Running Head: EXPERTISE AND CATEGORY-BASED INDUCTION

Expertise and Category-Based Induction

Julia Beth Proffitt[1], John D. Coley[2], Douglas L. Medin[1]

Abstract

This paper examines inductive reasoning among experts in a domain. Three types of tree experts (landscapers, taxonomists, and parks maintenance personnel) completed three reasoning tasks. In Experiment 1, participants inferred which of two novel diseases would affect “more other kinds of trees” and provided justifications for their choices. Experiment 2 used modified instructions, asking which disease would be more likely to affect “all trees.” Experiment 3 eliminated the conclusion category altogether, asking participants to generate a list of other affected trees. Among these populations, typicality and diversity effects were weak to non-existent. Instead, experts’ reasoning was influenced by “local” coverage (extension of the property to members of the same folk family) and causal/ecological factors. We conclude that domain knowledge leads to the use of a variety of reasoning strategies not captured by current models of category-based induction.

Expertise and Category-Based Induction

Cognitive psychologists are increasingly interested in conceptual functions beyond categorization (e.g., Barsalou & Hale, 1992; Markman, Yamauchi, & Makin, 1997; Pazzani, 1991; Ross, 1996, 1997; Wisniewski, 1995). Particularly, they have focused on the use of categories in reasoning and have proposed a number of formal models of category-based reasoning (e.g., Heit, 1998; McDonald, Samuels, & Rispoli, 1996; Osherson, Smith, Wilkie, López, & Shafir, 1990; Sloman, 1993; Smith, Shafir, & Osherson, 1993). These models have been developed to account for patterns of reasoning observed in experimental psychologists' favorite research population, American undergraduates. Whether these patterns, and hence these models, generalize to a broader population is open to question. In this paper, we are particularly interested in the effect of extensive domain knowledge on reasoning. To explore this issue, we examine expert reasoning on inductive tasks and evaluate the ability of existing models to account for experts' reasoning behavior.

In general, category-based induction requires that information about one set of categories be used to make inferences about another category. A set of premises establishes that one or more categories possess a certain property. The premises are followed by an assertion (the conclusion) that a target category also possesses that property. To ensure that such inference tasks do involve induction and not simply knowledge retrieval, researchers have typically employed novel, so-called "blank" properties. This puts the emphasis on reasoning about the categories. For example, people might be told that scientists have discovered a new disease that affects horses and be asked to judge whether it would affect all mammals.

Experiments frequently require participants to judge which of two complete arguments is stronger. Using this procedure, Rips (1975) observed typicality effects---inferences from typical category members were stronger than inferences from atypical category members (see also Osherson et al., 1990). For example, consider the following arguments:

(i) Robins are susceptible to disease A.

Therefore, all birds are susceptible to disease A.

(ii) Turkeys are susceptible to disease B.

Therefore, all birds are susceptible to disease B.

In the US, robins are seen as more typical than turkeys. Accordingly, when reasoning about unfamiliar properties, undergraduate subjects rate (i) as higher in inductive strength than (ii).

The diversity phenomenon involves inferences from pairs of premises. Consider the following pair of arguments:

(iii) Robins are susceptible to disease X.

Sparrows are susceptible to disease X.

Therefore, all birds are susceptible to disease X.

(iv) Cardinals are susceptible to disease Y.

Turkeys are susceptible to disease Y.

Therefore, all birds are susceptible to disease Y.

Cardinals and turkeys represent a more diverse set of birds than robins and sparrows do. For problems like this undergraduates usually rate arguments like (iv) as stronger; greater diversity among the premises leads to greater confidence in the general conclusion.

One very successful model of these effects is Osherson et al.'s (1990) similarity-coverage model (SCM). The SCM explains typicality effects in terms of coverage, defined as the similarity between the premise category and members of the lowest-level category that includes both the premise and conclusion categories. In evaluating arguments (i) and (ii), the average similarity (or coverage) of robin to sampled instances of the conclusion category bird is compared to turkey's coverage of bird. Because more birds resemble robins than turkeys, (i) is rated stronger than (ii). Thus, the typicality phenomenon results from the tendency of typical items to have high average similarity to other category members. Note that by this definition, good coverage corresponds to high central tendency: an item that is highly similar to other members of a category will have a high coverage score and will also be judged to be very typical. Typicality ratings have frequently been interpreted as estimates of central tendency, but see Barsalou (1985) and Lynch, Coley & Medin (in press) for another perspective.

The SCM also accounts for the diversity phenomenon in terms of coverage. In this case the principle is average maximal similarity; for each sampled instance, the most similar of the two premise categories gets entered into the calculation of average similarity. According to the SCM, argument (iv) is stronger than (iii) because cardinals and turkeys are highly similar to a wider range of birds than robins and sparrows are. In other words, in (iii) sparrow adds little coverage to robin because birds that are very similar to robins will also be very similar to sparrows. In contrast, turkey adds a fair amount of coverage to cardinal because turkeys are similar to large, edible birds (e.g., pheasants, chickens, ducks, etc.) that are quite different from cardinals. In brief, cardinal and turkey cover the category bird better than robin and sparrow. Thus, the SCM handles diversity effects with the same mechanism, coverage, that it uses to explain typicality effects.

Although this paper will be focusing on predictions made by the SCM, it is worth describing a second model that is also admirable in the level of detail it brings to the study of inductive reasoning. Sloman's Feature-Based Induction Model (FBIM, 1993) seeks to explain the same set of phenomena, but uses a connectionist mechanism. As just described, the SCM explains typicality and diversity effects by computing similarity between premise categories and sampled members of a more general category that encompasses both conclusion and premises. In contrast, the FBIM makes direct comparisons between features of the premise and conclusion categories. The argument with greater featural overlap between premise and conclusion is rated to be stronger. To decide between (i) and (ii) above, the FBIM would directly compare features associated with robin and turkey with those associated with bird. Presumably robin would share more features with bird, and so (i) would be a stronger argument than (ii). Although the calculation of argument strength for dual premise items is a bit less intuitive, it roughly amounts to the proportion of the conclusion category's features that the premise categories possess.

These models differ in terms of the precise mechanisms they use to explain induction, and in some cases make different predictions (see Sloman, 1993, 1998). However, they both rely heavily on the notion that similarity among categories--be it computed in a global way or based on featural overlap--is the principal mechanism driving inductive reasoning.[1] The importance of similarity in induction has garnered support from research with undergraduates from American universities (e.g., Osherson et al., 1990; Sloman, 1993, 1998; Smith et al., 1993). The generality of these findings has been implicitly assumed but rarely tested (see Choi, Nisbett & Smith, 1997, for an exception). However, recent findings with different populations have raised questions about the hegemony of similarity in induction.

López, Atran, Coley, Medin, and Smith (1997) studied category-based induction among the Itzaj Maya, an indigenous Amerindian population in central Guatemala. Among the Itaj, López et al. observed typicality effects in reasoning, but found no evidence of diversity effects. In fact, in some cases the researchers found reliable <u>negative</u> diversity, where the less diverse set of premises was preferred. This dissociation of typicality and diversity effects is surprising from the perspective of the SCM because it attributes both typicality and diversity to a common mechanism, coverage.

López et al. (1997) suggest that the knowledge base of the participants and the kind of properties used may account for the absence of diversity-based reasoning. In their study, López et al. asked the Itzaj, who subsist largely on hunting and farming, to make inferences about novel diseases affecting different kinds of mammals. In their responses, the Itzaj revealed a great deal of practical knowledge about local animal species. Their answers on the induction task frequently depended on arguments about mechanisms that might spread disease, rather than on similarity among categories. These results suggest that participants with extensive knowledge in a domain may employ a variety of strategies--involving causal knowledge of the properties and kinds involved as well as considering taxonomic similarity--when solving reasoning problems.

Thus, the findings of López et al (1997) raise important questions about the generality of models of induction that rely solely on similarity-based mechanisms. Of course, there are myriad cultural differences between the Itzaj and the undergraduates, which may account for the observed differences in reasoning. To evaluate the possibility that expert knowledge may lead to different patterns of inductive reasoning, the present studies focus on North American tree experts reasoning about trees. Like the Itzaj, this group has extensive category-relevant knowledge, but like undergraduates, they come from a mainstream American cultural background.

As an aside, some readers may object that the unfamiliar diseases used by López et al. (1997) did not constitute truly “blank” properties for the Itzaj. Their knowledge base may have enabled them to reason by analogy, using what they know about the epidemiology of real diseases in order to reach conclusions about hypothetical ones. A similar concern would hold for U.S. tree experts. We believe that this objection misses the point of what models of induction are trying to do, which is to account for how people go beyond their present knowledge to reason about novel properties, events, and categories. If people with knowledge in a domain extrapolate from what they know, then models of induction should describe the extrapolation process (see Heit, 1998, for an example). Surely inductive reasoning is not confined to knowledge-poor domains; rather, we expect it to be most powerful when used in conjunction with a rich knowledge base.

The present studies have two principal goals. First, we wanted to explore typicality- and diversity-based reasoning in another expert population in an attempt to expand upon and clarify the provocative results of López et al. (1997). It may be that their findings are due to some unidentified cultural artifact, and that U.S. experts will exhibit reasoning patterns quite similar to those of U.S. undergraduates. If so, we should expect that coverage will do a good job of predicting experts’ reasoning. Alternatively, experts may draw on a set of reasoning strategies completely unrelated to coverage. If extensive category-relevant knowledge leads to alternative content-based reasoning strategies, then tree experts may be more like the Itzaj Maya than like undergraduates. That is, coverage-based responding may be diminished or absent in tree experts, who may instead show causal/ecological reasoning.

A second question of interest was whether different kinds of tree experts would exhibit different reasoning behavior. Medin, Lynch, Coley, and Atran (1997) found different patterns of spontaneous sorting among three groups of tree experts: landscapers, taxonomists, and parks maintenance personnel. Landscapers differed from the other two groups by tending to sort in terms of goal-derived categories (good street trees, shade trees, specimen trees, etc.). Parks personnel and taxonomists used morphological properties, though the two groups differed in which properties were most salient. Medin et al. found that for parks personnel and taxonomists the sorting data was a good predictor of performance on reasoning tasks.[2] Given these findings, we were interested in the degree to which type of expertise affects inductive reasoning.

Based on prior work, we expect that individuals will draw on what knowledge they have to recruit a variety of reasoning strategies. We do not expect that a single type of strategy (i.e., coverage-based) will be the most natural solution for all problems. In addition, we expect that because experts vary in the knowledge they bring to bear on a problem, different kinds of experts will exhibit different reasoning patterns. Both outcomes pose potentially serious problems for existing models of category-based induction, which, although they can account for the diversity phenomenon, are unable to provide a satisfactory account for the diversity of reasoning strategies that experts bring to such tasks.

