Running Head: CATEGORY INFERENCES

Children’s Category Inferences and Mechanisms

Gayle Kaminsky

Hofstra University

Abstract

First and third graders, ages 6-9, from three low socioeconomic elementary schools and one after-school program were examined to investigate if they use sample size when they make estimates about the population. The research also examines reasons why children make different estimates depending on if they are looking at behavioral or biological categories. It was proposed that older children were more likely to make category-based inferences if they saw a greater number of animals with a characteristic. A repeated measures ANOVA showed differences in the inferences the different age groups reported. I found no relationship between whether children made more category-based inferences about biological or psychological characteristic. I found no significant difference developmentally in the ability for children to use sample size when I examined the different age groups. I also found children use psychological mechanistic answers more often than biological ones to explain why they made their category-based inference.

Children's Category Inferences and Mechanisms

Categories play an important role when making inferences in everyday life. When you see a cat you automatically assume it has whiskers, and when you see a person you presume that it has two eyes. Jacobs and Narloch (2001) emphasize children tend to make inferences based on a restricted amount of events and a narrow amount of information. They see one event and make assumptions about all things that belong to that event's category. Is the inference that the cat has whiskers created because you saw many cats with whiskers, and the inference made about two eyes based on the fact that most people have two eyes? Furthermore, how many cats with whiskers does a person have to see before they are convinced that the next cat they see will have whiskers? Will it matter if it is biological category or a behavioral one?

Osherson, Smith, Wilkie, López, and Shafir, (1990) found evidence that adults' category-based judgments are based on a variety of information including sample size and diversity. In comparison, as children move through elementary school they become more in tune with the understanding of sample size and variability (Kuhn, 1995; Schauble 1996). Mechanisms are also important to consider because everything has a cause and effect, and by learning about them it will contribute to having a better understanding of the category based inferences that are made in daily lives.

Jacobs and Narloch (2001) examined children's ability to employ sample size to categorical reasoning. They found that children from a very young age recognize some variability. The researchers observed first graders', third graders', fifth graders' and college students' ability to estimate about the population from a sample and their perceptions of variability. To measure population estimates and perceptions of variability the researchers asked subjects to estimate the percentage of population that share the same characteristic as the trait shown in the sample they observed. They assessed behavioral (e.g., color of shirts worn to school) and biological characteristics (e.g., number of eyes) and manipulated variability within each area characteristics by comparing categories that are highly variable (e.g., color of shirt worn to school) and those that have small variability (e.g., number of eyes).

Jacobs and Narloch (2001) found evidence that children use sample size and variability to make inferences. Sample size had little effect on characteristics with low variability (number of eyes), but a large impact was found with high varying characteristics (size). Hence, the researches found that children as young as first graders can take into account when things have common characteristics. Thus, they are more likely to infer that other things that have those characteristics belong to that same category. Furthermore, if there is less commonality among characteristics, children as young as first grade will be less likely to infer from it when making decisions about other things belonging to that category.

When Jacobs and Narloch (2001) examined scores from children estimating about the population from a sample, they found older children generalize less from what they saw than younger children. Nevertheless, they concluded there were no differences developmentally when using sample size. All developmental ages included sample size and variability when making population estimates. When making inferences from the population, children take into consideration whether the total number of things that share the trait being examined was regarded as high or low. When the total number of a certain trait is high, children infer a greater number of other things belonging to that category and when the total number of a characteristic is low, children will estimate a lower number. If children look at a characteristic that is highly variable (e.g., color of shirt worn to school), they are less likely to generalize from the data, but when variability is low (e.g., writing tools) size is more important when drawing a conclusion. When children estimate the percentage of the population that have similar biological characteristic as the one presented, they are more likely to use the size of the sample of biological characteristics, than with behavioral characteristics. Thus, children may see biological characteristics as less variable than behavioral characteristics. An illustration of this is no matter how many people a child sees wearing a blue shirt worn to school, the child may be less willing to make an inference that the next person to walk through the door will also be wearing a blue shirt. This is in comparison to a child having more confidence about inferring that the next person that walks through the door would have two eyes rather that one or five, even if he/she saw someone or even a few individuals with one eye rather than two. Their data implies that most of the developmental differences in the population estimates, especially in first graders, in biological and behavioral characteristics seem to be caused by sample variability rather than sample size.

Lopez, Gelman, Gutheil, and Smith (1992) examined children's inductive inferences which reflect basic reasoning processes that are not likely to have been learned explicitly. They found that by second grade children recognize how imperative it is to use sample size and diversity, which indicates that the skill emerges sometime between kindergarten and second grade.

Other research by Gutheil and Gelman (1997) stresses that based on a limited amount of data, cognitive systems infer by reasoning about the environment, combining probability and previous instances to a make a more general conclusion. The inductive knowledge is generally based on category information that is used to make inferences from that category to other things that belong to that category. People utilize category based inductions to reason beyond their experiences in addition to helping them generalize from what they know to the things that they do not know. In other words, people do not have to know everything about a situation; they use their preexisting schemas, plus the information supplied, to make a conclusion about a category. For example, an individual may take into consideration that people have tonsils, and that in the same category as humans are other mammals such as monkeys. Based on the fact that both people and monkeys are mammals, people may be likely to infer that if humans have tonsils then so do monkeys.

Gutheil and Gelman (1997) examined the basic-level category induction within the Osherson, et al. (1990) model. Gutheil and Gelman examined two types of information from the model; sample size and sample diversity. The researchers emphasize that fictional biological properties such as those used in previous research for lower -level induction are not appropriate for assessing children's understanding of sample size and diversity of information for simple inductions. This critique is because the properties that the researchers used at the basic level may not have been sufficiently variable (e.g. "have nesium inside"), and children may think that everything belonging to a basic category will have similar properties.

Gutheil and Gelman (1997) focused the subjects' attention on the importance of the premise characteristic at the basic level by employing types of properties that were highly variable in nature, to evaluate children's use of diversity. Gutheil et al. (1997)

presented university undergraduates animal properties (a fish with pink under its gills) that were considered to be a variable characteristic and assessed subjects' lower-level inductive decisions that were made about the properties. Participants read statements about different animals and observed the properties and estimated how common the properties would be to other animals of that type. The researchers found that adults use the amount of numbers when making judgments (e.g., if subjects were told that a sample of five had soft skin, then they reported a greater number of daskons turtles had soft skin on their stomachs than if they were told a sample of one had that same property). Subsequently, the researchers changed the format of the questions a little and added subordinate categories. The college students expressed if the sample characteristic generalized to one member of the category, not a percentage of the category as a total (e.g., You see five daskon turtles that have soft skin on their stomachs, and then you see a single daskon. Do you think it has soft skin on its stomach?). The results indicate that adults use sample size to make inferences with basic-level categories.

