UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Category Size and Category-Based Induction

Permalink

Journal

ISSN

Authors
Feeney, Aidan
Gardiner, David R

Publication Date
2002

Peer reviewed

Category Size and Category-Based Induction

Aidan Feeney & David R. Gardiner
Department of Psychology
University of Durham, Stockton Campus
University Boulevard
Stockton-on-Tees, TS17 6BH
United Kingdom
{aidan.feeney, d.r.gardiner}@durham.ac.uk

Abstract

In this paper we investigate the role of category size in category-based induction. In a series of three experiments we asked participants about the strength of inductive inferences from arbitrary subordinate categories to their superordinates. We show that people use both subordinate and superordinate category size as a cue in category-based induction (Experiments 1 & 2). However, the results of Experiment 3 show that the effect of subordinate category size is smaller when the categories are said to be similar than when said to be dissimilar. On the basis of this result we suggest that people use category size as an indication of how much uncertainty remains concerning the superordinate rather than as a means of assessing how representative the category is as a sample of the superordinate. We conclude with a discussion of possible strategies for inductive reasoning.

One of the functions of categories is to promote inductive inference. Knowing that one set of instances possesses a certain feature allows us to consider whether other sets are also likely to possess the same feature. For example, knowledge that all of the chairs in the lecture room we are currently in are made of plastic will assist us in making a prediction about the chairs in the lecture theatre next door, the cafeteria at the end of the corridor and the provost's office. The experiments to be reported in this paper were all concerned with the role played by information about category-size in such inductive inferences. They ask whether participants are more likely to project a property if it is possessed by instances of a larger category than of a smaller category and whether people are more confident about conclusions concerning large or small groups. Furthermore, if people do turn out to be sensitive to size cues in this manner, what kind of reasoning underlies their use of such cues? As we will see below, there are a variety of ways in which category size might influence people's judgements of inductive strength.

Other researchers have been interested in induction based on categories and there are several models of categorical induction in the literature all designed to capture between 12 and 15 phenomena (for an excellent review, see Heit, 2000). One factor that is common to all of these models is inter-category similarity (Osherson et al, 1990) or featural overlap between categories (Sloman, 1993). To illustrate how similarity might affect the strength of an inductive inference consider the arguments below where the statement above the line is a premise and the statement below the line is a conclusion. With arguments of this type participants are asked to assume the premise to be true and to evaluate the degree to which it supports the conclusion.

<u>Robins have an ulnar artery</u>
Thrushes have an ulnar artery   Argument 1

<u>Robins have an ulnar artery</u>
Flamingos have an ulnar artery   Argument 2

As the categories in Argument 1 are more similar than those in Argument 2, people will judge the former to be stronger than the latter. Where the conclusion category is superordinate to the premise category, as in Argument 3 below, the degree to which the premise category is typical of the superordinate category informs people's judgements of inductive strength (see Rips, 1975).

<u>Robins have an ulnar artery</u>
Birds have an ulnar artery   Argument 3

A second factor which, in at least one model of category-based induction, impacts upon judgements of argument strength is 'coverage' (Osherson et al, 1990). Coverage is the degree to which the premise categories are similar to instances of the conclusion category (or, in cases where the conclusion category is not superordinate to the premise categories, to instances of the nearest superordinate category containing both premise and conclusion categories). So, for example, Argument 4 below would normally result in greater ratings of inductive strength than would Argument 5.

German Shepherds produce phagocytes
<u>Poodles produce phagocytes</u>
All dogs produce phagocytes   Argument 4

German shepherds produce phagocytes
<u>Dobermans produce phagocytes</u>
All dogs produce phagocytes   Argument 5

As the premise categories in Argument 4 are similar to a greater range of instances of the conclusion category than are the premise categories in Argument 5, the former is judged to be stronger than the latter. In general, the more diverse are the premise categories, the stronger is the argument (although for exceptions see Sloman, 1993).

There are several things to note about much of the existing work on category-based induction. First, although rarely formally contrasted with normative models of induction (for an exception see Heit, 1998), many of the effects in the literature have an intuitively strong normative basis. For example, both effects of similarity and coverage might be expected under the assumption that participants are sensitive to the representativeness of the samples about which they have some information. Samples that are either similar to, or typical of, the population to which the property will be projected, are, intuitively at least, more representative of that population. Similarly, diverse sets of premises intuitively seem to be more representative of the premises than are non-diverse premises.

