Diverse Evidence for Dissociable Processes in Inductive Reasoning

Eoin Travers (etravers01@qub.ac.uk)
School of Psychology, Queen's University Belfast
University Road, Belfast BT7 1NN

Aidan Feeney (a.feeney@qub.ac.uk)
School of Psychology, Queen's University Belfast
University Road, Belfast BT7 1NN

Abstract

Previous work suggests that inductive and deductive reasoning may be accomplished by different processes. Here, we examine whether different phenomena of inductive reasoning, previously explained in the same way, may rely on different types of processes. In Experiment 1 we show that trials which examine sensitivity to sample size in inductive reasoning have greater effects on secondary task performance than do trials examining sensitivity to the diversity of the sample. In Experiment 2 we show that in a surprise recognition memory test, participants have significantly better memory for the content of diversity trials than for sample size trials. Both findings are consistent with the suggestion that some phenomena of inductive reasoning may be rule-based, whereas others may depend on feature-level processing.

Keywords: Reasoning; induction; diversity effect; law of large numbers.

Introduction

Not all thinking is the same. Because the same experimental manipulations affect them differently, it has been claimed that inductive and deductive thinking are dissociated (see Rips, 2001; Heit & Rotello, 2010). Heit and Rotello argue that deductive reasoning calls more on processes that are sensitive to logical validity, whereas inductive reasoning relies more on associative processes. However, a background assumption appears to be that inductive reasoning, for example, consistently draws on the same processes, and most theories of inductive reasoning attempt to capture different experimental phenomena in the same way (see Osherson et al, 1990; Sloman, 1993; Rogers & McClelland, 2004). Here, we will consider whether different processes underlie different phenomena that have been observed in people's inductive reasoning. Specifically, we will examine whether sensitivity to the size of the sample upon which an inductive inference is based may be due to rule-based processes, whereas sensitivity to the diversity of the evidence may call on more feature-based processing.

Sensitivity to the size and diversity of samples

Despite claims made by Kahneman and Tversky (1972), there is much evidence that adults, and sometimes children, are sensitive to sample size (see Piaget & Inhelder, 1975; Nisbett, Krantz, Jepson & Kunda, 1983). In experiments on category-based induction, where participants are typically taught that members of certain categories possess a novel property and are asked whether members of some other category also possess that property, the tendency to prefer arguments based on a larger sample of categories is known as the monotonicity effect (see Osherson et al., 1990). However, not everyone displays the monotonicity effect in such experiments (see Feeney, 2007).

Whereas sensitivity to sample size has been intensively studied in the literature on judgment and decision making, sensitivity to sample diversity has most often been studied in the literature on category-based inductive reasoning (for a review, see Heit, Hahn & Feeney, 2005). Although preference for more diverse evidential samples has been informally advocated by a variety of philosophers of science (e.g. Bacon, 1878; Carnap, 1950; Popper, 1963), attempts to formally justify a diversity principle are rarer and there are arguments against the existence of a general principle (see Lo, Sides, Rozelle & Osherson, 2002; Medin et al., 2003). Nonetheless, there are numerous demonstrations in experiments on category-based induction in which a majority of people consider arguments with more diverse premises to be stronger. People are sensitive to the diversity of the evidence, at least some of the time.

Accounts of sensitivity to sample size and diversity

Accounts of sensitivity to sample size are to be found in a variety of literatures whereas sensitivity to sample diversity is accounted for only by theories of category-based induction. Fong, Krantz & Nisbett (1986) claimed that sensitivity to sample size occurs because people possess intuitive but abstract rules that correspond to the law of large numbers, and showed that sensitivity to the law of large numbers can be enhanced by training. This account is similar in some respects to Piaget's (Inhelder & Piaget, 1975). In particular, both accounts stress the centrality of sensitivity to sample size to reasoning about probability more generally. Stanovich and West (1999) offer a dual process account of sensitivity to sample size, where such sensitivity when it is observed, is the result of effortful processes that draw on working memory in order to apply normatively justified rules or principles for reasoning.

Sensitivity to sample size, or adherence to the monotonicity principle, is explained very differently in models of category-based induction. For example, the similarity-coverage model (Osherson et al., 1990) holds that arguments are strong to the extent that the conclusion category is "covered" by the categories in the premises.

