

CHAPTER 2

METHOD

Participants

Participants included 30 children (15 seven year-olds and 15 ten year-olds) whose names were drawn from the birth records for Hampshire County held by the Department of Psychology at the University of Massachusetts. The average ages of the two age groups were seven years, two months, and ten years, two months. These ages were chosen to correspond roughly to Keil's (1989) groups of second graders (mean age seven years, six months) and fourth graders (mean age nine years, nine months). The 15 adult participants were volunteers from psychology classes at the University of Massachusetts. Adult participants received course credit for their participation.

Interview Procedure

Each session was held in a small lab room at the University of Massachusetts with the interviewer, the participant, and, if the participant was a child, his or her parent(s) present. The interviewer was an undergraduate psychology student at the University who received research credit for her assistance. The interviewer first explained the task to each participant. Adult participants signed consent forms, as did the parents of the participating children. The interviewer also asked each child's oral consent to participate. She then explained to the participant that he or she would be looking at a series of objects that would be transformed by a group of plastic surgeons. The interviewer finally proceeded with twelve story-question scenarios. These twelve scenarios were presented in a single, randomized order to the first half of the participants and in the reverse order to the second half.

Each story-question scenario proceeded as follows: The interviewer presented the first picture in the pair to the participant and verbally labeled the object. If the object in the picture was novel, the interviewer explained a little about the object. She then explained how the object was transformed by the plastic surgeons, and presented the picture of the transformed object (the second picture in the pair). Finally, with both pictures on the table before the participant, the interviewer asked the participant a series of questions to assess the participant's understanding of the transformed object. These questions, which resulted in four different scores (described below under scoring), were always of the same type, but had minor variations in wording to maintain the participants' interest in the questions. Humorous, extraneous questions (e.g. "Do you think it would like to watch TV?") were added to put the children at ease with the interviewer. These questions tended to elicit a smile, and also allowed the child to answer an easy question and avoid any frustration. A novel item example is given below; the complete stories and questions are found in Appendix A.

> This is a Soursop. It's a fruit that grows in the Caribbean, a bunch of islands south of Florida. It is big, and has a rough skin. It has lots of little black seeds inside, and tastes a little sour. We, in America, make lemonade out of sour tasting lemons and water and sugar, and they make drinks out of the soursop that way, too. They also make ice cream out of the soursop by adding milk, vanilla, water and sugar.
>
> The doctors took the soursop, and sucked all the insides out, so the skin shriveled up and got a little darker. They bleached the seeds so they were white, and they took most of the sour taste out of the fruit by soaking it in water for a long time. Then, they put everything back in, but the skin had shrunk, so they could only get half of the insides back in. It wasn't so round anymore. When they were all done it looked like this.
>
> 1) So now that the doctors are all done, and it looks like this, what kind of plant is it? A cucumber or a Soursop?
> 2) Uh huh, why do you think that?
> 3) Do you think it would like to sing?
> 4) If you took the seeds, and planted them, what would grow? A Soursop plant or a cucumber plant?

Stimuli

The objects presented to the participants were color pictures printed on 8 1/2″ x 11″ white paper. There were three types of picture pairs used in the story-question scenarios: familiar natural kinds, novel natural kinds, and artifacts. The list of pictures is found in Table 1.

Table 1. Pairs of stimuli used, grouped by object type.

Type of Object Pair	**Original Object**	**Transformed Object**
Familiar Natural Kinds	Horse	Zebra
	Raccoon	Skunk
	Potato	Apple
	Tiger	Lion
Novel Natural Kinds	Tazier	Koala
	Guitarfish	Shark
	Soursop	Cucumber
	Hoatzin	Goose
Artifacts	Soccer Ball	Bowling Ball
	Trash Can	Mailbox
	Umbrella	Flag
	Key	Coin

The pictures for familiar objects (natural and artifact) were taken from a commercially available clip-art package, and altered to insure the pictures within each pair were of approximately the same size and orientation. Three of the four pairs of familiar natural kinds (horse - zebra, raccoon - skunk, tiger - lion) were taken from Keil (1989). All references to altering the behavior of the animals were removed from Keil's scripts for these pairs. The fourth pair (potato - apple) was an inanimate natural object, and was included to help reduce the variance in the number of transformations performed on the objects. It replaced another example from Keil, in which a grapefruit was transformed to an orange in only two "steps" or actions performed by the plastic surgeons. The grapefruit was injected with sugar, and then it was dyed orange. It was possible that these two simple steps might make it more believable that the object had not changed, compared with the four steps in the tazier - koala script. Keil's grapefruit - orange pair was replaced by the potato - apple pair, allowing every pair in this experiment to be transformed in three to five verbal "steps."

Keil's (1989) original artifact pairs were also not used in this study because of limitations in the available artwork. Keil's artifact pairs (e.g. coffeepot - bird feeder) were hand-drawn illustrations; the artist could ensure that the first and last object looked similar. The available clip-art pictures showed large differences in the shapes of the two objects. As a consequence, Keil's artifact pairs would have been much more visually different from each other than the natural item pairs. To reduce this difference, other artifacts were selected so that (as judged by the author) the pictures within a pair had approximately the same perceptual similarity as the pairs of novel and familiar natural kinds.

Creating novel items from a given clip-art database presented a unique challenge, especially since adults and school-age children are familiar with many animals. The novel natural items were originally mythical-looking animals created with a computer by assembling a collage of parts from other clip-art objects. When these pictures were presented to a child in pilot testing, his or her first reaction was "That's not real!" Children may not have been including the unrealistic novel animal in their category of animals, and if so, would have no reason to generalize their knowledge from familiar animals to such unnatural stimuli.

To avoid the children's rejection of the objects as part of their categories of animals and plants, the novel natural kinds used in the experiment were selected because they were real but obscure items, and probably unfamiliar to the participants. The soursop is a rare island fruit, and its picture was found in a Caribbean cookbook (Sookia, 1994). The guitarfish, a member of the stingray family, was photographed for a Caribbean reef fish identification book (Humann, 1989) The tarsier[2] (which was inadvertently

---

[2] This animal was presented to all subjects as a tazier. At the end of the experiment, an error was discovered in copying the material from the source book, resulting in the mislabeling of the object. For the sake of consistency with the experimental materials, this animal will be referred to as a tazier within this thesis.

mislabeled in the scripts as a tazier) and hoatzin were found in a educational book on life found in jungles (Richards, 1970). The tarsier is a small animal, similar to a bush baby, that lives in the trees of tropical rain forests, and the hoatzin is a bird that inhabits the canopy of the rain forest. Stimuli for these novel objects were generated by using a computer drawing program to alter clip art pictures so they would closely match the photographs of the actual objects. Post-test questioning in piloting assured us that the children believed the obscure plant and animals were real, and that the children were indeed unfamiliar with the objects. Half-size, black and white copies of all pictures are reprinted in Appendix B.

The novel items were not transformed into different novel natural kinds, to parallel the other transformations for familiar natural kinds and artifacts. Instead, novel natural items were transformed into familiar natural items. This pairing hopefully constituted the strongest test of the participant's theory of the identity of the novel animals. Participants should know much more about the familiar object than the novel object. If the participants were biased to respond on the basis of familiarity with the object, they could be expected to have the most difficulty saying the object was still really the first, unfamiliar object.

Suggestibility

In the process of designing a script format, interviewers found it easy to lead a participant to the "correct" answer, especially when the participant was a child. This issue touches on the suggestibility of children, which has been investigated in psychology and has implications for child witnesses in our legal system. Ceci and Bruck (1993) reviewed eighteen studies of suggestibility and found that preschool children (even up to age six) are the most susceptible to suggestion, but that suggestion can also bias answers from older children to a smaller extent. Also, children will occasionally lie when they are motivated to do so. This fact has serious implications for the current study. If children could be properly motivated to lie about something they knew, it might be much easier to influence their answer for a question about which they were unsure.

Another area of concern was that our interviewer was unfamiliar to the children, and that fact might further motivate the children's suggestibility. Goodman and her colleagues (Goodman, et. al. 1995) investigated familiarity and bias as two factors that could influence recall. Four year-olds participated in play with a research assistant, and were then interviewed about the play session. They were interviewed by either their own mothers or by another mother. Also, half of the interviewers were told biased assumptions about the play session prior to the interview. They found that, especially during a free interview, children made the most errors in recall when questioned by a biased, unfamiliar interviewer. The children's own mothers elicited the best recall, regardless of their bias. Biased interviewers also questioned misinformation given by the children more strongly, and were less accurate in describing what they believed to have occurred during the play session.

Since half of our children were only slightly older than pre-school age, and they would be interviewed by an unfamiliar assistant, several steps were taken to minimize the possibility of suggestibility in the questioning: The interviewer was blind to the hypotheses of the experiment and was instructed not to try to help the child in any way. Also, the interviewer was given a strict script to follow. She was only allowed to add (a) Uh-huh's or OK's to make sure the child knew the experimenter was still interested, and (b) restatements of portions of the script to answer children's questions about the transformations. Finally, the same number of questions were asked during the natural and artifact transformation scenarios.

In contrast to these precautions, Keil's (1989) interview format was open-ended and run by an assistant familiar with the hypotheses of the experiment. Also, in the few transcripts provided as examples, interviews during natural object scenarios tended to be longer than interviews about artifact scenarios. These differences may have caused children to reevaluate their answers more often for natural items than for artifacts, leading them, in effect, to the "correct" answer. Despite these large differences in the interview format, we predicted that Keil's findings would still be present in an unbiased setting.

Scoring

Initial Judgment Score

Assistants scored the first statement given by the child in response to "What kind of _____ is it, a ____ or a _____?" In accordance with Keil's (1989) scoring procedure, a score of 1 meant that the child believed the object to be changed. Scores of 3 meant that she believed the object to remain essentially unchanged on the inside. Scores of 2 meant she was wavering, or was genuinely unsure about the current state of the object. A score of 3 also included all answers of the sort - "Well, it looks like the transformed animal, but really is the original animal." Using the soursop - cucumber pair as an illustration, the answers were coded 1 for when the response was "cucumber" and 3 for "soursop."

Final Judgment Score

In pilot testing for this experiment, children seemed to be biased to quickly name the transformed object in response to "What kind of _____ is it ?" It was possible that their first answer did not adequately represent their understanding of the transformed object. Therefore, the entire transcript of the scenario was examined for changes in the participants' judgments. Participants often changed their mind in response to the justification question (described below). A typical example from an adult was:

E.: What kind of animal is it?

S.: A koala.

E.: Why do you think that?

S.: Well, she looks like a koala, but she's probably a tazier.

The final judgment score used the same scale as was used for the initial judgment score. For illustration, the above example shows an initial score of 1 (for the changed animal), but a final score of 3 (for the original animal). The final judgment score may be more representative of the participant's understanding of the transformation.