A second characteristic of previous work on category-based induction is that researchers have been interested in investigating the effects of category knowledge on inductive judgements concerning natural kinds. As it is not normally possible to know the size of many naturally occurring categories (for example, how many members are there of the category 'bird'?) research has tended to concentrate on the role played by inter-category relationships. This may be contrasted with work on, for example, statistical judgement where both the a priori probabilities of the hypotheses as well as the probability of the evidence given each hypothesis has been manipulated. By presenting participants with problems concerning arbitrary categories in which category size was manipulated, the work to be described here attempted to address the role that category size plays in category-based induction.

Category Size and Category-Based Induction

Consider the following scenario:
> *672 people work in a 10 story office block. Of these, 313 work on floor 2 and 35 work on floor 7.*

Given this scenario, which of these arguments is the strongest?

All 313 people who work on floor 2 have an
identity number beginning with the letter Z
All 672 people who work in the office block
have an identity number beginning with the letter Z   Argument 6

All 35 people who work on floor 7 have an
identity number beginning with the letter Z
All 672 people who work in the office block
have an identity number beginning with the letter Z   Argument 7

There are at least two reasons for preferring Argument 6 to Argument 7. The first line of reasoning is that the sample size in Argument 6 is larger than that in Argument 7. As larger samples are held to be more representative of the populations from which they are drawn than are small samples, Argument 6 is stronger than Argument 7 (for a recent discussion of the psychological literatures on sensitivity to sample size see Sedlmeier & Gigerenzer, 1997, and Keren & Lewis, 2000). However, since Nisbett et al's (1983) work on statistical heuristics in induction, it has been known that the variability of the feature being projected interacts with sample size to determine inductive strength. For example, Nisbett et al found that only a very small sample was required for participants to project features for which there is little within category variability (e.g. colour in a specific species of bird) whereas a much larger sample was required for the projection of more variable features. In the scenario above, the information that people work on different floors may suggest variability in staff identity numbers. That is, if category structure is made salient by a scenario, sample size may not be considered relevant in determining the strength of the inference.

The type of reasoning described above relies on indirect inference. That is, an inference about characteristic of a population is made on the basis of evidence about the prevalence of that characteristic amongst members of a sample. A less sophisticated, but more direct, way of making the inference is to think about the sample as a proportion of the population. Thus, if a large proportion of the population is known to possess the characteristic, then there is less uncertainty about the remaining members of the population and hence, a greater probability that the characteristic is universally possessed.

If we find that participants are sensitive to category size when asked to evaluate category-based inductive inferences, then the question arises as to what form of statistical reasoning underlies that sensitivity. The first two experiments to be reported here were designed to investigate premise and conclusion categories as cues to inductive reasoning whilst the final experiment was designed to compare contrasting accounts of any category size effect.

Experiment 1

Method

A total of 40 participants from the undergraduate population of the University of Durham (Stockton campus) took part in this experiment. Of these, 11 were male and 29 were female. The average age of participants was 22 years.

Experiment 1 had an entirely within participants design. The dependent variable was the number of problems for which participants chose as strongest the argument

concerning a large premise category. Each participant received a set of instructions and eight reasoning problems. Each problem described a superordinate category and two subordinate categories. The absolute size of each category was described such that the *Large* subordinate category was 25-40%, and the *Small* subordinate category 5-8%, of the size of the superordinate category. Participants received problems such as the following:

> *Extensive research has shown that there are several strains of the dreaded, and always fatal, Xanthrax virus. 1,000 people are known to have died from the virus. One form of the virus is Strain 6 from which 300 people have died. Another form is Strain 3 from which 60 people have died.*

and were then asked to indicate which of two arguments was the stronger. These arguments consisted of a premise, concerning one or other of the subordinate categories, and a conclusion concerning the superordinate category:

Xanthrax Strain 6 produces a blotchy rash in sufferers
All 1,000 Xanthrax fatalities displayed a blotchy rash

Xanthrax Strain 3 produces a blotchy rash in sufferers
All 1,000 Xanthrax fatalities displayed a blotchy rash

The order in which the arguments appeared was controlled whilst the eight problems appeared in one of eight randomly determined orders. The other seven problems concerned books in a library, articles from several issues of a journal, the age of trees in a forest, houses sold by an estate agent, workers in an office block, characteristics of historical artefacts and works of art.