Experiment 1

The SCM explains both typicality and diversity effects in reasoning via the mechanism of coverage. Central to this paper is the question of whether coverage can adequately predict reasoning behavior across a variety of populations. López et al. (1997) found that these effects are not universal in human reasoning, but did not directly assess the role of coverage in Itzaj Maya induction. The goal of Experiment 1 was to assess the degree to which U.S. experts evaluated argument strength on the basis of coverage, as predicted by the SCM.

We use a forced-choice paradigm in which participants indicated which of a pair of arguments provided better support for a general conclusion. We presented participants with single- and dual-premise items (corresponding to typicality and diversity effects) to examine the impact of coverage on reasoning. In contrast to examples (i) - (iv) given above, participants were given two sets of premises, and asked which provided a better basis for generalizing to a common inclusive category. Specifically, we described two hypothetical diseases, which were known to affect different species of trees, and asked, “which disease would affect more of the other kinds of trees found around here?” The premise sets varied in their coverage of the category tree. We also asked experts to explain their reasoning; analyses of those justifications serves to illuminate the choice data. Of interest is the degree to which experts’ choices correspond to the coverage-based prediction, and whether types of experts differed in their patterns of choices and justifications.

To preview, we find that experts’ reasoning is not well predicted by coverage as construed by the SCM. Experts do not seem to be calculating coverage over the entire conclusion category (tree). Instead, it appears that experts view disease as spreading within smaller taxonomic groups, such as families of plants. As a result, they frequently base their reasoning on what we call “local coverage,” which roughly constitutes a form of coverage, but based on a subset of the conclusion category. To distinguish between these two senses of the term, from here on we will refer to coverage as described by the SCM (i.e., calculated over the category bridging the conclusion category and premises) as “global coverage.”

Method

Participants

Two women and 21 men (mean age 46 years, range 30 to 74 years old) having occupations related to trees completed Experiment 1. All were participants in an on-going research project investigating tree expertise (see Medin et al., 1997; Lynch et al., in press) and were paid for their participation. The average amount of experience dealing with trees was 22 years and ranged from 5 to 57 years. Seven participants had completed high school, four had some college work, nine had completed college, and three had advanced degrees (including two with doctorates). Almost all had some tree-related training.

Participants were drawn from a variety of sources, including the Morton Arboretum, several private tree maintenance companies, and two Chicago-area parks districts (Evanston and Skokie). Participants had a range of occupations, but earlier work with these tree experts (Medin et al., 1997) had identified taxonomists, landscapers, and parks maintenance personnel as showing distinct patterns of spontaneous sorting (see Medin et al., 1997, for details on differences among groups). Taxonomists are scientists (primarily working at institutions like the arboretum) who conduct research on trees, as well as doing some teaching and other educational activities. Landscape workers focus on selecting appropriate plants for installation based on the aesthetic and utilitarian characteristics of trees. Maintenance personnel are engaged in planting, removing, and generally caring for city trees. Although we use these groupings as a form of shorthand, this partitioning represents an oversimplification. In practice, the three categories may not be mutually exclusive.

For the purpose of this study, experts were classified based on their performance on two identical hierarchical sorting tasks, one run as part of a previous interview session with each participant, and a second performed immediately prior to experiment 1.[3] Both sorting tasks followed the same procedure, reported in full detail in Medin et al. (1997). Participants reviewed cards with the names of 48 trees printed on them. The set of trees was selected to be broadly representative of trees in the Chicago area and covering a broad spectrum of scientific taxa. Most were native trees, but some were common introduced species. The experimenter gave instructions to “Put together the trees that go together by nature into as many different groups as you’d like.” After sorting the cards, the participant provided labels or explanations for each of the groups created. The process (combining piles of cards and labeling the resulting category) was repeated until the participant was no longer willing to combine groups. The experimenter then recreated the piles resulting from the initial sorting and gave instructions to “Split as many of the groups as you’d like into smaller groups of trees that go together by nature.” Again, labels were offered for the resulting groups and the process was repeated until no further divisions made sense to the participant. This procedure yielded a different hierarchical taxonomy of trees for each expert. These sortings were used to classify the experts and yielded nine landscapers, eight parks maintenance personnel, and six taxonomists.

Materials

Materials consisted of 29 stimulus items, divided into 18 single-premise (SP) trials and 11 dual-premise (DP) trials (see Appendices A & B for a complete list). Each item consisted of a 4" x 6" index card with the name(s) of the tree(s) in the first argument printed on the left under the letter "A," and the name(s) of the tree(s) in the second argument printed on the right under the letter "B." Whether a tree or pair of trees occurred as set A or B was counterbalanced across participants. For all items, participants made inductions about an unspecified, and therefore novel, disease. Of central importance in this experiment is whether experts chose the stronger premise as predicted by global coverage. For each item, the A premise(s) and the B premise(s) differed in the degree to which they covered the category tree. Calculations of global coverage differed slightly for SP and DP items, as detailed below.

Calculation of Global Coverage for Single-Premise Items. We first estimated the psychological distance between each pair of trees using each expert's sorting data. This was done by counting the number of taxonomic levels that had to be crossed to reach the most specific category that included both trees. The results were then averaged across the two sorts to obtain a single distance matrix for each participant. The resulting pair-wise similarity ratings were averaged for each tree, yielding a single score for each tree reflecting its mean distance to the other 47 trees in the set. These were then averaged by expert group resulting in a single group score for each tree. This score provides an index of global coverage as defined in the SCM, based on experts' own sorting of the trees in question.

Calculation of Global Coverage for Dual-Premise Items. As for the single-premise items, this index of global coverage was based on the aggregated sorts for each of the three expert groups. However, the global coverage score for each pair of trees in the DP items was calculated using the minimum distance of tree 1 or tree 2 to each of the remaining 46 trees. Thus, global coverage for DP items was calculated using the minimum average distance (or maximum similarity) of the pair of trees to all other trees in the set. Again, this score provides an index of global coverage as defined in the SCM, based on experts' own sorting of the trees in question.

Note that because the three groups of experts had distinct sorting patterns, the calculated global coverage for a given pair of trees sometimes varied across groups. For SP items, the three groups perceived the same pair as having greater global coverage for half of the eighteen items. For the remainder, predictions differed by group (e.g., one premise was more typical based on landscapers' sortings, but the other was more diverse based on taxonomists' and maintenance workers'). For DP items, the three groups agreed on seven of the eleven items. The specifics are reflected in Appendices A and B; mean choices are entered on the same line as the tree (or pair of trees) with the greatest global coverage score for that group.[4]

Procedure

Participants were interviewed individually, usually at their place of employment. The reasoning task was presented along with a battery of other tree-related tasks, including tree identification from pictures and sorting. The reasoning task was always presented last in the session. Each participant was given all items, in random order. Phrasing was as follows:

> Suppose we discovered two new diseases that affect trees, Disease A and Disease B. All you know about these diseases is that Disease A affects Kentucky coffeetree, and Disease B affects river birch. Now, which disease do you think would affect more of the other kinds of trees found around here: A, which affects Kentucky coffeetree or B, which affects river birch?

For dual-premise items, each disease was described as affecting two types of trees. All trees were drawn from the set of 48 species used in the sorting task.

Asking about more trees represents a departure from the phrasing used in previous research (e.g., López et al., 1997; Osherson et al., 1990), which examined inductions to all members of a conclusion category. We were concerned that because, prima facie, very few diseases affect all trees, tree experts might question the credibility of our task and reject it as implausible. Thus, our formulation of the question focused on which premise set would generalize more broadly, rather than which one would be more likely to generalize universally (a patently implausible question given the background knowledge of our participants).

Once participants had indicated their choice they were asked to justify their answer. To minimize repetitiveness, the question was read in its entirety only for the first few items. After participants had become familiar with the question, they were simply shown the cards and asked for a choice and a justification. The full question was repeated occasionally to refresh their memories.

Results

For both SP and DP items, we first present group results, followed by an analysis of the kinds of justifications participants used to explain their choices.

Single-premise items

Global coverage. Each expert was given a score corresponding to the proportion of time they chose the premise with greater global coverage. Somewhat to our surprise, global coverage was not a good predictor of responses for any of the three groups. Table 1 gives the mean proportion of choices for the premise with greater global coverage, broken down by group. The grand mean was .45, marginally below chance ($t(22) = 2.04$, $p < .10$). In other words, a weak preference was shown for premises with less global coverage. This should be interpreted with some caution, as inspection of the individual probes in Appendix A reveals considerable variability across items. A one-way repeated measures ANOVA revealed no significant effect of expert type. Thus, we find no evidence that tree experts use global coverage to guide induction for SP items.

Family Size. While collecting the data, we were struck by the number of experts citing "family size" as a justification for their responses. The logic seemed to run along the following lines: disease is a property that spreads within a family of trees; therefore, the disease that infects the tree with the largest family will have the potential to infect more other kinds of trees.

To examine this observation more closely, we calculated family size for each tree in our set by counting the number of species found in the Chicago area that were members of the same scientific family as the premise tree.[5] For a number of items, family-size and global coverage make opposite predictions. This was particularly true for landscapers (10 items), somewhat less so for maintenance workers (6 items) and taxonomists (3 items). Responses to these critical SP items were rescored based on proportion of responses predicted by family size. Family size proved to be quite predictive of argument choice. For these items, the proportion of choices that agreed with family size (.72) was significantly higher than chance ($t(22) = 4.04$, $p < .05$). This held for both landscapers ($t(8) = 3.54$, $p < .05$) and maintenance workers ($t(7) = 3.55$, $p < .05$). A one-way ANOVA found no significant effect of expert type.

Insert Table 1 about here

Justifications. Experts were asked to justify their responses to each item. Justifications were classified into the following categories: (1) typicality, (2) family size, (3) distribution, (4) susceptibility, (5) resistance, (6) mechanism, (7) other, and (8) don't know. The first two types may be seen as broadly similarity-based. Typicality refers to an overt indication that the tree in one premise is more "typical" that the other. Although few justifications drew on this rationale (only 8) the clear *a priori* relevance of this category warranted its inclusion. Family size refers to justifications in terms of "more close neighbors" or "more trees in the same family." Although these usually focused on the genus or family level, there were a few justifications referring to categories as abstract as deciduous trees or evergreens. The other four strategies reflect thinking about the mechanisms involved in the spread of disease. Distribution refers to relative numbers of trees in the area (frequency) or to how widely distributed a species is (range). These two related aspects of tree distribution are collapsed because it sometimes proved difficult to distinguish between them. The strategies of susceptibility and resistance both concern how disease-prone the premise trees are, but with an important difference. Susceptibility addresses the situation in which the premise tree is very susceptible to disease, making it highly likely to catch, and therefore spread, a novel disease. In contrast, for resistance justifications, knowledge of how the tree responds to disease spurs inferences about the strength of the disease itself. For instance, the fact that a species like the gingko is highly resistant to disease might lead one to reason that if a disease manages to infect the gingko, that disease must be very virulent and therefore likely to spread. Mechanism justifications use analogies to a specific disease agent (e.g., "it must be like Dutch Elm Disease") or suggest a specific causal story about how tree diseases work. For example, a number of experts mentioned that big trees can produce a larger disease mass that could then spread more readily to other trees. Other was used for those justifications that could not be coded (e.g., "well worth saving") or that represented uncommon strategies. Finally, when an expert was unwilling or unable to explain an answer, it was coded as don't know.