They then asked some adults and some third-graders to determine the property that is most likely present in a specific picture by picking between two arguments (e.g, "Here is a butterfly. It has blue eyes. Here are five butterflies with grey eyes." They were then shown a dissimilar picture and asked "Will it most likely have blue eyes or grey eyes?"). At this point the results implied that children do not use sample size or sample diversity to make inferences as adults do. Furthermore, children did not make use of information about diversity when coming up with their inference. The researches felt that children may be using similarity to make judgments, so, they had some second and third graders perform the same procedure, but did not show them a conclusion picture so it would not distract them. Children then inconsistently applied sample diversity. The researches propose that children can make inferences using sample size and diversity, but they do not easily recognize the importance of utilizing it for inductive reasoning. They also propose that children only make use of sample size and sample diversity when they are together, and there are no other sources available for them to use. Furthermore, when children make inferences they may be preoccupied by the category information, distracting them from noticing important information such as the sample size and sample diversity. It supports Lopez et al. (1992)'s findings that by second grade children can recognize the importance of using sample size and diversity but, Lopez et al. adds to the body of research by supplying a reason why children do not appreciate the importance of sample size and diversity for inferring.

Rehder (2003) examined a causal-model theory which is based on the assumption that individuals' knowledge about categories is connected by two characteristics: category features (such as similarities) and causal mechanism. The theory is associated with Bayesian Networks which are networks that recognize that an effect variable is influenced by its immediate *parents* (a link that connects the direct causal relationship among the variables). Alone this network does not account for the details of the causal correlation that connect the network of intertwined variables. The Bayesian networks are incorporated into the causal-model theory, which makes specific assumptions about how people perceive causal relationship among binary numbers. The theory assumes that individuals see characteristics as being joined by probabilistically causal mechanisms. The theory also asserts that when people classify categories they make decisions by estimating the likelihood an exemplar was produced by the category's causal model (Rehder, 2003).

Rehder presented half of the adult subjects in the study with category features (e.g., two biological classes: non living natural things and artifacts) and their base rates (e.g., 75 % of Lake Victoria Shrimp have high body weight; in comparison 25% have normal body weight). The other half, the control group saw a chain condition with sentences with four mechanistic features (e.g. *high quantity of Ach neurotransmitter causes a long-lasting flight response, which causes accelerate sleep cycles…).* For a comparison Rehder also gave subjects the same information, but devoid of the percentages (features were weighted equally).

When the percentages were present researchers found a difference in the amount of influence put on the first mechanistic feature, but the features following it held less effect on ratings the characteristics, even though there was a causal effect added to the data. Furthermore, the causes that were directly connected had a stronger impact than the reasons that were not next to each other. Once the base rates were omitted, participants utilized the causal reasons while rating category memberships. A reason Rehder proposes for the differences between findings is that subjects were given information about the empirical base rates of features signifying that all characteristics were probably similar (i.e., 75%). They also found sizeable relationships among causal mechanisms that were directly linked pairs of the causal chain and a weak relationship among pairs that were not directly connected. These findings support the causal-model theory; category features and the causal mechanism are linked to individual's knowledge about categories.

The final part of Rehder's research followed the same procedure, but manipulated the data and presented a 100% occurrence for the first feature of the chain, and no base-rate was assigned to the rest of the 3 in the chain. He found, when making inferences about a non-existing defining feature (e.g., "???" was presented), subjects used the evidence of the existence of the causal mechanisms in addition to causal relationships. Therefore, whether the related characteristics supported or violated causal mechanisms the adult's category-membership was responsive based on the individual's causal knowledge.

Callanan and Oakes (1992) found that children ages 3-5 years of age are interested in the causes of things. Hence, because children are interested and attracted to the causes of things, this reveals the child's motivation to learn.

The research by Lopez et al. (1992) suggests that kindergartners and second graders use similarity, and they compute average coverage for category-based inferences. However, by second grade children recognize the importance of using data's diversity. Jacobs and Narloch (2001) found supporting evidence that children as young as first graders make use of the sample's size and variability to make inference and also base them on common characteristics. Both studies suggested that about second and third grade children become developmentally capable of evaluating sample size. Gutheil and Gelman.(1997) further supports their findings and add to it that third graders seem to use sample size or sample diversity, but inconsistently when making category-based inferences about information. They report that at this age children do not acknowledge the significance of using sample monotonicity and sample diversity for inductive reasoning, though. Gutheil and Gelman (1997) also found that children only employ it when it is the only resource available and sample size and sample diversity is combined together. In addition, category information may distract kids from realizing the importance of information like size and diversity of the sample. It is possible that children are aware of the importance of using data's sample monotonicity and diversity (Lopez et al., 1992), but inconsistently use it because they get distracted by the categorical information so they compute average coverage.

When looking at mechanisms Rehder (2003) found no correlation when variation in weights associated with features were present, but when the base rates were omitted, causal reasons were used to rate category memberships. It seems that the base rates distracted children from the mechanism and they inevitably take into account data size. Rehder concludes that because of causal mechanisms if the correlated features did or did not back up causal mechanisms, category-membership ratings were still responsive. They found evidence to back up the causal-model theory, so, an individual's knowledge about categories is related to the features of the category, in addition to the causal mechanism. Consequently, if a mechanism is added, adults employ it to make categorical inferences.

I have shown evidence that children use sample size and variables in different contexts for using inductive reasoning and also found support for the success of mechanisms being used for adults. These findings have important implications, because if adults respond to sample size, variability and also mechanisms, it may apply to children, too. If this is true and kids respond well to mechanisms, then it could be an important tool for schools to use and incorporate into their teaching techniques, for more efficient teaching. After reviewing the research about children's use of sample size when making category-based inferences, I believe that children that are older use sample size when making category-based inferences about a population more often than younger children. Furthermore, I will be testing if children use sample size more when making inferences about biological characteristics rather than psychological characteristics, particularly when it is used in the context of sample size. It is important to reexamine this topic because past research has only made a dent in the amount of information that can be obtained. Children are faced with biological and psychological characteristics all day, and it is important for us to add to this body of knowledge to have a better understanding of how children use these two unique categories to make inferences.