Results and Discussion

As expected, participants displayed a marked preference for arguments involving the large subordinate category. Out of a maximum of eight, the mean number of such arguments selected as being stronger was 6.13 (S.D. = 1.99). The difference between the number of large subordinate category arguments that were selected as stronger and the number that would be predicted by chance was statistically significant across all problem contents ($t(40) = 6.76$, $p < .001$). This preference for large premise categories was also statistically significant in all eight problem contents ($\chi 2(1) > 8$ in 7 out of the eight cases). Response frequencies, broken down by content, are displayed in Table 1.

The results of Experiment 1 confirm our intuition that participants are more likely to project a property to a superordinate category from a large rather than a small subordinate. In Experiment 2 we kept subordinate category size constant and, instead, manipulated the size of the

Table 1: Large and Small argument selection from Experiment 1.

Problem Content	Subordinate Category	
	Small	Large
Disease	13	27
Library	8	32
Housing	11	29
Forest	10	30
Journal	10	30
Office Block	10	30
Artefacts	7	33
Gallery	6	34

superordinate category. Our strong intuition was that participants would be happiest projecting a feature to a small, rather than a large, category. An analogous effect exists in the literature (Osherson et al, 1990) where participants have been demonstrated to prefer projections to lower level, and hence smaller, categories.

A second aim of this experiment was to demonstrate category size effects in a between participants design. The literature on base rate neglect (see Koehler, 1996) contains demonstrations that participants are more likely to take the base rate into account when base rate is manipulated within participants. Hence, in Experiment 2 we wished to investigate whether participants would take category size into account in a between participant design.

Experiment 2

Method

The experiment had a 2 x 3 mixed design. Population size was manipulated between participants whilst each participant received three different problems asking them to rate the strength of an inductive argument.

A total of 116 participants from the undergraduate population at the University of Durham (Stockton Campus) took part in this experiment. Of these, 58 were male and 58 were female. The average age of participants was 21 years.

Participants received a booklet containing a set of instructions followed by three reasoning tasks. These tasks asked participants to evaluate the strength of arguments projecting a feature possessed by a subordinate category to all members of its superordinate. The problems concerned sub-types of a disease, individual production lines in a factory, and different variants of a plastic. Participants were

requested to rate the strength of the arguments on a 1-10 scale (very weak – very strong).

In cases where the superordinate category was small, the subordinate category accounted for between 45 and 55% of the superordinate. When the superordinate category was large, subordinate categories accounted for between 5 and 8% of the larger category. Importantly, only the size of the superordinate category was altered in this experiment. Approximately equal numbers of participants attempted the problems in each of the six possible orders.

Results and Discussion

The means and standard deviations from this experiment are presented in Table 2. A 2x3 Anova analysis revealed a significant effect of population size on the strength ratings assigned to arguments ($F_{(1, 114)} = 5.32$, MSE = 9.98, $p<.03$). As expected, the mean ratings for large population arguments (mean = 2.81, S.D. = 1.68) were significantly smaller than for the small population arguments (mean = 3.59, S.D. = 1.96). Neither of the other effects tested by the analysis were significant.

Table 2: Means and standard deviations from Experiment 2.

Condition	Content		
	Disease	Factory	Plastic
Large	3.84 (2.41)	3.14 (2.23)	3.79 (2.71)
Small	2.74 (2.07)	2.84 (2.07)	2.84 (1.96)
	3.29 (2.31)	2.99 (2.14)	3.32 (2.40)

As expected, participants rated arguments projecting features to small conclusion categories more highly than they did arguments concerning large conclusion categories. However, one striking aspect of these results is that mean ratings of argument strength were very low. It would appear that, at least in a between participants design, making subordinate categories salient in the scenario causes participants to doubt the conclusion regardless of conclusion category size. One possible reason for this is that participants may expect there to be differences between subordinate categories. This expectation may contribute to their unwillingness to project a feature to the superordinate on the basis of evidence concerning only one subordinate. In Experiment 3 we investigated the effects of explicit information about similarities and dissimilarities between subordinate categories on ratings of inductive strength.

Experiment 3

Experiments 1 and 2 have demonstrated that participants are sensitive to both subordinate and superordinate category when evaluating the strength of inductive arguments. In Experiment 3 we attempted to contrast two possible accounts of people's use of category size as a cue for induction.