Individual responses could be coded as involving one or more types of justifications. Experts produced 350 analyzable justifications (excluding other and don't know) for SP items. Two experimenters independently coded the justifications; differences were reconciled by discussion. Overall (for both single- and dual-premise items), the inter-rater reliability was 85.8%. Frequencies of each justification type are presented in Table 2.

Insert Table 2 about here

Data were analyzed by comparing the goodness-of-fit of the observed data to an expected distribution under a null hypothesis of no differences using chi-square. We compared the "justification profile" for each expert group separately to the null that each type of justification should be equally frequent (e.g., that 1/6 of landscapers' justifications are typicality, 1/6 are family size, etc.; other and don't know were excluded from analysis). Rejection of the null suggests that the frequency of some justification type(s) was disproportionately large or small for the expert group in question. Large deviations were defined as post-hoc cell contribution ratios larger than 2. For each expert group, the observed frequency of justifications differ from the null expectation (χ^2s(5) > 43.05, ps < .0001). In line with the choice data, typicality was disproportionately rare for all groups, and family size was disproportionately common. Distribution was also relatively frequent for all groups, and resistance relatively infrequent. Susceptibility justifications were frequent for landscapers and maintenance workers, but infrequent for taxonomists. Finally, mechanism justifications were infrequent for landscapers and taxonomists.

Because of the importance of similarity-based coverage as an explanatory mechanism, we further investigated justifications by collapsing categories into similarity-based explanations (typicality and family size) and causal/ecological explanations (range, mechanism, distribution and susceptibility). We lumped together justifications invoking typicality and family size under the similarity heading because these kinds of justifications could be based on some calculation of similarity-based coverage. In the case of typicality, it might be that a tree's similarity to all other trees (its central tendency) is important. For family size, it may be that the number of highly similar matters. We compared the observed distribution of justifications into these collapsed categories to a hypothesized null distribution of 1/2 similarity-based justifications and 1/2 causal/ecological. All three groups drew on more causal/ecological justifications and fewer similarity-based justifications than expected, (χ^2s(1) = 21.12, 30.81, and 4.65 for landscapers, maintenance, and taxonomists, respectively, ps < .05).

To examine individual patterns of explanation, we classified each expert as consistently giving similarity-based (i.e., typicality and family-size) or causal/ecological (distribution, mechanism, resistance and susceptibility) justifications, or as being inconsistent.[6] Fourteen experts consistently gave causal/ecological explanations, whereas only 3 consistently cited similarity-based ones. Six experts showed no consistent preference. Again, this pattern was compared to a null uniform distribution (one third in each category) using chi square goodness-of-fit, which revealed that consistent similarity-based patterns were less frequent than expected, and causal/ecological patterns relatively more so, χ^2 (2, N = 23) = 8.43, p = .015.

In summary, patterns of justifications support the choice data by highlighting the importance of non-similarity based strategies in experts' reasoning. They also reveal some differences between experts with respect to the kind of information deemed most relevant to the problems at hand.

Dual-premise items

Global coverage. The dual-premise probes are the most relevant to López et al.'s (1997) failure to find diversity responding among the Itzaj Maya. For each expert, we calculated the proportion of choices in accord with global coverage, and compared the group means to chance (.50). Means are presented in Table 1. The overall mean (.534) was not significantly different from chance. Taxonomists' choices matched global coverage predictions more often than expected by chance ($t(5) = 2.79$, $p < .05$), and landscapers' choices were marginally higher than chance ($t(8) = 2.23$, $p < .06$). Maintenance workers responses did not differ from chance ($p = .35$). A one-way repeated measures ANOVA revealed a significant effect of expert type, $F(2,20) = 3.64$, $MSE = .019$, $p < .05$; Tukey post hoc tests revealed that taxonomists' choices fit with global coverage more than maintenance workers' did (HSD = 0.188, $p < .05$). This represents a partial replication of the López et al. (1997) finding; diversity-based reasoning, as indexed by global coverage, was only present among a subset of this expert population.

Family size. We again examined the impact of family size by computing the proportion of times each expert chose the premises with greater family size. Although the choice predicted by global coverage sometimes differed from the choice predicted by family size, this occurred quite infrequently. Therefore, we did not restrict the analysis to critical items. For dual-premise items, family size was computed by summing the number of trees in the one or two families represented by the two trees in a pair. Overall, family-size choices were higher than chance, $M = 0.575$, $t(17) = 4.51$, $p < .05$. An ANOVA indicated no reliable differences among groups with respect to choices based on family size. Individual tests showed that only taxonomists' choices differed reliably from chance ($t(5) = 5.83$, $p < .005$). In short, responses on dual-premise probes accorded slightly better with family size than with global coverage, but the nature of the probe items precludes a strong test.

Justifications. Justifications were coded using the same categories as for SP items, except that diversity replaced typicality. Justifications falling into this category explicitly used the term "diversity" or mentioned that one pair was less closely related or more different. Experts produced

228 analyzable justifications for DP items. Frequencies of different justification types are presented in Table 2. Justifications were analyzed as described above, using chi square goodness of fit tests under the null that justification types should occur with equal frequency. Profiles for each expert group differed from this expectation, (χ^2 (5) > 23.90, ps < .0003). Group profiles diverged more than for SP items. For landscapers, justifications involving susceptibility were disproportionately high, and those involving mechanism and resistance were disproportionately low. For maintenance workers, justifications involving mechanism, distribution, and susceptibility were high, and those involving diversity and resistance were low. Finally, for taxonomists, justifications involving diversity and distribution were high, and those involving mechanism, susceptibility and resistance were low.

We further investigated justifications by collapsing categories into similarity-based explanations (diversity and family size) and causal/ecological explanations (distribution, mechanism, resistance, and susceptibility). We compared the observed distribution of justifications into these collapsed categories to a hypothesized null distribution of 1/2 similarity-based justifications and 1/2 causal/ecological. Both landscapers and maintenance workers again used more causal/ecological and fewer similarity-based justifications than expected, χ^2s (1, Ns = 95, 67) = 11.46, 25.09, ps <.001. For taxonomists, this pattern was reversed. Although the distributions of justifications differed from the null for taxonomists (χ^2 (1, N = 66) = 3.88, p = .049), the deviations from expectation were not large, with post hoc cell contribution ratios less than 2.

Finally, to examine individual patterns of explanation, we again classified each expert as consistently giving similarity-based or causal/ecological justifications, or as inconsistent. Twelve experts consistently gave causal/ecological explanations, whereas only seven consistently cited similarity-based ones. Of the latter seven, four were taxonomists, two landscapers and one maintenance. This did not differ reliably from a null uniform distribution, (χ^2 (2, N = 23) = 4.26).

A look at the individual items in Appendix B reveals some of the variability across items and expert types as well as something of the role of strategies in choices. For example, the pair involving paper birch and river birch versus white pine and weeping willow seems to be an obvious choice in terms of diversity. Indeed, all of the taxonomists choose the latter pair. However, all but one of the maintenance personnel made the opposite choice, noting that birches are widely planted and highly susceptible to disease.

Discussion

Although the typicality and diversity phenomena have proven robust among undergraduates, they were weak to nonexistent among the tree experts studied here. For single-premise arguments, we found little evidence of typicality. Global coverage was a poor predictor of argument strength, while family size emerged as a clear and strong predictor of choices. Moreover, responses were frequently backed up by explicit appeals to family size. Overall, however, response justifications favored causal and ecological mechanisms of disease spread over similarity-based considerations of coverage.

For dual-premise items, expert groups differed in the degree to which they exhibited the diversity phenomenon (i.e., the degree to which responses were predicted by global coverage). Taxonomists' choices agreed with global coverage (and hence diversity) more than expected by chance, and more than maintenance workers, who did not differ from chance. Landscapers' choices were intermediate. These group differences were mirrored by experts' justifications. Overall, causal/ecological justifications were as frequent as similarity-based ones.

These results suggest that in an expert population, the role of coverage in directing induction is greatly diminished, if not completely overshadowed, by knowledge of the domain in question. Indeed, global coverage, when defined as taxonomic similarity to instances of a shared superordinate, seems not to predict patterns of category-based induction for single-premise arguments and to do so only weakly for dual-premise arguments. For both SP and DP items, justifications demonstrate that the task evoked a great deal of relevant knowledge beyond categorical relations. In addition to (or instead of) categorical or featural coverage, informants often thought in terms of mechanisms (frequency and range of tree distribution, susceptibility to disease, resistance to disease, mechanisms of disease transmission) that influenced how well facts about particular tree species served as an inductive base. For example, one expert mentioned that oaks are likely to spread disease through their roots and that their extensive root system made oaks a stronger base for induction.

In contrast to global coverage, family size did seem to play a role in guiding expert inferences, especially for the single-premise probes. Numerous participants justified their choices on the basis of the size of the "family" to which the tree belonged, and indeed, family size emerged as a reliable predictor of both SP and DP inferences. Participants expressed a belief that a disease is more likely to spread from one kind of tree to others if it belongs to a larger family. Implicit in this logic is the belief that diseases tend to affect trees of a given family, but are perhaps unlikely to spread beyond that limit. We believe this finding represents a kind of “local coverage.” Experts may be using similarity to evaluate coverage, but the category they are considering is not <u>tree</u>. Instead, a more specific group is considered, the "folk family," which can map onto the scientific genus, family, or other non-standard groupings. Note too, that coverage is not “calculated” in the exact way predicted by the SCM. No sampling is required; rather, the similarity of members the same folk family is taken as a given. We term this mechanism “coverage” because the group with the larger extension is favored, and “local” because the relevant category is subordinate to <u>tree</u>. This is consistent with other work on induction which suggests that the folk generic level is privileged and that inductive confidence drops when inferences are extended beyond this level (Atran, Estin, Coley, & Medin, 1997; Coley, Medin, & Atran, 1997; Medin et al., 1997).

