Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 4, 641–649

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.4.641

Development of Categorization and Reasoning in the Natural World: Novices to Experts, Naive Similarity to Ecological Knowledge

Patrick Shafto and John D. Coley
Northeastern University

Two experiments investigate the role of similarity and causal–ecological knowledge in expert and novice categorization and reasoning. In Experiment 1, university undergraduates and commercial fishermen sorted marine creatures into groups; although there was substantial agreement, novices sorted largely on the basis of appearance, whereas experts often cited commercial, ecological, or behavioral factors, and systematically subdivided fish on the basis of ecological niche. In Experiment 2, experts and novices were asked to generalize a blank property or novel disease from a pair of marine creatures. Novices relied on similarity to guide generalizations. Experts used similarity to reason about blank properties but ecological relations to reason about diseases. Expertise appears to involve knowledge of multiple relations among entities and context-sensitive application of those relations.

Similarity has been an extraordinarily useful—if somewhat recalcitrant—construct in the psychological literature (Goodman, 1972; Medin, Goldstone, & Gentner, 1993; Tversky, 1977). For instance, consider inductive reasoning. When told that "sparrows have property *x*" and asked whether cats or blue jays have property *x* as sparrows do, people generally use similarity to decide that blue jays probably have property *x*. Several models based on similarity—construed featurally or over taxonomic categories—have been proposed to explain such category-based inductive inferences (e.g., Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman, 1993). Research in this tradition has relied on *blank properties* such as "requires biotin for hemoglobin synthesis" (Osherson et al., 1990, p. 197), or "are susceptible to disease A" (Proffitt, Coley, & Medin, 2000, p. 811; see also López, Atran, Coley, Medin, & Smith, 1997). According to Smith, Shafir, & Osherson (1993), blank properties are those that "apply equally to all categories in a domain" (p. 69). The idea is that these properties would carry no a priori convictions about their validity in a particular category. However, it is rare that people encounter things in their adult lives that are truly novel. Indeed, even rather minimal knowledge may influence performance on these inductive tasks. For instance, Heit and Rubinstein (1994) showed that even when given properties for which they have no a priori expectations, people project novel anatomical properties (e.g., "its liver has two chambers that act as one") to anatomically similar animals (e.g., "from a whale to a bear rather than a tuna"), but they project novel behavioral properties (e.g., "usually travels in a back-and-forth, or zig-zag, trajectory") to behaviorally similar animals (e.g., "from a whale to a tuna rather than a bear").

Thus, knowledge—when it can be applied—may guide inductions about novel properties. Moreover, properties that are essentially blank for novices might serve to activate domain-specific knowledge for experts in a given area and thereby override general taxonomic similarity. Indeed, similarity-based models do not fare so well in explaining expert reasoning, as demonstrated by the diversity phenomenon. When given a problem such as "Cats and sparrows have property *x*. Robins and sparrows have property *y*. Do you think it is more likely that all animals have property *x* or property *y*?" people tend to choose property *x*. In other words, the argument with the more diverse premises is seen as stronger, ostensibly because the premises cover the more general conclusion category more completely. To examine the generality of this phenomenon, López et al. (1997) compared diversity-based reasoning in folk-biologically naïve U.S. undergraduates and folk-biologically sophisticated Itzaj Maya. Both groups were given a standard induction task involving two pairs of local mammals, told that each pair had a newly discovered disease, and asked to choose which disease was more likely to affect all local mammals. Results indicated that the U.S. undergraduates picked the more diverse premises 96% of the time. The Maya, however, picked the more diverse pair only 38% of the time. Because these groups differ in expertise, culture, and many other ways, it is difficult to attribute differences to expertise alone. Subsequently, Proffitt et al. (2000) looked at the use of diversity-based reasoning among U.S. tree experts. As in López et al. (1997) experts were given two pairs of local tree species, told that each pair had a new disease, and asked which disease was more likely to affect all trees. As a group, the experts did not choose the more diverse pair significantly more than chance. In both of these studies, diversity-based predictions were based on the participants' own sortings of the stimuli. In other words, it appears that unlike novices, Maya and U.S. experts'

Patrick Shafto and John D. Coley, Department of Psychology, Northeastern University.

We thank Neal Pearlmutter and members of the Categorization and Reasoning Lab for helpful comments. We extend our gratitude to Adam Nicolazzo for his help with data collection (and other miscellaneous things) and Liz Baraff for her help with data analysis. We also express our gratitude to Ted Shafto for his assistance with procuring our experts. And, most important, we are greatly indebted to the fishermen of Green Harbor for their generous contribution of time and expertise.

