When Sample Size Matters: The Influence of Sample Size and Category Variability on Children's and Adults Inductive Reasoning

Chris A. Lawson (clawson@andrew.cmu.edu)
Anna F. Fisher (fisher49@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15213 US

Abstract

A considerable amount of work has focused on the processes that underlie children's inductive reasoning. For instance, numerous studies explored the role of linguistic labels, perceptual similarity, and children's beliefs in generalization of properties to novel cases. The present studies investigated an aspect of induction that has received considerably less attention in prior developmental research, namely – the effect of the statistical properties of evidence on inductive reasoning. Studies presented below were motivated by the hypothesis that induction involves an evaluation of the statistical properties available in the evidence. From this perspective, sample size, or the amount of available evidence, should influence inductive reasoning. Sample size effects were investigated in three experiments with 90 5-year-olds and 90 adults. Results indicated that children made higher rate of projections for larger than smaller samples, particularly when samples were represented by a set of diverse exemplars.

Keywords: induction, development, sample size, reasoning

Introduction

Inductive reasoning is crucial for learning new information and making sense of novel experiences. From an early age children are able to generate inductive inferences in order to make predictions about novel exemplars. For example, given a sample of evidence about an animal with a specific property (e.g., "birds have omentum inside"), children as young as 2-years of age will predict that a new animal from the same category will also have the property (Gelman & Coley, 1990). Much of the developmental work has focused on how children's knowledge of object properties and categories promote induction (see Gelman, 2003 for a review). Children's attention to the statistical distribution of the properties in a sample has received considerably less attention in the literature. Studies that investigated this issue concluded that children are likely to generalize object properties that are the best statistical predictors of category membership regardless of the type of property (i.e., internal versus external properties and observable versus unobservable properties) (Lawson & Fisher, 2008; Rakison & Hahn, 2004). The present studies explored the influence of a particular statistical feature of evidence: specifically, we examined the influence of sample size on children's inductive inferences.

A natural hypothesis is that larger samples are a better source from which to make predictions than smaller samples. However, adults often do not consider this basic distinction between small and large samples (Kahneman & Tversky, 1972; Nisbett, Krantz, Jepson, & Kunda, 1983). At the same time, some studies reveal that adults are insensitive to differences in sample size only when evidence is presented in the form of proportions. In particular, Sedlemeier & Gigerenzer (2000) have shown that adults prefer large to small samples for making predictions when evidence is presented in the form of frequencies rather than proportions (see also Gigerenzer, 1994).

Few studies have explored the effect of sample size on children's inductive generalization, and all of these studies concluded that children are insensitive to sample size information. In one study children were equally willing to generalize properties attributed to samples of 1, 3, or 30 individuals, but children's responses were influenced heavily by property type (Jacobs & Narloch, 2001). For instance, 6-year-old children were less likely to generalize highly variable properties (e.g., size), than low variability properties (e.g., number of eyes), regardless of whether a sample was small (i.e., one object) or large (i.e., 30 objects). Studies using a category-based induction format also suggested that children's predictions are insensitive to sample size (Gutheil & Gelman, 1997; Lopez, Gelman, Gutheil, & Smith, 1992). In these studies 5-year-old children generalized properties associated with a single category member (e.g., a single butterfly with blue eyes) and those attributed to a group of several category members (e.g., five butterflies with gray eyes) to a novel instance from the same category (a new butterfly) at equal rates.

These findings are consistent with the argument that children might ignore statistical properties in evidence and instead rely on their conceptual knowledge or beliefs when performing inductive generalizations (Gutheil & Gelman, 1997; Rhodes, Brickman & Gelman, 2008). Specifically it has been argued that children ignore information about sample size and base their inferences on expectations about the variability of properties across category members (Jacobs & Narloch, 2001; Kahnman & Tversky, 1983 and Nisbett et al. 1983 offer a similar interpretation of adult projections).

However, there is mounting evidence that children are sensitive to statistical properties of evidence when generalizing object properties (Lawson & Fisher, 2008; Rakison & Hahn, 2004; Smith, 1989). Statistical learning supports the detection of structure via attention to redundancies in property-category and property-property correlations (Rogers & McClelland, 2004; Smith, 1989). An implication of this view is that sample size should have a strong influence on induction – larger samples should provide a broader set of evidence from which to detect the meaningful correlations in the data. Despite unsuccessful attempts to detect effects of sample on children's inductions (Gutheil & Gelman, 1997; Jacobs & Narloch, 2001; Lopez, Gelman, Gutheil, & Smith, 1992) it is possible that children are sensitive to sample size differences in the course of induction. We investigated three factors that may influence children's sensitivity to sample size information: (1) variability of the evidence exemplars, (2) variability of the conclusion exemplars, and (3) probability of occurrence of to-be-generalized properties.

