

The Effect of Causal Strength on the Use of Causal and Similarity-based Information in Feature Inference

Rachel G. Stephens (rachel.stephens@adelaide.edu.au)

Daniel J. Navarro (daniel.navarro@adelaide.edu.au)

John C. Dunn (john.c.dunn@adelaide.edu.au)

School of Psychology, University of Adelaide
SA 5005, Australia

Michael D. Lee (mdlee@uci.edu)

Department of Cognitive Sciences, University of California, Irvine
CA 92697, USA

Abstract

Category-based feature generalisations are affected by similarity relationships between objects and by knowledge of causal relationships between features. However, there is disagreement between recent studies about whether people will simultaneously consider both relationships. To help resolve this discrepancy, the current study addresses an important difference between past experimental designs: the strength of causal relationships between features. Participants were trained on a set of four different kinds of artificial alien animals (with a known perceptual similarity structure), and were taught about three novel features. Participants were taught that either: 1) there were no relationships between the three features; 2) the features shared weak causal relationships; or 3) the features shared strong causal relationships. After training, all participants then made predictions about the features of the four kinds of animals. As expected, it was found that the strength of the causal relationships influenced the degree to which participants' feature predictions were affected by causal and similarity considerations. Three probabilistic graphical models were fit to the participants' predictions, in a preliminary effort to predict participant responses.

Keywords: feature inference; causal information; similarity; graphical models.

Informants of Feature Inference

Inferring or predicting unobserved properties of objects is something that we do frequently in our daily lives, often with little conscious effort. We make decisions about inductive problems such as, "Is this coffee too hot to drink?" or "Is it safe to pat this dog?". Feature inferences of this sort involve generalising existing knowledge in light of the observed properties of the stimulus before us. For example, we may think that this cup of coffee looks a lot like the extremely hot one that we were given yesterday, and so today's coffee is also probably too hot. Alternatively, we may consider that this coffee was made 20 minutes ago, so has probably cooled by now. These two coffee considerations can be thought of as examples of two different types of information that may be used when inferring a feature: similarity relationships between objects (i.e., comparing today's coffee with those previously

experienced) and causal relationships between features of an object (i.e., the coffee's temperature is causally related to the "age" of the coffee).¹

Experimenters have studied the use of each of these types of information separately. There is experimental evidence that people do consider similarity relationships between objects (e.g., Carey, 1985; Rips, 1975). For example, if told that a mouse has some novel enzyme X, people generally agree that a rat is more likely to also have this same enzyme than is a sheep. It has also been shown that people do consider causal relationships between features (e.g., Heit & Rubenstein, 1994; Rehder & Burnett, 2005), such as if an animal has wings, it can probably fly.

It will often be rational and useful to simultaneously consider both types of relationships. For example, imagine seeing a fairy penguin for the first time, and trying to predict whether it can fly. It has wings, so normally you would predict that it can fly. However, this particular bird looks rather similar to other penguins that you have seen before, which you know cannot fly. Therefore, it would be reasonable to predict that this new bird is unlikely to fly. (For a similar example about predicting lung cancer, see Kemp, Shafto, Berke & Tenenbaum, 2007).

It is quite easy to imagine scenarios such as this where the two types of information should be integrated. However, there is mixed experimental evidence as to whether people will simultaneously consider both similarity relationships between objects and causal relationships between features, when such information is available. Recent studies by Rehder (2006) and Kemp et al. (2007) find seemingly contradictory results. Rehder's experiments suggest that people use only one or the other information types, and that if people have causal knowledge about properties, this will often draw attention away from similarity information. Rehder trained participants on novel categories, such as

¹ Note that referring to relationships between objects as "similarity" relationships and relationships between features as "causal" is mostly for descriptive convenience: relationships between objects can often be causal too, such as phylogenetic relationships between animals. The key distinction here is relationships *between objects* versus relationships *between features of an object*.

“Kehoe Ants”, which were presented as lists of characteristic category features such as “Blood high in iron sulfate”. In Experiment 3, for example, when participants were told of a causal relationship between a characteristic (cause) feature and a novel (effect) feature, decisions of whether this novel feature could be generalised from one category member to another were largely unaffected by the degree of similarity shared by two category members. Rather, generalisation decisions were based primarily upon whether the causal feature was or was not present in the second target category member.