Correspondence concerning this article should be addressed to Patrick Shafto, Department of Psychology, MS 0125 NI, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115-5000. E-mail: p.shafto@neu.edu

inferences are not well predicted by their own beliefs about general taxonomic similarity among items in their domain of expertise.

Justifications provided by experts give us some insight into this phenomenon. For the Itzaj, diversity items seemed to trigger causal/ecological reasoning. That is, on diversity items, the Maya picked the pair for which they could make the best reason as to why both had the novel disease. This often happened to be the less diverse pair. Likewise, Proffitt et al. (2000) found that experts preferred to justify generalizations of novel diseases with causal/ecological relations. Experts used causal/ecological relations for 56% of their justifications. In short, in explaining inductive generalizations, similarity took a back seat to domain-specific knowledge about causal/ecological relations among trees.

Therefore, people bring their experience to bear on novel situations, minimizing novelty. However, what kind of information do they rely on to do this? One possibility is that their experiences provide new ways in which items can be similar. There is some suggestion of this in the results of a sorting experiment done by López et al. (1997). Undergraduates from the United States and Itzaj Maya from Guatemala sorted cards corresponding to local animals on the basis of which creatures they believed "went together by nature." Although the sortings were remarkably similar, there was evidence of sorting on the basis of ecological considerations such as habitat, and of finer-grained distinctions among the Maya, but not among the U.S. undergraduates (see also Boster & Johnson, 1989). To extend this reasoning to induction, a flexible and multifaceted sense of similarity might be primed by the particular properties used in the inductive task. Properties may have primed behavioral versus anatomical similarity in Heit and Rubinstein (1994); disease may have activated ecological knowledge relevant to epidemiology for U.S. and Maya experts that was unavailable to novices.

On the other hand, it may be that experts develop a reliance on other kinds of relations—potentially orthogonal to similarity—not available to novices. Justifications from López et al. (1997) and from Proffitt et al. (2000) suggested that causal reasoning may play a role in experts' inductions; experts want to know why or how the property could get from a premise to a conclusion in the context of their environment. Indeed, over 10% of justifications recorded by Profitt et al. (2000) dealt with reasoning about some mechanism involved in the transmission of disease among trees. Relatedly, Medin, Coley, Storms, and Hayes (in press) showed that causal relations influence novice inductions when their relevance was made sufficiently salient.

Thus, evidence to date suggests that novices' generalizations are well explained by their notions of similarity, whereas experts' generalizations are not. Novices agree that more diverse premises create stronger arguments, whereas experts agree that causal/ecological relations are salient and useful in guiding generalizations and that they often trump diversity. Thus, it appears that folk experts and novices are reasoning in qualitatively different ways: experts are reasoning on the basis of causal/ecological reasoning steeped in the context of the respective environs; novices are using more decontextualized similarity-based reasoning. However, several issues remain unresolved.

First, López et al. (1997) looked at groups that differed in both experience and culture, whereas Medin et al. (1997) and Proffitt et al. (2000) looked only at experts. To learn the precise role of experience in categorization and reasoning, it is necessary to test experts and novices from the same culture on the same domain, although few studies have done so (see Bailenson, Shum, Atran, Medin, & Coley, 2002, for an exception). Second, the finding that experts rely on causal/ecological reasoning and that novices rely on similarity-based reasoning in diversity tasks has been somewhat confounded by the types of properties that have been used in the induction tasks. Although novice induction has been examined with varied properties (Osherson et al., 1990, used blank biological properties, e.g., "requires biotin for hemoglobin synthesis," [p. 197], whereas López et al., 1997, used novel diseases), diversity-based reasoning in experts has only been examined using novel diseases. As argued above, disease may prime ecological knowledge for experts but not for novices. Thus, it's not clear whether expertise results in a complete shift in the basis for induction, or in an increased sensitivity to the context of particular inferences. Finally, the induction tasks used have been somewhat artificial. Participants are generally given two competing sets of premises and are asked which provides a better basis for an inference about "all mammals" or "all trees." Although this phrasing does address the hypothesis at hand, it may push the envelope of ecological validity, especially for experts, who may be hesitant to generalize any property to such a broad superordinate.