Variability of Evidence Exemplars

One of the hypotheses motivating this research was that effects of sample size might become more pronounced as exemplars in the sample become more variable. There is evidence that variability in the sample influences induction in young children. For example, children appear unable to determine the breadth of conclusions supported by diverse samples of evidence (Rhodes, Brickman & Gelman, 2008) but able to do so for homogenous samples (Gutheil & Gelman, 1997). This difference in children's responses is often attributed to knowledge about the categories represented in an inductive problem; children are able to construct the appropriate conclusion category implied by homogenous but not diverse samples of evidence. However, it is possible these effects stem from sample size effects: tests of homogenous and diverse samples in previous research often involved small sets of evidence. Providing larger samples of evidence for diverse sets of exemplars might lead children to recognize the broad scope of the properties associated with such exemplars. The experiments reported here tested the prediction that as sample variability increases more evidence should be required to determine the scope of property projection.

Variability of Conclusions Exemplars

Unlike previous research, the present studies explored projections to a range of targets from the non—sampled category: how does evidence about a mammal influence projections to non-mammals? One possibility is that large samples of evidence about diverse exemplars might lead to high rates of projections about a wider range of exemplars, compared to large samples of homogenous exemplars. People might conclude that properties that are true of a wide range of exemplars generalize broadly to other kinds. Alternatively, given large samples of evidence people might expect that the evidence apply *only* to the range of sampled items, and not other instances. From this perspective increases in sample size should lead to decreases in non-sampled items.

Property Probability

The few studies that have explored the effects of sample size on children's induction measured projections about properties attributed to 100% of the evidence exemplars. However, few object properties are invariant in children's experience (for instance, most birds fly but some don't, most cookies are tasty but some aren't, etc.). Arguably, sample size information is most useful for generating predictions about properties that are variable rather than invariant. Therefore, in the present studies novel properties were attributed to most (75%) but not all of the exemplars in the evidence.

Experiment 1

Experiment 1 examined the influence of sample size for a highly variable range of evidence exemplars. Participants learned about a novel property ("omat inside") that occurred within a set of various mammals (e.g., rabbit, mole, elephant, mouse, etc.) and then asked to make predictions about novel instances from the *same* category (e.g., cows), *close* targets comprised of a collection of vertebrate animals from a different superordinate-level (e.g., birds), and *far* targets including a set of invertebrates (e.g., insects). Half of the participants learned about a small sample of animals (N = 4) while the other half learned about a large sample (N = 16). We predicted low rates of projections from small samples of evidence and high rates of projections from large samples of evidence. We also hypothesized that any developmental differences would reflect a greater effect of sample size for younger participants. While children may not generalize properties associated with small sets of diverse exemplars, we expected that increasing the amount of information associated with the sample would increase rates of projections.

Method

Participants. 30 Adults (undergraduates; 14 females, 16 males), and 30 children (M = 5;0; age range 4;5 – 5;6 years; 14 females, 16 males) participated. Adults received course credit for their participation, children were recruited from local preschools and received a small prize their participation. All participants were recruited from a moderately large, predominately white, eastern US city. No participants were involved in other studies reported here.

Design and materials. Sample size condition was manipulated between subjects. In the Small sample condition participants were presented evidence about four mammals, while participants in the Large sample condition were presented evidence about sixteen mammals. A set of sixteen mammals was created such that all of the items were presented in the Large sample condition and a randomly chosen subset of these items was presented in the Small sample condition. Each participant in the Small sample condition learned about 4 exemplars randomly selected from the set of 16.

Both conditions involved property attribution at a rate of 3:1, such that 75% of in the instances involved the presence of the property and 25% of the instances involved the absence of the property. The presence/absence of properties was pseudo-randomized, with the one constraint that in the Large sample condition the absence of a property was not attributed to consecutive instances. The presence or absence of the property was indicated in two ways. First, targets were verbally described as either having or lacking the property. Participants were shown a picture of the target and told, "This animal has/does not have omat inside of it". Second, targets appeared with an arrow pointing toward the animal referring to a circle containing a small red blotch when the property was present and an empty circle when the property was absent. As is typical in induction studies, properties were novel internal features that could be interpreted as biological (e.g., "omat inside").