In contrast, Kemp et al. (2007) found evidence that people do incorporate both causal and similarity information. The experimental design was quite different. Participants were trained on causal relationships between three novel enzymes, such as “percidase”. Participants then made generalisation decisions about the presence of each enzyme in four real animals of varied similarity. Participants’ feature generalisations were best accounted for by a model that captured both the tree-structured similarity relationships between animals and the causal relationships between the three features.

One possible explanation for this discrepancy of results is the apparent *strength* or *certainty* of the causal relationships between features that were available for consideration during the experimental tasks. It has been shown that people’s feature generalisations are sensitive to the strength of relevant causal relations between features (Rehder, 2009). Causal strength may explain the discrepancy here because similarity information may be considered relevant to informing feature predictions only when there is uncertainty about any causal relationships between features. It seems that the causal relationships taught by Rehder (2006) were strong, *deterministic* relationships. Rehder explicitly told participants that the novel feature under consideration for generalisation “is caused by” one of the characteristic category features (e.g., the novel feature may be “Has a venom that gives it a stinging bite”, and participants were told that “The stinging sensation is caused by the high concentration of sulfate in the venom.”). Rehder’s experimental results suggest that participants may have interpreted these causal relationships as deterministic: to the extent that a causal feature was likely to be present, participants tended to predict that an effect feature would also be present. Consequently, when participants made the feature inferences in Rehder’s experiments, perhaps the similarity information was considered to be of little importance, because the causal relationships appeared certain (see Schulz & Sommerville, 2006).

In contrast, Kemp et al. (2007) taught weaker, *probabilistic* relationships. Participants were shown that the three enzymes tended to cause each other, but the causal relationships were not deterministic. It may be the case that participants then went on to combine the causal and similarity relationships in the subsequent test phase because there was uncertainty about whether the causal relationships would hold, and thus similarity information was still useful.

The aim of the current study is to demonstrate that the strength of causal relationships between features affects whether people integrate both similarity relationships between objects and causal relationships between features. The goal is to further our understanding of the conditions under which people integrate both types of information when making category-based feature inferences. Furthermore, this will help to resolve the different findings of Rehder (2006) and Kemp et al. (2007).

The plan of this paper is as follows. First, three models of feature inference will be presented and then the experiment and its results will be outlined. Finally the predictions of the three models will be compared with the experiment’s data.

Models

The predictions of three preliminary probabilistic graphical models of feature inference will be compared. The simplest model considers only the causal relationships between features. A second model considers only the similarity relationships between objects, and a final model combines both relationship types. These models are based on those used by Kemp et al. (2007). The models infer parameters that control feature transmissions within or between objects. Each model can then make posterior predictions that correspond to the feature inferences that people should draw if they are using the same structure as the model to inform their inferences. The models were implemented in WinBUGS (Lunn, Thomas, Best & Spiegelhalter, 2000) and are qualitatively described here. Further details of the models can be found in Lee & Wagenmakers (2009).

Causal Model

The causal model involves simple feature transmission along a causal feature structure. For this experiment, there are three binary features that are linked in a causal chain, so that the presence or absence of earlier features in the chain influences the presence or absence of later features (see Figure 1, right panel). The causal model (just like the participants) is given this basic causal chain structure, then learns three probability parameters that control the causal feature transmission process for each object. The presence or absence of the first feature for an object is determined by a base-rate γ . If this first feature is absent, then the second feature for the object becomes present with probability α . If the first feature is present, then the second feature is also present with probability β . Similarly, the third feature for the object is present with probability α if the second feature is absent, or with probability β if the second feature is present. These parameters are inferred from training observations of the presence or absence of the features for a set of objects. These are the same training observations that participants see during the experiment. The model “observes” one of the three sets of training data shown in Table 1 (see the Procedure section).

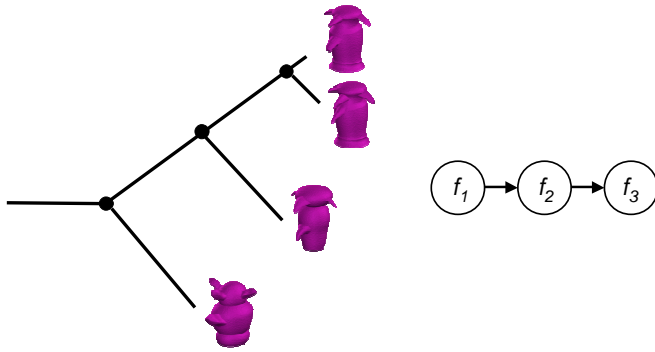


Figure 1: Left panel: the similarity relationships between the four “alien animal” objects, modelled as a taxonomic tree structure. The black lines are the “branches” of the tree, and the black dots are the “internal nodes”, which are both discussed below. Right panel: Causal chain relationships between three features, f_1 to f_3 .