I also examined children's use of mechanisms. I think that if children deem a mechanism supplied by the experimenter as making sense, it would be paired with children being more likely to say the next thing in that category will be like the rest of the sample. I also believe that children use mechanisms more to explain their answers after they hear a mechanism. The present research examines if before children receive a mechanism if they would more likely say psychological mechanistic reasons rather than biological mechanistic reasons to support why they came up with their population estimate. I think that this will be true because I deem biological mechanisms as something that children understand less. Furthermore, I will examine the same thing for after the children heard a mechanism, and I believe that children would more likely again say a psychological mechanism over a biological mechanism for why they came up with their population estimate. Mechanisms are important to examine because past research has not looked at both sample size and mechanisms together when looking at children's inferences.

Method

Participants

In the current study the sample consists of first and third graders from three Long Island elementary schools and one after school program. Only those students whose parents gave permission to take part in the study participated. The sample consisted of 61 students; 26 male participants and 35 female participants. The sample consisted of 38 first grade participants and 23 third graders. The ages of the participants ranged from 6 to 9 years old. The students were from a low socioeconomic bracket. Furthermore, there was a small percent of bilingual (ESL) participants.

Materials

Participants received varying information about the same five novel animal categories (e.g., rascal bear*)* that were invented for the purpose of the experiment. All categories contained a psychological characteristic of the animal (e.g., these types of bears are very nervous) and a biological characteristic of the animal (e.g., these animals have grey bellies). Participants were told about a forest far away that contains these animals.

Additionally, demographic information was attained from the participants such as age, grade and sex. See appendix for more details about the categories, questions about the categories and demographic questions.

There were toy animals shown to the children that had the same the biological characteristic that they were being told about to help the children visually conceptualize what they were being told. There were poster-board squares that contained varying numbers of stars (1, 3, 10, 20 and 30) scattered evenly around the board. Other materials used to interact with the children were also matching poster-board pieces that had written numbers that were compatible with the five boards with the stars. There was also a toy car and a road that had the numbers 1 through 10 written on them, plus another road that contained the words does not make sense, kind of makes sense, mostly makes sense and completely makes sense.

Procedure

Child participants, teachers and principals were given a consent form to be filled out by a legal guardian. Subsequently, participants were interviewed individually. Verbal consent was obtained from the children. First children were tested on their knowledge of numbers by examining five different small poster-boards that contained varying numbers of stars on the board and matching poster-board pieces with written numbers that were compatible boards that contained the stars. The participants were told they were playing a matching game, and they were to place the appropriate number on top of the board that showed that number of stars.

Participants were informed that they were allowed to stop helping the researcher at any point in time by simply stating that they no longer wanted to be apart of the study. It was also made known that their school grade would not be affected in any way, and their teachers, parents and even the researcher would not know who said what after their answers were recorded it would be put in a pile of other participants without any names on it (explained confidentiality). It was also disclosed that I was only looking for their opinion in order to understand first and third graders better, therefore, there were no right or wrong answers. Subsequently, participants were informed that if they did not understand what they were being asked to tell the interviewer because in order to get their true opinion was important that they understood the questions.

Participants were then shown how to use the toy car and how to drive it to the appropriate number that they thought was the answer. They were then informed how to answer the questions that asked if something made sense. For the answer "does not" was told that *it means it makes no sense at all, "*kind of makes sense" means that *it makes little sense or close to no sense*, "mostly makes sense" means *makes a lot of sense, but not total sense*, and "completely makes sense" means *makes total sense, 100 percent sense.*

The participants were then informed about the make-believe trip that they will be taking to a forest far away. The verbal directions that were administered contained a scenario that they were taking a trip to Africa with family, to take a nature walk and learn about different animals in the deep forests (See appendices A and B for more information).

The participants were randomly assigned to three conditions varying the amount of animals (1, 3 or 30) that were encountered on the trip. For example, participants in the n = 1 condition will was informed that they spot one rascal bear in the forest. Participants in the n=30 condition were informed that they spotted thirty rascal bears in the forest.

The participants were then asked to make inferences about other animals they may come across (e.g., rascal bears) in that same category (e.g., Why do you think it would have a grey belly?). Each participant answered questions about four animals' characteristics, 2 biological and 2 psychological. The participants only answered questions about one category for each animal (biological or psychological). For example, half the participants were asked to make inferences about psychological characteristics of bears and cats, (e.g., If you encounter another rascal bear, do you think it will act nervous?) and biological characteristics of giraffes and zebras (e.g., If you were to continue on in the forest in Africa and came across a tyson giraffe, do you think that it will enjoy eating sweet berries?). The other half of the participants were asked to make inferences about the biological characteristics of bears and cats (e.g., If you encounter another rascal bear do you think it will have a grey belly?) and psychological characteristics of giraffes and zebras (e.g., If there were ten tyson giraffes in the African forest you are traveling through, how many do you think would have no spots?).

The participants subsequently were asked to give a percentage of animals (number out of 10) they would expect to have that characteristic. Participants were asked if they came across ten of the particular animal in the African forest, how many they thought will have the same characteristic (e.g., If there were ten bears in the African forest you are traveling through, how many do you think would be nervous?). Additionally, I asked the participants to explain why they answered the question the way they did, in order to examine how they surmised their category-based inference.

Subsequently, I gave the participants a mechanism for the characteristic and then probed to see if they deem it plausible. For example, participants were told the rascal bear(s) have grey fur on their belly to try and blend in with its environment, then they were followed up with questions such as if they thought it made sense where they responded on a four point scale: does not make much sense, kind of makes sense, mostly makes sense, or completely makes sense. Another question that was asked was if they saw another animal from that category, if they thought it would have that characteristic they saw the other animals had (e.g., If you came across another rascal bear in the forest that you are traveling through, do you think it will have a grey belly like the other 3 rascal bears you saw?). See Appendices A and B for examples of what children were asked.

The design was 3 x 2 x 2 (*sample size* x *biological or psychological* x *mechanism or no mechanism*). The independent variables that were tested are age of participant, sample size, biological verses psychological characteristic and mechanism provided. The dependant variable is the amount of the inference made and mechanistic answers from the participants.