In sum, the results of Experiment 1 suggest that the inductive phenomena reported with undergraduate students do not hold as consistently for expert populations. Global coverage does not predict responses on single-premise items, nor does it uniformly predict responses on dual-premise items. Although we did find some evidence of “local” coverage, it is but one of a handful of strategies that were clearly evident. Before we consider the import of these findings, it is important to examine methodological differences between Experiment 1 and previous research.

Experiment 2

In Experiment 1, participants were given two premises (or pairs of premises), and were asked which disease would affect more kinds of trees. Previous research on category-based induction has asked about inferences to all members of a conclusion category (e.g., López et al., 1997; Osherson et al., 1990). As stated previously, we were concerned that experts might reject this conclusion as implausible because few if any diseases affect all trees. Still, we must acknowledge the possibility that asking about "more" rather than "all" trees might have changed the task enough to encourage use of local, rather than global, coverage. Because of the interest of this point to theory, in Experiment 2 we put our reservations aside and asked directly about which premises gave stronger reason to believe that a novel disease would affect ALL trees. The probe items were identical to those used in Experiment 1. Note that direct comparisons of Experiments 1 and 2 are hindered by the fact that all participants completed Experiment 1 before Experiment 2 (though the interval between experiments was about two years). Nonetheless, the findings would be of great interest if the difference in instructions led to a dramatic increase in the number of typicality- and diversity-based responses.

Method

Participants

Eighteen men having occupations related to trees completed Experiment 2. These included seven landscapers, six parks maintenance personnel, and five taxonomists. All were participants in an on-going research project investigating tree expertise (see Medin et al., 1997; Lynch et al., in press) and were paid for their participation. All had participated in Experiment 1; an additional 5 participants who participated in Experiment 1 were not available to participate in Experiment 2.

Procedure and Materials

Experiment 2 followed Experiment 1 by an average of 25 months. Participants were interviewed individually, usually at their place of employment. The reasoning task was presented along with a battery of other tree-related tasks, but was always given first in a testing session.

The materials and procedure were identical to that used in Experiment 1, with one exception. To more closely approximate previous research, participants were asked which of two diseases would more be likely to generalize to <u>all</u> trees. Specifically, for DP items they might be told:

> Suppose we discovered two new diseases that affect trees, Disease A and Disease B. All you know about these diseases is that Disease A affects horsechestnut and Ohio buckeye, and Disease B affects silver maple and amur maple. Now, which disease do you think would be more likely to affect <u>all</u> trees: A, which affects horsechestnut and Ohio buckeye or B, which affects silver maple and amur maple?

For SP items each disease would be described as affecting only one species of tree.

<u>Results</u>

Responses were analyzed exactly as in Experiment 1. For single- and dual-premise items, we first present analyses of choice data, followed by justifications. In each section, we also evaluate the effect of changing the instructions.

<u>Single-premise items</u>

<u>Global coverage and Family Size.</u> Overall, the proportion of choices made in accord with global coverage did not differ from chance, although taxonomists' choices were marginally above chance ($\underline{t}(4) = 2.25$, $\underline{p} < .10$). Experts did not differ reliably. Analyses of family size included only those items on which family size and global coverage made different predictions. Unlike Experiment 1, family size did not reliably predict responses to SP items overall, although maintenance workers were above chance ($\underline{t}(5) = 3.24$, $\underline{p} < .03$). Group differences, although marked, were not statistically reliable (see Table 1).

<u>MORE versus ALL.</u> Performance under MORE versus ALL instructions was directly compared using a 2 (Instructions: MORE v ALL) x 3 (Expert Type) mixed ANOVA on global-coverage-based responses. This analysis only included data from critical items for experts who

had participated in both experiments. Results revealed that global-coverage-based responses were lower for MORE instructions (M = .31) than for ALL instructions (M = .48),[7] $F(1,15) = 4.93$, MSE = .053, $p < .05$. Note however that this change represents a shift from above-chance preference for family size to chance performance.

Justifications. Justifications were coded and analyzed as in Experiment 1. Experts produced 318 analyzable justifications (see Table 2). For each expert group, the observed frequency of justifications differed from the null expectation (χ^2s(5) > 68.62, ps < .0001). Again, each group showed a unique profile. For landscapers, justifications based on family size, distribution, and resistance were relatively frequent, and those based on typicality, mechanism, and susceptibility were relatively rare. Taxonomists were similar in that justifications based on family size and distribution were relatively frequent, and those based on typicality, mechanism, and susceptibility were relatively rare. In contrast, for maintenance workers, justifications based on mechanism were frequent, and those based on typicality and resistance were rare.

As in Experiment 1, we further investigated justifications by collapsing categories into similarity-based explanations (typicality and family size) and causal/ecological explanations (range, mechanism, distribution and susceptibility). We again compared the observed distribution of justifications into these collapsed categories to a hypothesized null distribution of 1/2 similarity-based justifications and 1/2 causal/ecological. Landscapers and maintenance both used more causal/ecological and fewer similarity-based justifications than expected, (χ^2s(1, Ns = 130 and 104) = 35.57 and 49.85, respectively, ps < .0001). The distribution of taxonomists' justifications did not differ from the null (p = .76).

To examine individual patterns of explanation, we again classified each expert as consistently giving similarity-based or causal/ecological justifications, or as inconsistent. As in Experiment 1, the vast majority of experts (13) consistently gave causal/ecological explanations, whereas only 4 consistently offered similarity-based ones; a single expert showed no consistent preference. This pattern differed reliably from the null (χ^2 (2, N = 18) = 13.00, p = .0015)

MORE versus ALL. In order to assess whether the change in instructions had affected the pattern of justifications, we compared frequency of different types of justifications for those experts who participated in both experiments. Overall, when asked about ALL trees, experts provided reliably more mechanism-based justifications ($t(17) = 2.18$, $p = .04$) and reliably more resistance-based justifications ($t(17) = 2.37$, $p = .03$) than they had in Experiment 1. However, after Bonferroni corrections, neither of these differences remains statistically reliable, so they must be considered as marginal.

Appendix A reveals item differences associated with the change in instructions that are consistent with strategy reports. Note from Table 2 that landscapers showed a large reduction in susceptibility justifications and a large increase in resistance justifications relative to Experiment 1. In Appendix A landscapers also show a sharp drop in choices for river birch and paper birch (susceptible trees) and an increase in choices for the gingko (the most resistant tree in our sample).

Dual-premise items

Reasoning about DP items under ALL instructions was analyzed in the same way as for Experiment 1.

Global coverage and family size. Mean proportions of choices according to global coverage and family-size are presented in Table 1. Again, the overall mean agreement with global coverage (0.57) did not differ significantly from chance. The general pattern was strikingly similar to that noted in Experiment 1. Only taxonomists' choices differed from chance ($t(4) = 2.79$, $p = .05$). The ANOVA indicated only a marginal effect of expert type, $F(2,15) = 2.93$, $MSE = .028$, $p < .10$); the differences were in the same direction as in Experiment 1.

Responses in accordance with family size are also shown in Table 1. Averaged across expert type, family-size-based choices exceeded chance levels ($t(17) = 2.59$, $p = .02$). Both taxonomists ($t(4) = 6.16$, $p < .005$) and landscapers ($t(6) = 4.51$, $p < .005$) responded to DP items as predicted by family size at levels exceeding chance. Maintenance personnel agreed with family size less than 50 percent of the time. An ANOVA indicated that these group differences were reliable, $\underline{F}(2,15) = 11.83$, $\underline{MSE} = .018$, $\underline{p} < .001$. Tukey post hoc comparisons showed that both landscapers (HSD = 0.194) and taxonomists (HSD = 0.211) chose the premise with a larger family-size more often than maintenance workers did.

MORE versus ALL. Finally, to assess the effect of the change in wording across experiments, we compared choices based on global coverage and family size for the subset of experts who participated in both experiments. Separate 3 (Expert Type) x 2 (Instructions) ANOVAs were run on number of global-coverage- and family-size-based choices. No reliable differences as a function of instructions were found.

Justifications. Justifications were coded and analyzed as before. Experts produced a total of 200 analyzable justifications for DP ALL items (see Table 2 for frequencies). For each expert group, the observed frequency of justifications differ from the null expectation (χ^2 (5) > 68.62, $\underline{p}$s < .0001). Again, each group showed a unique profile. For landscapers, justifications based on diversity and resistance were relatively frequent, and those based on family size, mechanism, and susceptibility were relatively rare. Taxonomists were similar in that justifications based on diversity were relatively frequent, and those based on mechanism, and susceptibility were relatively rare, but unlike landscapers, resistance justifications were also rare for taxonomists. Maintenance workers presented a contrasting profile, in which justifications based on mechanism were frequent, and those based on family size were rare.

We again collapsed categories into similarity-based explanations (diversity and family size) and causal/ecological explanations (distribution, mechanism, resistance and susceptibility). As in Experiment 1, observed distributions were compared to a null 50/50 distribution for each expert type. Maintenance workers again used fewer similarity-based and more causal/ecological justifications than expected (χ^2 (1, $\underline{N}$ = 63) = 21.73, $\underline{p} < .0001$), while taxonomists used more similarity-based and fewer causal/ecological justifications than expected (χ^2 (1, $\underline{N}$ = 57) = 7.74, $\underline{p} < .01$).Landscapers' justifications were evenly split between the two categories (χ^2 (1, $\underline{N}$ = 80) = 0, ns).

To examine individual patterns of explanation, we again classified each expert as consistently giving similarity-based or causal/ecological justifications, or as inconsistent. Six experts consistently gave causal/ecological explanations, seven consistently offered similarity-based ones, and five were inconsistent. This did not differ from the null uniform distribution, χ^2 (2, N = 18) = 0.33, ns.

MORE versus ALL. Although choices of the global coverage-consistent pair did not increase significantly with the change in instructions, there was a noticeable increase in diversity-based justifications. Selective attrition may be a partial factor in this finding. Four of the five experts who were unable to participate in Experiment 2 had consistently offered causal/ecological explanations in Experiment 1. In order to assess changes in explanations attributable to instructions, we compared frequency of explanations for those experts who participated in both experiments. When asked about ALL trees, experts provided reliably more diversity-based justifications ($t(17) = 3.12$, $p = .006$) and resistance-based justifications ($t(17) = 2.52$, $p = .02$), but marginally fewer susceptibility-based justifications ($t(17) = 2.01$, $p = .06$). After Bonferroni corrections, only the difference in diversity-based justifications remained reliable, although the analyses presented above suggest that this increase in diversity-based justifications may have been largely confined to landscapers.