Similarity Model

The similarity model captures similarity relationships between objects, using a taxonomic tree structure, as shown in Figure 1 (left panel). This tree represents the perceptual similarity between the four “alien animals”, but for real objects, such a tree could capture the intuitive similarity between objects, or even the shared evolutionary history of fauna or flora. In the similarity model, four objects lie at the terminal nodes or “leaves” of a tree, and a mutation process across the tree determines how the objects obtain or “inherit” their features. The idea is that a feature begins at the root node of the tree, then is transmitted down the tree, via internal nodes, to the objects. Along any branch of the tree, a feature can “mutate” or be switched from present to absent or from absent to present. As defined by Kemp et al. (2007), the probabilities of these switches depend on both the length of the branch between nodes, and on a mutation rate parameter. This model can produce the expected generalisation gradient for the “alien animal” objects in Figure 1, whereby if a feature is present for the topmost object, it will be predicted that the second (most similar) object will be more likely to share this feature than the third and fourth (more dissimilar) objects.

Integrated Model: Root Variables Hypothesis

The final model augments the first model that captures causal relationships between features with the similarity model that captures the taxonomic structure shared by objects. This integrated model is based on the idea that similarity relationships are used for predicting unobserved root cause features (f_i in Figure 1) in a causal structure. (This idea is similar to the *root-variables* hypothesis, suggested by Rehder; see Kemp et al., 2007.) In this model, the similarity tree-plus-mutation model is used to predict the presence of the first feature for each object and then the feature transmission chains from the causal model follow from these first features, to predict the second and third features. Note that this model is a simplified version of the

full integrated model used by Kemp et al. (2007), which integrates the taxonomic similarity and causal feature structure for the second and third features. The scope of this paper is limited to examining a simpler way of integrating the similarity and causal structures.

Experiment

In this experiment, the strength of causal relationships between features was directly manipulated, to test the effects on feature inferences. The main conditions of interest were the *weak causal relationships* and *strong causal relationships* conditions, which involve feature relationships that correspond to those used by Kemp et al. (2007) and Rehder (2006), respectively. There was also a *no relationships* condition that served as a baseline. The design of the current study was based on that of Kemp et al. (2007), with new extensions that are highlighted below.

Method

Participants. Participants were 62 people², recruited from the general community (30 males, 32 females). Ages ranged from 18 to 41 years. Participants went in a draw to win gift vouchers.

Materials. Stimuli were selected for this experiment with two goals in mind. The first goal was to match the objects and features as closely as possible in terms of both their novelty and inherent structure. Meeting this goal helped to ensure a fair test as to whether people would consider the similarity or causal relationships. While Kemp et al. (2007) used real animals, but novel enzymes for the features, the current study used both novel “alien” animals, and novel features. Images were used to present the animal objects, so that the objects had an immediate perceptual similarity structure. Three features were also chosen that were more meaningful than arbitrarily related novel enzymes, with the potential to share believable causal relationships, as described below. Thus the features, like the animal objects, also had some inherent structure. The second goal was to use a set of objects and a set of features that were both semi-realistic and as engaging as possible for participants, in an attempt to encourage sensible and ecologically valid feature inference behaviour.

For the novel animals, four “greeble” images were selected, as shown in Figure 1, which are made available courtesy of Michael J. Tarr (<http://www.tarrlab.org/>). The four greebles were selected on the basis of objective similarity data collected from an odd-one-out task (Stephens & Navarro, 2008), to have a tree-based similarity structure similar to that of the animals used by Kemp et al. (2007). The lengths of branches in the tree were found using counts of the times each greeble was chosen by participants as the

² Nine participants were removed from the weak cause condition, for failing to demonstrate an understanding of the causal relationships in the causal pre-test and the first simpler test phase task. This did not alter the overall conclusions.

Table 1: Counts of training charts shown to participants in each condition, to reinforce the causal relationships taught. In column 1, “1” marks the presence of a feature and “0” marks the absence, for the three features: f_1 = spanastete molecule, f_2 = oily surface substance and f_3 = pungent smell.