After the introduction of evidence participants were asked to make predictions about a range of conclusion exemplars. There were four instances in each of the three conclusion exemplar types: Same superordinate-level category conclusion exemplars (e.g., other mammals), Close conclusion exemplars (e.g., birds, lizards), and Far conclusion exemplars (e.g., insects, crustaceans). Participants responded to twelve items in total. For each conclusion exemplar, participants were shown a picture of the animal and asked, "Do you think this animal has omat inside of it?" After responding "yes" or "no" the next item was then introduced. All conclusion exemplars were presented in random order.

Procedure. Children were interviewed in a quiet location at their preschool. The task was administered on a laptop and designed using Psyscope. Adult participants completed the task on individual computers in a computer room. Adults were told the task was part of a larger project aimed at understanding how children use evidence to make predictions and that their responses would serve as a measure of adult reasoning. They were then given the same instructions as children. Children were told that the "game" involved two parts – one in which they would learn something about an animal and a second in which they would be asked some questions. The game involved a scene with a tree which participants were told needed to be clicked in order to see the animals. When participants clicked on the tree a single exemplar emerged and they were told about the animal (or, in the projection phase, asked to predict the property of the animal). The method of selecting and discovering the animals was designed to sustain children's interest in the task and to assure the items were presented individually. The task lasted approximately 10 minutes.

Results and discussion

The first set of analyses was conducted on rates of projections to the three sets of conclusion exemplars. Responses to the four items in each set were averaged to get a composite score for Same, Close, and Far targets. The proportion of projections for each of the three targets in both conditions is presented in Figures 1 and 2. Scores that were different from chance are indicated with an asterisk

The first analyses involved a mixed analysis of variance (ANOVA) with Age (Adults, Children) and Sample size (Small, Large) as between subjects variables, and Target (Same, Close, Far) a within subjects variable. The analysis revealed a main effect of Sample size, $F(1, 56) = 18.01$, $p < .0001$, with post hoc analyses (Tukey's) confirming higher rates of projections in the Large rather than the Small sample condition. The analysis also revealed effects of Age, $F(1, 56) = 10.50$, $p < .01$ and Target, $F(2, 112) = 21.94$, $p < .0001$. Both effects were mediated by an interaction, $F(2, 112) = 14.23$, $p < .0001$. The interaction revealed that children made a higher rate of projections than adults for Close, $F(1, 58) = 9.07$, and Far targets, $F(1, 58) = 20.26$, both p's $< .01$.

The hypothesis of higher rates of projection in the Large than the Small condition was explored further with separate target comparisons for each age group. These tests revealed that children made a higher rate of projections in the Large than Small sample condition for all three targets: Same, $F(1, 28) = 14.33$, $p < .001$, Close, $F(1, 28) = 5.61$, $p < .05$, and Far, $F(1, 28) = 5.69$, $p < .05$. Adults showed a higher rate of projections to the close targets in the Large than Small condition, $F(1, 28) = 4.79$, $p < .05$.

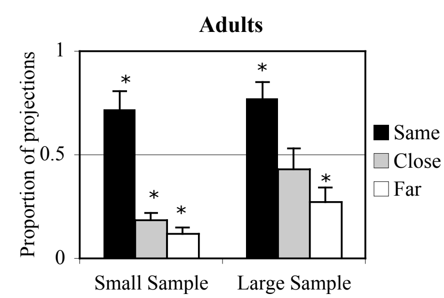

Figure 1: Adults projections across target types in both sample conditions in Experiment 1. Error bars represent the standard error of the mean.

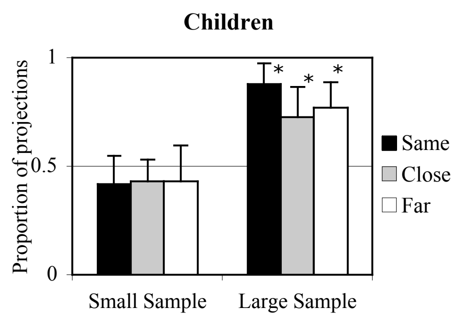

Figure 2: Children's projections across target types in both sample conditions in Experiment 1. Error bars represent the standard error of the mean.