In the present study, we seek to address these weaknesses in previous work while extending knowledge on how experience impacts conceptual structure and reasoning. We targeted the domain of marine life; our experts were commercial fishermen, and our novices were university undergraduates. This domain represented a coherent ecosystem that has been minimally impacted by mankind (relatively speaking). Creatures interact in ways that have been dictated by the mutual development of the ecosystem. Also, most people have some familiarity with marine life but relatively little in-depth knowledge of it. In contrast, our experts had extensive experience with many facets of the local marine ecosystem. Thus, while novices were likely to have enough knowledge to complete the task, the domain was sufficiently complex for expertise to matter.

Specifically, we focused on two questions. First, we employed a categorization task to examine how experts and novices differed with respect to their use of general taxonomic similarity and causal/ecological knowledge in categorization. Second, we employed an inductive reasoning task to examine the role of taxonomic similarity and causal/ecological relations in guiding experts' inferences about different properties (novel disease vs. novel unspecified property). We also introduced what we take to be a more ecologically valid induction task. Rather than ask participants to generalize to broad superordinates such as "all marine creatures," we instead asked about the plausibility of inferences from premise pairs to a wide range of individual target species, which allowed us to examine the specific basis for induction by ascertaining the relationship between the premises and the targets of induction. By systematically testing experts and novices from the same general population, systematically varying the property being projected, and using a more sensitive and ecologically valid induction paradigm, we hoped to clarify and extend our knowledge on how experience impacts conceptual structure and reasoning.

Experiment 1

Method

Participants. Fifteen undergraduates at Northeastern University and 16 commercial fishermen participated in this study. Undergraduates participated in exchange for course credit. Commercial fishermen participated in exchange for cold, refreshing beverages. None of the undergraduates had previous experience with either commercial fishing or college-level biology courses. Commercial fishermen were required to have a minimum of 5 years experience. Overall, the fishermen had an average of 21.8 years of experience in commercial fishing. None of the commercial fishermen had previous experience with college-level biology courses. Of the 16 fishermen, 3 had achieved college degrees, with the remainder having completed at least high school (some had college experience as well).

Materials. Twenty-eight laminated 6 × 5-in. cards were used. Cards had drawn pictures and colloquial names of a variety of marine creatures. Pictures were culled from various texts (Paxton & Eschmeyer, 1995; Whitefield, 1998). The creatures depicted ranged from bivalves (clams and mussels), crustaceans (lobsters, krill), and fish to sharks and whales. Creatures were selected to be indigenous to the areas fished by the experts and to be representative of the diversity of that ecosystem (see Appendix A for a complete list of creatures).

Procedure. Undergraduates were tested in a university lab on campus. Commercial fishermen were tested on the docks from a mobile testing station. All participants were tested individually. Participants were told that we were interested in how they thought about the creatures and that there were no right or wrong answers. The experimenter shuffled the cards and the participant was given the set of 28 cards and asked to "put them into groups that go together." In the rare instance that an expert was not familiar with one of the items, that item was removed from the set for that participant. If they inquired further, participants were instructed to make as many or as few groups as they liked and that they should base the groups on whatever they thought was important. Justifications were requested for each group. The experimenter recorded the groupings and justifications by hand. The participant was then asked to perform a successive pile sorting by "putting the groups that you have created into groups that go together." The experimenter recorded the new groupings and the pile sort was repeated. This process was repeated until the participant either refused to continue grouping or all of the cards were in one pile. Participants were free to combine as many groups as they liked at any level of sorting. At this point, the experimenter returned the first groupings that the participant made. They were then asked to "split these groups into things that go together." The experimenter recorded the subgroupings, and the participant was asked to split the groups again. Participants were free to split as many subgroups as they liked at any given round of sorting. This was repeated until the participant refused or there were no more subdivisions to be made.

Results and Discussion

How do experts and novices differ in their use of similarity and causal/ecological knowledge? To start to shed light on this matter, we first looked at overall agreement across all participants and within groups of experts and novices. We then turned to a more specific consideration of patterns of sorting coupled with justifications, in order to begin to elucidate the changing roles of similarity and causal/ecological knowledge in structuring the natural world across development.

Scoring. We derived taxonomies from individual sorts by converting the initial sortings, successive superordinate, and successive subordinate sortings into a hierarchy. An example of one expert's sorting is provided in Figure 1. Each taxonomy was then converted into a pairwise distance matrix. In each cell of the 28 × 28 matrix was a number that corresponded to the distance between two creatures. Lower numbers meant greater similarity between creatures. Distance between any item and itself was assumed to be 0. Creatures placed together at the lowest level of the taxonomy were assigned a distance of 1; if they were placed together at the second level they received a distance of 2, and so on. Any creatures that were unfamiliar to an expert were thrown out, and the corresponding cells were treated as missing data.