Results

The second story (psychological) that the participants heard (e.g., *You are hiking through the forest in Africa, and you see 1 cat that is called roothy cat. It likes being petted. If you came across ten more roothy cats, how many do you think would like to be petted?* or *As you continue on your journey you spot 1 giraffes called tyson giraffes. They are sad. If you came across 10 more, how many do you think would be sad?*) had a significant difference based on story using a univariate ANOVA, F (1, 54) = 6.851, p = .011. The means were consistently lower for giraffes (for sample size 1, M = 7.2, sample size 3, M = 6.3 and sample size 30, M = 5.56) than compared to cats (for sample size 1, M = 7.9, sample size 3 M = 7.5 and sample size 30, M = 7.25). To amend this problem I only used the first story (biological). To balance the stories I also used the fourth story (psychological) which kept all four animals in the testing (giraffe/cat and bear/zebra). I did not include the second set, where the story effect was found (which was psychological). To make it even I excluded another set (biological), keeping all four animals in the study.

To test the hypothesis that older children use sample size when making category-based inferences about a population more often than younger children, I used a general linear model. I tested the affect that age and sample size has on the number amount that children answered out of 10 on the prevalence estimate. For children’s use of sample size there was no significant difference found in inferences based on the sample size seen, $F(2, 57) = .286$, $p = .752$ (see Figure 1 and 2). I did find an age difference, $F(1, 57) = 4.362$, $p = .041$ (see Figure 3 and 4). We found no interaction $F(1,57) = 35.110$, p = ‹ .001.

The coding for children’s responses was divided into six major categories. These categories consisted of 1) don’t know (e.g., “*I don’t know)*, 2) mechanistic based which had five subcategories, which consist of mechanistic based (e.g., *because the rascal bears sleep on the rocks so they can blend in),* psychological mechanism (e.g., *they are happy when they are by themselves)*, biological mechanism (e.g., *it might be a baby and babies have grey stomachs*) environmental mechanism (e.g., *they may all live close together)* and relationship mechanism(e.g., *they may be brothers*), 3) Category (e.g., *they are all rascal bears),* 4) sample size (e.g., *The other 30 rascal bears I saw had a grey belly)* 5) difference (e.g., *not all animals look the same)* and 6) no difference (e.g., *all rascal bears have grey bellies*).

When I examined first graders’ responses to why they answered the way they did I added up all the times the participants answered sample size (could have answered sample size up to 8 times). I was examining why children came up with their answers and if sample size was a major reason why. I found 16 out of 37 (43%) said sample size, compared to 13 out of 24 (54%) third graders. When I examined if first graders' use of sample size as an explanation was associated with the amount of the sample size they saw (1,3, or 30), I found no association , χ^2 (2, N= 37) = 1.054, *p* = .595. Third graders' use of sample size as a reason was also not associated with the sample size they saw χ^2 (2, N= 24) = 3.253, *p* = .197. No significant association was found for children in the first and third grade using sample size for explanations.

To examine if children would make use of the sample size more when making inferences about biological characteristics rather than psychological characteristics I used a repeated measures ANOVA and found Λ = .985, *F* (1, 58) = .885, *p* = .351. No significant difference was found when testing this (see figure 5). The means for psychological characteristics (*M* = 7.34) was slightly greater than for biological characteristics (*M* = 6.99). This difference between whether children would make use of the sample size more when making inferences about biological characteristics rather than psychological characteristics was not significant. Furthermore, when I examined if children make more inferences about biological characteristics than psychological characteristics when making inferences based on the sample size, I found *F* (2, 58) = .121, *p* = .886. I found no significant difference between biological and psychological characteristics.

I also examined the children's use of mechanisms. I tested our theory that if children deem a mechanism supplied by the experimenter as making sense, then children would be more likely to say the next thing in that category will be like the rest of the sample. I found *t* (31.439) = 3.439, *p* = .002 for the first story (biological- giraffe/ cat) that I examined, a significant association was found. For the second story (Psychological- cat/giraffe) I found t (58) = 1.148, p = .56, no significant association was found. In the third story (biological-zebra/bear) I found t (59) = 3.016, p = .004, a significant association was found. Lastly, for the fourth story (psychological- bear/zebra) I found t (15.336) = 1.904, p = .076 which is marginally significant.

To examine if children used a mechanism more to explain their answers after they received a mechanism from the experimenter, I collapsed across mechanistic answers given by children by adding up all the possible mechanistic answers that the participants supplied (biological mechanism, psychological mechanism, environmental mechanism, relationship mechanism and mechanism based on the one the experimenters supplied). I examined the use of mechanistic answers before the supplied mechanism and following it using a paired samples t test. I found a significant association; children supplied more mechanistic answers after they heard a mechanism (M = 2.02) than before a mechanism was supplied (M = 1.21), t (60) = -5.133, $p < .001$.

To further examine the difference in mechanistic answers before and after children are supplied with one, I broke the conditions down further and examined the difference between biological and psychological mechanisms. First, I examined before children received a mechanism if children would more likely say psychological mechanistic reasons rather than biological mechanistic reasons why they came up with their population estimate. I found a significant association when I tested the affect of the mechanism that was supplied on the likelihood of children then saying they responded the way they did because of a psychological mechanism, rather than a biological one using a paired samples t test and found t (60) = -2.124, p = .038. Children did use a psychological mechanistic answer (*M* = .74) more often than biological (.48) as predicted. I tested the same thing for after the children heard a mechanism, and I believed that children would more likely again say a psychological mechanism over a biological mechanism about why they came up with their population estimate. I found a marginally significant relationship when using a paired samples *t* test. I found t (60) = -1.751, $p = .085$. Children supplied more psychological mechanisms (*M* = 1.11) than biological mechanisms (*M* = .90) as expected.

Discussion

I found no significant difference developmentally in the ability for children to use sample size when I examined the different age groups. There was no evidence to support our hypothesis that younger children take into consideration the size of the sample less often then the older children when making category-based inferences. This is not what I expected and it does not support the past research.

Past research by Jacobs and Narloch (2001) found no differences developmentally when children use sample size and variability to make inferences about a population. They take into account the degree of the sample size and variability (high or low amount) of the traits that are shared. Furthermore, Lopez et al. (1992) found that by second grade children recognize how imperative it is to use sample size and diversity, suggesting that the ability appears sometime between kindergarten and second grade. They further add that children do not appreciate the importance of sample size and diversity for making category-based inferences. Gutheil and Gelman (1997) support Lopez et al (1992) findings and that children can make inferences using sample size and diversity, but they do not easily recognize the importance of utilizing it for inductive reasoning.