That is, to the extent that instances sampled at random from the conclusion category are similar to the categories in the sample. As a larger sample is more likely to better cover the conclusion category than a smaller sample, people judge arguments based on larger samples to be strong. Sloman (1993) predicts that arguments will be judged strong to the extent that there is overlap in the features of the conclusion category and the features of the categories in the sample. This account predicts sensitivity to sample size on the grounds that larger samples, on average, lead to greater featural overlap. Notably, all accounts of category-based induction, including Bayesian models (Tenenbaum, Kemp & Shafto, 2007) explain sensitivity to sample size and diversity in the same way.

In summary, different explanations of sensitivity to sample size posit different types of process. Early developmental and decision making accounts posit the existence of abstract and intuitive rule-like representations which, according to some accounts (see Stanovich & West, 1999) are effortfully applied. On the other hand, accounts of sensitivity to sample size in the literature on category-based induction appeal to processes operating over the relations between specific members of the sample. Some accounts (e.g. Sloman, 1993; Rogers & McClelland, 2004) hold that the application of these processes is relatively effortless. Accounts of sensitivity to sample diversity appear only to be found in the literature on category-based induction, and are similar to the accounts of sensitivity to sample size that are to be found in the same literature.

Dissociating the effects: Two different paradigms

The goal of the experiments to be described below was to examine whether similar or different processes underlie the sample diversity and sample size effects in induction. To achieve this goal we derived hypotheses about possible differences between the two phenomena in terms of the effort required by each and about the side effects of the underlying reasoning processes.

Effort and secondary tasks To the extent that models of category-based inductive reasoning are correct in assuming that sensitivity to sample size and diversity require the operation of the same processes, we should expect to find no differences between the effort required in order to demonstrate each effect. However, if sensitivity to sample size requires the operation of a rule-based process (Fong et al, 1986) that draws on working memory (Stanovich & West, 1999) then we might expect to be able to show that sample size trials require more cognitive effort than do diversity trials. To test this hypothesis we presented reasoning trials (the primary task) concurrently with a memory task (the secondary task). Such designs have previously been employed to test hypotheses about the effort required by particular types of thinking (see De Neys, 2007). If sample size materials require more effort than diversity materials, we should expect to observe (a) a greater effect of the secondary task on sensitivity to sample size than on sensitivity to diversity; or (b) greater effects of the sample size task than the diversity task on the secondary task; or (c) both effects. The first experiment to be described below tested these hypotheses.

Induction then recognition A contentious claim in the literature is that the processes applied during reasoning may have consequences for the type of representation which reasoners construct of the problem material, and hence for their ability to accurately recognize the materials they reasoned about (see Sloutsky & Fisher, 2004). There is evidence that following a simple inductive reasoning task, children have better recognition memory for the problem materials than to adults, although they perform equally well on the reasoning task. Sloutsky and Fisher (2004) claimed that this recognition memory effect was a consequence of children and adults using different processes to reason. They claim that adults reason on the basis of category membership and therefore construct category-level or gist (Brainerd, Reyna & Forrest, 2002) representations of the reasoning stimuli. Children, on the other hand, reason on the basis of correspondences or similarities between the entities in the reasoning problem. This leads them to construct a verbatim (Brainerd et al., 2002) representation of the materials. When all participants are subsequently presented with old pictures and new critical lures, it is children with their more detailed representation of the original materials who are better able to discriminate between old and new items.

Although there has been disagreement about whether the original induction-then-recognition experiments necessitate conclusions about developmental changes in reasoning processes (Wilburn & Feeney, 2008; Hayes, McKinnon & Sweller, 2008), the paradigm may be a very useful tool for determining whether different reasoning phenomena are caused by different reasoning processes. For example, if sensitivity to sample size in category-based induction is due to the application of an intuitive rule, then we would not expect participants to encode verbatim representations of the reasoning stimuli. On the other hand, if sensitivity to diversity requires representation of the relations between the entities in the reasoning problems, then participants should be more likely to construct verbatim representations of the entities in those reasoning materials. This difference in the type of representation that is constructed might have consequences for participants' ability to subsequently recognize the entities that they have previously reasoned about. Specifically, participants may have better recognition memory for diversity materials than for sample size materials. On the other hand, if the same processes are involved in sensitivity to both phenomena, we would expect no differences due to reasoning phenomena in recognition accuracy. Experiment 2 below will test these hypotheses.