Offspring Score

A third score, the offspring score, was based on the response to the question, "What kind of babies will it have?" Answers here were coded to parallel the scores in the initial and final judgment scores. A score of 1 was given when the participant believed the offspring would be the same kind as the transformed object. A score of 3 was given when he or she believed the offspring would be like the original object, and a score of 2 was again for unsure answers. Thus, if a raccoon was transformed into a skunk, and a participant said that the animal would now have skunk babies, the score was a 1. If he said the animal would have raccoon babies, the score was 3.

Justifications

The second question asked of the participants was, "Why do you think that?" or "Why isn't it a ______ (object in the pair, not named by the participant)?" The responses can be considered justifications of their answers to the judgment question. These justifications were categorized (post-hoc) into thirteen varieties and then lumped more broadly into three major explanatory groupings (see Table 2). The groups of justification types represent three types of reasoning: perceptual, behavioral/functional, and biological.

Table 2. List of justifications, grouped by type of reasoning and listed in order of decreasing frequency. The examples are taken from the experimental sessions

NAME	DEFINITION	EXAMPLE
Perceptual		
Features	Lists one or more features.	"Because [the zebra] is black and white."
Whole	Cites the perception of the whole object.	"Because that looks like a koala, not a tazier"
Definition	Cites a feature or whole object as defining	"Because horses don't have stripes."
Told	Cites interviewer's definitions.	"Because you told me so."
Behavioral	**/Functional**	
Transform	Cites the transformation.	" 'Cause they...changed everything."
Function	Explains the function for a particular feature.	"[It's a mailbox] because you put things in it...and those are things you need."
Subj's Behavior	Explains their behavior or reactions.	"[The shark] looks meaner now."
Behavior	Explains how the animal would now behave.	"[It's a goose] because it doesn't climb trees [like a hoatzin], it just flies."
Family	One adult used this comment, difficult to categorize.	"A horse is in the same family as a zebra, so it could be changed."
Biological		
Looks/Is	The object looks like X, but really is Y.	"It's still a guitarfish, it just looks like a shark."
Final	You can change an artifact easily, but not a natural object.	"It's easier to change a ball than the internal structure of an animal to something else.
Out/In	The outside is changed but the insides remain the same	"It's not a zebra 'cause [they didn't] change like anything internal ..."
Parts	The parts are the same, just the configuration is different	"It's made out of everything that was in the original one."

The perceptual and biological types of reasoning were identified first, and they parallel the scoring system used on the judgment scores. The perceptual category, was hypothesized to closely match Keil's description of the children's answers to "What is it now?" that resulted in a score of 1 in his experiment. As stated in the introduction, Keil believed a score of 1 meant the child was being led by the characteristic perceptual features. The biological category was hypothesized to closely match Keil's description of the children's answers to "What is it now?" that were scored as a 3. A score of 3 meant (for Keil) that the child was overriding the features with non-perceptual explanations or essences.

The behavioral/functional category, evolved from an examination of the answers that did not fit in the perceptual or biological categories. These answers refer to properties not visible in the picture, but are also not as essential to the objects as biology or DNA.

Participants often did not limit themselves to only one of the thirteen different varieties of justifications when responding. The coder recorded every justification that was given by the participant, but within a transformation scenario each individual variety was only recorded once. To illustrate the scoring procedure, consider the following answer given by a participant: "That's the koala... that's what it looks like(a), it's got its ears, small eyes, and claws(b), and climbs trees(c)." This answer was scored as having three separate varieties of justification (a = whole percept, b = features, and c = behavior). The fact that three separate features were mentioned was not recorded as three instances of the "features" justification.

CHAPTER 3

RESULTS

Preliminary Analyses

Two independent assistants coded 31 (68%) of the 45 participants on the initial judgment, final judgment and offspring score. The scores given by the two coders agreed 98.0 percent overall, with 98.7% agreement on the initial score, 97.3% on the offspring score, and 98.0% on the final score. If any discrepancies occurred in the scoring of these 31 participants, the scores given by the two coders were then averaged. Just one of the assistants scored the remaining 14 participants. He also was the sole coder for all justifications.

A 2 between (order, forward or reverse) by 3 within (type of object: familiar, novel or artifact) ANOVA was run on the initial judgment score to test for order of presentation effects. There was no significant effect of order ($F(1,43) = .58$, $p = .45$) and no significant interaction ($F(2,86) = .44$, $p = .64$) between order of presentation and the type of items. The same type of ANOVA was run on the offspring score (there were only two types of objects in this score: familiar and novel natural kinds). Neither main effects of order ($F(1,43) = .16$, $p = .69$) nor interactions with item type (F(1,43) = 1.8, p = .19) were significant. Since there were no order of presentation effects, the data for the two different orders of presentation were combined for all subsequent analyses.

Initial Judgment Score

The mean initial judgment scores for seven year-olds, ten year-olds, and adults are depicted in Figure 1. On their initial judgments, the youngest children treated all items nearly identically. Their mean familiar natural kinds score (hereafter, fam), mean novel natural kinds score (hereafter, nov), and mean artifacts score (hereafter, art) were all 1.0. The ten year-olds' responses reflected more variability in responding, where their fam was 1.35 and nov was 1.27, but art remained near one (1.02). The data from adults was similar to the date from ten year-olds. Fam and nov were 1.6 and 1.5, but art was again 1.0. This data shows an increasing separation, as a function of age, in the scores of the natural items, both familiar and novel, from the artifact scores.

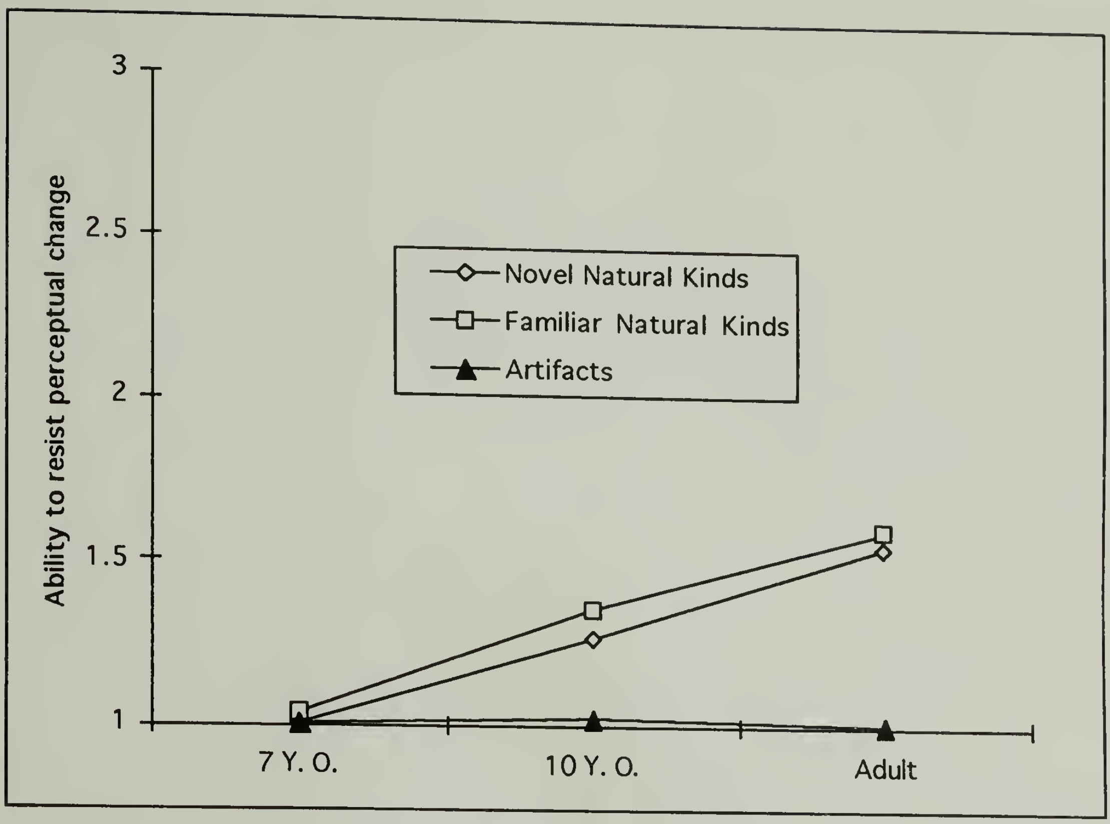


Figure 1. Initial judgment score. This score represents the first answer given to "What is it now?" 1 = judgment that the object has changed, 2 = indecision in judgment, and 3 = judgment that the object is still the same as before the transformation.

A 3 between (age) by 2 within (type: familiar or artifact) ANOVA was run on the initial judgment score. The main effect of the type of object presented was significant ($F(1,42) = 13.61$, $p = .0006$). The main effect for age was not significant ($F(2,42) = 3.18$, $p = .05$), but this analysis includes the performance on artifact items, which did not change with age. There was a significant interaction of kind and age ($F(2,42) = 3.51$, $p = .04$). This interaction replicates Keil's (1989) finding that younger children respond similarly to the two types of objects, and that older children's and adults' responses show a differentiation between the types of objects.

A second analysis of the initial judgment score was carried out to test whether the participants were treating familiar and novel natural kinds similarly. A 3 between (age) by 2 within (type, familiar or novel) ANOVA found a significant increase in the scores due to age ($F(2,42) = 3.77$, $p = .03$). Although the mean score for familiar natural items was higher than for the novel natural items this difference fell just short of significance ($F(1,42) = 3.66$, $p = .062$). There was no interaction effect of novelty with age ($F(2,42) = .25$, $p = .77$). The lack of a novelty main effect is probably due to a restriction in the range of scores, since this experiment most of the scores were 1's, and the means are very low. It is also possible that no real differences exist between the performance on the novel and familiar natural items.

Final Judgment Score

The analysis of the final judgment score presented another challenge. All of the participants finally decided that all artifact items had changed; there was no variation in all 180 scores. Due to this obvious floor effect (or ceiling effect, since the lowest score is the expected answer), there was no variance on which to base an ANOVA that included the artifact items.

The data from the initial and final judgment scores present the same picture. The youngest group did not respond differentially to the three item types as can be seen in Figure 2 (fam was 1.02, nov was 1.02, art was 1.00). In contrast, older participants again differentiated between natural objects and artifacts. Ten year-olds showed small differences between their natural (fam = 1.35, nov 1.33) and artifact (art = 1.00) scores, while the differences were larger in the adult group (fam = 1.78, nov = 1.68, art = 1.00).

The similarity between the initial and final judgment scores is evident in the number of transformation scenarios where subjects changed their minds. To calculate this, the number of questions that received different scores for the initial and final judgments were counted. If a subject's responses were coded by two assistants, each coder's scores were calculated separately and then averaged. Seven year-olds received different initial and final judgment scores only 2 times (out of a possible 180) or 1.1 percent of the time. Ten year-olds received different scores 5 times, or 2.8 %, and adults 9.5 times, or 5.3 %. Most of the data from adults resulted from one participant who consistently (seven times) replied that the natural object really had changed, but when replying to the offspring question stated that it was still really the first object. Even including this one participant, subjects only changed their minds 3.1 percent of the time, resulting in highly similar initial and final judgment scores.