Discussion

In Experiment 2, we examined wording as a possible explanation for the discrepancy between the results of Experiment 1 and previous research. The results of Experiment 2 clearly show that this difference in wording only partially explains the absence of global-coverage-based reasoning in tree experts in the first experiment. Under the new instructions, global-coverage-based responses did not differ from chance for SP items. Generally, use of justifications did not differ between the two studies; asking about ALL trees still led experts to appeal to family size and causal/ecological factors. For DP items, responses to the ALL instructions were again driven more by family size than global coverage; no reliable differences between choices in Experiments

1 and 2 emerged. Moreover, justification profiles for maintenance workers and taxonomists were very similar in the two experiments. Thus, we can be confident that the pattern of results seen in Experiment 1 was not driven by the discrepancy in wording between Experiment 1 and previous research.

Nevertheless, there were some systematic differences in reasoning patterns between the two experiments. For SP items, global-coverage-based responses increased reliably, although as noted this increase was from a preference for family size to no preference. Another trend was for decreasing reliance on susceptibility and increasing reliance on resistance. This reflects the impact of domain knowledge on expert reasoning. A disease-resistant species may act as a strong premise because if it is susceptible it may mean that the disease itself is virulent and therefore likely to spread (see Smith et al., 1993, for a related observation and a model for this form of reasoning about properties). Finally, for DP items, the ALL instructions led to more diversity-based justifications, primarily among landscapers.

An increase in diversity justifications suggests that the salience of global coverage was increased by the ALL instructions. Nonetheless, even under ALL instructions, causal/ecological justifications largely predominated.

A look at the individual items in Appendix B suggests that specific knowledge about individual tree species might help explain why a large increase in diversity justifications was only accompanied by a modest increase in diversity- (i.e., global coverage-) based choices. This is seen most clearly for the landscape group, which showed the largest increase in diversity justifications. For items like the probe involving the river birch, paper birch pair, the ALL instructions led to more choices of the alternative with greater diversity. There are two notable exceptions to this trend towards choosing the more diverse pair. The first involves a contrast between the American mountain ash, green ash pair and the Siberian elm, tree of heaven pair. Medin et al., (1997) reported that a modest number of landscapers (and some maintenance personnel) treated the two ash trees as members of the same folk generic, even though only the green ash belongs to the Fraxinus genus. Therefore, some individuals might well choose the Siberian elm, tree of heaven pair as more diverse (which it is, according to science), even though their consensual sorting favored green ash and American mountain ash as more diverse. The second exception, which pits a hawthorn, linden pair against an amur maple, American hornbeam pair, may represent a related phenomenon. Note that the probe involving the hawthorn and linden is one of the tests where the maintenance consensual sort leads to a different prediction from the landscaper consensual sort. Medin et al. found that landscaper reasoning could actually be better predicted by maintenance consensual sorting than by landscaper consensual sorting. In short, landscapers may have been giving diversity justifications, but picking the more diverse pair according to a morphologically-derived taxonomy (not their own). The problem would be that their consensual sorting conforms to a goal-derived taxonomy. In any event, this analysis suggests less discrepancy between choices and justifications than might first appear.

Given how often causal domain knowledge was invoked, it is possible that the hypothetical disease, while novel, did not function as a purely blank property for our experts. To examine how the property might affect the results, we ran a brief follow-up study with a subset of eight experts (primarily maintenance workers) who had shown no diversity justifications in Experiment 1. This consisted of five dual-premise items involving mammals, five dual-premise items involving tree pairs and novel enzymes as the blank property, and five dual-premise items involving tree pairs and novel diseases. They were presented in the order just described with the goal of determining whether the property being projected would affect use of diversity and if diversity-related strategies could be "primed" by experimental manipulations. The mammal stimuli did not invoke diversity responding---tree experts used causal/ecological strategies for the mammal probes as well as the tree probes. We found that global coverage-based responses were reliably higher for tree-enzyme items than for tree-disease items (60% versus 35%), but even for tree-enzyme probes global coverage-based responding was not reliably above chance. Diversity justifications were also higher for tree-enzyme items than for tree-disease items, but causal/ecological justifications still predominated. We take these preliminary results as weak evidence that diversity strategies and reasoning may vary somewhat with the particular property employed, but not enough to challenge our pattern of findings.

Our results help to clarify López et al.'s (1997) findings that US undergraduates employ diversity when reasoning about familiar mammals, though the Itzaj Maya of Guatemala do not. Differences of culture or amount of expertise alone do not explain the discrepancy in Itzaj-US reasoning patterns; present results show that some US experts use diversity reliably whereas others do not. Thus, Americans are not uniform in their use of diversity, nor are experts uniform in their rejection of diversity (see also Coley, Medin, Proffitt, Lynch & Atran, 1999). Instead, almost all experts rely on a broader repertoire of strategies with the exact mixture varying even in response to small changes in wording.

We found that experts often used specific knowledge about trees and diseases to guide their inferences. For maintenance personnel, this was common enough to override any attention to information about taxonomic diversity. Landscapers and taxonomists also frequently used causal knowledge of this sort when solving the problems. Diversity effects may be stronger among undergraduates than among experts (and Itzaj Maya) because novices do not have the knowledge that would support these alternative strategies

Note the critical role that strategy assessment plays in our experiment. On the surface we observe diversity effects among some experts but not others and weak typicality effects, neither of which are consistent across type of expertise. This appears to be evidence against the use of global coverage, the primary explanatory component of the SCM. However, justifications reveal reliable use of <u>local</u> coverage (family size) on single-premise items and both local coverage and diversity on dual-premise items. Of course, we also obtain clear evidence for causal, knowledge-based reasoning on many items. Yet without the justification data, both the local coverage and causal/ecological strategies would have gone undetected.

In sum, Experiments 1 and 2 reveal that global coverage, as described by the SCM, plays a minor role, at best, in predicting experts' category-based inductions in the domain of trees. Instead, a strategy that we have called local coverage, which seems to synthesize ideas of coverage with beliefs about disease transmission, seems to be more important. However, our results show that experts have multiple reasoning strategies available, including causal/ecological ones. These strategies may differ in their salience as a function of type of expertise and, perhaps, the nature of the inference being made. In Experiment 3, we take another approach to examining these phenomena.

Experiment 3

Although experts generally appeared willing to make inferences to “all trees” as a conclusion category in Experiment 2, we still wondered how natural that was as a conclusion. In this experiment, we obtained a direct measure of property extension by asking experts for open-ended projections from premises. We presented experts with pairs of trees said to share a hypothetical disease, and asked them what other trees might be susceptible. We believe that this kind of inductive task---deciding how far to extend a property based on limited evidence---has greater ecological validity than the typical reasoning task, which involves choosing between two arguments in which the conclusion categories are explicitly stated.

This design also allowed us to assess whether the premise pairs were plausible. Results from López et al. (1997), as well as occasional comments from experts during Experiments 1 and 2, suggested that expert participants sometimes answer in terms of the plausibility of the premises rather than the strength of the argument. The more diverse the premises, we feared, the less plausible they might be to our experts. Thus, the very conditions that would be most likely to lead to diversity-based reasoning might be implausible to experts, in effect blocking the appearance of diversity as a reasoning strategy. If so, we would expect that for premise pairs that were extremely taxonomically dissimilar, participants would be more likely to express disbelief about the premises or perhaps would not extend the property to other trees.

A second advantage of this methodology is that it enabled us to obtain direct judgments of "coverage." If local, rather than global, coverage is frequently driving inductions, as suggested by

Experiments 1 and 2, then open-ended projections should reflect the importance of families as inductive units (see also Coley et al., 1997; Atran et al., 1997). Specifically, trees from the same family should license projection to the entire family, whereas trees from different families may license inductions to two different families rather than to a higher-order taxonomic class. By varying the similarity of the pairs of trees in the premise, we directly test this prediction.

Remember, however, that in the first two experiments, justifications invoking local coverage were overshadowed by those calling on causal/ecological explanations. Therefore, we also expected to see some non-taxonomic extensions. We were interested in whether the preponderance of causal/ecological justifications meant participants would extend a property to trees that were causally or ecologically, but not necessarily taxonomically, related. For example, an expert may generalize from river birch and paper birch, not to the birch family but to other water-loving trees or other trees with thin bark. This would still represent a form of category-based induction, only using non-taxonomic categories.

Method

Participants

Participants were recruited at the 1996 annual meeting of the Illinois Arborist Association, held in St. Charles, Illinois. Experimenters sat at a display booth in the exhibitor section of the conference, and asked passers-by if they would be interested in participating. The thirty-nine participants who completed the study were each paid a small sum. The typical informant was 40 years old (range = 23 to 71), had an Associates (N = 7) or Bachelor's degree (N = 18) and had 15 years experience working with trees (range = 1 to 46). Most participants completed at least one other task related to tree expertise.

Materials

Participants were given a booklet containing 10 different items. Each item contained two different trees, e.g., Austrian pine and eastern red cedar. (See Appendix C for a list of these items.) Pairs varied in terms of the taxonomic distance the two trees bridged. Items were presented in random order on separate pages in the booklet.

Procedure

Instructions included in each booklet directed participants that they would read about a new tree disease, but that the only information they would receive would be a listing of two trees “that can be infected by the disease.” They were told that their task was to decide what other kinds of trees would be susceptible to the disease, and to indicate why. Participants wrote their responses in the test booklet.

Results and Discussion

The responses were coded into one of five categories based on patterns of extension and the nature of justifications: (1) bridging to the lowest level category shared by the two species (e.g., for paper birch/river birch, “all birch trees would be affec extending beyond the lowest level linking category to a higher level shared taxonomic category or to all trees (e.g., for the same item, “willow, cottonwood, sycamore”---three trees that are linked at the level of class), (3) splitting the projection to separate classes related to each individual tree in the pair (e.g., for eastern white pine/weeping willow, “plants in the Salix and Pinus genus”), (4) projecting on the basis of ecological or physiological properties (e.g., "all thin-barked trees" or “any trees that prefer wet soil”), (5) giving no response, i.e., failing to project and/or questioning whether the premises could be true. Ecological and physiological were grouped because they represented a qualitatively different kind of projection from the other taxonomically-based categories. In addition, ecological and physiological points were frequently combined, making the categories difficult to separate. If an expert made multiple projections (e.g., "all oaks, or all deciduous trees, or all trees"), we counted only the most general one.

Recall that local coverage reflects a reluctance to compute coverage for the abstract category tree, but instead depends on a presumption that trees in the same lower-order category (e.g., folk-generic or folk-family) will share important properties. If local coverage is a salient reasoning strategy, then pairs of trees that are closely related should evoke responses that appeal to a common category (i.e., bridging or extended-taxonomic responses). In contrast, pairs of more distantly-related trees should evoke responses that appeal to each family individually (split), or perhaps lead to rejection of the premise as implausible.