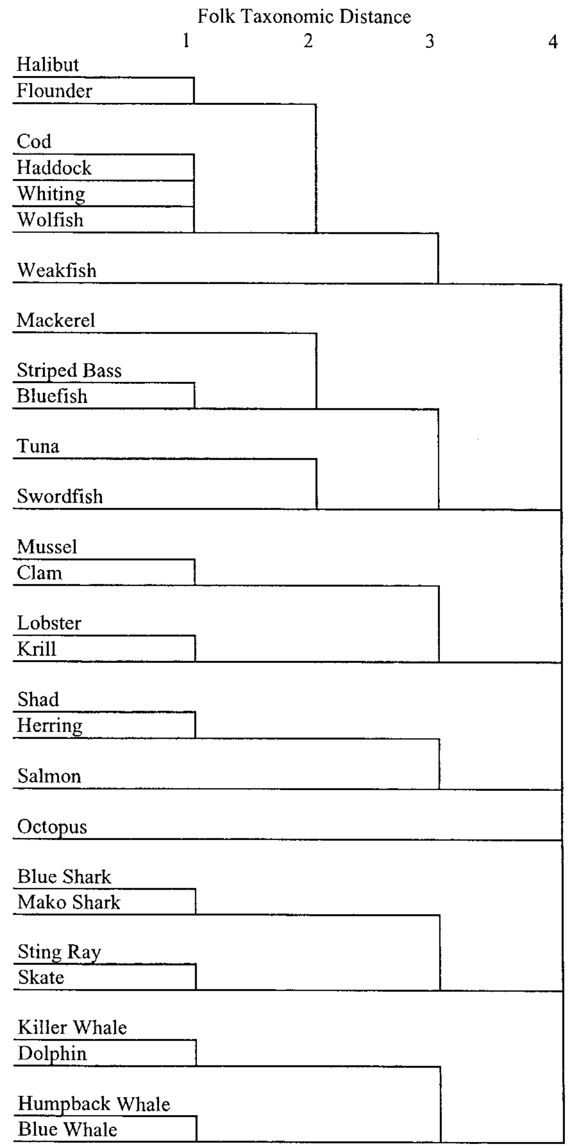

Figure 1. Example of one expert's sorting of marine creatures.

Cultural consensus model. To investigate agreement, we used the cultural consensus model as described in Romney, Weller, & Batchelder (1986) and Weller (1987; for similar applications see Boster & Johnson, 1989, and Medin, Lynch, Coley, & Atran, 1997). The cultural consensus model assumes that the amount of agreement between respondents is indicative of the presence (or lack) of an underlying common body of knowledge among the respondents. Respondents who give sufficiently similar answers to a set of questions are defined as a group because their many individual response patterns are effectively summarized by a single group pattern. This is useful in that it allows us to investigate the agreement between our two groups and gives us a base from which to investigate the possibility of group-specific response patterns.

Agreement is quantified in this model by correlating each participant's pairwise distances with every other participant's pairwise distances. In our case, combining experts and novices resulted in a 31×31 intersubject correlation matrix representing agreement between each pair of participants. The intersubject correlation matrix is then submitted to a principal-components analysis. Under the specifications of the cultural consensus model, consensus is indicated when three criteria are met: (a) The first latent root is relatively large compared with the other latent roots, (b) all scores on the first factor are positive, and (c) the first factor accounts for the majority of the variance. A strong fit of the model indicates that the participants' responses are summarized by one response pattern and that responses represent variations on a single underlying body of knowledge.

Our overall results showed a strong consensus. The first eigenvalue was 17.35 and accounted for 56% of the variance. The subsequent two eigenvalues were 3.45 and 1.57, and they accounted for 11.1% and 5.1% of the variance, respectively. All scores on the first factor were positive, with a mean of 0.74 and a range of 0.48–0.89. These scores suggest a strong fit of the overall model. However, the second factor accounted for 11.1% of the variance. On inspection, second factor loadings corresponded perfectly to our a priori groups: All 16 fishermen loaded positively on the second factor, and all 15 novices loaded negatively on the second factor. This strongly suggests the presence of subgroup agreement within our overall consensus.