Furthermore, when children make inferences they may be preoccupied by the category information, distracting them from noticing important information such as the sample size and sample diversity.

The design of the study may have had an effect on our lack of findings. It is possible if I performed a within subjects design where all participants in the study were to see a 1, 3, and 30 condition in order to make a comparison within each participant, I may have been able to have a better idea about how children use sample size to make inferences about the population the characteristic belongs too. The fact that children only saw one condition (1, 3 or 30) may have an effect because children did not have a comparison to make before they made an inference. This may not be enough information for a young child to take sample size or whether or not the characteristic was biological or psychological into consideration. Children may need more information than was supplied to use the reasoning that I examined.

Furthermore, I saw that some children do use sample size when supplying reasons why they answered the way they did, but it was not enough of a difference to be significant. This indicates that some children do recognize the implications of using sample size when making category based inferences about a population but not consistently.

I did find a noticeable difference between ages in the answers children supplied when estimating about the population. I found 9 year olds make considerably lower estimates than 6-8 year olds. It seems that the ability to reason and not just say any number when making population estimates appears somewhere between 8 and 9 years old. Further research should examine if this continues on an upward trend as the child enters 10 years of age and so on. When this development tapers off should also be examined in the future.

I found these results when I excluded the second (psychological) and third (biological) set because I found a story effect in the second set. To balance out the biological and psychological sets I also excluded the third set, which contained two different animals than the second set did, ensuring that all animals would still be included in the study. One of the stories mentioned that the animal was elderly; this could have caused the responses to the stories to be different, which could have affected the data. The story effect may have been caused by the difference between ages introduced by the giraffe story I included, making the conditions slightly different from each other. I only found an interaction between the giraffe and the cat in one set though (psychological). The participants gave lower estimates for the giraffe than for the cat.

The finding that 9 year olds make considerably lower estimates than 6-8 year olds is an important finding for teachers and developmental professionals because it helps us understand young children’s capabilities better. It can help teachers and other school administrators design class assignments more age appropriately. This finding also can help developmental specialists because it helps them have a better understanding of the developmental process.

One question that I asked to find out the percentage that the child thought would have the same characteristic was: After you see 1 cat called a *rascal bear* that is nervous, if you were to see ten more bears in the African forest you are traveling through, how many do you think would be nervous? This is a lot of information at once, which may have overloaded the children’s ability to understand what were asking. The interviewer had to re-ask the questions many times before the child said they fully understood the questions being asked. It is possible that the participants only finally said they understood to satisfy the interviewer. This could have affected the results enough that the findings were skewed, which may not have examined what I intended to look at.

Lopez, et al. (1992) demonstrated that when children make inferences about a population that has similar biological characteristics to a sample, they make greater use of sample size of biological characteristics than they do with behavioral characteristics. Therefore, children may perceive biological characteristics as less variable than behavioral characteristics. However, our findings suggest that there is no existing relationship. The means showed that children (both first and third graders) made slightly higher estimates for psychological characteristics (e.g., children said they would see more rascal bears that are nervous rather than rascal bears that have grey bellies), which would contradict the past research that I examined, but our findings were not significant.

Many factors can be taken into account to explain our difference in findings from previous research. I used a small sample size in our study which may have affected our results. In addition, I tested a different population that may not be representative of the entire population. Participants were from low socioeconomic communities, and many were ESL, therefore, our results can only be generalized to that population. It is possible that some children may not have understood what I was asking them because they did not want to seem unintelligent or dissimilar to other children. This may have contributed to their inability to express their lack of understanding of the material being presented. In the future if I test this population again, it would be in the researches best interest to bring an interpreter to the interviews.

I further found that if children deemed our mechanism as plausible, then they were more likely to make category-based inferences based on the size of the sample. I did not find a relationship in the story that contained the story effect though. This inconsistency may be caused by the fact that the story had a story effect. Therefore, I do not hold much weight to this inconsistent finding and infer more from the other three stories that consistently show an existing relationship. This evidence suggests that if children see a mechanism as being reasonable, then they may take it into account when making inferences.

I also examined if children use mechanisms more to explain their answers after they are supplied with a mechanism and found a relationship. The data supplied evidence that children are less likely to supply a mechanism before they hear one. After hearing a mechanism, children are more likely to supply one as their own answer. This suggests that children use mechanisms but do not tend to answer using a mechanism unless they hear about one first. Young children do use mechanisms as answers which supports the Callanan and Oakes (1992) findings that young children are interested in the causes of things. This is manifested in their use of mechanisms in the present study. However, children supplied more mechanistic answers after they heard a mechanism than before a mechanism was supplied. This demonstrated children's interest in the cause of things adding to the body of research.

This finding is important because it gives us a better understanding of the cognitive development of children. The fact that children use mechanisms more after they hear one helps us understand how children learn. If children examine the world looking at the causes of things, it gives them a better understanding of their environment.

Further research should examine if children were to receive mechanisms from parents whether or not it would affect the discipline positively. If the mechanism is used more after they receive one, then maybe it is possible that mechanisms have a great effect on how the child reacts to disciplining.

I further examined the difference between biological and psychological mechanisms. Firstly, I examined children prior to receiving a mechanism. I found that children did use a psychological mechanistic answer more often than a biological one. The same thing was examined for after the children heard a mechanism. I found a marginally significant relationship for children supplying more psychological mechanisms than biological mechanisms. This suggests that children may understand and relate to psychological characteristics better. The fact that children supplied psychological mechanisms more often, before and after they received a mechanism, suggests that their understanding of what they inferred about the sample size for biological category-based inferences is minimal especially when compared to their understanding of category-based inferences about psychological characteristics. This is interesting because children see biological characteristics and talk about them many times a day, but this may mean that they may not understand completely the bases of what they are seeing. Children may understand psychological characteristics more, which is something that is not physically visible, but an easier thing for them to understand.

Children are around categories all day long. They make inferences based on these categories. I have examined how children make category based inferences, and some of our findings supported our hypotheses while others did not. Nonetheless, I have added to the body of knowledge of how developmentally children may not see the importance of taking into consideration the size of the sample when making inferences about categories. I also found no difference between whether children make more population estimates based on biological characteristics rather than psychological characteristics; in fact I found no relationship at all. This has left us with more questions than answers. I plan to further investigate these hypothesis using different methods to see if our findings are replicated, which would give professionals more confidence in the body or research about children's use of sample size when making category-based inferences (also biological and psychological).