The results are summarized in Table 3, averaged across pairs linked at the level of scientific genus (e.g., river birch, paper birch) versus pairs linked at higher levels in the scientific taxonomy (e.g., American elm, Colorado spruce). Proportions do not sum to one because not all justifications fell into the categories mentioned above.

Insert Table 3 about here

Overall, the pattern of projections is consistent with the idea of local coverage. As can be seen in Table 3, responses to genus-level pairs differed greatly from responses to higher-order pairs. Bridging responses were much more common for genus pairs, $t(38) = 8.79$, $p < .001$. In contrast, extended taxonomic responses ($t(38) = 2.67$, $p < .05$), splits ($t(38) = 6.06$, $p < .001$), and "none" responses ($t(38) = 2.60$, $p < .05$) were more common for higher-order pairs. For closely-related trees, participants overwhelmingly extended properties to the lowest-level superordinate---the folk family. In contrast, for more distantly related trees, participants extended the property to two disparate groups (split) as often as they bridged or extended, and sometimes even rejected the premise as implausible or refused to respond to the problem. Even when trees were as taxonomically dissimilar as possible (i.e., linked at the division level), experts extended the property to all trees only 25% of the time. Instead, as the linking level became more abstract, the experts were increasingly likely to question the premises or not extend beyond the examples.

As we saw in the earlier studies, many projections were based on physiological or ecological relations that do not fall along strict taxonomic lines. Experts depended on this type of projection equally for genus and higher-order pairs. These responses also fit with the ecological/causal reasoning reported by López et al. (1997). Apparently, experts' knowledge gives them access to reasoning strategies beyond those based on similarity and taxonomic relations.

Finally, these results parallel previous research suggesting that the folk generic level (e.g., maple, oak) is privileged for inductive reasoning. Coley et al. (1997, see also Atran et al., 1997) have shown that both US undergraduates and Itzaj Maya have strong expectations that biological entities from the same folk generic category share properties; when the bridging category is more abstract, however, inductive confidence drops sharply. The open-ended projection task used here provides a converging measure of inductive privilege. For trees linked at the genus level, bridging is the predominant pattern; when the pair of trees was linked at the level of order or higher, there was more projection from individual trees (splitting) to other members of the same folk generic. Local coverage is consistent with these findings in that it reflects an unwillingness to project a property beyond a limited set of closely related species.[8]

General Discussion

By asking tree experts to reason about hypothetical diseases, we have replicated some previous work on category-based induction and identified a number of challenges for future research. Throughout these experiments, examining justifications along with choices of relative argument strength proved to be crucial in understanding patterns of reasoning. In the following sections, we summarize our results and then consider their implications for models of inductive reasoning.

Summary of Results

This research developed out of two principal goals. The first was to examine inductive reasoning in U.S. tree experts in order to consider the relative contribution of culture and knowledge to observed differences between the reasoning patterns of U.S. undergraduates and the Itzaj Maya of Guatemala. Specifically, we were interested in whether global coverage, as specified by the SCM, did a good job of predicting expert reasoning. The second goal was to determine whether type of expertise led to differences in patterns of reasoning.

We would like to begin by highlighting four major findings that emerged. The first general result was that, overall, the experts' patterns of reasoning were more similar to those of the Itzaj Maya than undergraduates. Although experts occasionally used taxonomic relations to guide their inductions, reasoning about ecological relations among trees and other kinds of causal knowledge consistently played a more important role. Second, even when experts did use similarity-based taxonomy, they tended to rely on local, rather than global, coverage. Third, experts with different areas of specialization showed different patterns of category-based reasoning. Finally, changes in the instructions only led to subtle differences in patterns of reasoning.

A consistent finding throughout the three experiments was the use of causal, domain-specific knowledge about trees in addition to knowledge about taxonomic relations. The justifications revealed a high frequency of strategies based on explicit reasoning about mechanisms of property dispersion, such as range and abundance of tree species, susceptibility, physiological processes, and external agents. In the case of maintenance personnel this was the dominant strategy, but all groups employed causal/ecological reasoning with considerable frequency. Likewise, 20% of extensions in Experiment 3 were based on ecological/physiological considerations. Rather than viewing this as an "intrusion" of world knowledge into "pure" category-based reasoning, we see this as a primary phenomenon that models of category-based reasoning must address; relevant world knowledge is recruited to inform induction.

In Experiment 1, there was virtually no evidence of global coverage-based reasoning on SP items, though landscapers and taxonomists did use it some of the time for DP items. Instead, we found more consistent evidence of reasoning on the basis of family size, a phenomenon we refer to as "local coverage." Justifications highlighted the importance of causal mechanisms in experts' responses. However, while all expert types called on causal/ecological strategies to some extent, the degree to which they did so, as well as preferences for specific kinds of justifications, depended on expert type.

Despite the shift in conclusion category from "more kinds of trees" to "all trees," the results were fairly similar in Experiment 2. Phrasing the question in terms of all trees did not reduce the importance of causal/ecological reasoning, though it did yield shifts in which specific justifications were preferred. For SP items, changing the instructions increased global coverage-based reasoning to chance levels. For DP items the new instructions had little effect on choice patterns, but diversity-based justifications increased, especially for landscape personnel. Thus even under the most favorable conditions, causal/ecological explanations were as or more important than similarity-based ones.

The open-ended projection task in Experiment 3 provided further evidence of both causal/ecological reasoning and use of local coverage. Experts projected from closely related trees to a superordinate, but were much less willing to do so for distantly related trees.

These results may shed some light on the López et al. (1997) results with the Itzaj Maya. Like the Itzaj, our tree experts report causal/ecological reasoning strategies and like the Itzaj, some of our experts fail to show diversity effects. In some conditions the Itzaj show less than chance diversity-based responding. We see the same pattern for most of our experts on some of the probes. The justifications help explain this pattern of results. The experts are not choosing the less diverse option because it is less diverse, rather "negative diversity" is an artifact produced by the chance correlation of alternative reasoning strategies (e.g. susceptibility, distribution) with the less diverse pair. We suspect that this same factor is at work in the studies of the Itzaj Maya.

In short, coverage-based strategies are found among tree experts, but so also are causal/ecological ones. Typicality and diversity effects may be much more prominent among undergraduates than among experts or Itzaj Maya because undergraduates lack the knowledge needed to mediate ecological/causal reasoning. This contrast echoes the distinction made by Newell and Simon (1972) between strong and weak (general) problem solving methods. The causal/ecological reasoning strategies used by the experts are clearly knowledge-based and at least partly domain specific. However, while the coverage-based strategies may be more general, they are still dependent on knowledge about the similarity and taxonomic relationships among objects in the domain. These findings are also consistent with observations about the importance of theory in driving cognitive processes (Murphy & Medin, 1985; Wisniewski & Medin, 1994).

Implications for Models of Inductive Reasoning

Our results present three challenges for models of category-based induction. First, can they handle the local coverage phenomenon? Second, how do they explain differences among populations? And finally, can they account for the causal/ecological reasoning we observed? We now consider the implications of each of these issues for existing models.

Local Coverage

As noted, we did not observe strong typicality or diversity effects in Experiment 1. Instead, experts used local coverage based on the implicitly high similarity of species within taxonomic categories more specific than tree. Even when the instructions called for inferences to "all trees," as in Experiment 2, experts continued to rely more on local than global coverage. Under the open-ended instructions of Experiment 3, experts were reluctant to extend properties to trees outside the genus groups of the premise trees. It appears that for this population, the category tree is too abstract to be covered by a single premise.

Can current models account for the findings in Experiment 1? For models such as the SCM and FBIM, the first problem is how to deal with questions about "more" as well as "all" members of a category. Our concern that arguments of the latter form have limited ecological validity (in the real world, few properties are true of all members of a category) initially led us to ask which property would be true of more members of the category. This reframing amounts to a question about cardinality: Which of two arguments will generalize to more instances? But note that use of local coverage even under Experiment 2's "all trees" instructions suggests that this approach is not dependent on the "more trees" framing. Neither the SCM nor the FBIM could address this kind of problem without substantial modification.

Other models, such as McDonald et al.'s Hypothesis-based Model (HBM, 1996) and Heit's (1998) Bayesian model, may have an easier time accommodating questions of cardinality. The HBM takes a hypothesis-testing approach to induction, wherein the premises are considered data and the conclusion, a hypothesis. Argument strength is a negative function of the salience of alternative hypotheses. If experts hold a theory that tree diseases spread within the genus or family groups and no further, then (1) coverage with respect to the category tree is irrelevant to the problem, and (2) the hypothetical conclusions to each problem will involve the families of the trees mentioned in each premise. Potentially, the sizes of these can be compared easily to determine which disease would affect more trees. Heit's (1998) Bayesian model also provides a nice explanation of the local coverage phenomena. According to this model, novel properties will follow the same distribution as known properties. If the diseases that experts know tend to cluster in families, that same clustering phenomenon will be expected of a novel disease, resulting in what we call local coverage.

Group Differences

Neither the SCM nor the FBIM provides a natural account of group differences in strategy use. It is true that the models could posit that different groups or individuals use different metrics to compute similarity (or represent categories with different feature vectors). However, it seems less clear how each model might explain the finding that one group prefers coverage-based responses while another avoids them. Our results suggest that group differences arise because diversity is but one strategy among many available to experts, and these strategies may vary in their salience and appeal to different groups of experts.

Ecological/Causal Reasoning

This availability of various reasoning strategies appears problematic for any model that posits a single mechanism to determine inductive strength (i.e., coverage). Although the idea of competing strategies is compatible with the HBM, at a finer level of detail the implications are less clear because the HBM is less constrained. Indeed, by far the most problematic aspect of our data for current models is the presence of strategies based on ecological and causal mechanisms. Sloman (1994) demonstrates the importance of explanations in induction, and argues that an argument is strengthened if an explanation of the premise (i.e., why the premise property applies to the premise category) also applies to the conclusion, and weakened if different explanations come to mind. Some of our findings are consistent with this view. For instance, resistance-based justifications may represent explanations that apply to both premise and conclusion: a disease might infect gingko (normally a disease-resistant species) because it's highly virulent, and this also explains why the disease would be more likely to affect all trees. Note, however, that the majority of the causal/ecological mechanisms used by our experts were not based on any kind of similarity between premise and conclusion, be it taxonomic, featural, or in terms of shared explanations. Rather, they were based on causal/ecological links based on prior knowledge, computed over the premise and conclusion categories. For maintenance workers, inferences from two birches to all trees were stronger than from a pine and a willow not because the two birches are more similar in any way to all trees, nor because the same explanation (above-average susceptibility) applies to premise and conclusion, but because of a complex causal chain involving knowledge about susceptibility, distribution, varied habitat, and beliefs about epidemiology.

