Inductive Reasoning Tasks Revisited: Object Labels Aren't Always the Basis of Inference Within Taxonomic Domains

Jonathan Loose (J.J.Loose@exeter.ac.uk)
Department of Psychology, University of Exeter; Perry Rd
Exeter, EX4 4QG, UK.

Denis Mareschal (D.Mareschal@exeter.ac.uk)[1]
Department of Psychology, University of Exeter; Perry Rd
Exeter, EX4 4QG, UK.

Abstract

This study is designed to investigate the predictions of a connectionist model of the development of inductive inference (Loose & Mareschal, 1997). We demonstrate that adults sometimes use perceptual as opposed to label information when reasoning about a taxonomically structured domain (biological kinds). Thirty six participants were taught the names of a set of tropical seeds. Participants believed that they were learning about real seeds, however the stimuli were constructed after the predictions of the model. Participants were taught that one seed had a particular non-perceptual property, and that a second did not. The task was to infer whether a third seed would have this property. In some cases, the third seed was given the appearance of one seed type, but the name of another. The results supported the model's prediction that participants would make *perceptually* based inferences in this condition ($N = 32$, $t=2.18$, $p<0.05$). These results stand in contrast to previous work using this experimental paradigm (e.g. Gelman & Markman, 1986). The results challenge previous interpretations of inference behavior to recognize that the use of perceptual information as a guide depends in part on the perceptual structure of the category in question, and is not simply explained by an appeal to conceptual representation in terms of causal "theory" structures.

Introduction

This paper on inductive reasoning reports a study investigating whether adult participants will make inductive inferences on the basis of perceptual similarity as opposed to shared category labels when learning about new objects within a taxonomically structured domain (that of biological kinds).

Early work in cognitive development led to the view that children's representations move from a phase of being perceptually based to being conceptually based. More specifically, studies of categorization and induction suggested that judgements regarding an object's category membership, or the likelihood of its sharing a property with another object, are made on a different basis depending on the child's age. Younger children apply a new fact to perceptually similar objects, whilst older children or adults utilize more profound conceptual information. Thus this traditional view claims that the younger child is perceptually bound, and only after entering a subsequent stage of development can he or she utilize abstract, categorical information (Inhelder & Piaget, 1964).

More recent studies have challenged this notion, demonstrating an early competence in the use of conceptual information. For example, when perceptual similarity and category membership are pitted against one another in an inductive reasoning task, children as young as 3½ will respond on the basis of shared category labels when making inferences about non-perceptual properties. This label reliance occurs despite a highly salient perceptual similarity which would suggest a different inference (Gelman & Markman, 1986, 1987).

These more recent findings have been explained through an appeal to knowledge as being structured primarily as some form of theory (Carey, 1985, Keil, 1989, Murphy & Medin, 1986). A theory is a structure in which causal relations are primary. Development is then seen as consisting of changes made to causal connections between concepts. Since the names of objects in a taxonomically structured domain place them in categories which cover some essential similarity between members, we would expect a theory-based representation to weight object names strongly as compared with other kinds of information such as perceptual similarity. This is especially true for concepts taken from taxonomic domains like biological kinds.

Both the traditional and theory based view would predict that the mature cognitive system of an adult should utilize labels when making inferences in a taxonomically structured domain. It may be, however, that even in adults, a preference for one kind of information over another will also be affected by the perceptual structure of the categories involved. We seek to investigate this possibility here.

[1] Now at the Department of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

We have previously reported a connectionist model (Loose & Mareschal, 1997) designed to simulate the inductive inference paradigm used by Gelman and Markman (1986) in their investigation of the development of inferential reasoning. A prediction of the Loose and Mareschal model is that even adults may prefer to make inferences on the basis of perceptual features *as opposed to* category labels. There are two conditions for perceptually based reasoning as predicted by the model. First, categories with distinct perceptual prototypes compared with other represented categories will promote perceptually based responding. Second, categories with a narrow perceptual extension around the prototype will also promote perceptually based responding. These predictions support the previous speculation that perceptual similarity may have a significant impact on inductive inference despite other factors, such as the taxonomic structure of the domain from which the objects are taken. It also states explicitly what these perceptual constraints might be. This study is designed to investigate the model's predictions by examining the performance of adult participants on a task analogous to that of Gelman & Markman (1986).

Our study differs from theirs in a number of ways – one significant difference is that here the objects about which participants must make inferences have been created by the experimenter. Unlike other studies using novel stimuli (e.g. Florian, 1994), we require that subjects have some knowledge of the variance of the exemplars of the novel categories and thus include a training phase in the study. It is only once the various perceptual categories and object labels have been learned that an investigation of inference can begin. We are therefore in a good position to know the perceptual structure of the categories that are represented.