Experiment 1

The aim of this experiment was to test for differential effects of a secondary task on sensitivity to sample size and diversity in inductive reasoning, and to test for differential effects of these reasoning phenomena on performance of a secondary task.

To facilitate the use of the Induction then Recognition paradigm in Experiment 2, across both experiments we adopted a paradigm recently used to test for diversity effects in children (Rhodes, Brickman & Gelman, 2008) in which participants are asked to select between a diverse and non-diverse sample of category members in order to help them decide whether all members of the category possess a novel property.

Method

Participants Sixty students (29 males) were recruited in a quiet area of the library at QUB, and paid £2 each to take part in the study. The mean age was 28.63 years.

Materials In each reasoning task, participants were told about a novel property that might be possessed by all members of a category, alongside two samples of members of that category, and were asked which sample they would like to test in order to decide whether all members of the category possess the property. On the five trials assessing sensitivity to diversity, the diverse sample consisted of pictures of two category members of different coloration, species, or breeds (in the case of dogs), while the non-diverse sample consisted of one of the diverse sample members, and another similar category member. On the five trials assessing sensitivity to sample size, the small sample consisted of two category members, and the large sample consisted of the same two category members plus one additional member. Unique categories, images, and properties were used on each trial.

Because of the possibility that participants might complete the sample size trials without processing the content of the images, we included five control trials at the end of the experiment which asked participants to choose between a small diverse sample and a larger non-diverse sample. If some participants complete the sample size trials without processing the content of the images in the samples, we should find that participants choose the large sample in the control trials as often as in the sample size trials. In addition, there should be a strong correlation between the tendencies to choose the large sample in both types of trial.

The secondary task (see De Neys, 2006) required participants to memorize an array of dots on a 3x3 matrix before each reasoning task, and recreate it immediately afterwards.

Procedure All participants completed the experiment on a laptop computer running E-Prime software. They were told before beginning that the experiment would investigate how people make judgments about category members and their properties. On each trial, participants were presented with a statement at the top of the screen, with the two possible samples below it on either side. They were instructed to press the ‘1’ button to choose the left sample, and the ‘2’ button to choose the right sample. There were two practice reasoning trials before the experimental trials began. The first ten trials tested for sensitivity to sample size and diversity and their order was randomized separately for each participant. The final five trials pitted a two-member diverse sample against a three-member homogenous sample.

Before the beginning of each trial, participants were presented with a 3x3 grid for 1000 ms, containing either four dots in random positions (complex condition), or three dots in a straight or diagonal line across the grid (simple condition). After given a response in each reasoning trial, they saw a blank grid, and were required to recreate the pattern seen previously. Participants were instructed to remember the dot pattern as well as they could, while still paying attention to the reasoning task.

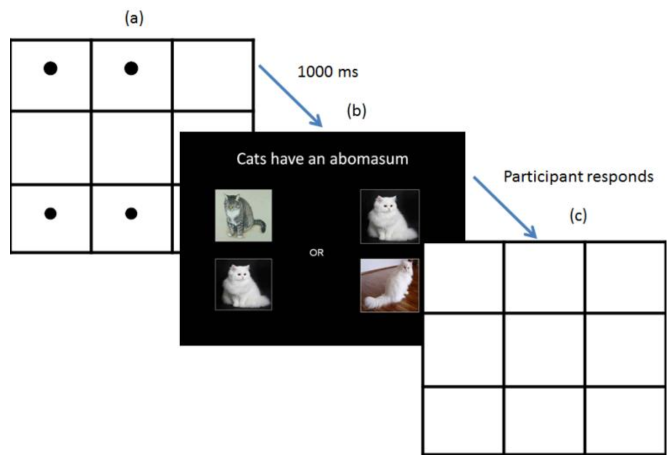

Figure 1: Timeline for trials in Experiment 1. At (a) participants were presented with a simple or complex spatial array to memorize; after 1000 ms they were presented with a reasoning problem (b) requiring them to choose one of two samples; and (c) once they chose a sample they recreated the spatial array.