The frequency of these non-similarity-based inductive reasoning strategies suggests that existing models of inductive inference are not capturing a major component of inductive reasoning. Our goal is not to refute current models, but rather to point out that they are incomplete and to motivate the extension of these models to explain the kind of phenomenon we observe when experts reason in a domain of expertise. Real world knowledge informs induction in ways that transcend similarity-based accounts. We believe that this use of knowledge is not confined to experts, but is the rule rather than the exception in everyday reasoning. It is worth noting that even non-experts may be using such strategies some of the time. The standard experimental task---which involves forced choice or argument ratings, but no collection of justifications---will not reveal such strategies.

The overall picture that emerges from this work is one in which coverage constitutes one of many strategies that experts may use in category-based reasoning. Although it is not yet clear what triggers the use of one strategy versus another, our follow-up study with a subset of experts suggested that the property (e.g. enzyme versus disease) may play some role. We also suspect that the premise categories may play a role in potentiating different strategies. For instance, our informants were much more likely to mention susceptibility (or lack thereof) for trees that are distinctive with respect to susceptibility (e.g., the highly susceptible birch or the particularly resistant gingko) than for others. Similarly, they were more likely to mention distribution for trees that appear in a range of settings (e.g., “weed trees” like the boxelder). In short, we believe that category-based reasoning is far more variable than what would be licensed by models relying on a single strategy, uniformly applied.

It would be wrong to conclude that the variety of strategies we have noted and the differences as a function of type of expertise (and perhaps property) constitute a prescription for chaos. Rather, the range of justifications we observed is both relevant and normatively defensible. The most relevant thing to be said about the gingko with respect to disease is that it is extremely disease-resistant. Likewise, the boxelder is weedy and ubiquitous, and the Colorado spruce is often planted in sub-optimal growing conditions and therefore susceptible to disease. Nor is taxonomy irrelevant--we have analyzed the distribution of known tree diseases in the Chicago area and taxonomic distance is a good predictor of whether two kinds of trees will be susceptible to the same disease. Many known tree diseases do not cross genus boundaries (although some of the most common ones are widespread). Thus, experts' responses are driven by an attempt to draw on relevant knowledge. What is needed is a model of category-based induction that is similarly driven by relevance.

One possible objection to our results is that we did not use "blank" properties. Smith et al. (1993) suggest the following criterion for property blankness: “A rough test of whether a predicate is blank or not is whether it applies equally to all categories in a domain, or instead characterizes some categories better than others” (p. 69). For our tree experts, the property of a novel disease seems to have failed this test; based on the justifications provided, they clearly knew (or at least believed) that trees differ in their susceptibility to diseases. We have three responses to this concern.

First, it is crucial to point out that the principal reason to use blank properties is so that a priori beliefs about the veracity of a conclusion do not bias responses. Given our design, we do not believe that this is an issue. Although asking about hypothetical diseases clearly activated experts' causal background knowledge, experts could not have had a priori beliefs about whether "disease A" or "disease B" was more likely to affect more kinds of (or all) trees.

Second, while a novel disease may not be a "truly blank" property, it does not seem to be significantly less blank than properties used in previous research. For example, Osherson et al. (1990), produced typicality effects using the property "has a higher potassium concentration in their blood than humans" and diversity effects with "use norepinephrine as a neurotransmitter." To our knowledge, no one has argued that these properties are insufficiently blank, yet certainly creative or motivated individuals can still reason about them. Expectations about what kinds of things have potassium in their blood are probably influenced by beliefs about what kinds of things have blood, how they might be exposed to potassium, and why they might need potassium. Even the most extreme case, "has property X," may not be immune to such expectations. For instance, consider an argument with the premises "Grass has property X; Cows have property X." Careful consideration of the premises alone can promote certain expectations about this purportedly blank property (e.g., that it can be transmitted through digestion). Hence, even the most "blank" of properties can generate expectations about what kind of thing the property is and what kind of information is relevant. In turn, these expectations can guide what kind of reasoning strategies get used (see Heit & Rubinstein, 1994). When considering disease, our participants clearly engaged their theories about how disease spreads.

Finally, as mentioned earlier, we believe that it is necessary for models of induction to strive for greater ecological validity and that as part of that goal, they need to address how people use knowledge as a base for induction. Too much psychological research has attempted (and failed) to "build up" from artificial content-free experimental tasks.

Smith et al.'s gap model (1993) does take on the challenge of non-blank properties directly. According to this model, non-blank properties "potentiate" a subset of attributes. For example, the predicate "can bite through wire" activates features such as powerfulness and size. When the premises violate expectations about the predicate (e.g., "poodles can bite through wire"), the critical values for the relevant features are adjusted. In other words, if poodles can bite through wire, it must require less power and size than previously believed.

Although this model marks some progress, we think a number of extensions are needed. First, it is not clear how the gap model would account for blank properties. As noted by Sloman (1993) a model of category-based induction should be able to address blank and nonblank properties within the same general framework. Second, the gap model only provides for adjustments to expectations about the predicate. It is conceivable, however, that beliefs about the premise categories might be adjusted instead (e.g. "Poodles must be more powerful than I thought"). Third, the model only accounts for specific conclusions, e.g., German Shepherds, not dogs. Nor can it accommodate open-ended reasoning tasks, which may in fact constitute a good share of real-world reasoning. Our final concern may require a more fundamental rethinking of the model. The gap model assumes that the only element driving feature activation is the property. Yet multiple elements may determine which features are activated: (1) the property, (2) the categories in question (both premise and conclusion, if there is one), (3) the influence of knowledge and experience which may drive attention to certain elements, based on theories about the interactions among the first two components (property and categories), and (4) priming. We believe all of these factors may contribute to what knowledge is deemed relevant to a particular reasoning task in a particular context. We expect that computational solutions to these problems can be found but that they will require efforts to incorporate each of these factors.

Expertise and Reasoning

To what extent do experts and novices differ in terms of their reasoning? There is certainly ample evidence in the expert/novice literature to support such differences (e.g., Larkin, McDermott, Simon, &. Simon, 1980). Yet the picture here is somewhat more complex than a simple dichotomy wherein experts use one set of strategies and novices another. As discussed earlier, it is possible that the design of previous research may have masked the use of alternate reasoning strategies by novices. In addition, though similarity-based strategies were outnumbered by causal/ecological ones, we did find that experts used diversity and typicality some of the time. In fact, one expert group, the taxonomists, seemed to prefer such similarity-based strategies, at least under some conditions.

We believe that both experts and novices have a variety of reasoning strategies at their disposal. Typicality and diversity, as described by Osherson et al. are two strategies that are powerful because they may be used in a wide range of situations. To some extent, they are domain general, however they do depend on knowledge about the similarity relationships among categories. Even judgements of similarity are context bound, depending, at a minimum, on expectations about what kinds of similarity are relevant to the properties in question (Heit & Rubinstein, 1994). The use of local coverage requires a bit more expertise in a domain, since it depends on knowledge about the size of family groupings. The causal-ecological strategies exhibited by the experts are even more dependent on an elaborated knowledge base, however, and for that reason may be applicable only in certain domains. Indeed, some strategies may be so driven by domain knowledge that they may only apply in certain, specific instances.

Summary and Conclusions

In these experiments, we have shown that that experts use considerable domain-specific causal knowledge when solving induction problems in their domain of expertise. We have also shown that local coverage predicts patterns of inductive reasoning about trees and hypothetical diseases for tree experts better than does global coverage or overall central tendency. In addition, the use of diversity-based reasoning differs reliably among experts. Taken together, these findings challenge current models of inductive reasoning and make it clear that models which fail to take relevant knowledge and experience into account may have little to do with how people reason on a daily basis.

References

Atran, S., Estin, P., Coley, J. D., & Medin, D. L. (1997). Generic species and basic levels: Essence and appearance in folk biology. Journal of Ethnobiology, 17, 22-45.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 629-654.

Barsalou, L. W., & Hale, C. R. (1992). Components of conceptual representation: From feature lists to recursive frames. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), Categories and concepts: Theoretical views and inductive data analysis (pp. 97-144). San Diego: Academic Press.

Choi, I., Nisbett, R. E., & Smith, E. E. (1997). Culture, category salience, and inductive reasoning. Cognition. 65(1), 15-32.

Coley, J. D., Medin, D. L., & Atran, S. (1997). Does rank have its privilege? Inductive inferences within folkbiological taxonomies. Cognition, 64 (1), 73-112.

Coley, J. D., Medin, D. L., Proffitt, J. B., Lynch, E. B., & Atran, S. (1999). Inductive reasoning in folkbiological thought. In Medin, D. L. & Atran, S. (Eds.), Folkbiology. Cambridge, MA: MIT Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), Rational models of cognition (pp. 248-274). Oxford: Oxford University Press.

Heit, E. & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. Journal of Experimental Psychology: Learning , Memory, and Cognition, 20, 411-422.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. Cognitive Science, 4, 317-345.

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. Cognitive Psychology, 32, 251-295.

Lynch, E. B., Coley, J. D., & Medin, D. L. (in press). Tall is typical: Central tendency, ideal dimensions and graded category structure among tree experts. Memory and Cognition.

Markman, A. B., Yamauchi, T., & Makin, V. S. (1997). The creation of new concepts: A multifaceted approach to category learning. In T. Ward, S. Smith, & J. Vain (Eds.), Creative Thought. Washington D. C.: APA.

McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis-assessment model of categorical argument strength. Cognition, 59, 199-217.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? Cognitive Psychology, 32, 49-96.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. Psychological Review, 92(3), 289-316.

Newell, A. & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

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

Pazzani, M. (1991). Learning to predict and explain: An integration of similarity based, theory-driven, and explanation-based learning. The Journal of the Learning Sciences, 1, 153-199.

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

Ross, B. H. (1996). Category representations and the effects of interacting with instances. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1249-1265.

Ross, B. H. (1997). The use of categories affects classification. Journal of Memory and Language, 37, 240-267.

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

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

Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. Cognition, 49, 67-96.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 449-468.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. Cognitive Science, 18, 221-281.