Training examples: $f_1 f_2 f_3$	Condition		
	No relationships	Weak cause	Strong cause
1 1 1	3	7	12
0 0 0	3	7	12
1 0 0	3	3	0
0 1 0	3	1	0
0 0 1	3	3	0
1 1 0	3	3	0
0 1 1	3	3	0
1 0 1	3	1	0

odd one out, for each triple shown during the odd-one-out task (see Figure 1, left panel). In the current study, the greeble images were presented to participants with the explanation that each image represents a different kind within a class of alien animals. For each participant, each greeble image was randomly assigned one of the names shown in Figure 2, to the left of the greebles.

Finally, three features were chosen that were novel, but could share sensical, uni-directional, causal chain relationships. The three features were: 1) whether a greeble has the spanastete molecule; 2) whether a greeble has an oily substance on its surface; 3) whether a greeble has a pungent smell.

Procedure. Participants were given a cover story asking them to imagine they were a biologist specialising in alien life forms. Participants were then trained on the greebles and features, and answered pre-test questions to check that the greebles’ similarity structure was noticed, and that the causal property structures were understood. Participants were first introduced to the “four different kinds” of greebles, then learned (by trial and error) to correctly name all greebles, in order to become familiar with the four images. Participants then answered four similarity pre-test questions.

Next, participants were trained on the three novel properties. This phase was where the experimental manipulation occurred, in a between-participants design. For all participants, the three properties were listed. Participants in the *no relationships* baseline condition were not given any further description of relationships between the properties. In contrast, participants in the *weak causal relationships* condition were given a description similar to that used by Kemp et al. (2007): “Your team has found that a greeble’s pungent smell is produced by several pathways. The most common pathway begins with greebles having the spanastete molecule, which can lead to greebles having an oily substance on their surface, which can in turn lead to





		spanastete molecule	oily surface substance	pungent smell
anavod		100		
ertese			0	
suzole				
macana				

Figure 2: Task 2 of the test phase. Participants were told that feature 1 was present for greeble 1, and feature 2 was not present for greeble 2. Participants then completed the remaining unknown cells of the object-by-feature matrix, using a confidence scale of 0 to 100.

greebles having a pungent smell.” Finally, participants in the *strong causal relationships* condition were given a description with phrasing to match that used by Rehder (2006): “Your team has found the cause of a greeble’s pungent smell. If greebles have the spanastete molecule, this causes greebles to have an oily substance on their surface, which in turn causes greebles to have a pungent smell.”

The causal property structure of each condition was then reinforced with training charts showing how often the three properties occur together (see Table 1). The charts were presented as “test results for a set of different kinds of greebles”. Participants saw a set of simple charts (one at a time in random order) that each displayed whether or not each of the three properties had been observed for a greeble. A different training set was used for each condition. For the strong cause condition, the presence of the three properties always “agreed” and for the no relationships condition, the properties agreed 50% of the time and were uncorrelated. The weak cause condition was intermediate, with properties 1 and 2 and properties 2 and 3 having a .71 probability of agreeing. Participants then answered three causal pre-test questions.

After training, participants completed three test phase tasks, the second of which will be the focus of this paper. Participants were asked to reason about the presence of the three features in the set of four greebles. Participants were told, “Each day, as new information is obtained about the greebles’ properties, your colleagues will ask for your opinion about all remaining unknown properties, for all four of the greebles.” Participants reported their feature predictions by completing an object-by-feature matrix, as can be seen in Figure 2. The greebles were presented in random order, but are shown in order of decreasing similarity in Figure 2. For Task 2, participants were told, “On Day 2 of 3, your team found that in addition to the

[name of greeble 1] having the spanastete molecule, the [name of greeble 2] did NOT have the oily surface substance.” The matrix cell corresponding to greeble 1 and feature 1 was thus completed with “100”, and the matrix cell for greeble 2 and feature 2 was completed with “0”. Participants completed the remaining empty cells. Participants’ predictions were reported as confidence ratings between 0 (“very unlikely”) and 100 (“very likely”), which were selected from a confidence rating bar on the screen.

All experimental tasks were completed individually on a personal computer. The entire experiment took around 15 to 20 minutes.