Results

Primary task performance Across secondary task conditions, participants showed sensitivity to diversity on only 52.3% of trials (SD = 24%), and sensitivity to sample size on 72% of trials (SD = 28%). A 2 (secondary task: complex vs simple) x 2 (trial type: montonicity, diversity, & control) mixed ANOVA revealed a main effect of trial type only, $F(1, 58)$ 21.44, $p < .001$. Neither the effect of load nor the interaction between trial type and load achieved statistical significance.

Secondary task performance Participants’ ability to correctly recall the dot arrays broken down by complexity condition and trial type is to be seen in Figure 2. A 2x2 mixed ANOVA revealed a significant main effect of complexity condition, $F(1, 58) = 50.90$, $p < .001$, and a significant interaction between complexity condition and trial type, F(1, 58) = 12.88, p = .002. Post hoc tests on the means involved in this interaction revealed that reasoning about diversity trials had a significantly greater effect on simple secondary task performance than did reasoning about sample size trials, t(29) = 2.92, p < .01. However, performance on the complex secondary task was affected to a greater degree by sample size trials than by diversity trials, t(29) = 2.48, p < .02.

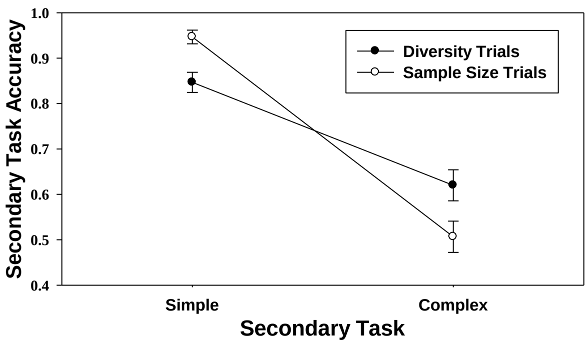

Figure 2: Interactive effect of trial type and secondary task on secondary task performance in Experiment 1.

Control performance One potential issue with interpreting the results of this experiment and the next is that participants may complete the sample size trials by simply counting the number of images in each sample without processing the content of the samples. One finding that suggests this did not happen is that the mean inspection time for sample size trials was almost identical (6377ms) to the mean inspection time for the diversity trials (6380ms). In addition, analysis of the control trials revealed that participants selected the large sample in the control trials 60% (SD = 30%) of the time which is significantly less often, t(59) = 2.36, p < .03, than in the sample size trials. If participants had not been processing the content of the samples but only their size, we would have expected the rate at which the large sample was chosen to be virtually identical in these two conditions. In addition, there was almost no association between the tendency to select the large sample in the sample size and control trials, r(60) = .02.

Discussion

Participants in Experiment 1 demonstrated less sensitivity to sample size than to sample diversity, and they performed better on the simple than on the complex secondary task. Furthermore, performance on the complex secondary task was significantly worse when the primary task required sensitivity to sample size than when it required sensitivity to sample diversity. On the other hand, performance on the simple secondary task was worse when the primary task required sensitivity to diversity. These results are consistent with the claim that different processes underlie the sample size and diversity effects. The findings for the complex secondary task, in particular, suggest that participants who are sensitive to sample size may possess a simple rule. Because the operation of such a rule requires general cognitive processes related to working memory (see Stanovich & West, 1999), performance of a complex secondary task which also requires working memory, is particularly impaired. Fong et al. (1986) suggested that the sample size rule is abstract but intuitive. Its intuitiveness may explain why sensitivity to sample size was observed on a relatively high proportion of trials, and why performance on the simple secondary task was barely impaired when the primary task required sensitivity to sample size.

Notably, performance on the primary task was not affected by the nature of the secondary task and it is not clear why this was the case. One possibility is that participants prioritized the reasoning task.

Experiment 2

The aim of Experiment 2 was to provide further evidence for dissociation between sensitivity to the size and diversity of the sample in inductive reasoning. To do this we asked participants to complete a surprise recognition memory test once they had completed the reasoning items. If sensitivity to sample size involves the application of an intuitive rule, then we might expect participants to build a gist rather than a verbatim representation of the content of the samples. This representation should lead to relatively poor recognition memory for the entities in the samples. Memory for the entities presented in the diversity trials should be more accurate, if sensitivity to diversity depends on more feature-based processing of the images in the samples. Such processing should be more likely to result in verbatim representations of the pictures in the samples which will better support accurate recognition of those entities.