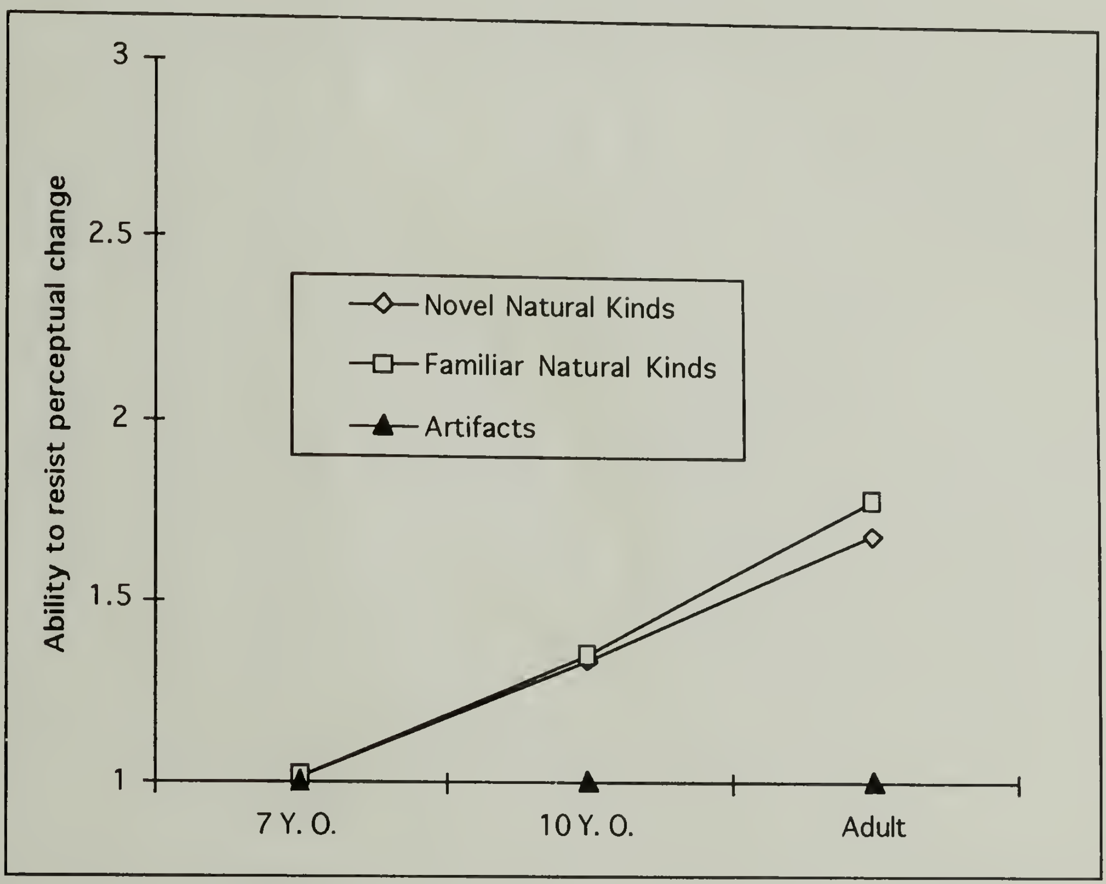


Figure 2. Final judgment score. The final score re-codes the participant's judgment, accounting for any changes in the participants' answers.

A 3 between (age) by 2 within (type: familiar or novel) showed a main effect of age ($F(2,42) = 5.31$, $p = .009$) and no effects for novelty ($F(1,42) = 1.71$, $p = .20$) or interactions with novelty and age. ($F(2,42) = 1.71$, $p = .19$). These results are consistent with those found in the initial judgment score. The novelty effect was not even marginally significant here, however. Participants did not treat the novel and familiar natural objects differently.

Offspring Score

The offspring scores are found in the familiar and novel natural kinds scenarios only. Figure 3 shows that the mean for both types of items increased as a function of age. This increase was also found for familiar and novel natural items in the initial and final judgment scores. However, the offspring question produced higher scores overall than the previous two scores. Seven year-olds were again the lowest scorers, but their mean responses were not near one for the first time (fam = 1.48, nov = 1.45). Adults scored higher, as expected (fam = 2.27, nov = 2.14). Ten year-old children scored between these two, but it is also interesting that their novel scenario mean (1.79) is higher than their familiar scenario mean (1.73). This is different from the previous two (initial and final judgment) scores, where all age groups scored higher on the familiar items than the novel items.

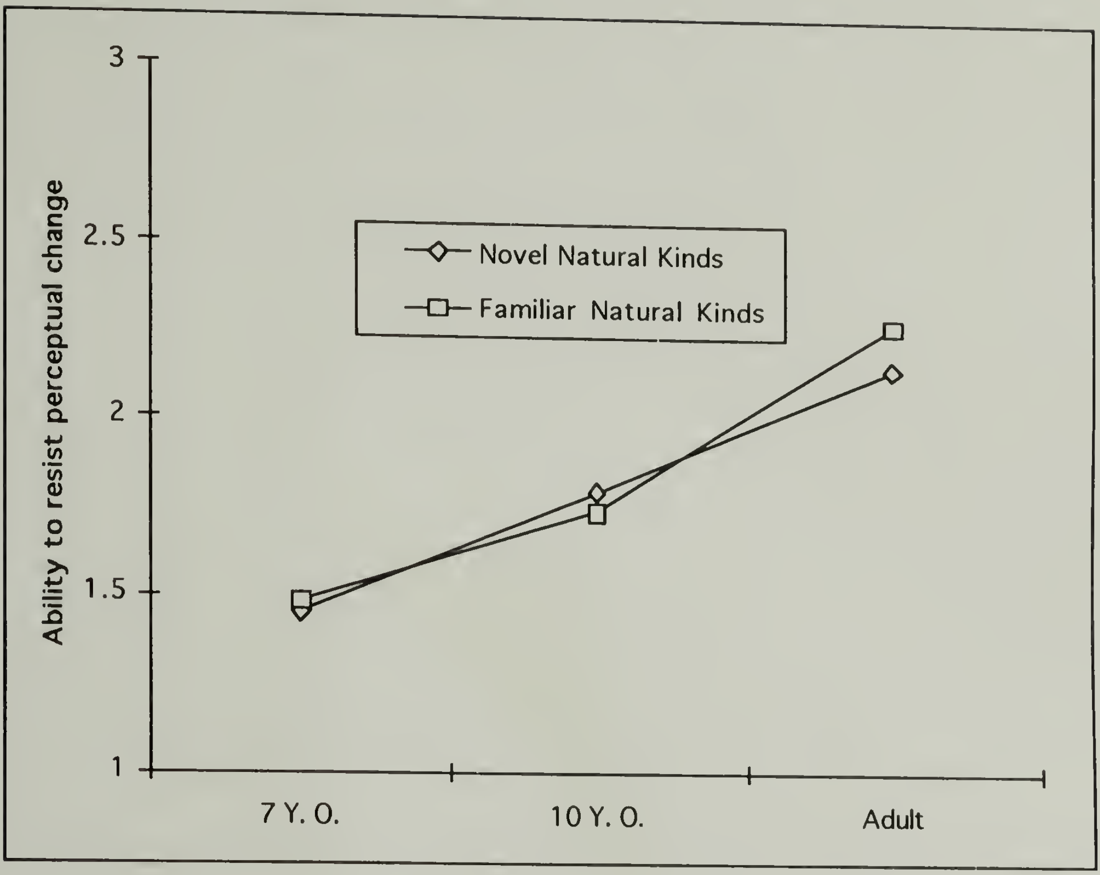


Figure 3. Offspring score. This score represents answers to "What kind of babies will it have?" 1 = the baby was judged to be the same kind as the changed object, 2 = the participant was unsure, 3 = the baby was judged to be the same kind as the original object.

A 3 between (age) by 2 within (type: familiar or novel) ANOVA showed a nearly significant effect of age ($F(2,42) = 3.01$, $p = .06$). The fact that responses to the offspring question did not significantly increase as a function of age was unexpected. It is possible, however, that participants of all ages better understand the biology involved in procreation, increasing scores at all age levels for this question.

As in the initial and final judgment scores, there was no interaction effect for novelty and age ($F(2,42) = 2.36$, $p = .11$). However, unlike the initial score, a main effect for novelty fell far short of significance ($F(1,42) = .30$, $p = .59$). The lack of an effect for novelty replicates the findings from the final judgment score analysis. Furthermore, as stated above, the novel score was higher than the familiar score for the ten year-olds. The fact that these two scores could change their ordinal relationship to each other further suggests that there are no overall effects of novelty, even accounting for the restricted range of scores found in this experiment. Participants used information from their knowledge of familiar natural items on nearly all the novel items.

Justifications

Frequency Count

Three analyses were carried out on the justifications participants had provided for their responses to the transformation scenarios. First, a simple tabulation of the frequencies of the justifications, as a function of age, is given in Table 3. Older participants might have been expected to give more varied answers because of their greater verbal skills. However, the total number of different varieties of justifications did not vary with age. Both the seven and ten year-olds gave a total of 210 different justifications in the experiment, and adults gave only slightly more (220).

Table 3. Simple frequency count of all justifications given, summed over trial type.

Perceptual	Seven Y. O.	Ten Y. O.	Adults	Sub-totals
Features	127	113	68	308
Whole	38	25	50	113
Definition	3	2	0	5
Told	0	0	1	1
Perceptual Sub-Totals	168	140	119	427
Funct./Behav.	**Seven Y. O.**	**Ten Y. O.**	**Adults**	**Sub-Totals**
Transform	35	34	41	110
Function	3	7	10	20
Subj's Behave	3	3	0	6
Behave	0	2	3	5
Family	0	0	1	1
Funct./Behav. Sub-Totals	41	46	55	142
Biological	**Seven Y. O.**	**Ten Y. O.**	**Adults**	**Sub-Totals**
Looks/Is	1	13	28	42
Final	0	8	5	13
Out/In	0	2	9	11
Parts	0	1	4	5
Biological Sub-Totals	1	24	46	71
All Types Sub-Totals	210	210	220	**GRAND TOTAL 640**

Individual Justification Type Analyses

To further analyze the justifications, the data had to be transformed since a participant often responded with more than one type of justification within one transformation scenario. To eliminate problems with having different numbers of answers per question, the justifications were collapsed so that each subject was given perceptual scores, functional/behavioral scores, and biological scores. Each score was calculated in the following manner. A participant scored one "point" for each transformation scenario where he or she had provided a justification from that type. These "points" were summed within each transformation scenario type (familiar, novel and artifact). This summing resulted in three scores that ranged from zero to four. An example from the familiar natural kinds will illustrate this data consolidation. One child answered with both features (perceptual) and transformation (functional/ behavioral) justifications on one question, just the transformation (functional/ behavioral) justification on a second scenario and the looks/is (biological) justification on the final two scenarios. Thus the participant's responses resulted in a perceptual score of 1, a functional/behavioral score of 2 and a biological score of 2 for the familiar natural kinds.