The results were consistent with the primary hypotheses. Large samples yielded higher rates of projections than small samples. This pattern was particularly true for young children who responded at higher rates for each of the targets in the Large than the Small sample condition from which projections were no different from chance.

Experiment 2

The goal of Experiment 2 was to examine the effects of sample size for a less diverse sample of evidence. Experiment 2 used the same design as Experiment 1, but rather than learning about the properties of mammals, participants learned about a collection of animals from the same basic-level category (i.e., rabbits). The conclusion categories included a novel set of animals from the *same* basic-level category (e.g., other rabbits), *close* matches comprised of a set of mammals (e.g., cows), and *far* matches including non-mammal vertebrates (e.g., birds). Thus, the *far* and *close* exemplars were the *close* and *same* categories, respectively, from Experiment 1. We expected the results would replicate the findings from Experiment 1: large samples of evidence would lead to higher rates of projection and broader generalization. As in Experiment 1, we expected sample size to have the strongest effect on young children with large samples leading to a higher rate of projections to sampled and non-sampled exemplars. Though, we speculated that we may fail to observe sample size effects for conclusion exemplars most similar to the evidence exemplars (e.g., Same matches) because small samples of evidence may be sufficient to warrant the prediction that the property generalizes to instances from the same basic-level category.

Method

Participants. 30 Adults (undergraduates; 14 females, 16 males) and 30 Children (*M* = 5;2, age range, 4;7 years – 5;8 years; 13 females, 17 males) participated. Participants were recruited from the same population using the same methods as in Experiment 1. No participants were involved in other studies reported here.

Design, materials, and procedure. All facets of the design were the same as in Experiment 1 with the exception of the exemplars used. The evidence exemplars were comprised of a set of rabbits. Also, the conclusion exemplars were the same as in Experiment 1, except that the insect items were eliminated, and a set of basic-level matches (i.e., other rabbits) was included. We chose this strategy in order to keep the amount of conclusion items constant across the experiments. Thus the conclusion exemplars included rabbits (Same basic-level category matches), other mammals (Close matches), and other living kinds (Far matches). In all other respects the design, materials and procedures were identical to Experiment 1.

Results and discussion

Mean projection scores for the three types of conclusion exemplars in both sample size conditions are presented in Figures 3 and 4. Scores that were different from chance are indicated with an asterisk.

The primary analysis involved a mixed ANOVA with Age and Sample size between subjects variable and Target a within subjects variable. The analysis revealed the same effects as in Experiment 1: Age, $F(1, 56) = 30.32$, $p < .001$, with children making higher rates of projections than adults. There were also effects of Target and Age, which were mediated by an interaction, $F(1, 112) = 13.34$, $p < .0001$; simple effects showed that children made a higher rate of projections than adults to Close, $F(1, 58) = 22.59$, $p < .0001$, and Far targets, $F(1, 58) = 35.27$, $p < .0001$.

The effect of condition was mediated by age, $F(1, 56) = 7.91$, $p < .01$; higher rates of projection in the Large than the Small sample condition was evident in children's, $F(1, 28) = 13.39$, $p < .01$, but not adult responses, $F(1, 28) > 1$, *ns*. Children's higher rates of projection in the Large sample condition were evident for Close, $F(1, 28) = 12.52$, $p < .001$, and Far targets $F(1, 28) = 8.57$, $p < .01$, but not Same targets, $F(1, 28) > 2$, *ns*.

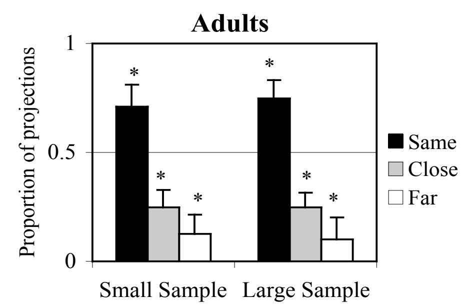

Figure 3: Adults projections across target types in both sample conditions in Experiment 2. Error bars represent the standard error of the mean.

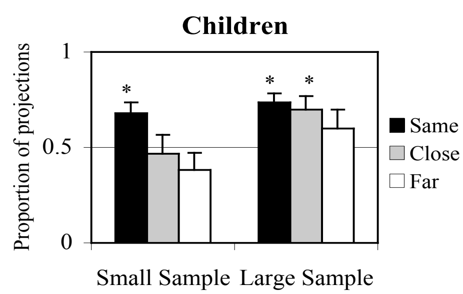

Figure 4: Children's projections across target types in both sample conditions in Experiment 2. Error bars represent the standard error of the mean.