An important similarity between this study and previous ones is that the stimuli to be used are derived from what is probably the archetypal taxonomic domain – that of biological kinds. This is a further factor which would lead us to expect label-based inference . If information from this domain is represented in a primarily theoretical manner (and if any domain is represented in this way, then surely biological kinds is) then we should expect labels to be of critical importance in attributing non-perceptual properties to the objects. Use of stimuli from this domain therefore biases the study in favor of a label - based response, and against the predictions of the model.

In summary, our hypothesis (based on the model's predictions) is that given a set of objects that are taken from the domain of biological kinds and that have a perceptual structure with a distinct prototype and tightly clustered exemplars, adult participants will tend to base their inferences about non-perceptual properties on the basis of the object's appearance and not its label.

Method

Participants

Participants included 36 male and female students and young professionals from the University of Exeter, UK and surrounding locality. Participants varied in age from 16;0 to 25;4, with a mean age of 21;8. All participants were either in education (studying for A levels / degree), or graduates.

Procedure

The study consisted of three phases. First, a training phase in which participants were taught to name instances of each category. Second, a test phase in which participants were asked to make inferences about the newly learned objects. Third, a category test, in which participants demonstrated that the initial learning of object categories had persisted through the second phase. These phases are described in turn.

Training Phase In the training phase, participants were taught to name instances of each category. Five exemplars were randomly selected from five of the ten seed types, and both the image and label of each exemplar was presented to the participant. Once these had been studied, three further images, randomly sampled from the five were presented. The participant was asked to choose the appropriate label for each seed type. Initially, some mistakes were tolerated, however as the participant progressed through this phase, tolerance of mistakes was reduced. Performance at less than criterion level led to extra training. Once the participant had successfully completed the 3 item test of the first five seeds, exemplars were presented from the remaining five seed types.

This process of presentation/test continued until participants had seen, and been tested on all of the seed types. At this point the error tolerance was reduced, and the learning procedure repeated. The learning phase was completed once participants had been successfully tested on all seed types with zero error tolerance. Thus no subject was able to enter the next phase until they had learned all of the categories and had performed perfectly when tested.

Once a participant had completed this phase, we considered them to have constructed a perceptual category and associated it with an appropriate label for each of the seed types. The participant was congratulated on finishing the procedure (which typically took around 15 minutes), and was prepared for the next phase.

Test Phase Testing was the most important phase within the study. It consisted of asking participants to make inferences about the objects that they had just learned to recognize in the previous phase. It is in this phase that perceptual similarity and category label were sometimes pitted against one another in order to test for a bias towards perceptual / label information in inference making.

Participants were presented with two images of seeds, each of a different type. Beneath each seed was written a fact. The facts beneath the seeds were mutually exclusive. For example, a picture of a 'calliax' would be presented next to a picture of a 'davus'. Beneath the 'calliax' might be written "This calliax contains the chemical *auxin*". Beneath the 'davus', "This davus does not contain the chemical *auxin*". The reason for using the presence / absence of a particular attribute rather than separate attributes was to reinforce the either/or nature of the question that followed.

Once the participant had spent time studying these images and statements, a third image of an exemplar of one of the selected seed types was shown. Beneath this image was a question, asking the participant to infer whether the previously taught fact also applied to this new exemplar. The question might be, "Does this davus contain any auxin?".

Answers to these questions were recorded. In half of the trials, the label and image used in the question conformed to prior learning, thus the question "Does this davus contain any auxin" was written beneath a picture of an object looking like a "davus" as initially trained. In these trials, the inference question had a 'correct' answer. Both perceptual similarity and category label would indicate one particular answer to the question. In the example, the correct answer would be 'no', since the participant was previously informed that another davus did not contain any auxin. The reason for asking participants to make inferences when there was *no* conflict between perceptual and label information was to see whether participants would make appropriate responses when the question being asked did indeed have a correct answer. This meant that a participant's data could be excluded from the analysis if, for whatever reason, he or she was not making appropriate inferences.

In the other half of the trials, the perceptual information presented to the participant was taken from one of the seed types previously presented, and the label used in the question was taken from the *other*. Thus, to continue the previous example, the participant might be asked, "Does this davus contain any auxin?", however this time the image beneath which the question was written would be one looking like an instance of the previously trained 'calliax' seed type. Thus, the participant, had the choice to respond on the basis of perceptual similarity, or on the basis of shared category label. To the participant, this example looked like a davus, but was nevertheless correctly labeled 'calliax', and some judgement needed to be made about which of these sources of information was most significant in determining which other seeds it shared hidden (non-perceptual) properties with.