Expected Pattern of Responses

The different causal relationships learned in each condition should lead to different overall patterns of responses in the test phase task. First, to help explain the expected responses, consider what a sensible pattern of responses would have been if participants had known only that greeble 1 had feature 1 (i.e. if the presence of feature 2 for greeble 2 was unknown, rather than absent)³. Across all conditions, we would expect to see a standard similarity gradient for feature 1. Since the first feature was present for the first greeble, participants should have been most willing to generalise this feature to the more similar greeble 2, and least willing to generalise the feature to the dissimilar greeble 4 (as can be seen in the large white and grey bars for feature 1 in the first panel of Figure 3). However, if participants’ judgements were affected by the strength of the causal relationships between features, responses for features 2 and 3 would have differed across conditions: we would have expected a “causal gradient” across features that was appropriate for the causal strength that was learned. Firstly, in the no relationships condition, participants were shown that the features were uncorrelated. Therefore, regardless of whether the first feature was likely or not to be present for a greeble, these participants could only sensibly respond with a rating of 50 (the base rate of the features) for features 2 and 3, across all greebles. In contrast, in the weak cause condition, we would have expected to see the pattern of responses found in Kemp et al. (2007). Here, participants’ predictions for features 2 and 3 could be informed by the weak causal relationships between features. If feature 1 had been present for a greeble, feature 2 should also have been predicted as quite likely to be present, as should feature 3. However, due to the weak causal relationships, the probability of feature 3 being present would be smaller than that of feature 2, which would be smaller than that of feature 1 (similar to that seen in the large white and grey bars for greeble 1 in the second panel of Figure 3). Finally, in the strong cause condition, since the causal relationships were deterministic, we would have expected a flat causal gradient across features: to the extent that the root cause (feature 1) was present, the other

“effect” features should also be present, with the same probability, similar to Rehder’s results (2006).

What effects should we expect to see when participants are additionally told of the absence of feature 2 for greeble 2? With this additional knowledge, depending on condition, considerations of similarities between objects and considerations of relationships between features make some opposing predictions for greeble 1 and greeble 2. Firstly, for the no relationships condition, the main effect of knowing that feature 2 is absent for greeble 2 should be for predictions of whether feature 2 should be present for the most similar object, greeble 1. Knowledge that features 1 and 2 are generally uncorrelated should lead to a predicted rating of 50 for feature 2 for greeble 1, but knowledge that greebles 1 and 2 are similar should lead participants to predict that since feature 2 is absent for the second greeble, it should also be absent for greeble 1. Also, while knowledge of the (lack of) relationships between features should lead to a prediction of 50 for the first feature for greeble 2, similarity knowledge predicts that greeble 2 should probably have the first feature.

Secondly, for the weak cause condition, the additional knowledge that greeble 2 does not have feature 2 should also affect predictions for greebles 1 and 2. Knowledge of the weak causal relationships between features suggests that the three features should probably be present for greeble 1, but absent for greeble 2. However, the similarity shared by greebles 1 and 2 suggests that greeble 1 should not have feature 2 either, and greeble 2 should have the first feature (since greeble 1 has feature 1).

Thirdly, in the strong cause condition, since the causal relationships are deterministic, the known presence of feature 1 for greeble 1 should lead to very high confidence that features 2 and 3 should also be present for greeble 1. Similarly, the absence of feature 2 for greeble 2 should lead to very high confidence that features 1 and 3 should also be absent for this greeble. However, as in the other conditions, the similarity shared by greebles 1 and 2 suggests that greeble 1 should not have feature 2 either, and greeble 2 should have the first feature.

For each condition in this study, we can thus examine whether participants’ predictions seemed to rely on the similarities between objects, the causal relationships between features, or whether indeed participants integrated both sources of information and compromised between competing predictions. We expected that participants would rely more on similarity in the no relationships condition, where the presence of one feature for an object was uninformative about the presence of the other two features. We expected that in the weak cause condition, just as was found by Kemp et al. (2007), participants would integrate both types of information because the causal relationships were useful but uncertain, so similarity still contributed useful information. Finally, we expected that in the strong cause condition, we would reproduce the effect found by Rehder (2006), whereby the certain deterministic causal knowledge would dominate over similarity considerations.

³ This expected pattern of responses across conditions was generally demonstrated in Task 1 of the test phase.