Appendix A

Single Premise Items, Experiments 1 and 2

Item	Landscapers		Maintenance		Taxonomists	
	Exp. 1: "More"	Exp. 2: "All"	Exp. 1: "More"	Exp. 2: "All"	Exp. 1: "More"	Exp. 2: "All"
Kentucky Coffeetree	0.11	0.71	0.25	0.00		
River Birch †					0.50	0.80
London Planetree	0.44	0.57	0.25	0.67	0.33	0.60
E. White Pine †						
Gingko	0.22	0.57				
E. Black Alder †			0.63	0.50	0.67	0.80
Tuliptree	0.22	0.43				
Boxelder †			0.50	0.67	1.00	0.60
Grey Dogwood						
Star Magnolia †	0.44	0.43	0.50	0.40	0.17	0.60
Catalpa	0.33	0.71	0.13	0.00		
Paper Birch †					0.83	0.60
Am. Elm	0.22	0.14				
Norway Maple †			0.50	0.67	0.83	0.60

(table continues)

Black Cherry †			0.50	0.17	0.83	0.60
Little Leaf Linden	0.33	0.71				
White Oak †	1.00	0.71	0.63	0.67	0.83	0.80
Black Walnut						

	0.33	0.43	0.13	0.50		
					0.83	0.80
Honeylocust	0.89	0.86				
Ohio Buckeye			0.63	0.33	0.00	0.00
Hackberry	0.56	0.57				
Am. Beech †			0.38	0.17	0.33	0.40
Sugar Maple	0.44	0.17	0.71	0.83	0.17	0.60
Silver Maple						
Horsechestnut †	0.56	0.29	0.75	0.67	0.67	0.60
Catalpa						
Bur Oak †	0.11	0.29	0.63	0.67	0.33	0.80
White Poplar						
Colorado Spruce †						
Star Magnolia	0.22	0.71	0.75	0.20	0.33	0.60
American Mountain Ash	0.25	0.71	0.38	0.50	0.33	0.40
Colorado Spruce						

(table continues)

Eastern White Pine †						
Catalpa	0.44	0.86	0.00	0.17	0.17	0.60

The proportion of each group that chose the tree is entered for the trees with greater coverage.

Where family size made a prediction, (†) indicates which tree comes from the larger family.

Appendix B

Dual Premise Items, Experiments 1 and 2

Item	Landscapers		Maintenance		Taxonomists	
	Exp. 1: “More”	Exp. 2: “All”	Exp. 1: “More”	Exp. 2: “All”	Exp. 1: “More”	Exp. 2: “All”
Sib. Elm, Tree of Heaven					0.50	0.60
Am. Mtn. Ash, Green Ash †	0.67	0.14	0.75	0.33		
Amur Maple, Am. Hornbeam †			0.25	0.00	0.00	0.20
Wash. Hawthorn, L-leaf Linden	0.89	0.29				
London Plane, Sycamore	0.33	0.14				
Catalpa, Bl. Cherry†			0.38	0.50	1.00	1.00
Horsechestnut, Ohio Buckeye						
Amur Maple, Silver Maple †	0.78	0.71	0.75	0.83	1.00	1.00
E. Cottonwood, Wh. Poplar						
Am. Elm, Col. Spruce†	0.67	0.71	0.63	0.67	0.83	1.00
Norway Maple, Boxelder	0.22	0.43	0.38	0.67	0.33	0.20
Grey Dogwood, Crabapple						
Wh. Oak, Bur Oak						
Pin Oak, Star Magnolia†	0.67	0.86	0.38	0.33	0.83	1.00

(table continues)

Shgbrk Hickory, Blk Walnut						
Honeylocust, E. Bl. Alder†	0.78	1.00	0.38	0.17	0.83	0.80
Austrian Pine, E. Redcedar						
Scotch Pine, Gingko	0.22	0.86	0.13	0.50	0.50	0.20

White Birch, River Birch						
E. Wh. Pine, Weeping Willow†	0.56	1.00	0.13	0.17	1.00	0.80
Hackberry, Wh. Mulberry	0.44	0.57				
Am. Beech, N. Red Oak†			0.63	0.67	0.00	0.60

The proportion of each group that chose the pair of trees is entered for the pairs with greater coverage. Where family size made a prediction, (†) indicates which set of trees have more trees in their family (or families).

Appendix C

Items used in Experiment 3

Taxonomic relationship	Premise pair
Same genus	Eastern Cottonwood, White Poplar
Same genus	London Planetree, Sycamore
Same genus	Paper Birch, River Birch
Same genus	White Oak, Bur Oak
Same subclass	Catalpa, Black Cherry
Same subclass	Pin Oak, Star Magnolia
Same class	Scotch Pine, Gingko
Same order	Austrian Pine, Eastern Redcedar
Same division	American Elm, Colorado Spruce
Same division	Eastern White Pine, Weeping Willow

Author Note

[1]Department of Psychology, Northwestern University, Evanston, IL

[2]Department of Psychology, Northeastern University, Boston, MA

This work was partially funded by NSF Grant 9110145 and NIH Grant MH55079 to Douglas Medin. We thank Elizabeth L. Lynch, Connie Pankratz, Hillarie Schwartz, Lori Hansen, and Jessenia Magua for assistance collecting and coding the data. We also thank Evan Heit, Barbara Malt, Daniel Osherson, Trex Proffitt, Pat Shafto, and two anonymous reviewers for helpful comments on the paper.

Correspondence concerning this article should be addressed to Julia Beth Proffitt, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60207-2710. E-mail: juliabeth@nwu.edu.

Footnotes

Author Note

[1]Department of Psychology, Northwestern University, Evanston, IL

[2]Department of Psychology, Northeastern University, Boston, MA

This work was partially funded by NSF Grant 9110145 and NIH Grant MH55079 to Douglas Medin. We thank Elizabeth L. Lynch, Connie Pankratz, Hillarie Schwartz, Lori Hansen, and Jessenia Magua for assistance collecting and coding the data. We also thank Evan Heit, Barbara Malt, Daniel Osherson, Trex Proffitt, and two anonymous reviewers for helpful comments on the paper.

Correspondence concerning this article should be addressed to Julia Beth Proffitt, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60207-2710. E-mail: juliabeth@nwu.edu.

Footnotes

[1] Note that Sloman might object to our characterizing his model as similarity-based. "A second difference between the two models is that only the category-based model [referring to the SCM] assumes that subjects explicitly compute similarity. The feature based model does assume a feature-matching process, but not one that computes any empirically valid measure of similarity per se." (p. 233, 1993)

[2] Landscapers did not appear to use their goal-derived categories in reasoning; instead, their reasoning patterns were predicted best by parks personnel sorting. The Medin et al. (1997) tasks differed from those presented here in that they involved specific arguments (the premise and conclusion categories are at the same level of the taxonomic hierarchy) rather than general arguments (the conclusion category encompasses the premises).

[3]Several experts entered our study after the first sort had been administered; therefore they were classified on the basis of the second sort only. The classifications were stable for all but one of the participants who had completed both sortings. The result of the second sort was used for that individual.

[4] When constructing the items, we relied on the heuristic of dissimilarity, rather than coverage. While these two measures are highly correlated, they are not synonymous (see Osherson et al., 1990, pp. 199-200, for analysis and an example). As a result, the test items varied in how strong a contrast in coverage they provided. To take differences in prediction strength into account, we ran logististic regression analyses in addition to the analyses reported here. The regressions largely support the pattern described in the results section.

[5] Admittedly, the experts seemed to be using the term “family” in a fairly loose way, sometimes referring to the scientific family, but occasionally suggesting more abstract groupings. For example, a claim about “all maples” or “the maple family” could be referring to all species in the genus acer, or to the larger set of species encompassed by the family acerae (of which maples

are the principal members). For almost all of the items here, the trees (or pairs of trees) from larger families also represented larger genera.

[6] Individuals were classified as preferring a justification type if it outnumbered the other type two to one. In addition, the preferred type of justification must have been offered for at least half of the items. This second condition was required since codable justifications were not always provided for every item (some were left blank, some fell into the “Other” category).

[7] The mean for the ALL condition (.31) does not equal one minus the average for family size from Table 1 because data from five participants who missed E2 were omitted.

[8] Whether properties are projected to a folk-generic or folk-family group remains unclear because of ambiguity in participants’ responses, as discussed earlier.

Table 1

Proportions of Coverage- and Family-Size Choices on SP and DP Items, Experiments 1 and 2

	Exp. 1: "More"			Exp. 2: "All"		
Expert Group	Coverage	Family Size	Family Size	Coverage	Family Size	Family Size
SP Items	(all items)	(all items)	(critical items)	(all items)	(all items)	(critical items)
Landscaper	.39*	.64**	.69**	.55	.42	.42
Maintenance	.45	.63**	.75**	.43	.60*	.73**
Taxonomist	.51	.67*	.72	.60*	.61	.40
Average	.45*	.65**	.72**	.53	.53	.52
DP Items	(all items)	(all items)		(all items)	(all items)	
Landscaper	.56*	.60		.61	.72**	
Maintenance	.43	.472		.44	.41	
Taxonomist	.62**	.67**		.67**	.76**	
Average	.53	.58*		.57	.62**	

Scores represent the proportion of items for which participants chose the tree with greater coverage or family size. */** Indicates that the mean was different from chance (.50) by t-test (p < .10/p < .05).

Table 2

Justification Frequencies by Expert Type and Experiment

Justifications for Single Premise Items

Expert	Typ	Fam	Mech	Dist	Susc	Resis	Other	DK	Total
Exp 1									
L (9)	7	33	10	30	41	12	21	21	175
M (8)	1	26	19	32	29	6	26	13	152
T (6)	0	41	1	54	4	4	4	8	116
Total	*8*	*100*	*30*	*116*	*74*	*22*	*51*	*42*	*443*
Exp 2									
L (7)	0	31	15	31	7	46	4	2	136
M (6)	0	16	46	14	18	10	10	6	120
T (5)	0	38	0	31	3	12	8	2	94
Total	*0*	*85*	*61*	*76*	*28*	*68*	*22*	*10*	*350*

Justifications for Dual Premise Items

Expert	Div	Fam	Mech	Dist	Susc	Resis	Other	DK	Total
Exp 1									
L (9)	20	11	8	16	37	3	18	11	124
M (8)	3	10	17	17	18	2	15	9	91
T (6)	29	12	0	23	1	1	4	1	71
Total	*52*	*33*	*25*	*56*	*56*	*6*	*37*	*21*	*286*
(table continues)									
Exp 2									
L (7)	38	2	7	13	0	20	0	1	81
M (6)	12	1	23	12	7	8	11	3	77
T (5)	28	11	1	12	2	3	3	0	60

Total	*78*	*14*	*31*	*37*	*9*	*31*	*14*	*4*	*218*

Table 3

Proportion of Different Types of Projection From Premise Pairs as a Function of Lowest Level Linking Category in Experiment 3

Level of Linking Category	Type of Extension				
	Bridge	Extend	Split	Ecological	None
Genus Level (n = 4)	.66	.12	.00	.18	.00
Above Genus Level (n = 6)	.18	.24	.30	.20	.07