It is possible that participants might respond on a perceptual basis during the test phase because they had forgotten which label went with which category. For this reason the subjects were again tested on their knowledge of the categories *after* the test phase.

Category Test In this final phase, participants were shown an image depicting an exemplar of each seed type which they had learned in the first phase. They were required to label each one consistent with learning in that phase. During the second phase, participants had seen objects looking like one object, but with a different label. This process may have upset the original learning, and if it had, then we would expect errors in this final phase. Thus, data can be excluded from participants whose learning was deemed to have been confused by the inference questions in the second phase.

Apparatus

The apparatus used consisted of a computer program written specifically for the purpose of the study, and a set of stimuli drawn and labeled to be plausible types of tropical seed.

Computer Program for Stimulus Presentation The task demanded a learning procedure with three parts: (1) Images and names must be taken from a pool and presented to participants, (2) participants' responses must be recorded, and (3) future presentations of stimuli must be dependent on these responses. Such interactive presentation suggested a computer based approach. The " JavaScript" language was used. An overview of the program's operation follows, providing a detailed account of the training and testing. Initially, instructions were presented to explain the program's operation and reinforce the participants' belief that they were about to learn about real biological organisms

Having read and understood the instructions, participants entered the training phase described above. The algorithm used by the computer during this phase was as follows:

1. Set the number of required correct answers to 1(out of 3)
2. Present in succession exemplars of five of the ten seed types, with labels.
3. Present pictures of three exemplars of categories taken from the five just trained.
4. If the number of correct answers given to (3) is below the required number, then go back to (2), else continue.
5. Repeat (2)—(4) with the other five categories.
6. If the number of correct answers required is 3, then go to (8), else go to (7)
7. Increase the number of required correct answers and go back to (2)
8. The participant has now answered all questions correctly... In order to make sure that all categories are fully learned, now test the participant's recognition of one exemplar from each category.
9. If (8) was accomplished without any errors, then the participant has learned the categories. If there were errors, then go back to (2).

The algorithm gradually tolerates less error in the participant's responses until the set has been well learned. The final testing phase ensures that all categories are well known.

Next, the program moved on to present sequences of two images of seeds, with facts about them written underneath. Having pressed a button to indicate that they should move on, the two images were replaced with a single test image, along with a question.

The algorithm by which the presentations / questions were constructed was as follows:

1. Randomly order the list of 10 properties.
2. Randomly select 5 of the properties, and mark ready for use in the "consistent" condition.
3. Mark the remaining categories for use in the "conflict" condition.
4. Randomly construct a list of pairs of seed types, with the constraints that each seed type must be used exactly twice, and that each pair must consist of two different types.
5. Take a pair of categories, and a property (initially starting at the top of the list). Construct the appropriate sentences about the two seed types.
6. Select an exemplar of each of the two seed types, and present them to the participant with the appropriate fact sentence printed underneath. Wait for the participant to confirm that we can move on.
7. Construct a question to be asked of a third seed type. Select (randomly) one of the two categories currently in use. If in the consistent condition then use *this* category label in the question. If in the conflict condition, then use the *other* category label in the question.
8. Display an exemplar of the selected category, and underneath print the question previously constructed (*Note: This will either contain the label of the displayed seed, or another label, dependent on condition.*)
9. Record the participant's answer to the question.
10. Go back to (5) above, and repeat until all properties have been used in asking questions.

Once the participant had moved through the 'test' phase, the 'final category test' phase was entered. A single exemplar of each category was presented sequentially, and participants were required to identify each one in turn using the same presentation steps as earlier. The number of correct responses was recorded.

Stimuli The stimuli used needed to satisfy a number of constraints. Firstly, participants should believe that what they see are representations of objects that actually exist in the natural world. Since natural kind domains are typically highly structured, this belief is designed to trigger reasoning in terms of category labels as opposed to perceptual similarity. Secondly, the objects should in fact be artificially created such that the perceptual structure of the category is carefully defined—exemplars should have a distinct prototype, and the category should have a narrow extension around that prototype.