Perceptual scores, as illustrated in Figure 4, did not seem to vary as a function of the type of scenario. These scores were the types of justifications that encoded perceptual information, and perception was equally accessible for all objects. They did, however, slightly decrease as a function of age. For example, familiar natural kinds averaged 3.4 questions answered with a perceptual justification for seven year-olds, 3.1 questions for ten year-olds, and 2.3 questions for adults. A 3 between (age) by 3 within (type of scenario) ANOVA showed only a significant main effect for age ($F(2,42) = 3.229$, $p = .05$).

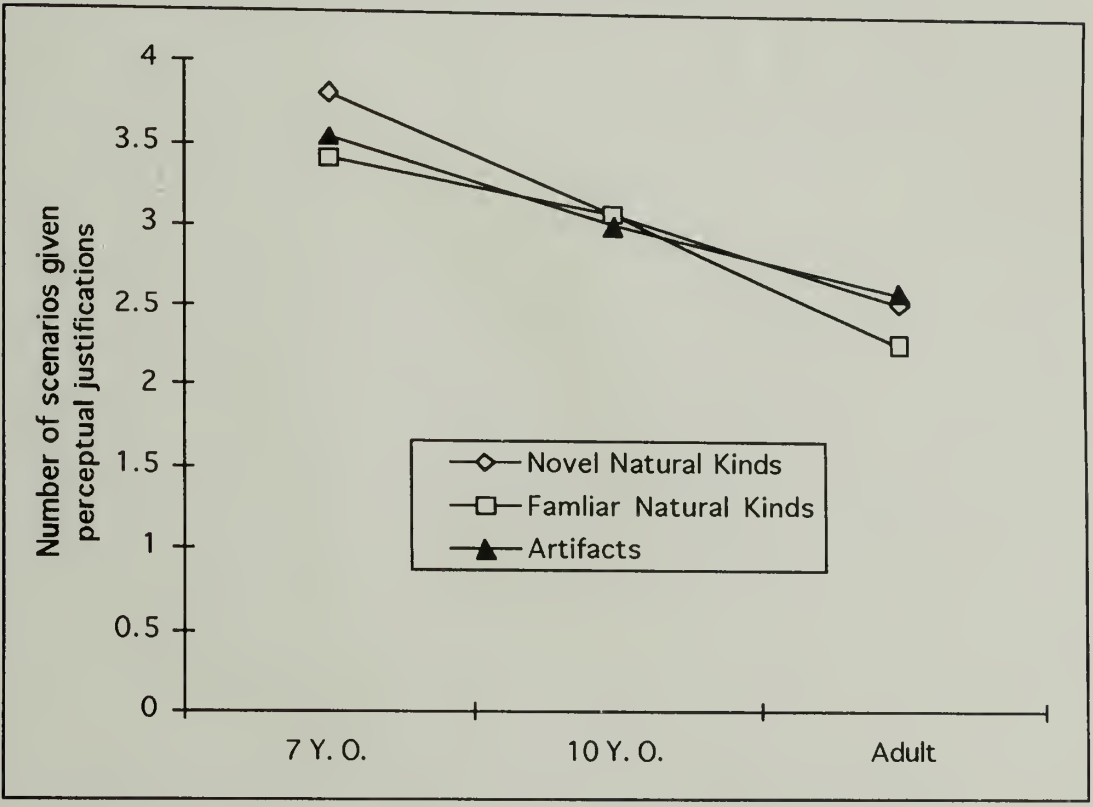


Figure 4. Perceptual justifications score. Each score was the number of scenarios where participants gave perceptual types of justifications. Maximum score = 4 in each category.

As can be seen in Figure 5, functional/behavioral scores showed a very different pattern than the perceptual scores. There was no obvious age trend for the natural kinds of items. The mean for familiar natural kinds answered with functional or behavioral justifications, for example, was 0.6

questions in seven year-olds, 0.8 questions for ten year-olds, and 0.67 for adults. Artifacts, on the other hand, do show an increasing age trend, where the seven year olds' mean was 1.13, and the adults' mean was 1.93. A 3 X 3 ANOVA found that the only significant effect was the main effect for type of scenario ($F(2,84) = 15.132$, $p < .0005$).

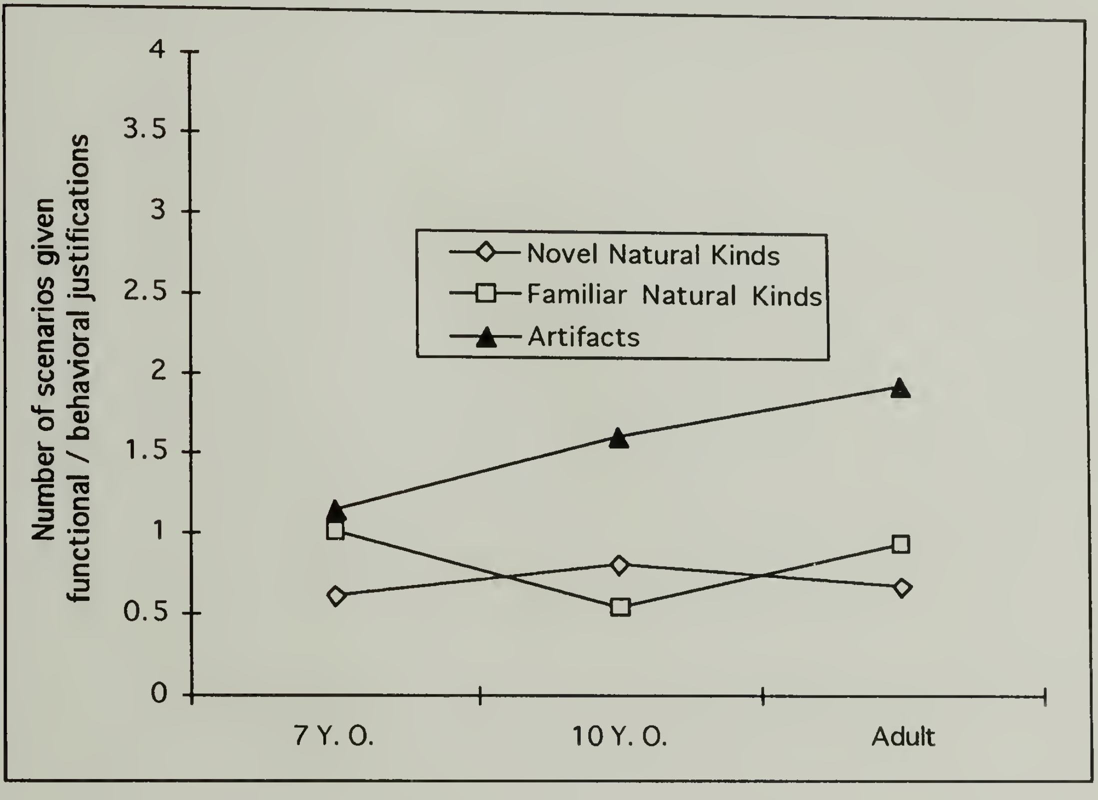


Figure 5. Functional and behavioral justifications score. Each score was the number of scenarios where participants gave functional or behavioral types of justifications. Maximum score = 4 in each category.

Biological justifications would be less relevant to the artifact scenarios than the natural kinds scenarios, and they seem to be the most mature level of explanation. All means for biological scores are illustrated in Figure 6. There was, as in the initial and final judgment scores, an increasing separation between artifact and natural scores as a function of age. Seven year-olds answered no familiar natural kinds or artifacts questions with biological types of justifications and only averaged 0.07 questions on novel natural kinds. The mean for adults, on the other hand, was 0.27 artifacts questions answered with biological justifications, and 1.27 and 1.33 questions on novel and familiar natural scenarios, respectively. Again, a 3 x 3 ANOVA was carried out on the biological scores. Unlike the perceptual and functional/behavioral scores, all effects were significant: main effect of age ($F(2,42) = 9.807$, $p = .031$), main effect of scenario type ($F(2,84) = 12.418$, $p < .0005$) and interaction of age and type ($F(4,84) = 4.774$, $p = .002$).

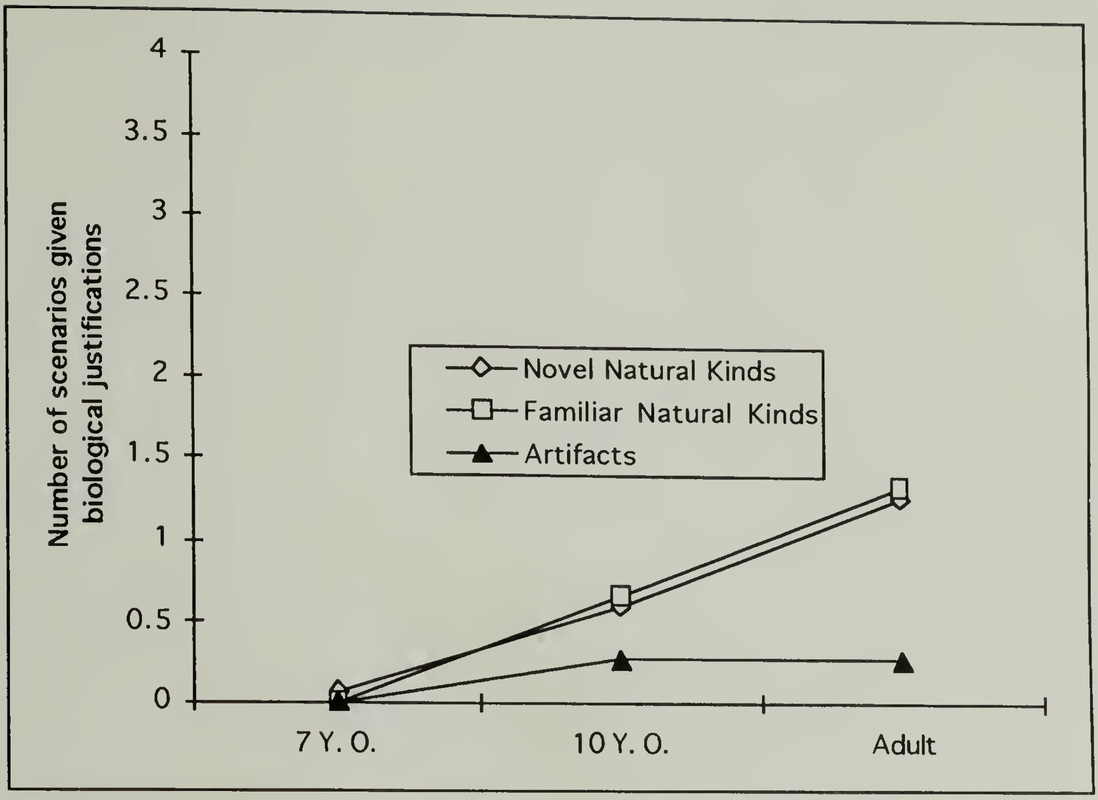


Figure 6. Biological justifications score. Each score was the number of scenarios where participants gave biological types of justifications. Maximum score = 4 in each category.