To investigate this possibility, we performed an analysis of residual agreement (Nakao & Romney, 1984). We created a subject-by-subject residual agreement by subtracting agreement predicted by the consensus (represented by first factor loads) from the observed agreement in the data. The predicted agreement matrix is calculated by multiplying first factor loadings of each pair of participants, resulting in an index of agreement predicted by each participant's knowledge of the consensus. Subtracting the predicted agreement matrix from the observed agreement matrix resulted in a matrix of agreement that was not accounted for by the consensus, as represented by the first factor. We then standardized the residual agreement matrix. A matrix of ones and zeros corresponding to group membership was created. We compared the residual matrix to the group matrix via Monte Carlo simulations (Hubert & Shultz, 1976). If agreement was greater within our subgroups than between randomly chosen pairs of participants, we would conclude that there was subgroup agreement that was not accounted for by the first factor consensus. The agreement within our subgroups was greater than expected by chance as assessed by the quadratic assignment program (Hubert & Schultz, 1976), $z = 21.23$, $p < .001$. This indicates subgroup agreement among experts and novices that is not accounted for by the overall consensus. These analyses indicated that although there was a sizable amount of agreement between experts and novices, there was a marked divergence in the respective groups' responses.

Sorting and justifications. Our stimulus set contained a number of salient groups of marine creatures that may correspond to basic level categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), including sharks, whales, and fish (see Appendix A). Tanaka and Taylor (1991) presented evidence that experts tend to make finer distinctions in their domains of expertise than novices do; this suggests that fishermen and novices may differ in the degree to which these groups are maintained or divided in the sorting task. In the following analyses, we investigated the degree to which experts and novices respect the named groups sharks, whales, and fish. We coded each individual for whether he or she placed all members of the particular group together at any level of their taxonomy below the all-inclusive level. We performed separate 2×2 chi-square analyses for each named group, comparing the number of experts and novices who maintained these groups to the number that did not. Sharks and whales were always grouped together at some point by both fishermen and novices, and because of this perfect relationship, we can conclude that these groups are overwhelmingly salient. However, novices and experts sorted fish differently. Specifically, 10 of 15 novices grouped all fish together, whereas only 4 of 16 experts did so, $\chi^2(1, N = 31) = 5.42$, $p < .05$. Thus, at some point, most novices lumped all fish together into a single category, whereas most experts did not.

To further examine the bases for sorting, two independent raters coded justifications into one of seven categories. The seven categories were taxonomic ("sharks"), commercial value–edibility ("fish you would eat"), environment–migration ("warm water pelagic"), appearance ("external shell"), behavior–feeding ("predators"), insides ("greasy"), or other ("weird ones"). Categories were not treated as mutually exclusive; justifications such as "large, meat-eating" would be scored as appearance and behavior–feeding. Each category could be scored only once for each justification. If a participant said, "large, blue ones" the justification would score only once for the appearance category. Raters were in agreement on 91% of justifications; disagreements were resolved via mediated argumentation.

On average, experts produced marginally more justifications per sorting session than novices did (*M*s = 21.1 and 16.3, respectively), $t(29) = 1.94$, $p = .062$. Given this difference in production, relative frequencies of each type of justification were computed for each participant. Figure 2 shows the mean proportion of each category of justification given by experts and novices. Experts and novices used strikingly different kinds of justifications to explain their sortings. For novices, appearance and taxonomic justifications predominated; no other class of justification exceeded a mean of 10%. Novices used appearance more than experts did, $M = .43$ vs. .10, $t(29) = 5.17$, $p < .0001$, but rates of taxonomic justifications did not differ reliably, $M = .35$ vs. .25, $t(29) = 1.48$, $p > .15$. For experts, environmental and taxonomic justifications were more common, although they also used justifications based on commercial value and behavior with some frequency. Experts were more likely than novices to justify groupings on the basis of commercial considerations, $M = .13$ vs. .02; environment, $M = .33$ vs. .06; and behavior, $M = .14$ vs. .06, $t(29) > 2.67$, $p < .02$. In sum, both groups used taxonomic justifications with some frequency. Beyond taxonomic concerns, novices relied almost solely on appearance, whereas experts rarely appealed to appearance, and instead cited commercial, environmental, or behavioral considerations.