Table 1: Dimensions of Stimuli

Dimension	Description
Elliptical.	x-axis / y-axis ratio
Split.	Presence/thickness of split.
Background Spots.	Presence / size of surface spots
Facetedness.	# faces, from cube - sphere
Colour	Range - light brown - deep red.
Texture	Roughness of the surface.
Tail	Presence and length of "tail"
Large lump	Seed is a composite object
Small lumps	# small spheres on surface
Spikes	# small spikes on surface

In order to satisfy these constraints, the domain of tropical seeds was chosen. An abstract object can be constructed on the basis of a set of perceptual dimensions, the object can be called a "tropical seed", and participants should agree that this is a genuine example. The word "tropical" is intended to prepare western participants for the fact that they won't have seen anything quite like this before. The program informs participants that the stimuli are taken from a "biological seeds database, © Weydan Labs, Cat kr9q". Such fictitious information enhances the participants' view that they are seeing veridical objects.

The dimensions used in the construction of the seed stimuli are shown in Table 1. Table 2 shows the names and prototype composition of each of the seed types. Prototypes were generated first by constructing a 3-D sphere, deforming its shape, adding texture, color and other features as required. Each prototype consisted of a single salient feature along with a number of other minor features. Exemplars were generated via small modifications of the amount of the principal feature present, coupled with larger varia-

Table 2: Composition of Different Seed Types

Name	Principal Feature	Other features
Pitrium	Elliptical	Large lumps, color, spikes.
Apriona	Split	Elliptical, background spots, faceted
Calliax	Background Spots	Tail, small lumps, split
Davus	Facetedness	Color, Background spots, small lumps
Mellimus	Color	Spikes, small lumps, large lump
Covis	Texture	Tail, small lumps, facetedness
Sictor	Small lumps	Texture, facetedness, split
Faevius	Tail	Background spots, elliptical, large lump
Glodia	Small lumps	Split, color, texture
Distores	Spikes	Texture, tail, elliptical

tions in the presence of minor features. Each of the stimuli had an immediately recognizable prototype, and the exemplars that were used did not vary far from it. One important kind of similarity to be manipulated was that of overall shape. Shape is a salient dimension for categorisation, and has been an important determinant of perceptual similarity in previous studies of this kind. Inspection of the stimuli by a number of colleagues and students confirmed that the stimuli were distinct and the variance low.

The seed types were given Latin-sounding non-words as names (Table 2). These names were intended to be distinct, memorable and short, as well as sounding like 'scientific' names.

Design

In the test phase, participants were taught contradictory facts about two seeds of different types, and were then asked to judge which of those facts applied to a third seed. In the 'consistent' condition, the third seed presented consisted of an image and a label taken from the same seed type. In the 'conflict' condition, the image was of a seed taken from one category type, but the label was that of the other type. All participants completed both conditions, which were randomized without replacement, each participant making five inferences in each of the two conditions with the total of ten inference questions being counterbalanced across subjects.

In the consistent condition, participants could make correct or incorrect inferences. If the participant claimed that the test seed shared its properties with the other seed with the same name / perceptual appearance, then the question could be considered to have been answered correctly. There is no basis in this condition for giving any other answer. The data provided by participants who did not score highly in this condition could be treated as suspect.

In the conflict condition, each participant's answer must depend on the information they consider to be most important in determining whether two seeds are to be considered similar. If perceptual information is more important, then the participant will assume that the test seed shares hidden properties with the similar looking seed. If label information is more important, then the participant will assume that the test seed shares hidden properties with the similar looking seed. Should the two kinds of information be equally important, then responding at chance level would be predicted.

The question of interest is not the difference in performance across the two conditions. Strong performance in the 'consistent' condition is merely a prerequisite for consideration of results in the 'conflict' condition. What is of interest is whether there will be a significant bias towards percept / label information in 'conflict' condition responses from those whose data are not excluded due to their performance in the 'consistent' condition.

Results

Experimental Context

Having listened to the instructions, and read them from the screen, participants indicated that they were ready and able to take part in the study. No participant questioned the explanation that they were interacting with a database held in the biological sciences department. On debriefing, participants were surprised, or even disappointed to discover that they had not been learning about real tropical seeds. We have no doubt that subjects took our study at face value, and assumed that they were learning about genuine natural kinds.

Retention of Learning

All subjects moved successfully from the first phase of the study, thus indicating that they had learned the categories appropriately. Of the 36 subjects, 71% made no errors at all in recognizing all seeds at the *end* of the study, indicating that the initial learning was both thorough and persistent through the inference phase. Of those that did make errors, only 3 made more than two mistakes. It would not seem appropriate to exclude subjects on the basis of only two misidentifications, especially since this might be attributable to a single confusion of two stimuli. Therefore, a criterion was set that subjects must score > 7/10 correct on the final category recognition phase for their data to be included in the final analysis.

Inference Behaviour in the Consistent Condition (Understanding of the Task)

Performance in this phase was good. As expected, subjects made few errors in inference when perceptual and label information were consistent. This implies that subjects generally understood the task. Those making more than a single error were excluded from the analysis.