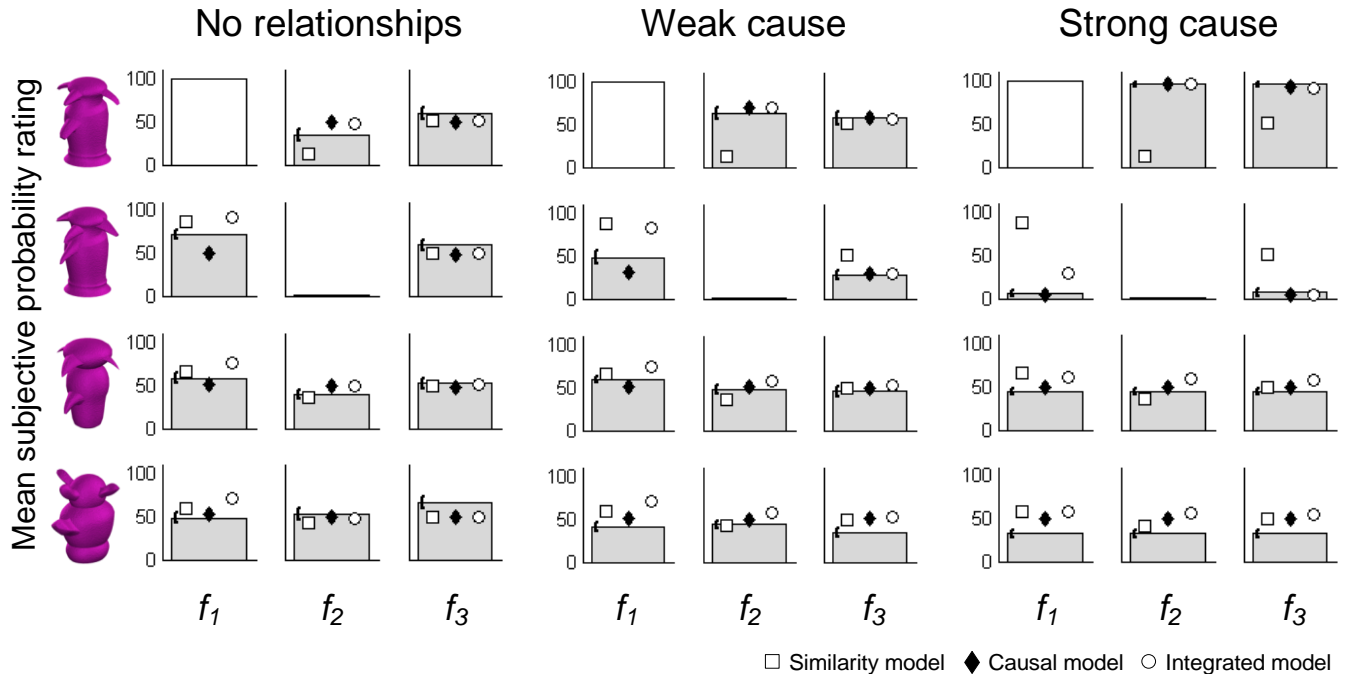


Figure 3: Test phase data and model predictions for the three conditions, presented in object-by-feature matrices: white bars represent the information given to participants (feature 1 is present for object 1, and feature 2 is absent for object 2), grey bars are mean participant responses and points are model predictions. The small black marks at the top left corner of each grey bar are error bars. Here the greeble objects are presented in order of decreasing similarity.

Results

Examination of Pattern of Responses. As can be seen in the grey bars presented in Figure 3, the experimental manipulation on the strength of causal relationships between features influenced participants' predictions. In the no relationships baseline condition, participants seemed to rely more on similarity information. As expected, since feature 1 was known to be present for greeble 1, participants in the no relationships condition predicted that feature 1 was much more likely to be present for greeble 2 than did participants in the other conditions. Similarly, since feature 2 was known to be absent for greeble 2, participants in this condition thought that feature 2 for greeble 1 was less likely to be present. However, the average participant response here *could* be a compromise between a rating of 50, as predicted by the (lack of) relationships between features, and a much lower “absent” rating, as predicted by the similarity between greebles 1 and 2. Nonetheless, the similarity information appeared to be relied upon the most in this condition.⁴

Now focussing on the strong cause condition, it is obvious that the deterministic causal feature knowledge dominated over similarities between objects, successfully reproducing the effect found by Rehder (2006). To the extent that one feature was present (or absent) for a greeble, participants

predicted that the other two features would also be present (or absent). Predictions for greebles 1 and 2 appear not to be tempered by the close similarity shared by those greebles.

Finally, participants' predictions in the weak cause condition appear to be somewhere in between that of the no relationships and strong cause conditions. Knowledge that feature 2 was not present for greeble 2 pulled down predictions of whether feature 2 was present for greeble 1 (compared to Test Task 1), but not as low as occurred in the no relationships condition. Similarly, predictions of whether feature 1 was present for greeble 2 were intermediate between the no relationships and strong cause condition. Participants do appear to have integrated both similarities between objects, and causal relations between features. This replicates the pattern of responses found by Kemp et al. (2007), but with the novel greeble objects and new, more meaningful features.