We have seen that whereas experts and novices agreed on the grouping of two basic level categories, sharks and whales, they

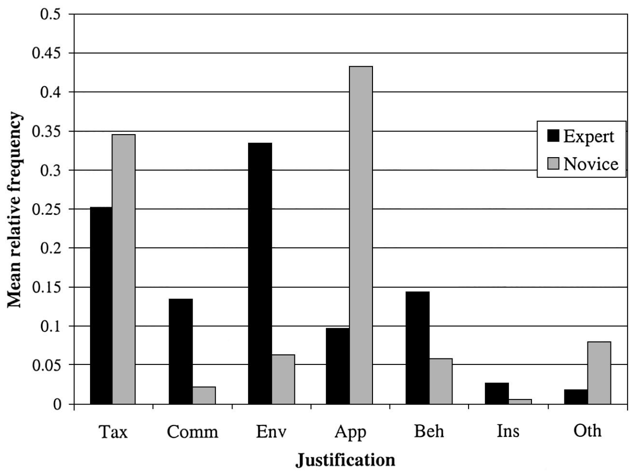

Figure 2. Average relative frequency of sorting justifications for experts and novices. Tax = taxonomic; Comm = commercial; Env = environmental; App = appearance; Beh = behavior; Ins = insides; Oth = other.

differed in the grouping of another basic level category, fish. In examining justifications for these salient subgroups, we found that both experts and novices overwhelmingly justified their whale and shark groups taxonomically (88% and 81% taxonomic justifications, respectively, for novices, and 78% and 74% for experts). Fifty-eight percent of novices also justified their fish grouping on taxonomic grounds (17% appealed to commercial concerns or edibility, and 17% appealed to appearance). In contrast, only 9% of experts' highest level groupings, which included fish, were taxonomically justified; instead, environmental–migratory justifications predominated (57%), with 17% involving behavior–feeding. In summary, novices and experts relied on taxonomic information when grouping sharks and whales; novices extended this approach to grouping fish, whereas experts relied on more specific contextual information when splitting fish into separate groups.

On informal inspection, experts' environment–migration justifications seemed to divide fish into two distinct subgroups: migratory–pelagic creatures and nonmigratory–bottom-dwelling creatures. Agreement among experts on this division was nearly unanimous. Migratory–pelagic fish included bluefish, herring, mackerel, salmon, shad, striped bass, swordfish, and tuna; nonmigratory bottom fish included cod, flounder, haddock, halibut, whiting, and wolfish. In addition, by these criteria the sharks should also be considered a member of the migratory–pelagic group. To test the validity of these groups defined by experts' justifications, we compared the average pairwise similarity of creatures within the ground fish and pelagic fish groups to the average pairwise similarity of randomly selected like-numbered groups of fish by means of Monte Carlo simulation. The result of this process was a test statistic that can be used to compare our experimentally observed within-group similarity to chance by way of a z score. If observed within-group similarity is significantly greater than expected by chance, z scores would be less than -1.96. For our novices, the within-group similarities for ground fish and for pelagic fish did not differ significantly by chance, $z\text{s} = -0.56$ and -1.38, respectively. In contrast, for experts, the within-group similarities for both groups were significantly higher than expected by chance: for ground fish, $z = -5.8$, and for pelagic fish, $z = -3.7$. Clearly, experts reliably sorted fish into subgroups and used ecological explanations to justify these groupings; novices did not.

In summary, despite similarities between experts' and novices' performance on the sorting task, we also documented striking differences. Similarities included an overall consensus on sorting across groups, similar grouping of whales and sharks with taxonomic justifications for those groups, and similar relative frequency of taxonomic justifications. Differences were also clear. First, the cultural consensus model shows that experts and novices agree more with their respective experience groups than across groups. Second, novice justifications heavily favored appearance, and secondarily, taxonomic factors, whereas experts relied more heavily on environmental factors and taxonomy but also invoked commercial and behavioral factors to explain groupings. Third, most novices classified fish into a single group, justified predominantly on taxonomic grounds, whereas most experts did not lump all fish together, reliably formed two subgroups, and justified their splits on the basis of ecological factors. Taken together, these results suggest that novices, having little experience with the species in question, largely rely on visual cues and taxonomic similarity to organize marine creatures. Experts, with much greater first-hand experience, do not abandon taxonomic considerations but augment them with knowledge of commercial concerns, environment, migratory patterns, and behavior. In Experiment 2, we examined the consequences of experts' use of more varied knowledge for inductive reasoning.

Experiment 2

In Experiment 1, we found that both experts and novices used taxonomic information when categorizing marine creatures, whereas experts alone relied on additional causal/ecological information. In Experiment 2, we set out to investigate whether experts and novices differed in the way they generalize information. We focused on two related questions. First, do similarity and causal/ecological relations play differing roles in guiding the inductive inferences of experts versus novices? Second, do experts and novices differ in their generalizations of these two kinds of properties, and does the content of the property being generalized influence the basis for the generalization?