Inference Behaviour in the Conflict Condition

Participants were asked to make five inferences about objects with conflicting perceptual / label information. Thus each subject could provide from 0 – 5 perceptually based responses. The mean across all subjects included in the study was 3.19 / 5 perceptually based responses. This indicates a significant bias in favor of perceptually based responding as compared with a response at the chance level of 2.5 / 5 ($N=32$, $t=2.18$, $p<.05$).

When we include those participants who did not meet the criteria for persistence of learning / accurate inference, we find an even stronger bias towards perceptually based responding (compared with chance), with a mean score of 3.3 / 5 perceptually based inferences ($N=36$, $t=2.87$, $p<.01$). The size of the effect actually increased when including all subjects. Thus the results cannot be explained in terms of the strategy for exclusion of a participants' data. The strategy actually *reduced* the size of the effect. These data support the hypothesis that appropriately manipulated perceptual categories will extinguish the dominance of labels as predictors of shared properties.

Discussion

The study was designed to bias participants towards a reliance on shared object labels in making inductive inferences, *against* the predictions of the model which we were seeking to support. This bias was provided primarily by participants' belief that they were learning about categories from a taxonomically organized domain, i.e. tropical seeds. Despite this, we have found that any label bias is effectively extinguished when considering suitably constructed perceptual categories. By suitable categories, we mean those which consist of perceptually similar exemplars with a prototype which is highly distinctive as compared with other represented perceptual categories. This result confirms the predictions of our model of the development of inductive inference (Loose & Mareschal, 1997). This is an important result, since it runs counter to the findings of previous inference studies of this kind (e.g. Gelman & Markman, 1986), and challenges the explanation of those findings.

Gelman and Markman showed their subjects pictures of two animals, teaching them mutually exclusive facts about each. Subjects were then shown a third animal looking like one of the first two, but sharing its category label with the second. They found that subjects reliably used *the shared category label* as a predictor of shared non-perceptual properties.

It is possible to reconcile our findings with theirs. The reconciliation is to be found in the choice of categories about which subjects must reason. In an example taken from their preliminary study with adults, subjects are asked to make inferences about the categories "bird" and "bat". The "bird" is a category, containing many subcategories (from garden birds to waders to birds of prey, etc). It contains everything from albatross to humming bird to emu to penguin to parrot. This is a category with *extreme perceptual variability*. It is also a category with a prototype which is not very dissimilar to the other category probed - "bat". Thus both of the model's conditions for appearance-based reasoning are violated in this example. This allows us to predict both the outcome of their study and ours on perceptual grounds. In our study, the predictions of the model are not violated, and the inference behavior is of the opposite kind. Their result fits our emerging picture nicely.

This reconciliation of the two studies leads to an alternative explanation of inductive inference – one in which inference behavior depends not on the assumption that concepts are represented in causally structured clusters – but rather assumes that all the available information will be exploited as far as possible. If a category consists of exemplars with a wide perceptual variability, then similarity of appearance is a weak predictor of shared non-perceptual properties inherent to the category. Thus the remaining ground for inference is category label. Alternatively, if a category is narrowly defined (perceptually), and distinct from other categories, perceptual information is better able to carry the burden of prediction of non-perceptual properties.

Finally we can consider the developmental story with which we began. The focus of the model is an understanding of the development of inference phenomena. This study demonstrates that the mature cognitive system is not restricted to reasoning on the basis of object labels even in domains which would seem to imply such reasoning – however the next step must be to examine *children's* performance on these tasks when perceptual factors are systematically varied. A study testing this is currently underway.

References

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

Florian, J. (1994) Stripes do not a Zebra make, or Do They? Conceptual and Perceptual Information in Inductive Inference. *Developmental Psychology*, **30**:1, **88-101**.

Gelman, S. & Markman, E. (1986) Categories and Induction in Young Children. *Cognition*, **23**, 183-209.

Gelman, S & Markman, E. (1987) Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, **57**, 1532-1541.

Inhelder, B. & Piaget, J. (1964) *The Early Growth of Logic in the Child.* New York, Norton.

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

Loose, J. J. and Mareschal, D. (1997). When a word is worth a thousand pictures: A connectionist account of the percept to label shift in children's reasoning. In M.G. Shafto & P. Langley, *Proceedings of the nineteenth annual conference of the Cognitive Science Society, (pp. 454-459)*. London: Erlbaum

Murphy, G. & Medin, D (1985) The Role of Theories in Conceptual Coherence, *Psychological Review*, **92**, 289-316.