Each of the justifications scores reveals a story consistent with the data found in the judgment scores. Perceptual justifications simply decrease as a function of age. They are available to participants of all ages and are also the most common types found in all ages. However, these justifications decline in frequency with age and are replaced with functional justifications for artifacts and biological justifications for both types of natural items.

Functional/behavioral scores are indeed valid as a separate designation, especially in the artifact scenarios.

While the mean judgment scores on artifact scenarios remain stable as a function of age, the functional/behavioral justifications in these scenarios become more frequent with age. The functional/behavioral justifications seem to provide a richer source of information by which to analyze the data from the artifact scenarios. The fact that older participants responded more frequently with these functional justifications may indicate that functional justifications represent a higher level of reasoning than the perceptual justifications used by younger participants.

Biological justifications mirror the initial and final judgment scores. Biological reasoning on natural items is more frequent in older participants for both the familiar and novel kinds, but not for the artifact scenarios. This high degree of similarity is to be expected, since the biological justifications category was designed to mirror the examples given by the participants in Keil's experiment who received scores of three. In both the judgment scores and the justifications, more biological reasoning gives higher scores.

Correlating Justifications and Judgments

A final analysis of the justification scores verifies that the perceptual and biological justifications correspond to judgment scores of one and three, as they were designed to do. Four Pearson R correlations compared participants' scores on the initial and final judgments with the number of natural item questions answered with perceptual or biological justifications.

Participants' initial judgment score and their perceptual justifications score showed a strong negative correlation ($r = -.89$). A strong negative correlation was also found between the final judgments and perceptual scores (-.93). These correlations could also be seen as a strong positive correlation between the participants' likelihood of choosing the transformed animal as what it is now, and the number of perceptual justifications they gave.

The correlations between the biological justifications score and the judgment scores show the opposite pattern, as expected. Initial judgments correlated highly positively with biological scores ($r = .95$) as did final judgment scores ($r = .94$). These positive correlations mean that if participants chose the original animal in their judgments, they were extremely likely to give biological justifications for their answers.

The previous correlations were highly predictable, but comparing the justification scores with offspring scores provides an interesting and unexpected picture. Figure 7 graphically represents the data for correlation between perceptual justifications score and offspring score (each diamond may represent more than one data point). The correlation between the perceptual scores and offspring scores is not as strong, but is in the same direction as the correlations with the judgment scores (r = -.66). Figure 7 shows that the reason for the reduction in the strength of the correlation is that there are data points located in the upper right of the graph.

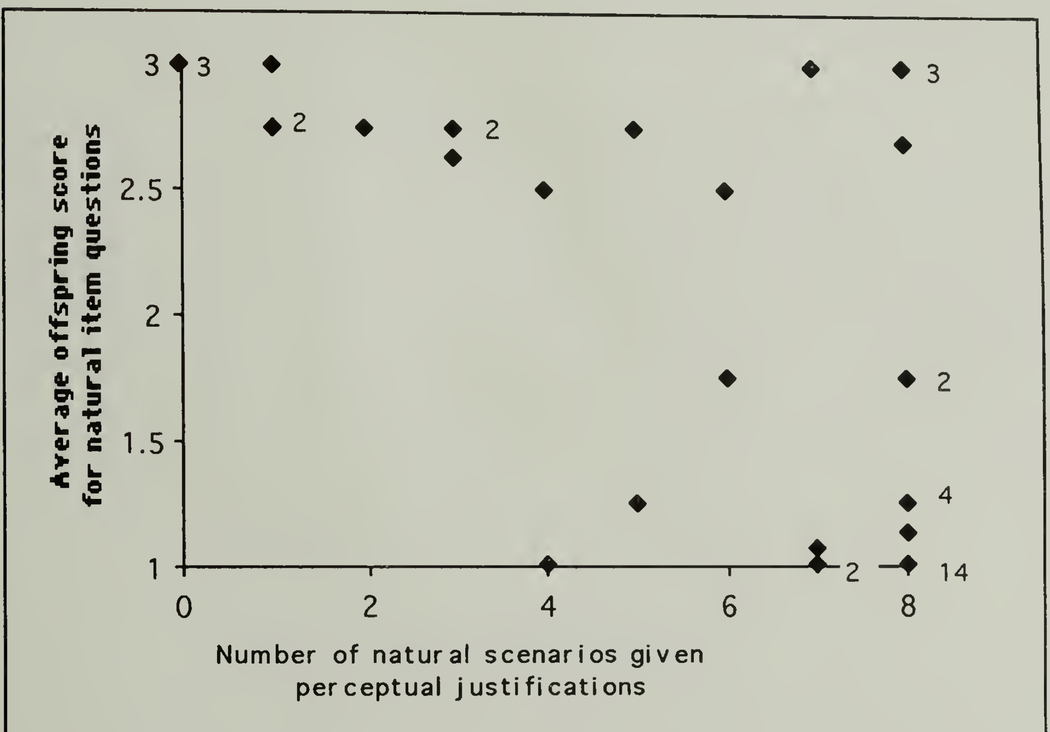


Figure 7. Relationship between perceptual justifications and offspring scores. If a diamond represents more than one data point, the number of data points it represents is given immediately to the right of the diamond.

The correlation between biological scores and offspring scores is again not as strong as, but in the same direction as, the correlation with the judgment scores (r = .66).

Figure 8 shows this positive correlation, and again the points may represent more than one item in the data. The reason for the smaller strength of this correlation is very different. It stems from the fact that the number of biological justifications given by participants was very low, and therefore most of the data is clustered in the bottom left corner of Figure 8.

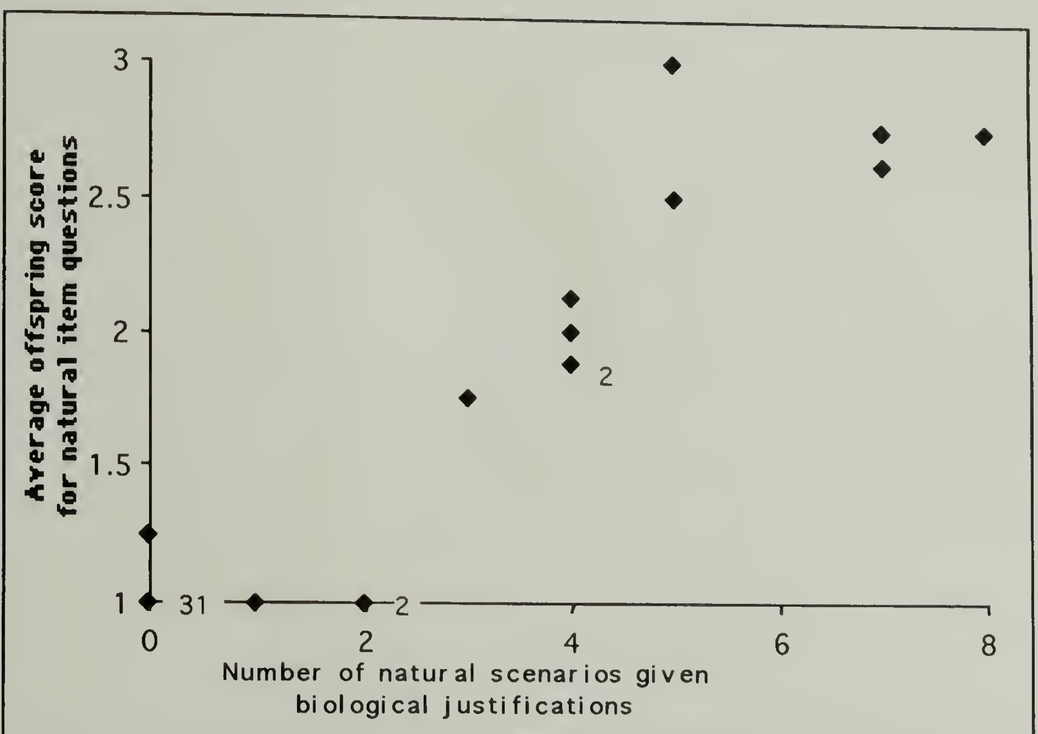


Figure 8. Relationship between biological justifications and offspring scores. If a diamond represents more than one data point, the number of data points it represents is given to the right of the diamond.

When comparing these two graphs, it seems that participants were able to correctly answer the offspring question while still giving perceptual justifications for their judgments of the current identity of the object. The reverse was not true for the biological justifications. If a participant could use biological justifications for his answers, he would then be able to come up with the correct answer for the offspring question.

A few subjects seemed to be in an intermediate stage of understanding about the implications of these transformations. Three seven- year-olds, one ten-year-old, and four adults consistently answered with the changed animal for judgment questions and with the original animal for offspring questions, while other participants echoed this trend less consistently. The eight participants who consistently answered in this pattern also gave high levels of perceptual justifications. Their mean combined natural items perceptual score was 6.6, and four of these participants gave perceptual justifications in all eight natural scenarios. In contrast, the mean for the number of natural scenarios answered with biological justifications for these eight was less than one. These participants, along with others who showed this trend to a lesser degree, make up the outlying data points in Figure 7, while not straying from an ideal 1.0 correlation line in Figure 8.

CHAPTER 4

DISCUSSION

As stated earlier, there were three goals to this thesis. The first was to replicate Keil's (1989) transformation study. The second was to show that examining justifications can more effectively measure the types of intuitive theories people might use in categorization. The final goal was to show that children and adults can generalize what they know from other objects within a category to make inferences about what variable properties a new member of the category might have.

Replication of Keil (1989)

The first goal was only partially met. Keil's basic pattern of findings was replicated. Younger children treated artifacts and natural kinds equally, and older children and adults started to differentiate their answers based on this category difference. The main limitation in this replication of Keil's work is that Keil's mean scores were much higher than the ones reported here. The means Keil reported have been estimated from his graphs and placed in Table 4. As a comparison, Keil's fourth graders' mean score on natural kinds was about 2.8, but the ten year-olds in the present study had mean scores of only 1.35. This large difference in mean scores does not, however, eliminate the replication of the effects originally found by Keil.

Table 4. Scores estimated from Keil's (1989) transformations study

Grade (Mean Age)	Biological Kinds Mean	Artifacts Mean
Kindergarten (5:8)	1.75	1.4
Second Grade (7:6)	2.5	1.2
Fourth Grade (9:9)	2.8	1.3

What could cause this large difference in the range of scores between this study and Keil's? One explanation lies in the fact that maintaining genetic heritage despite perceptual transformation may be a more suitable measure of biological understanding than maintaining identity. The scores obtained in this experiment for the offspring question are more similar to those reported by Keil. Keil also reports that "several kindergartners, though ultimately deciding that kind membership was changed, were nonetheless troubled that the transformed animals had babies of the original kind of animal. ... They therefore may have had some inkling that the type of baby matters, but not enough to override the salient characteristic features." (Keil, 1989) The offspring question seems much more valid and pivotal in understanding biological essences.