Method

Participants. Twenty-three undergraduates at Northeastern University and 20 commercial fishermen took part in this study. Of the 20 fishermen, 3 took part in the first study. The participants were demographically similar to the participants in the first study.

Materials. Twelve pairs of marine creatures from Experiment 1 were chosen as premises for the induction task, with the goal of providing a wide range of interpair similarity (see Appendix B for a list of premise pairs). The remaining 26 stimulus items from Experiment 1 were used as potential inductive targets for each pair, and 4 additional items were added to increase the representation of some of the already-present subgroups. These additional items are noted in Appendix A.

Design. This study had two conditions: the blank property condition and the novel disease condition. These conditions were varied between subjects. In the blank property condition, participants were told that the premise pair had "a property in common called ____." In the disease condition, participants were told that the premise pair had "a disease in common called ____." Inserted into the blanks were a series of biological or Latin roots. The same words were used for each condition, but the order of words was randomized for each participant. The order of premise presentation was blocked, so that participants saw all items for each premise pair before proceeding to the next premise pair. The order of premise pairs was randomized for each participant. Also, the presentation order of potential conclusion items was randomized anew for each premise pair.

Procedure. Undergraduates were tested in a university laboratory. Fishermen were again tested on the docks from a mobile testing station. Participants were told "I am going to tell you about a (property/disease) that two creatures have in common, then I am going to ask you if you think other fish could have the same (property/disease). There are no right or wrong answers. I am just interested in what you think. Okay?" Participants were then told, for example, that "lobster and wolfish have a property in common called sarca." They were then shown each member of the set individually and asked, "Do you think that haddock could have sarca like lobster and wolfish?" (the participant responds "yes" or "no"). "Do you think that dolphin could have sacra like lobster and wolfish?" As this was being asked, the participant's attention was drawn to each labeled picture card. The question was repeated intermittently throughout the task. Otherwise, participants were simply asked, for example, "Do you think tuna could have sarca? How about killer whales?" This process was repeated for each target over the 12 premise pairs. The two cards that represented the premises were left out for the duration of the time the participant was being asked about them. With their permission, participants' responses were recorded on tape for later transcription. This procedure was repeated for each of the 30 items on each of the 12 premise pairs. The result was 360 "yes" or "no" responses by each participant.

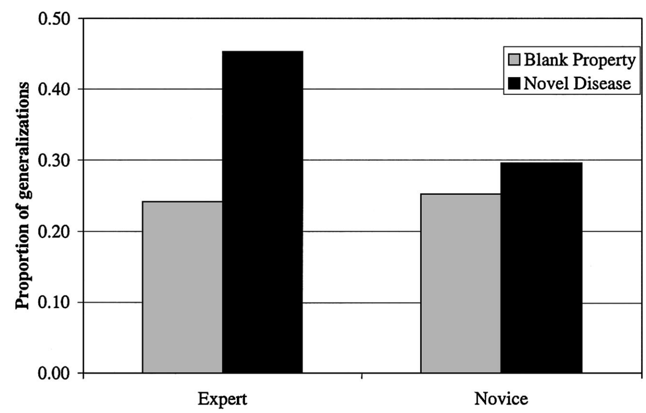

Figure 3. Mean proportion of generalizations.

Results and Discussion

For each premise pair, we noted which target items the property was projected to. This allowed us to look at overall patterns of inference by examining frequency of projections and also to infer the basis of projection by examining in detail which pairs licensed projections to which targets. Below we examine overall patterns of projection for blank properties and novel diseases for experts and novices, and then we examine the role of similarity and causal relations in guiding these projections.

Do experts and novices differ in their generalizations of blank properties and novel diseases? To investigate this question, we computed the mean proportion of projections across items and analyzed the data by experience and property in a 2 $\times$ 2 analysis of variance (ANOVA). The results show a main effect for property, $F(1, 37) = 9.77$, $MSE = 0.159$, $p < .01$, and a marginal effect for experience $F(1, 37) = 3.16$, $MSE = 0.051$, $p < .10$. Most important, these main effects are qualified by an interaction between experience and property, $F(1, 37) = 4.32$, $MSE = 0.070$, $p < .05$. Experts in the novel disease condition simply generalized more than did experts in the blank property condition and more than novices in both conditions (see Figure 3). In other words, the nature of the property being projected influenced experts, but not novices, and specifically, experts projected a novel disease more widely than they did a blank property. This finding raises several possibilities about how knowledge guides induction. First, all participants could be using similarity to guide their generalizations, with experts simply finding diseases to generalize more widely than blank properties. On this account, expertise does not influence relations between similarity and reasoning, but rather it influences the plausibility of different kinds of inferences. Alternatively, experts and novices may rely on different knowledge to guide inferences about novel diseases. Specifically, experts may replace or augment similarity relations with causal/ecological relations, as seen in previous research (e.g., López et al., 1997; Proffitt et al., 2000).