One last point of interest is that in the no relationships baseline condition, there was a trend that participants unexpectedly seemed to predict that the third feature should be present for the most dissimilar greeble 4. One possible explanation for this is that participants simply had a prior bias to expect that the “pungent smell” feature would be present for greeble 4. However, if this explanation is true, this bias was overridden in the other two conditions, which would need to be explained. An alternative, more interesting explanation, perhaps worthy of further investigation, is that participants saw that greeble 4 was different to the other greebles and thought it should also be differentiated in

⁴ This was also supported by a comparison with responses in Task 1 of the test phase.

Table 2: R^2 values for the correlation between average participant responses and predictions of the three models. The value for the winning model/s is shown in bold type. Negative signs in parentheses indicate where the correlation was negative.

	Similarity	Causal	Integrated
No relationships	.69	.06	.24
Weak cause	.04 (-)	.44	.42
Strong cause	.39 (-)	.92	.86

feature generalisations. With the absence of useful causal information, participants in this condition may have assumed that for feature 3 to be interesting in this small toy greeble world (and maybe even for feature 3 to be worth including in this study in the first place), feature 3 could be a unique feature of greeble 4.

Comparison With Model Predictions. R^2 values are shown in Table 2 for the correlations between the model predictions and the average participant responses. This is a simple assessment of whether the models capture the trends in the average participant responses. In the no relationships condition, the similarity model best captured the response trends, since this model could utilise similarities between objects, and produce similarity gradients for features 1 and 2 across greebles. The causal model could not use similarity information, and the current integrated model could not adequately adjust predictions around the known absence of feature 2 for greeble 2 (see Figure 3).

In the weak cause condition, the causal and integrated models performed equally well overall. However, as can be seen in Figure 3, the causal model missed the similarity gradient for feature 1, and the integrated model overestimated predictions for feature 1. The similarity model failed because it could not use the causal information. Kemp et al. (2007) found that their integrated model best predicted responses with the weak causal relationships that were used. However, the integrated model presented in this paper is simpler than that used by Kemp et al. (2007), so their full model must be implemented before direct comparisons can be made.

Finally, in the strong cause condition, the causal model accounted for participants' responses slightly better than the integrated model, and both performed much better than the similarity model. Again, the similarity model failed because it could not use the causal information. As shown in Figure 3, the causal model well captured the responses for greebles 1 and 2, but the integrated model again set predictions too high for feature 1 across greebles.

In summary, at present, the similarity model can best account for participants' feature generalisations when features are uncorrelated. If Ockham's razor is employed, at present, it seems that when there are weak or strong causal relationships between features, the causal model should be preferred over the more complex integrated model. However, as a complete test, the full integrated model by

Kemp et al. (2007) needs to be considered, as does performance across a range of tasks. For example, the causal model may have an advantage in this particular test Task 2 (for the weak and strong cause conditions) because the position of the absent feature reduces the contribution that can be made by similarity between objects for predicting the first feature.

Discussion

The results of this study suggest that the strength of causal relationships between features affects whether feature generalisations are informed by similarity relationships between objects, by causal relationships between features, or perhaps by both. When causal relationships are weak and uncertain, as in Kemp et al. (2007), people may base their feature inferences on both the causal information and similarities between objects. Alternatively, when causal relationships are deterministic, as in Rehder (2006), these relationships can be trusted, and similarity information becomes less important. Similarity information may be more heavily considered when the presence of one feature is uninformative about the presence of another, as when features are unrelated. Note that this (no relationships) condition used features that are closest to the "blank" features that are typically used in experiments to demonstrate similarity effects in feature generalisation (e.g., Carey, 1985; Rips, 1975).

It was possible that in the experiments by Kemp et al. (2007), people considered similarities between objects (beyond considering only causal feature relationships) only because real objects were used that were familiar to participants, and perhaps easy to reason about. However, this study helps to rule out that possibility, since participants seemed to consider similarities between objects as well as the weak causal relationships, despite the fact that the objects were novel and artificial.