The first account we considered was that people's judgements are affected by subordinate category size because they use a sample size heuristic (see Nisbett et al, 1983). That is, people realise that larger samples are more representative of the population from which they are drawn than are smaller samples. Alternatively, people may reason that a large subordinate leaves a smaller proportion of the superordinate unaccounted for than a small subordinate and hence, makes for a stronger argument.

To test these alternative accounts we gave one group of participants reason to believe that all subordinate categories in the domain were similar whilst telling another group that they were dissimilar. Our reasoning was that 'similar' problems should result in fewer attributions of variability to the superordinate category. If participants were cautious in using a sample size heuristic in Experiment 1 due to worries about the representativeness of the sample, then for 'similar' problems participants in this experiment should be less cautious and a greater effect of category size should be observed. On the other hand, 'dissimilar' problems should produce even more caution and a smaller effect of category size.

An alternative hypothesis is that the effect of category size in Experiment 1 may have been due to an assumption that the subordinate category was unrepresentative of the superordinate. That is, participants may have been unwilling to project a property to the superordinate because the subordinate category for which information was available may not have 'covered' the superordinate category. Accordingly, participants may have endorsed inferences from large premise categories more strongly because large premise categories leave fewer cases unaccounted for in the conclusion category. Thus, telling participants that the subordinate categories in the scenario are similar may cause them to rely less on category size as a cue and to assign higher ratings of inductive strength regardless of category size. Explicitly telling them that members of the superordinate are dissimilar may lead them to rely even more heavily on sample size.

Method

Eight male and 52 female undergraduate students (average age 27 years) from the University of Durham's Stockton Campus took part in this experiment which had a 2 x 2 entirely within participants design. The factors manipulated were the similarity said to hold between subordinate categories in each of the problems (similar vs. dissimilar) and the size of the subordinate category (large vs. small).

Each participant received a nine-page booklet comprising a set of instructions and eight inductive reasoning problems.

In each of these problems participants were told that instances of a subordinate category possessed a feature and were asked to evaluate the strength of an argument projecting this feature to all members of the superordinate category. The problems concerned workers in an office block, symptoms of a disease, features of handmade chairs, materials used by an engine manufacturer, properties of a type of plastic, the type of material used in indigenous art, production lines in a factory, and stock at a fish farm. Each participant received one version of each problem and two problems in each condition of the experiment. Each problem content appeared equally often in each condition of the experiment.

The problems were designed so that the large subordinate category was 45-55% of the size of the superordinate category, whilst the small subordinate category contained 5-8% of the superordinate population. To achieve the similarity manipulation participants were explicitly told that some similarity/dissimilarity existed between all members of the superordinate category. For example, in the office block problem participants were told that the workers either worked for the same, or different, divisions of a company.

Results and Discussion

As participants completed two problems per condition of the design, we calculated a mean score per condition for each participant. We carried out a 2 x 2 entirely within participants ANOVA on this data, the means from which may be seen in Figure 1. Our similarity manipulation had a highly significant effect upon the ratings of argument strength ($F (1, 59) = 24.70$, MSE = 10.38, p<.001). Mean ratings of argument strength were higher for categories containing similar members (8.15) than for categories containing dissimilar members (6.08). This finding replicates previous work (e.g. Nisbett et al, 1983) suggesting that within category variability significantly affects people's willingness to project a property from a sample to a population.

Figure 1: Interaction between Category Size and Similarity from Experiment 3
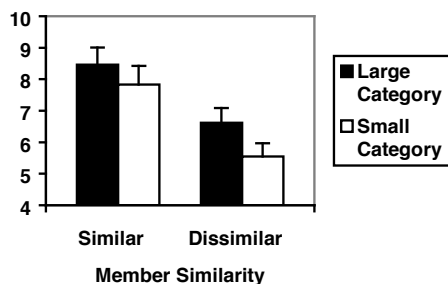
Our category size manipulation also had a significant effect on ratings of argument strength ($F (1,59) = 6.03$, MSE = 7.19, p<.02). Arguments involving the large subordinate category were rated as stronger (7.54) than those involving the small subordinate (6.69). Although the interaction between these factors did not approach significance ($F(1, 59) = .34$, MSE = 8.21, p > .5), planned comparisons revealed a significant effect of category size when category members were said to be dissimilar but not when they were said to be similar.