The individual data from participants in this study also support the offspring question as a better measure of biological understanding. As stated earlier, at least eight participants believed that a transformed animal could be one kind of object and have a baby of the other kind. These participants seemed to have no hesitation in generating these conflicting answers. Their theory of transformations and biology allowed both of these answers simultaneously.

Another possibility for the difference in responses between this study and Keil's original experiment, is that Keil's experiment had a flexible interview format, with numerous challenges to the children's thinking. The interviewers asked not only, "What is it now?" and, "Why do you say that?" but also asked "Even though" questions that further challenged their thinking. Beyond even that, the interviewer frequently told the children directly what kind of babies an animal would have, and what kind of parents it had. Our study left the question of offspring for the participants to answer.

In pilot testing for the current experiment, a more flexible interview format was used. This interview was similar to Keil's; there were more challenges to the children's thinking, and the interviewer was possibly biased to influence the responses, since she was the author. There is no direct evidence of bias, however. The many different methods attempted in piloting make it difficult to directly compare the pilot data with the current data. However, the single judgment score was in general higher than the means obtained in the current experiment for the initial and final judgment scores. Since most of the items were the same in the pilot and actual experiments, these differences in interview format seem the likely cause for any differences between the means reported in this thesis and those reported by Keil.

What in the format difference might have contributed to the differences between the responses in the two experiments? I believe that differences in interview format had two effects. First of all, Keil's free-form, challenging interview forced the participants to think deeply about the question during the interview. No simple answer went unchallenged, and the participants had to more fully explain their understanding of the transformation. Secondly, the free-form interview format and the informed interviewer may possibly have led the participants to give the expected pattern of responses. By not allowing participants to rest on any simple answer, the challenging interview format may have led the participants to believe that their first answer was false and that they should reexamine and consider a change in their answer. Also, biased interviewers can lead children to inaccurate recall of experienced events (Goodman, et. al., 1995).

The next procedural step is to try to encourage deep contemplation of the questions while still insuring that the experimenter is not leading the participants to the right answer. There are several possible methods that might produce these ideal conditions. The interviewer might have the children listen to a puppet make mistakes in answering the questions, and then judge whether or not the puppet was correct. The children would first be informed that the puppet is from the moon and is often wrong about earth questions. This procedure has been used successfully by linguists who have found that younger children can correct puppets who are making mistakes that they themselves had produced in a language elicitation task (Hiramatsu & Lillo-Martin, 1998).

Another possible method (which might be more appropriate for the older children and adults) is to try to challenge the participants to think deeply about their answers within a fixed script format. This challenge could be accomplished by designing a set of follow-up questions that are arranged in a tree-diagram. Each response could then be specifically challenged with a "even though..." question. For example, in the raccoon to skunk scenario, if the participant responded that the object was a skunk, the experimenter could reply, "So, even though the animal used to be brown with black stripes on its tail, you think it is a skunk?" Or, if they responded that it was still a raccoon, the reply could be, "So, even though the animal now smells like a skunk, you still think it's a raccoon?" This method seems to be the most likely to challenge subjects of all ages to deeply consider their answers, and yet it still retains the unbiased structure of a strict script format.

Analysis of Justifications

The second goal of the experiment met with success. In accordance with their design, the perceptual justifications corresponded to the response that the animal had changed during the transformation, and the biological justifications corresponded to the opposite response. The justification scores and the judgment scores are certainly not independent, and should not be. The importance of these types of justifications lies in the fact that it can show a more complete picture of the participants' intuitive theory of natural kinds and objects. As stated in the introduction, Keil mentions that functional justifications seem to be more evident at later ages for artifacts, but he stops his analysis there. The current experiment successfully replicated that brief comment with a full analysis.

Justification levels also produced a more complete picture of the participants who answered that the object was changed, but the offspring wouldn't be. The fact that these participants usually used perceptual justifications reinforces the idea that they are in an intermediate level of understanding concerning these transformations.

Inference Within Categories Using Rules or Variables

In regard to the third goal, there were several possible outcomes for the subject's responses to the novel objects. The first outcome is that no participant would be able to generalize what she know about natural kinds to the new natural objects. This outcome, of course, is not what we would predict based on intuition. However, all models of categorization before the Theory Theory would predict this outcome. As stated earlier, Marcus (1997; in press) has argued that these models have no mechanism for dealing with inferring new properties (hoatzin babies) to new objects (hoatzins) within known categories (animals).

The second possible outcome is that all participants would treat the novel exactly the same as the familiar. There would then be no differences in their answers to familiar and novel natural objects. Once children and adults have included the object into a category, they would be able to generalize properties to all members of the category equally. This outcome would show that the inference mechanism for the theory of a category was something similar to a rule. The children would just substitute one instantiation of a variable (or exemplar of a category) for another in the rule (Marcus, 1997).

The evidence from this experiment would seem to support the second possibility. There were large differences between the older participants' responses to novel items and the artifacts, but no significant differences were found between the novel and familiar natural kind items on any of the measures. The only possible problem with this evidence is the restriction in the range of answers. If the range were larger, would the small difference between novel and familiar on the initial judgment score have been significant? The replications proposed above would hopefully remove any doubt about this question. I predict, however, that even with larger ranges of responses there will be no significant effect of novelty. This prediction is based partly on the fact that the differences between familiar and novel scenarios were not nearly significant on many measures, and partly on the fact that the means had switched ordinal position as the highest average for the ten year-olds on the offspring question.

This thesis shows that the ability to infer novel properties to novel examples of known categories is not developing, at least within the age range tested. All participants scored approximately equally on the novel and familiar natural items, since there were no age by novelty interactions on any of the measures. If the information was available to the participants, they used it equally well on the novel items, regardless of age.

Thus it seems that Keil's experiment, with a few modifications, was able to shed new light on the structure of the theories behind categories like natural kinds. Since all participants, children and adults alike, generalize novel properties to novel animals, they are probably using some variant of a rule system. Also, the Theory Theory, unlike any of the early models of categorization, can incorporate such a rule system into it's current model of categorization. To account for children's and adults' ability to generalize variable properties to novel items, the structure for the Theory Theory, or any future models of categorization, must include a rule or variable manipulating mechanism.

APPENDIX A

SCRIPTS

Scenarios are presented in this appendix according to type of object pair. Numbers after the title correspond to placement in the order of forward presentation.

Preamble

For adult participants only
This experiment is designed for seven and ten year old children. We are testing adults on the same concepts. Please take the questions relatively seriously, even though they may seem a little simplistic to an adult.

For all participants
What we're going to do today is, I'm going to show you some pictures, and tell you some stories about them, and then ask you some questions. OK?

The stories are all about a group of very good doctors who perform special operations. Have you ever heard of operations called plastic surgery? That's where a doctor can change how a person's face looks so they look like someone totally different ... well, that's the kind of operations these doctors are going to do. They are going to change the way things look.

Familiar Natural Objects

Horse - Zebra - 2

The doctors took a horse and did an operation that dyed black and white stripes all over its body. They braided its tail, and they operated on its mane to make it stiffer, so it would stand up like the bristles of a toothbrush. When they were all done, the animal looked just like this.
1) When the doctors were finished, what kind of animal was it? Was it a horse or a zebra?
2) OK, why isn't it a ______?
3) What kind of babies do you think it would have? Would it have zebra or horse babies?
4) Do you think it would like to live in a bathtub?

Potato - Apple - 3

The doctors took a potato and did surgery that took the eyes (those black spots) from the outside, and put them in middle in the inside. They polished the outside until it was shiny and smooth. They dyed the outside red. They gave it shots of sugar to make it sweeter. Finally they pushed a stem into the top, and when they pushed it in, the top and bottom were dented. Now it looks like this.

1) What is it now, is it a potato or an apple?
2) Why isn't it a ______?
3) If you look the whole thing and planted it in the ground, what would grow, an apple tree or a potato plant?
4) Do you think it would make a good baseball?

<u>Tiger - Lion - 7</u>

The doctors took a tiger, like this, and bleached his fur so that the black stripes went away and the orange parts looked yellow. Then they gave him a lot of thick hair around his head, and a little puff of fur at the end of his tail. When the doctors were done, he looked like this.

1) Now, when the doctors were all done, and the animal looks like this, what kind of animal is he? Is he a tiger or a lion?
2) OK, why isn't he a _______?
3) Do you think he would like to sleep in your bed?
4) If he had babies, what kind of babies would they be? Lions or tigers?

<u>Raccoon - Skunk - 8</u>

The doctors took a raccoon and shaved away some of its fur. They put the extra fur on its tail so it was big and fluffy. They dyed the fur all black. Then they bleached a single stripe all white down the center of its back. Then, with surgery, they put in its body a sac of super smelly odor. When they were all done the animal looked like this.

1) After the operation, what kind of an animal was it? A skunk or a raccoon?
2) OK, why do you think that?
3) Do you think it would like to watch TV?
4) If it had babies, what kind of babies would it have? Raccoons or Skunks?

<u>Novel Natural Objects</u>

<u>Tazier - Koala - 1</u>

This is a tazier (TAH-ZEER). She lives in the trees in tropical rain forests. She has long thin fingers to get a good grip on the tree branches when she climbs. She is also nocturnal, which means that she sleep during the day, and comes out to eat bugs at night. Her big eyes help her see well in the dark.

The doctors took the tazier and made her eyes smaller. They gave her claws instead of those skinny fingers. They also did surgery to make her body thicker and furrier. Finally, they dyed her fur gray. Now she looks like this.

1) Now that the operation is all done, and she looks like this, what kind of animal is she? Is she a koala or a tazier?
2) OK, why do you think that?

3) If she had babies, what kind of animal would they be? Would they be taziers or koalas?
4) Do you think she could play the guitar now?

<u>Guitarfish - Shark - 5</u>

This is a guitarfish. It lives in the ocean near the Caribbean. It likes to lie flat on the bottom of the ocean and hide. Its mouth is on the bottom of its head, so it eats scraps of food left on the sea floor.

The doctors took the fish, and made its head narrow, not like a triangle. They made its mouth bigger, and gave it a few more fins. When they were all done, it looked like this.
1) Now that the doctors are all done, and the fish looks like this, what kind of fish is it? A shark or a guitarfish?
2) Why do you think that?
3) Do you think it would tell funny jokes?
4) What kind of babies would it have, guitarfish or sharks?

<u>Soursop - Cucumber - 6</u>

This is a Soursop. It's a fruit that grows in the Caribbean, which is a bunch of islands south of Florida. It is big, and has a rough skin. It has lots of little black seeds inside, and tastes a little sour. We, in America make lemonade out of sour tasting lemons and water and sugar, and they make drinks out of the soursop that way, too. They also make ice cream out of the soursop by adding milk, vanilla, water and sugar.