I investigated mechanisms and found that children use them when giving reasons for explanations, especially after they hear one. It should be further examined to see if a mechanism added to the story would change the population estimate given for the inference about the category. Future research should focus on adding a mechanism to determine if children would change their reasoning when making category-based inferences. It would be interesting to see what adults and children would prefer when they are presented with sample size and variability in addition to a mechanism.

These findings are important in helping people learn more about the cognitive development and abilities of children. Children are complex and unique and we can learn a lot about each developmental age by examining children and their behaviors. The present research has added to our knowledge about the complex mind of the child.

Set 1 and 4 Inferences based on the Size of Sample	
Set 1 for a sample size of 1	6.8
Set 1 for a sample size of 3	7.2
Set 1 for a sample size of 30	7.0
Set 1 for a sample size of 1	7.2
Set 1 for a sample size of 3	7.4
Set 1 for a sample size of 30	7.1

Inference based on the Size of the Sample and if it is a Psychological or Biological Characteristic

Variable	**Mean**
Biological Inference Amount for a sample size of 1	7.17
Biological Inference Amount for a sample size of 3	6.88
Biological Inference Amount for a sample size of 30	6.98
Total Biological Inference for a all sample sizes	6.99
Psychological Inference Amount for a sample size of 1	7.48
Psychological Inference Amount for a sample size of 2	7.30
Psychological Inference Amount for a sample size of 3	7.23
Total psychological Inference for a all sample sizes	7.34

References

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: causal thinking in everyday activity. *Cognitive Development, 7,* 213-233.

Gutheil, G., & Gelman, S.A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64,* 159-174.

Jacobs, J.E., & Narloch, R.H. (2001). Children's use of sample size and variability to make social inferences. *Applied Developmental Psychology, 22,* 311-331.

Kuhn, D. (1995). Scientific thinking and knowledge acquisition. *Monographs of the Society for Research in Child Development, 60,* 12-157.

Lopez, A., Gelman, S.A., Gutheil, G., & Smith, E.E. (1992). The development of category-based induction. *Child Development, 63,* 1070-1090.

Osherson, D.N., Smith, E.E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based inductions. *Psychological Review, 97,* 185-200.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1141-1159.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102-119.

Figure Captions

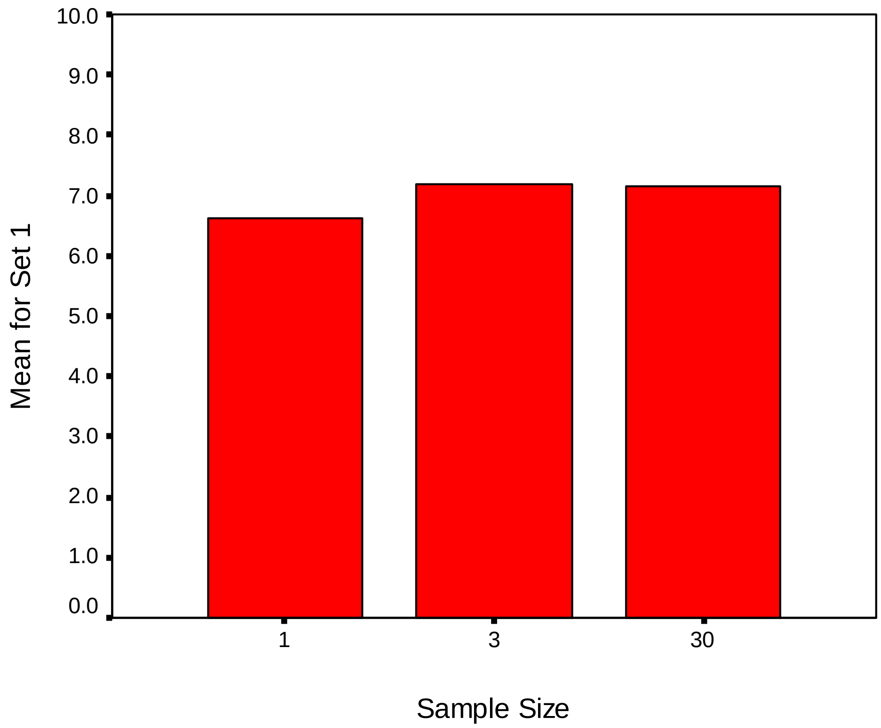
Figure 1. Population estimates based on sample size mean and set (story) 1.

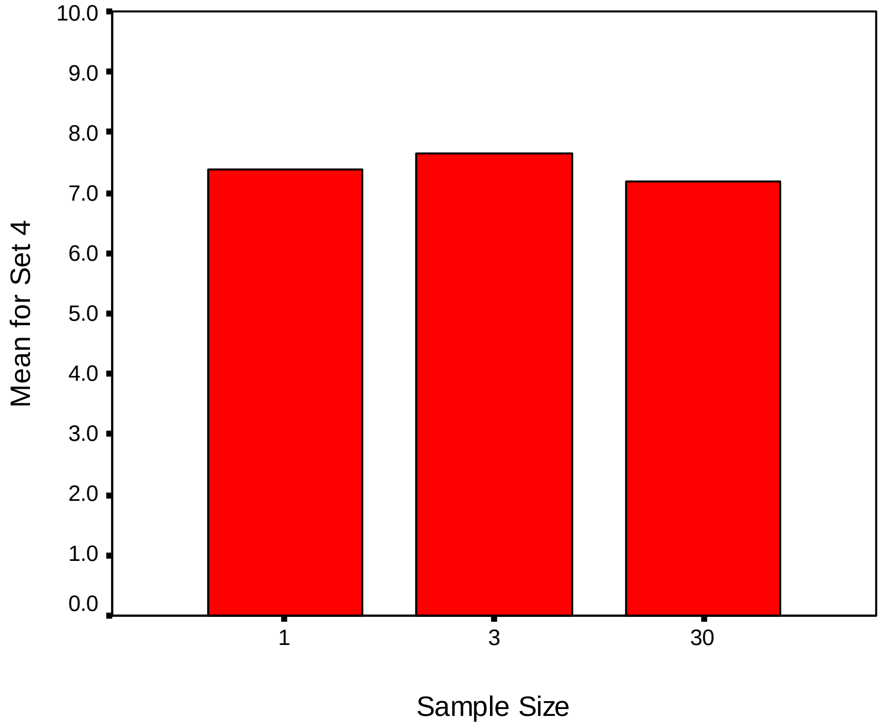
Figure 2. Population estimates based on sample size mean and set (story) 4