Based on the present results, several directions for future research are apparent. The first concerns the nature of learning. In the present study, participants were trained on the objects and on relationships between features in separate phases. However, it is likely that people learn about real objects and relationships between their features in a more simultaneous (or at least alternating) fashion. Future research could study feature inferences after more naturalistic combined training of objects and features. Can people learn these structures simultaneously? Does this enhance or hinder the learning of each structure type?

Second, the present study used simple perceptual similarity relations between objects. Yet, people are also able to consider causal relationships between objects, when such information is available, such as food chain relationships between animals (Medin, Coley, Storms & Hayes, 2003; Shafto & Coley, 2003). Further work can attempt to incorporate these kinds of relations between objects into probabilistic graphical models of feature inference.

A third issue is that for people to use knowledge of causal relations between features, people must first understand and be aware of the relations. To what extent do people really understand and think about cause and effect with features of real objects? The experiments by Heit and Rubinstein (2004) suggest that people can reason about causal relationships between features for natural animals: for people to generalise a property from one category to another, the categories need to share the causal mechanisms responsible for the property. Participants were more willing to generalise an anatomical property, such as a “liver with two chambers that act as one”, from bears to whales than from tuna to whales. This was presumably because whales are more likely to share the responsible biological mechanisms with other mammals than with fish. In contrast, for a behavioural property, such as “travels in a zig-zag trajectory”, people were instead more willing to generalise from tuna to whales. This was apparently because whales are thought more likely to share a survival behaviour with tuna, another prey animal in the same ecology, than with bears. However, Heussen and Hampton (2008) found that when asked to explain features of natural kinds and artefacts, people often do not provide complete or thorough explanations. Expertise will be important. For example, Shafto and Coley (2003) demonstrated that in a feature induction task, fisherman utilised food chain relationships for decisions of whether a disease would generalise across species of marine creatures, but novices did not. Context can also activate knowledge of different causal relations, as when it was found that expert fire fighters made opposing predictions for the spread of fire based on either wind or terrain slope, depending on whether the task was presented as a bushfire to be fought, or a controlled back burn fire (Lewandowsky & Kirsner, 2000). How do people decide which causal relations are useful for the current task?

In experiments with novel causal feature relations, such as those by Rehder (2006), Kemp et al. (2007) as well as the current study, the causal relationships between features are explicitly presented to participants. Demand characteristics within the experiment probably encourage participants to use this information. These experimental designs can test whether people can use the causal information, and test the impact this has on use of other information, such as similarities between objects. However, there will still remain an open question for further research about when people will spontaneously consider particular sources of information, or combinations of sources.

There are many possible directions for future work, and the modelling approach used by Kemp et al. (2007) can be used to make precise predictions and help to test which types of information inform people’s feature generalisations.

Acknowledgments

RGS was supported by an Australian Postgraduate Award, DJN by an Australian Research Fellowship (ARC grant DP-0773794) and JCD by the ARC through grant DP-0877510. We thank Amy

Perfors and Charles Kemp for their helpful comments, and Michael J. Tarr, Brown University, for making the grebbles available.

References

- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press, Bradford Books.
- Heit, E. & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 411-422.
- Heussen, D. & Hampton, J. A. (2008). Ways of explaining properties. In V. Sloutsky, B. Love, & K. McRae (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 143-148). Austin, TX: Cognitive Science Society.
- Kemp, C., Shafto, P., Berke, A. & Tenenbaum, J. B. (2007). Combining causal and similarity-based reasoning. *Advances in Neural Information Processing Systems*, 19, 681-688.
- Lee, M. D. & Wagenmakers, E.-J. (2009). *A course in Bayesian graphical modeling for cognitive science*. Unpublished manuscript. <http://users.fmg.uva.nl/~ewagenmakers/BayesCourse/BayesBook.pdf>
- Lewandowsky, S. & Kirsner, K. (2000). Knowledge partitioning: Context-dependent use of expertise. *Memory & Cognition*, 28, 295-305.
- Lunn, D. J., Thomas, A., Best, N., Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.
- Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, 10, 517-532.
- Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition*, 34, 3-16.
- Rehder, B. (2009). Causal-based property generalization. *Cognitive Science*, 33, 301-344.
- Rehder, B. & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, 50, 264-314.
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665-681.
- Schulz, L. E. & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers’ causal inferences. *Child Development*, 77, 427-442.
- Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 641-649.
- Stephens, R. G. & Navarro, D. J. (2008). One of these grebbles is not like the others: Semi-supervised models for similarity structures. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1996-2001). Austin, TX: Cognitive Science Society.