The doctors took the soursop, and sucked all the insides out, so the skin shriveled up and got a little darker. They bleached the seeds so they were white, and they took most of the sour taste out of the fruit by soaking it in water for a long time. Then, they tried to put everything back in, but the skin had shrunk so they could only get half of the insides back in. It wasn't so round anymore. When they were all done it looked like this.
1) So now that the doctors are all done, and it looks like this, what kind of plant is it? a cucumber or a Soursop?
2) Uh huh, why do you think that?
3) Do you think it would like to sing?
4) If you took the seeds, and planted them, what would grow? A Soursop plant or a cucumber plant?

<u>Hoatzin - Goose - 11</u>

This is a special bird that lives in the jungle of Guatemala. It's called a Hoatzin (WAT-ZEEN). When it is young, it has extra claws on its wings and uses all four sets of claws to climb trees. When it is older, like this one, the claws fall off, but it still uses its wings to climb trees like arms.

The doctors took the hoatzin, and dyed its belly white. They gave it a longer neck, and a long, skinny head and beak. Then they gave it webbed feet instead of claw feet. Finally, they made its tail shorter. Now it looks like this.

1) Now that the doctors are all finished, what kind of bird is it? A hoatzin or a goose?
2) How do you know that?
3) What kind of babies do you think it will have? Geese or Hoatzins?
4) Do you think it would like to dance?

Artifacts

<u>Key - Coin - 4</u>

The doctors took a key, like this, and heated it until it was so hot that they could mold it into a circle. Then they put the circle into a big machine that pressed the circle between two pieces of metal that had pictures on them. When the circle came out, it looked like this.
1) Now, when the doctors were all done, and the stuff looks like this, what kind of thing is it? a coin or a key?
2) Uh huh, how do you know that?
3) Do you think it would be heavy?
4) Could you break it in half?

<u>Soccer Ball - Bowling Ball - 9</u>

The doctors took a soccer ball that looked like this. They painted it all black, and cut three little holes in it. Then, they filled it with some really dense plastic, so it was really heavy. When the doctors were done, it looked like this.
1) So now that the doctors are all done, and it looks like this, what kind of ball is it? A bowling ball or a soccer ball?
2) Uh huh, how do you know that?
3) Would it still be able to roll down a hill?
4) Could you plug it in and listen to the radio?

<u>Umbrella - Flag - 10</u>

The doctors took an umbrella, like this, and they took the cloth parts off. They cut up the cloth, and re-sewed it to look like a rectangle. Then they dyed the rectangle in three big stripes of red, yellow and blue. Then they bent the metal handle so it was a straight pole. They used the extra metal from the spines of the umbrella to attach the cloth to the pole. When they were done with the operation, it looked like this.
1) When they were finished, what kind of thing was it? was it an umbrella or a flag?
2) OK, why isn't it a ___________?
3) Do you think it will taste good?
4) Do you think it looks like a flag of a country?

<u>Trash can - Mailbox - 12</u>

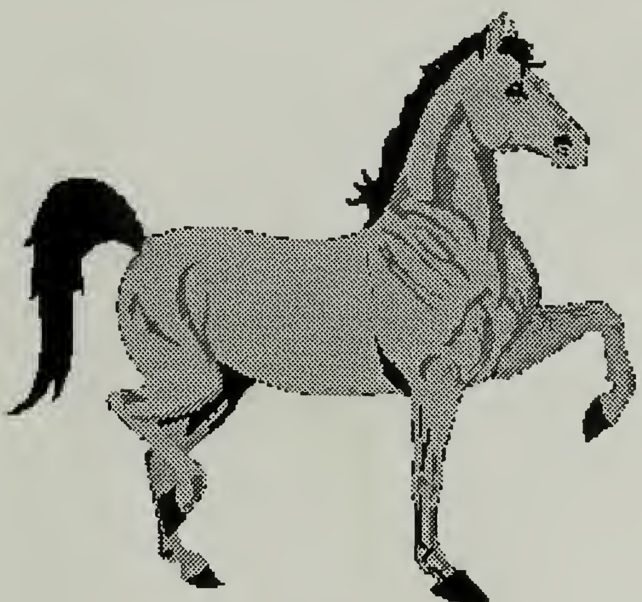
The doctors took a trash can that looked like this. They slit it down the side, and took some of the metal out so it would go straight up, not flare out at the top. They pounded the can flat on three sides, and left one side curved. They took the extra metal and made a lid for the can, and a little flag to put on the side of the can. They painted the flag red. When they were all done, they put the thing on top of a tall piece of wood, so that it looked like this.

1) When the doctors were done, what kind of thing did they have? A trash can or a mailbox?
2) OK, why do you think that?
3) Do you think it could feel sad?
4) Does it look like your mailbox at home?