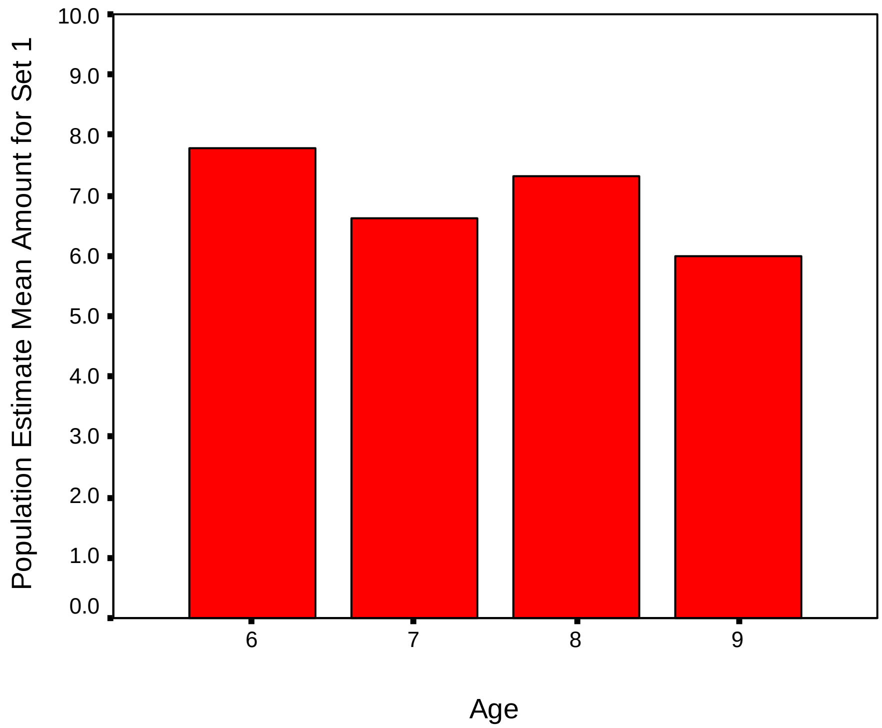
Figure 3. Population estimates for each age in set (story) 1.

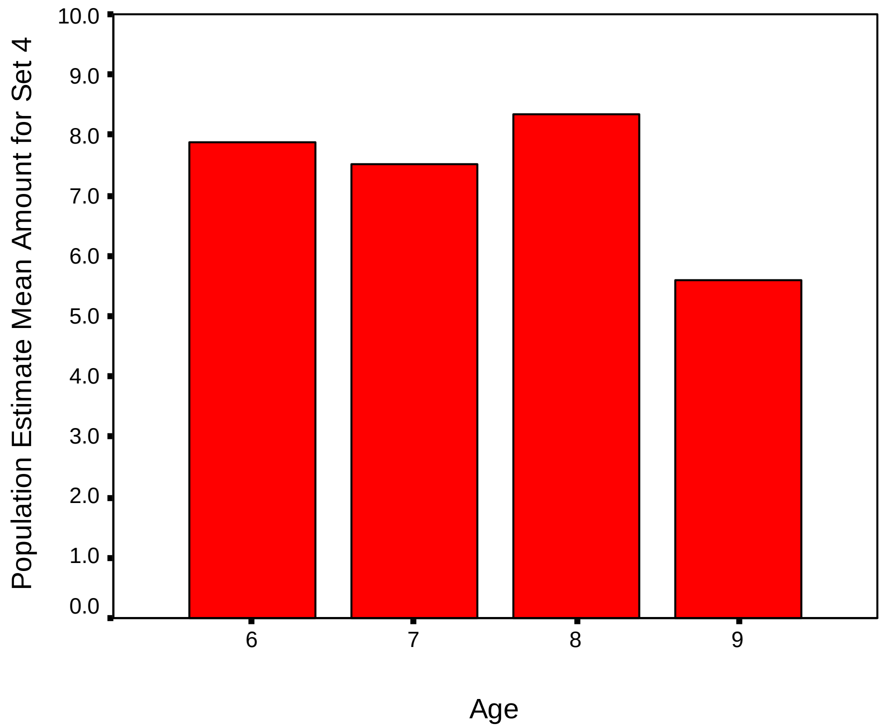
Figure 4. Population estimates for each age in set (story) 4.

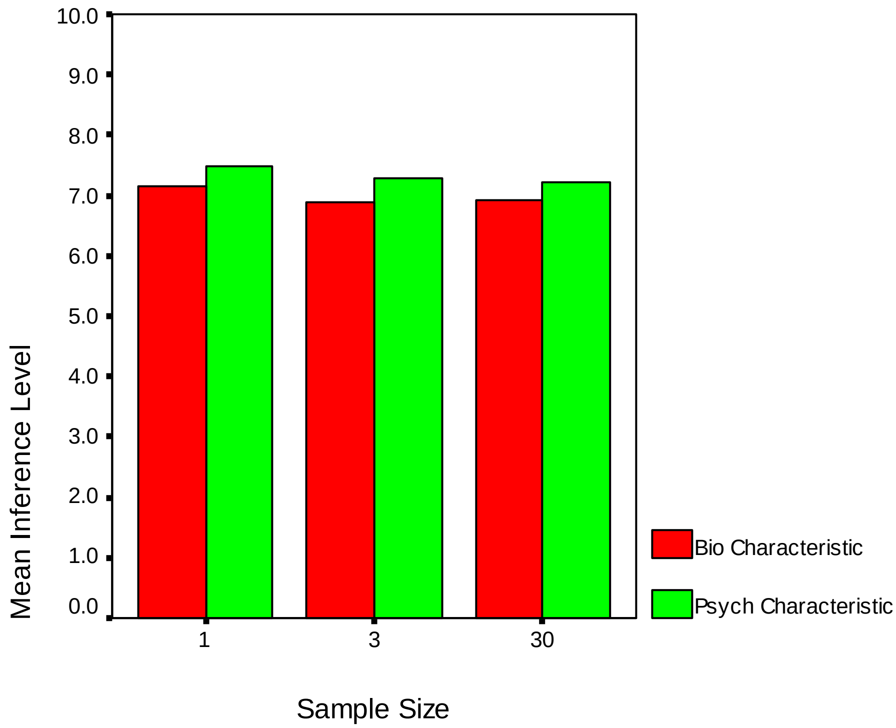
Figure 5. Population estimates of psychological characteristic and biological characteristics based on sample size.