The results were consistent with the main finding from Experiment 1: children, but not adults, made higher rates of projections for large than small samples of evidence. The one exception was responses to the same targets, for which children made an equal rate of projections across both samples. Thus, the effect of sample size was pronounced for non-sampled items. It could be argued that these results show that children simply overgeneralize large samples of evidence. Experiment 3 was designed to examine whether the effect of sample size depends on variability in the evidence, or if larger samples always yield higher rates of projections.

Experiment 3

In Experiment 3 participants were presented evidence about the same individual (i.e., the same white rabbit). The conclusion exemplars included the *same* individual, instances from a *close* match (e.g., other rabbits), and *far* matches (e.g., mammals). We predicted that sample size would have a limited influence on projections in Experiment 3 because the evidence presented to participants came from a homogeneous sample: evidence about four instances of an individual might be as informative as evidence about sixteen instances. Similarly, the narrow sample might not promote projections to close and far matches, and more evidence might not lead to a greater willingness to generalize to non-sampled items. Experiment 3 tests the degree to which sample size, alone, supports broad projections. If large samples always lead to higher rates of projections, then children should show a higher rate of projections to non-sampled items in the Large sample condition.

Method

Participants. 30 Adults (undergraduates; 14 females, 16 males), and 30 children (M = 5;0; age range, 4;3 – 5;10 years; 16 females, 14 males) participated. Participants were recruited from the same population using the same methods as in Experiment 1. No participants were involved in other studies reported here.

Design, materials, and procedure. All facets of the design were the same as in Experiment 1 with the exception of the exemplars used. The evidence exemplar was a single white rabbit presented four times (in the Small sample condition) or sixteen times (in the Large sample condition). Also, the conclusion exemplars were the same as in Experiment 2, except that the "individual" items were included as the Same matches, a collection of rabbits (used in Experiment 2) served as the Close matches, and a set of mammals (used in Experiments 1 and 2) were the Far matches. In all other respects the design was identical to Experiment 1.

Results and discussion

Mean projections to each of the items for both conditions and age groups are presented in Figures 5 and 6. Scores that were different from chance are indicated with an asterisk A mixed ANOVA with Age and Sample size between subjects and Target within subjects revealed an effect of Target, F (2, 112) = 43.3, $p < .0001$, which interacted with age, F (2, 112) = 4.40, $p < .05$; simple effects showing the effect was caused by a higher rate of projections to the far targets for children than adults, $F(1, 58) = 19.25$, $p < .001$. There was no effect of condition or age.

Unlike Experiments 1 and 2, results from Experiment 3 revealed no effect of sample size. Participants in all age groups used small and large samples of evidence to make inferences about the same individual in the future. Overall children made higher rates of projections, but the effect was not mediated by sample size. Overall, the results indicate that sample size had common effects on children and adults' projections when the evidence sample was maximally homogenous (Experiment 3) but not when the evidence sample was variable (Experiments 1 and 2).

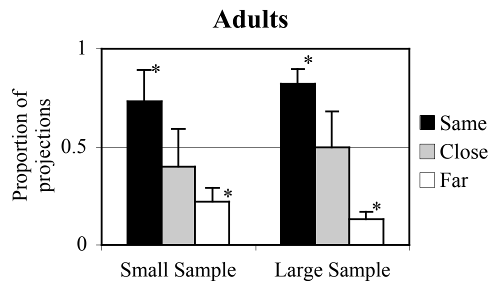

Figure 5: Adults projections across target types in both sample conditions in Experiment 3. Error bars represent the standard error of the mean.

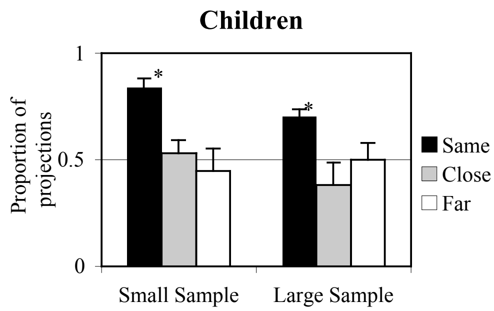

Figure 6: Children's projections across target types in both sample conditions in Experiment 3. Error bars represent the standard error of the mean.