When participants are explicitly told that the members of the superordinate category are similar, premise category size ceases to have a significant effect on judgements of inductive strength. Instead, it would appear that category size is more important under conditions where 'indirect' inductive inference from a subordinate to a superordinate is likely to be unsafe. This suggests that category size functions as an indicator of the number of cases for which uncertainty remains.

General Discussion

Our finding, that category size acts as a cue in category-based inductive inference, is entirely novel (if not entirely unexpected). Likewise, the finding that the effect of category size decreases when participants are explicitly told that subordinate categories are similar is also novel We will discuss possible interpretations of these findings and their implications for the question of strategy use for induction.

The results of Experiment 3 might be regarded as being consistent with Nisbett et al's findings as in their study participants were found to be insensitive to sample size when the property that they were required to project was unlikely to vary. Similarly, our participants were also relatively insensitive to sample size when told that there were similarities between subordinate categories. It might even be argued that participants' greater sensitivity to sample size in the dissimilar condition is also consistent with Nisbett et al's results and is evidence that participants realised that in conditions of variability, a large sample is safer.

We are unsure about this reading of the results because we find it implausible that participants told that there were dissimilarities between subordinate categories should consider a bigger subordinate category more likely to be representative of the population than a smaller subordinate. We find it much more plausible that participants attributed greater strength to arguments that accounted for the most population members. It is plausible, however, that in the similar condition, participants projected features to the population on the basis of characteristics of the sample. Even in the small category size condition, premise categories always had more than 20 members. This figure

was chosen as 20 was the largest sample size given to participants in Nisbett et al's experiment. Given the existence of a subordinate category structure, participants may have considered the increase in inductive strength from small to larger premise categories to be insubstantial.

We contend that our results suggest that people are flexible in the strategies that they adopt for inductive inference. In conditions of low variability, people will project properties to a population on the basis of the characteristics of a sample. However, where high variability exists, people may be more likely to base their judgements of inductive strength on the number of cases outstanding.

The 'direct' strategy is a very interesting one, partly because it has been somewhat neglected in the literature (although see Evans & Dusoir, 1977). The conditions for its application are population size being both finite and approximately known. In addition, we suspect that a 'direct' strategy will be used in conditions where the population is small. This is because indirect induction only becomes necessary where the population is large or infinite and it is difficult, or impossible, to check all members. With small populations a direct strategy based on checking as many members as possible is more tractable.

Evidence that group size can affect how people perform induction comes from Wang (1996) who showed that the demonstration of classic framing effects depends on the size of the group being reasoned about. His explanation for this finding is that we possess social-group domain-specific reasoning abilities. As Wang only gave scenarios concerning social groups to his participants, we are unwilling to subscribe to the notion of social-group domain-specific abilities. However, we agree that the human species is likely to have made inferences about relatively small populations throughout most of its history. In addition, formal notions of induction evolved relatively recently (see Gigerenzer et al, 1989), most probably in response to a need for safe inferences about population whose size was unknowable or infinite. Although the 'direct' strategy is primitive when compared to more 'indirect' forms of induction, such a strategy may work very well with small populations.

References

Evans, J.St.B.T. & Dusoir, A.E. (1977). Proportionality and sample size as factors in intuitive statistical judgement. *Acta Psychologica*, 41, 129 - 137.

Gigerenzer et al (1989). *The empire of chance*. Cambridge: Cambridge University Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In Oaksford, M. & Chater, N. (Eds.) *Rational models of cognition*. Oxford, UK: Oxford University Press.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin and Review*, 7, 569-592.

Keren, G. & Lewis, C. (2000). Even Bernoulli might have been wrong. *Journal of Behavioral Decision Making*, 13, 125-132.

Koehler, J.J. (1996). The base-rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, 19, 1-53.

Nisbett, R.E. et al (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339-363.

Osherson, D.N., *et al* (1990). Category-based induction. *Psychological Review*, 97, 185-200.

Rips, L. (1975). Inductive judgements about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665-681.

Sedlmeier, P. & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, 10, 33-51.

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

Wang, X.T. (1996). Domain specific rationality in human choices: Violations of utility axioms and social contexts. Cognition, 60, 31 - 63.

Acknowledgements