10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Mean for Set 1
1
3
30
Sample Size


10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Mean for Set 4
1
3
30
Sample Size


Population Estimate Mean Amount for Set 1
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
6
7
8
9
Age


Population Estimate Mean Amount for Set 4
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
6
7
8
9
Age


10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Mean Inference Level
1
3
30
Sample Size
Bio Characteristic
Psych Characteristic

Appendix B

Age_________

Grade_______

Sex_________

Pretest:

How many wrong?

Which one(s)?

Posttest:

Forest in a place far away:

You and your family take a trip to Africa. You decide that you want to go on a nature walk and find as many animals as possible. You bring a journal and take notes on all the animals you see, so when you get home you can describe them to your teachers, friends and other family members.

- ✓ Biological:

You spot 1 elderly giraffe called a *tyson giraffe*. It is yellow with no spots.

- After you see 1 elderly *tyson giraffe* that is yellow with no spots, if you were to see ten more elderly *tyson giraffes* in the African forest you are traveling through, how many do you think would have no spots?

Why?

Some scientists think that the elderly *tyson giraffe* is yellow with no spots because it lives in the sun all day and the sun fades its spots.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that the elderly *tyson giraffe* has no spots because it is in the sun all day and the sun fades the spots? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense.
- After you see the1 elderly *tyson giraffe* that is yellow with no spots, if you were to continue on in the forest in Africa and came across an elderly *tyson giraffe*, do you think it would have no spots? Why

✓ Psychological:

You are hiking through the forest in Africa and you see 1 cat that is called *roothy cat*. It likes being petted.

- After you see 1 *roothy cat* that likes being petted, if you were to see ten more *roothy cats* in the African forest you are traveling through, how many do you think would like to be petted?

Why?

Some scientists think that the *roothy cat* likes being petted because when it was a kitten it was always petted by its parents.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that the *roothy cat* likes being petted because its parents always petted it when it was a kitten? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense
- After you see the1 *roothy cat* that likes being petted, if you kept walking in the forest in Africa and came across another *roothy cat*, do you think it would like to be petted?

Why?

✓ Biological:

As you go a little farther into your walk you come across 1 *koda zebra.* It has long ears.

- After you see 1 *koda zebra* that has long ears, if you were to see ten more *koda zebras* in the African forest you are traveling in how many do you think would have long ears?
- Why?

Some scientists think that the *koda zebra* has long ears so it could hear other animals in the forest when they come near.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that *koda zebras* have long ears to help it hear other animals in the forest when they come near? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense.
- After you see the1 *koda zebra* that has long ears, if you came across another *koda zebra* do you think it would have long ears? Why

✓ Psychological:

In another part of the forest you spot 1 *rascal bear*. It is nervous.

- After you see 1 cat called a *rascal bear* that is nervous, if you were to see ten more bears in the African forest you are traveling through how many do you think would be nervous?

Why?

Some scientists think that the *rascal bear* is nervous because it is always worried about being seen by predators when it is not ready.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that the *rascal bear* is nervous because it is always worried about being seen by predators when it is not ready? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense.

- After you see the1 *rascal bear* that is nervous, if you came across another *rascal bear* in the forest, do you think it would be nervous?

  Why?

Appendix B

Age_________

Grade_______

Sex_________

Pretest:

How many wrong?

Which one(s)?

Posttest:

Forest in a place far away:

You and your family take a trip to Africa. You decide that you want to go on a nature walk and find as many animals as possible. You bring a journal and take notes on all the animals you see, so when you get home you can describe them to your teachers, friends and other family members.

- ✓ Biological:

You are hiking through the forest in Africa and there are 30 cats that are called *roothy cats*. They have two long pointy teeth.

- After you see 30 cats called roothy cats that have two long pointy teeth, if you were to see ten more *roothy cats* in the African forest you are traveling through, how many do you think would have two sharp teeth?

Why?

Some scientists think that *roothy cats* might have two long pointy teeth to help them catch prey (food).

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that *roothy cats* have two long pointy teeth to help them catch prey (food)? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense.
- After you see 30 *roothy cats* that have two long pointy teeth, if you kept walking in the forest in Africa and came across another *roothy cat*, do you think it would have two sharp teeth?

Why?

✓ Psychological:

As you continue on your journey you spot 30 giraffes called *tyson giraffes*. They are sad.

- After you see 30 elderly *tyson giraffes* that are sad, if you were to see ten more elderly *tyson giraffes* in the African forest you are traveling through, how many do you think would be sad?

Why?

Some scientists think that *tyson giraffes* are sad because they can not run as fast as the other animals.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that *tyson giraffes* are sad because they can not run as fast as the other animals? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense.
- After you see 30 elderly *tyson giraffes* that are sad, if you were to continue on in the forest in Africa and came across a *tyson giraffe* do you think that it will be sad? Yes/No

Why?

✓ Biological:

In another part of the forest you spot 30 *rascal bear*s. They have grey stomachs.

- After you see 30 *rascal bears* that have grey stomachs, if you were to see ten more *rascal bear* in the forest, how many do you think would have grey stomachs?

Why?

Some scientists think that *rascal bears* have grey stomachs because they sleep on rocks while laying on their back and the grey bellies protects them from being seen by other animals.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that *rascal bears* have grey stomachs to try and blend in with its environment? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense.

- After you see 30 *rascal bears* that have grey stomachs, if you came across another *rascal bear* in the forest, do you think it would have a grey stomach?

Why?

✓ Psychological:

As you go a little farther into your walk you come across 30 *koda zebras.* They are independent and like to be by themselves.

- After you see 30 *koda zebras* that are independent and like to be by themselves., if you were to see ten more *koda zebras* in the African forest you are traveling in how many do you think would be independent and like to be by themselves?

Why?

Some scientists think that *koda zebras* are independent and like to be by themselves because they are taught by their parents that they have to fend for themselves.

- Sometimes explanations make a lot of sense. Sometimes they do not make very much sense. Do you think it makes sense that *koda zebras* are independent and like to be by themselves because they are taught by their parents that they have to fend for themselves? Does not make sense, kind of makes sense, mostly makes sense or completely makes sense

- After you see 30 *koda zebras* that are independent and like to be by themselves, if you came across another *koda zebra* do you think it would be independent and like to be by itself?

Why?