General Discussion

The studies reported here explored a fundamental yet understudied component of induction, the influence of sample size. Overall, there were several novel findings. First, unlike previous studies, we observed strong effects of sample size on induction in young children, particularly when evidence exemplars had high category variability: 5-year-old children were more likely to generalize novel properties to all types of conclusion items (i.e., same, close, and far matches) from larger than from smaller samples (Experiment 1). Second, children made narrower projections (i.e., projecting properties to same and close matches but not to far matches) when category variability in the evidence exemplars was reduced (Experiment 2). Third, children used sample size information to make property projections to the non-sampled items: children were more likely to generalize properties to non-sampled items from larger than from smaller samples (Experiments 1 and 2). Fourth, sample size did not influence projections when the sample was comprised of a homogenous set of exemplars (Experiment 3). Evidence from Experiment 3 suggests that conclusions about sample size effects should be reserved to cases in which samples include variability among the instances. Perhaps the most surprising result from the present studies was that sample size effected children's but not adult's responses.

Results reported above are inconsistent with the idea that children ignore sample size information (Jacobs & Narloch, 2001). However, the results support the account of early induction that treats the evaluation of available evidence as a version of statistical learning. Because homogenous evidence is invariant, less evidence is needed to determine which inferences are warranted from the sample. Increases in the diversity of samples yields samples that are more variable; as a result, more evidence may be necessary to determine which inferences are warranted.

In fact, the present results indicate that children rely *more* on sample size than adults. It is possible these age differences were caused by differences in the ability to generate a representative set of conclusions from a sample of evidence. In each of the experiments adults appeared to recognize that evidence identified a specific category, while children never constrained their inferences to the most specific category that included all of the exemplars. However, children were the only group whose conclusions fluctuated as a function of sample size. Certainly more work is needed to examine the nature of the developmental differences in the use of sample size to support induction.

This is the first set of studies to conclude that children as young as 5-years of age are influenced by sample size information in the course of induction. Studies that have reported no sample size effects are different from the present design in several ways. First, previous studies have used familiar properties (e.g., eye color, size), while the present studies used novel properties. Familiarity with content may cause children to rely on prior knowledge about property variability. Second, past studies attributed properties to evidence exemplars at a rate of 100%. It is possible that sample size information is more useful for making projections about properties that are variable rather than invariant.

Third, previous work explored children's ability to choose between two different samples of different size, the present studies manipulated sample size between participants. It is possible that children did not choose larger samples over smaller samples in previous studies because they had difficulty making the appropriate comparisons (for instance, due to working memory demands). Finally, previous studies have presented evidence as collections of exemplars, while the present studies introduced each instance of the sample as an individual. It is possible that the method of presentation used here may have highlighted the frequency of instances in the sample thus making it easier for children to attend to sample size.

Indeed, there are a number of important differences between the present studies and existing studies on sample size effects in children. Future studies should clarify the conditions under which sample size influences children's induction. Ultimately, any comprehensive theory of induction will have to consider the effect of statistical properties, such as the amount of evidence and variability in the evidence.

Acknowledgements

This work was supported by an NIH postdoctoral training grant (T32 MH019102). Special thanks to the teachers, parents, and children of Carnegie Mellon University's Children's School, Beginnings, and Overbrook Head start for their generous support of this research.

References

Gelman, S.A. (2003). *The essential child*. Oxford University Press.

Gelman, S.A., & Coley, J.D. (1991). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796-804.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129-161). Chichester: Wiley.

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. *Journal of Applied Developmental Psychology, 22*, 311-331.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Lawson, C.A., & Fisher, A.V. (2008). Children's attention to property likelihood as a guide to property projection. In B.C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

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

Nisbett, R.E., Krantz, D.H., Jepson, D., & Kunda, Z. (1983). The use of statistical heuristics in everyday reasoning. *Psychological Review, 90*, 339-363.

Rakison, D.H., & Hahn, E. (2004). The mechanisms of early categorization and induction: Smart or Dumb Infants? In R. Kail (Ed.), *Advances in Child Development and Behavior*, Vol 32. New York: Academic Press.

Rogers, T.T., & McClleland, J. (2004). *Semantic Cognition*: The MIT Press.

Rhodes, M., Brickman, D., & Gelman, S.A. (2008). Developmental changes in the consideration of sample diversity in inductive reasoning. *Journal of Cognition and Development, 9*, 112-143.

Sedlmeier, P., & Gigerenzer, G. (2000). Was Bernoulli wrong? On intuitions about sample size. *Journal of Behavioral Decision Making*, *13*, 133-139.