APPENDIX B

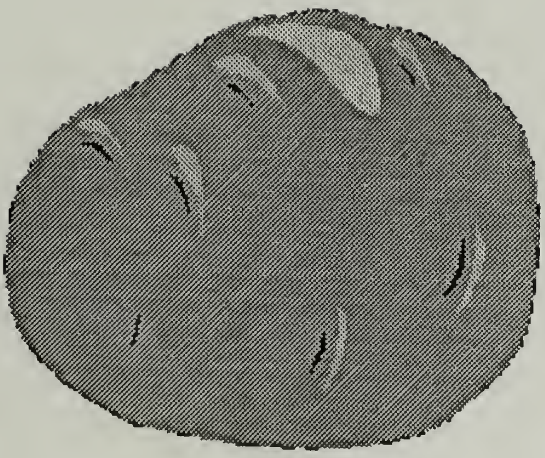
PICTURES

Familiar Natural Kinds Object Pairs

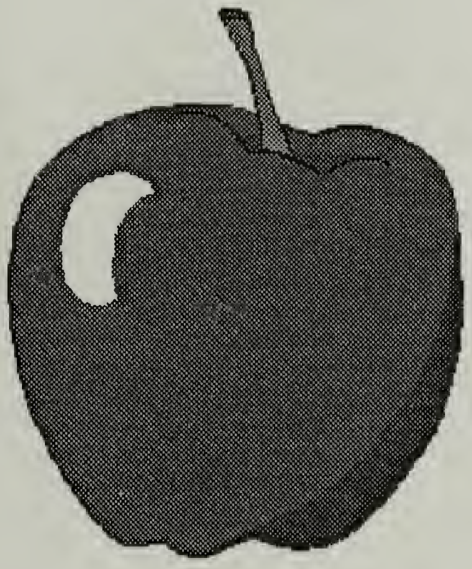


Horse

Zebra

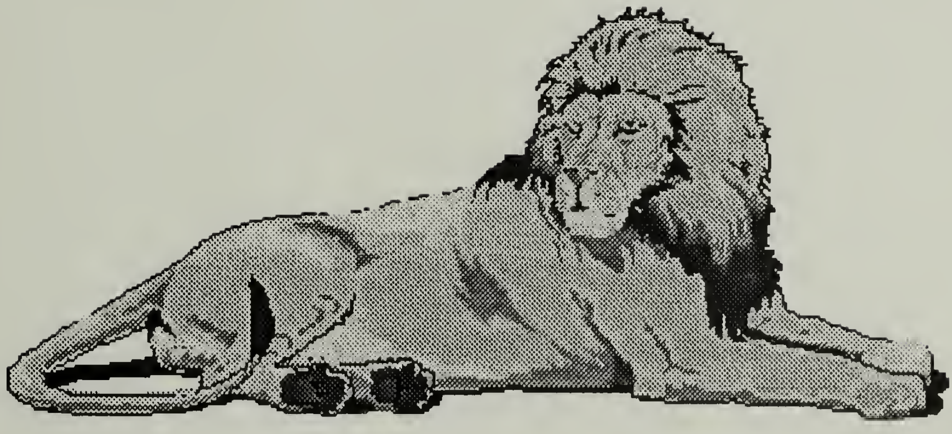


Potato



Apple

Tiger

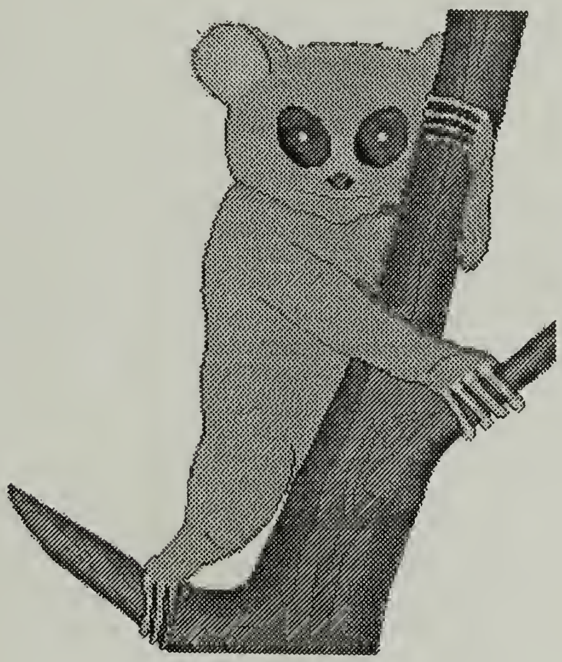


Lion

Raccoon

Skunk

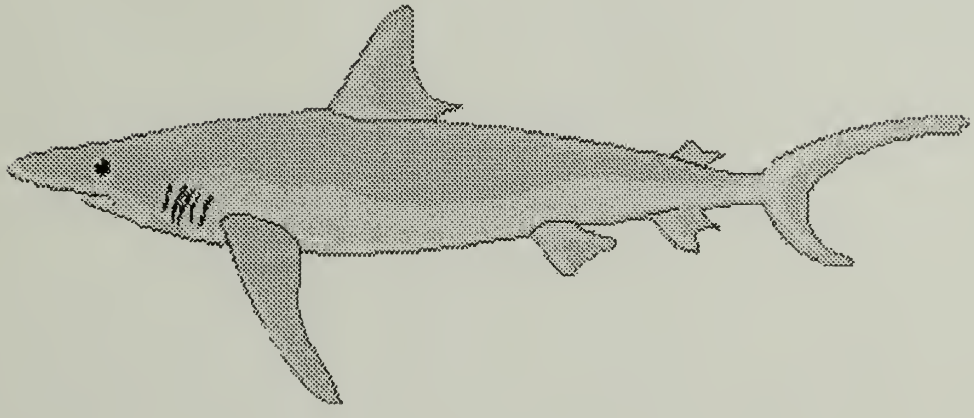
Novel Natural Kinds Object Pairs



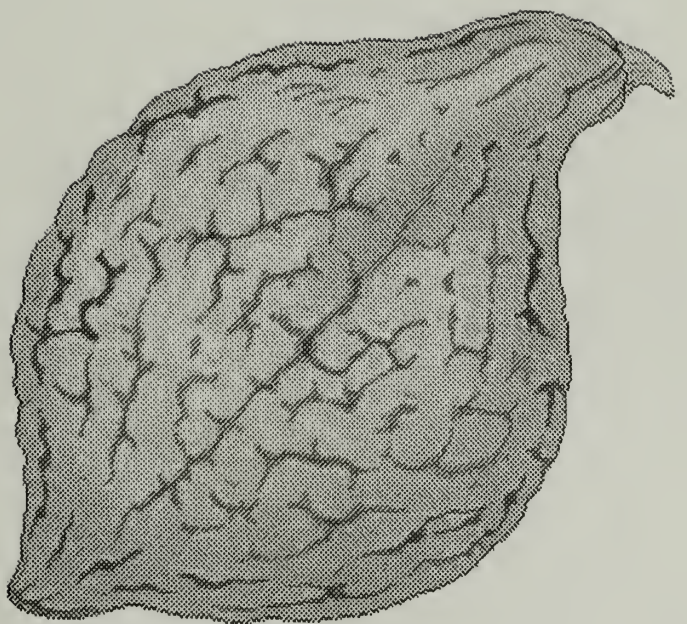
Tazier

Koala

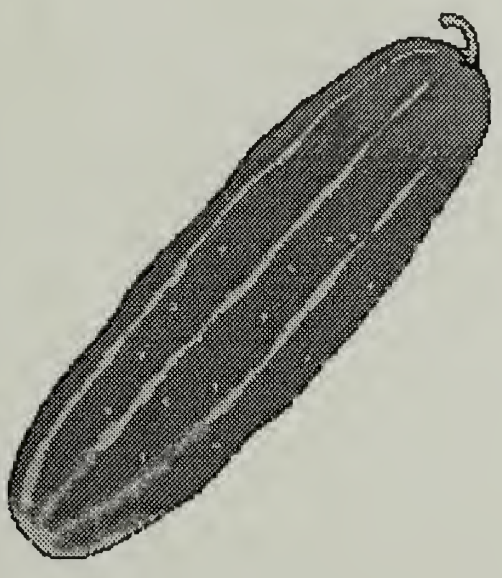
Guitarfish



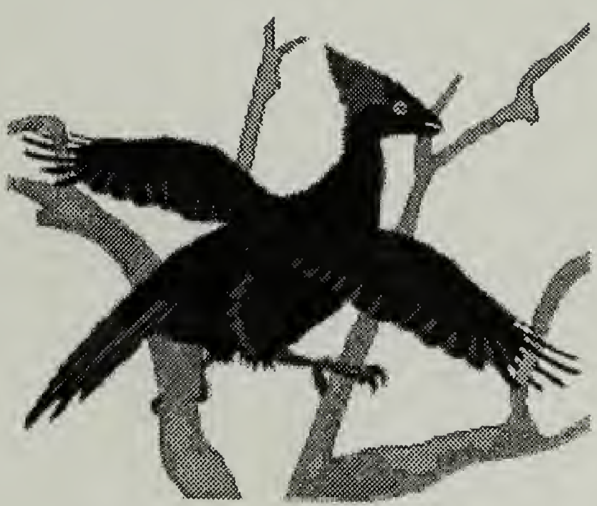
Shark



Soursop



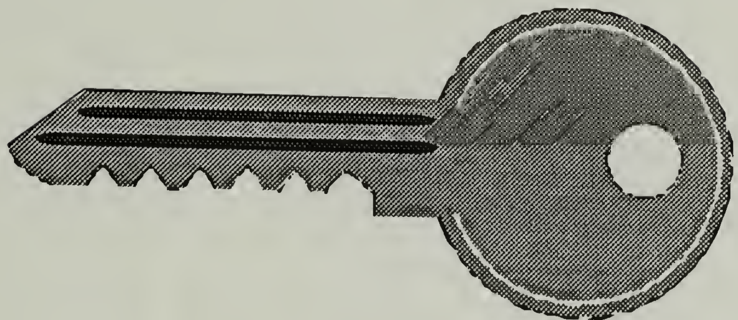
Cucumber



Hoatzin

Goose

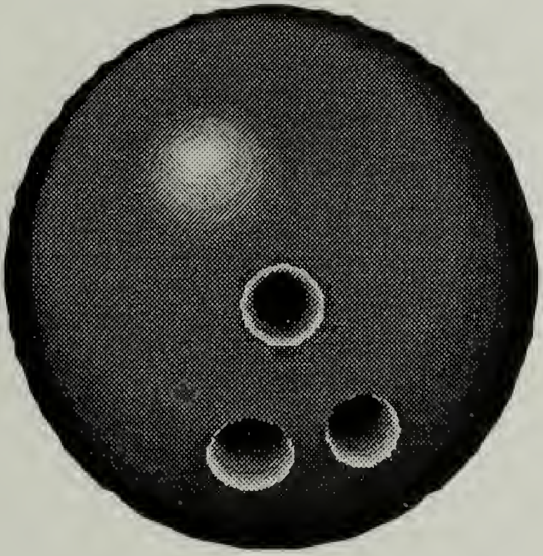
Artifact Object Pairs



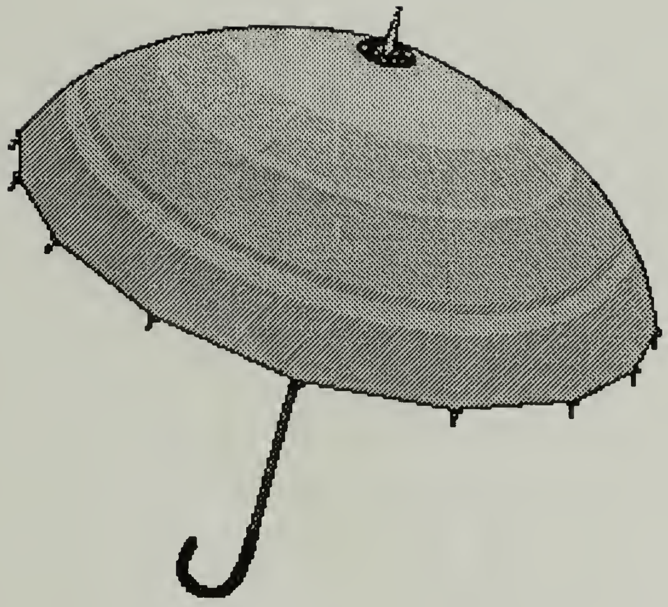
Key

Coin

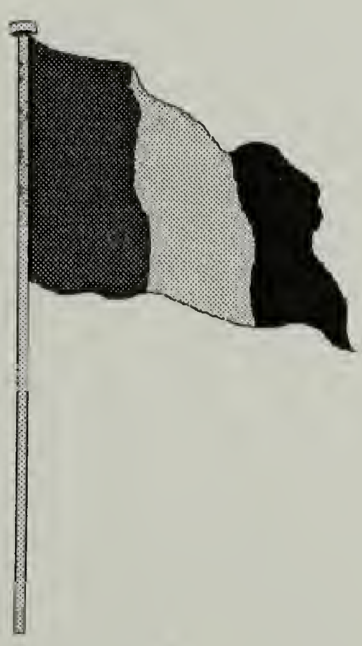
Soccer Ball



Bowling Ball

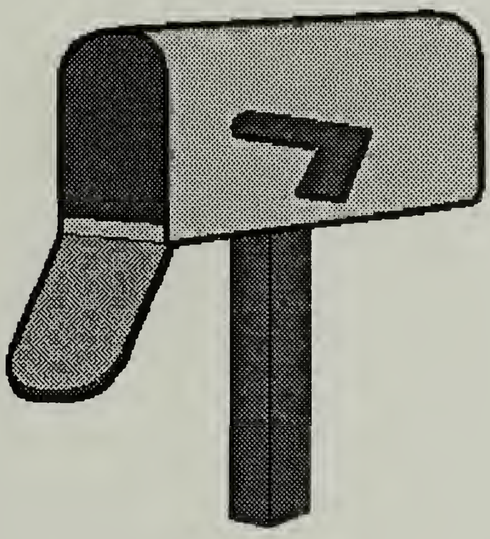


Umbrella



Flag

Trash Can



Mailbox

BIBLIOGRAPHY

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. Cognition, **13**: 263-308.

Carey, S., (1985). Conceptual Change in Childhood. Cambridge, MA: MIT Press.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. Psychological Bulletin, **113**: 403-439.

Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. Cognitive Development, **5**: 151-176.

Gelman, S. A., & Hirschfeld, L. A. (1997). How biological is essentialism? Manuscript at the University of Michigan.

Goodman, G. S., Sharma, A., Thomas, S. F., & Considine, M. G. (1995). Mother knows best: Effects of relationship status and interviewer bias on children's memory. Journal of Experimental Child Psychology, **60**: 195-228.

Gopnik, A., & Meltzoff, A. N. (1997). Words, Thoughts, and Theories. Cambridge, MA: MIT Press.

Hiramatsu, K., & Lillo-Martin, D. (1998). Children who judge ungrammatical what they produce. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh, (Eds.). Proceedings of the 22nd Annual Boston University Conference on Language Development, Vol. 1. Sommerville, MA: Cascadilla Press.

Humann, P. (1989). Reef Fish Identification: Florida, Caribbean, Bahamas. Orlando, FL: Vaughan Press.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. Psychological Review. **104**, 427-466.

Keil, F. C. (1989). Concepts, Kinds and Cognitive Development. Cambridge, MA: MIT Press.

Keil, F. C. (1989). The emergence of theoritical beliefs as constraints on concepts. in S. Carey & R. Gelman (Eds.) The Epigenesis of Mind: Essay on Biology and Cognition. Hillside, NJ: Erlbaum.

Keil, F. C. (1994). Explanation, association and the acquisition of word meaning. Lingua, **92**: 169-196.

Kuhn, T. S. (1962). The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Marcus, G. F. (in press). The Algebraic Mind: Reflections on Connectionism and Cognitive Science. Cambridge, MA: MIT Press.

Marcus, G. F. (1997). Concepts, features, and variables. Manuscript at the University of Massachusetts.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. Cognitive Psychology, **20**: 158-190.

Morton, A. (1980). Frames of Mind. Oxford: Clarendon Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. Psychological Review, **92**: 289-316.

Putnam, H. (1975). The meaning of meaning. In H. Putnam (Ed.), Mind, Language and Reality, Vol. 2. London: Cambridge University Press.

Richards, P.. (1970). The Life of the Jungle. New York, NY: McGraw-Hill.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd, (Eds.), Cognition and Categorization. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.) The Psychology of Learning nad Motivation. San Diego: Academic Press.

Smith, E. E., & Medin, D. L. (1981). Categories and Concepts. Cambridge, MA: Harvard University Press.

Sookia, D. (1994). Caribbean Cooking. Seacaucus, NJ: Chartwell Books.

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infant's knowledge of object motion and human action. In D. Sperber, D. Premack, and A. J. Premack (Eds.) Causal Cognition. Oxford: Clarendon Press.

Wellman, H. M. (1990). The Child's Theory of Mind. Cambridge, MA: MIT Press.