Method

Participants 59 QUB students (25 males) were tested in a quiet area of the university library, and paid £2 each to take part in the study. The mean age was 26.5 years.

Materials Materials were the same as used in Experiment 1, except that there were seven diversity and seven sample size trials. We did not include control trials in this experiment. The recognition memory task consisted of 63 images: 28 pictures previously seen in the reasoning tasks (2 from each trial, one of which was featured twice in the trial), 28 previously unseen pictures of members of the previously featured categories, and 7 pictures of categories not featured at any stage in the experiment.

Procedure The procedure for the reasoning part of the experiment was broadly similar to the procedure followed in Experiment 1. However, the secondary task was omitted, trial type was blocked and block order was counterbalanced.

The order in which trials were presented within blocks was randomized.

Once they had completed the reasoning trials, participants were told that the second part of the experiment would consist of a surprise recognition test, and instructed to try to identify which images had been seen previously in the reasoning tasks. Images were presented one at a time and participants pressed the '1' button for pictures seen before, and the '2' button for new pictures.

Materials check Our hypothesis is that recognition memory for the contents of diversity trials will be better than for the contents of sample size trials. We carried out a check to ensure that the materials used in the diversity trials were no more memorable than those used in the sample size trials. We presented the materials used in the reasoning part of the experiment to 34 participants. The information about properties was not included and instead of asking participants to make a choice between the samples, we instructed them to memorize the images for a subsequent memory test.

Results

Reasoning task Participants selected the diverse sample on 73.6% of trials (SD = 25%), and the larger sample 81.8% of the time (SD = 26%). Participants were significantly more sensitive to sample size than they were to sample diversity, $t(58) = 2.105$, $p = .04$.

Recognition memory Performance on the recognition memory test was analyzed with the A' statistic (Snodgrass & Corwin, 1986), a non-parametric analogue of the d' signal detection measure. An A' of .5 corresponds to chance

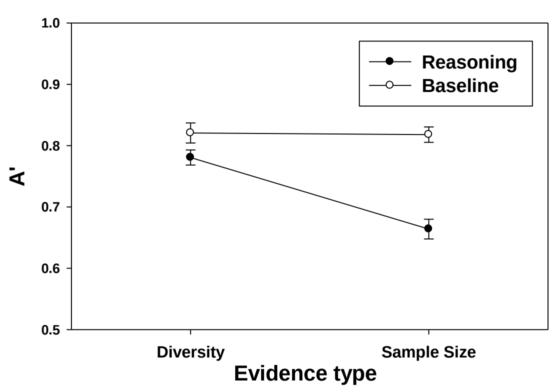

Figure 3: A' scores from Experiment 2 broken down by trial type and whether participants reasoned about the materials or studied them for memory.

discrimination between old and new stimuli, while a score approaching 1 indicates perfect discrimination. In Figure 3, A' scores for the main experiment are presented alongside scores from the materials check. While A' scores for diversity and sample size materials were almost identical for participants in the baseline memory condition (A' = .82, and .81 respectively, SDs = .09 and .07), $t(33) = .177$, amongst participants in the main reasoning condition, recognition was much better for the diversity materials (A' = .78, SD = .13) than the sample size materials (A' = .66, SD = .13), $t(58) = 6.343$, $p < .001$.

Inspection times We measured the time between presentation of each reasoning item and participants' responses. The average of this inspection time was 5579 ms (SD = 2055 ms) for diversity trials, and 5256 ms (SD = 2273 ms) for sample size trials. This difference was non-significant, $t(58) = 1.003$. Thus, the difference due to reasoning phenomenon in the recognition memory data cannot be attributed to differences in how long participants looked at the materials for each trial type.

Discussion

As we predicted, participants had better recognition memory for the entities they reasoned about in the diversity trials than they did for the entities in the sample size trials. Additionally, the results of our materials check confirmed that the diversity entities were not more memorable than the sample size entities. These results suggest that different processes underlie sensitivity to diversity and sensitivity to sample size. Whereas the former requires feature-based processing of the entities, resulting in a verbatim-type representation which supports accurate recognition memory, the latter is driven by the application of a rule, which leads to a gist representation of the samples and significantly less accurate recognition memory.

General Discussion

Both experiments reported here show evidence of a disassociation between the processes underlying sensitivity to the sample diversity and size. In Experiment 1, sensitivity to sample size and to diversity differentially impacted upon the secondary task, indicating a dissociation of the underlying mental processes. Similarly, in Experiment 2 materials used in diversity trials were remembered significantly better, suggesting a greater degree of feature-based processing. Taken together, these findings are problematic for single-process accounts of inductive reasoning (e.g. Osherson et al, 1990; Sloman, 1993; Rogers & McClelland, 2004).

Recent findings have challenged the classical view that inductive inference is the product of similarity-based or associative processes, while deduction relies on the application of abstract logical rules (Evans, 2012). On one hand, similarity-driven processes have been shown to underlie some deductive phenomena (Sloman, 1998). On the other, Heit and Rotello (2010; see also Rips, 2001) have shown that both similarity and logical validity determine inductive and deductive argument strength, but with induction drawing more heavily on similarity-based or associative information. With the blurring of the boundaries between the processes underlying the two forms of reasoning, it has become somewhat unclear what is distinctive about induction. Heit (2007) offers two views on defining induction: the process view, which relates to the processes by which we make an inference, and the problem view, relating to the structure of the inference to be made. While from the problem view deduction and induction remain discrete, our findings suggest that, from the process view, reasoning cannot be so easily partitioned. Our results, from two diverse paradigms, suggest that there is a disassociation between the processes underlying sensitivity to sample size and to sample diversity in category-based induction, and by extension, that inductive reasoning cannot be captured by single process accounts.

Acknowledgements

Eoin Travers is funded by a postgraduate studentship from the Department of Enterprise and Learning in Northern Ireland. We are grateful to Michelle Leckey for her help with the development of the materials used in these experiments.

References

Bacon, F. (1878). *Novum organum*. Clarendon press.

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2003). Are young children susceptible to the false–memory illusion? *Child Development, 73*, 1363-1377.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

De Neys, W. (2006). Automatic-heuristic and executive-analytic processing in reasoning: Chronometric and dual task considerations. *Quarterly Journal of Experimental Psychology, 59,* 1070-1100.

Evans, J. S. B. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, *18*, 5-31.

Feeney, A. (2007). How many processes underlie category-based induction? Effects of conclusion specificity and cognitive ability. *Memory & Cognition*, *35*, 1830–1839.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253-292.

Hayes, B. K., McKinnon, R., & Sweller, N. (2008). The development of category-based induction: Reexamining conclusions from the induction then recognition (ITR) paradigm. *Developmental Psychology*, *44*, 1430-1441.

Heit, E. (2007). What is induction and why study it? In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 1–24). Cambridge: Cambridge University Press.

Heit, E. & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 805-812.

Heit, E., Hahn, U. & Feeney, A. (2005). Defending diversity. *Categorization inside and outside of the laboratory: Essays in honor of Douglas L. Medin*, (pp. 87–99). Washington: APA.

Inhelder, B., & Piaget, J. (1975). *The origin of the idea of chance in children*. New York: NW Norton and Company.

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

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 26*, 181-206.

Medin, D. L., Coley, J. D., Storms, G. & Hayes, B. L. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, *10*, 517–532.

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

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

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge.

Rhodes, M., Brickman, D. & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition*, *108*, 543–556.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*, 129–134.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT press.

Sides, A., Osherson, D., Bonini, N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory & Cognition*, *30*, 191-198.

Sloman, S. A. 1993. Feature-based induction. *Cognitive Psychology*, *25*, 231–280.

Sloman, S.A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, *35*, 1-33.

Sloutsky, V. M., & Fisher, A. V. (2004). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science*, 15, 553-558.

Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50..

Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, *38*, 349-385

Wilburn, C. & Feeney, A. (2008). Do development and learning really decrease memory? On similarity and category-based induction in adults and children. *Cognition*, *106*, 1451-1464.