To examine the degree to which similarity predicted expert and novice inferences, we calculated the mean similarity between the members of each premise pair and each target item separately for novices and for experts (based on data from Experiment 1). Then, for each item, we calculated correlations between similarity ratings and proportion of generalizations; separately for experts and novices, and for blank property and novel disease. If similarity was guiding generalizations, all groups should show strong, positive correlations. All correlations were positive, and as can be seen from Table 1, mean R^2 (averaged across all 12 items) was high in all four conditions, suggesting a clear role of similarity in guiding inductions. However, if nonsimilarity-based relations were informing expert inferences about disease, then similarity should show a lower correlation for those items. Indeed, mean R^2 for experts was lower in the disease condition than in the blank property condition, one-tailed $t(22) = 1.87$, $p = .037$. No such difference was observed for novices, one-tailed $t(22) = 0.42$, $p = .34$. Considering the multitude of potential interpretations of "has a property called sarca," these correlations showed that similarity derived from sorting is a viable default guide for induction. Importantly, novices relied on this default whether they were reasoning about blank properties or about novel diseases. In contrast, similarity was less important for experts when reasoning about novel diseases than when reasoning about blank properties.

This finding raises the possibility that experts were selectively relying on causal/ecological relations to guide inferences about novel diseases. One candidate for such a relation is the food chain. Clearly, such relations are especially salient for the commercial fishermen, whose goal is to catch fish, and indeed, our participants commonly mentioned predation as a mechanism for disease transmission. What emerged from experts' comments was an informal folk theory that a species could contract a disease by eating another species that already had the disease. Moreover, this is a particularly good example for our purposes because food chain relations are largely orthogonal to similarity; it may be true that you are what you eat, but (fortunately) we do not usually look the part.

To test whether experts were relying on this specific causal relation to guide induction, we formed two subgroups of target

Table 1
Mean R^2 for Correlations Between Similarity and Proportion of Projections

Group	Blank property	Novel disease
Novices	.593	.549
Experts	.603	.459

items for each premise pair: creatures higher in the food chain (i.e., creatures that prey on the premise items) and creatures lower in the food chain (i.e., creatures that the premise items prey on). These groups were derived from informal reports given by experts. For each of our four participant groups (experts and novices projecting blank properties and novel diseases), the average proportion of generalizations from each premise pair to all items was subtracted from the average proportion of generalizations from each premise pair to items in the target food chain groups. These were then averaged across items, resulting in a score that represented the average deviation from the condition mean for the target subgroups. These are shown in Figure 4; a score of zero represents projections to the subgroup that matched overall projections to all targets, whereas a score above zero indicates disproportionately high projections to a given subgroup. If participants were using food chain relations to guide generalizations, then the subgroups' average deviation should be significantly greater than zero. Based on one-group t tests, the only deviation that was reliably above zero was for experts in the novel disease condition when projecting to items higher in the food chain, $t(11) = 3.59$, $p < .01$. No other effects were significant.

This suggests that experts (but not novices) believe that creatures are likely to pass a disease (but not a blank property) up (but not down) the food chain. Novices, meanwhile, show no evidence of basing inferences on causal relations. The asymmetric, unidirectional use of causality by experts is a particularly interesting finding. One plausible explanation for this bias is that it reflects the statistical nature of causal relations embedded in the food chain itself. If a creature eats an animal known to be diseased, it may be very likely to become sick. However, if told that a creature is diseased, it is not at all clear where the disease originated, as any creature is likely to feed on many kinds of creatures.[1] Sensitivity to such relations may have led our experts to readily generalize diseases up the food chain, but not down it.

On the whole, results of Experiment 2 indicate that expertise brings about complex and content dependent changes in induction. Novices' inductive generalizations were not noticeably sensitive to differences in property at the gross or specific levels. Despite having sufficient knowledge to differentiate properties and diseases, novices treated them as largely the same, applying some convergent form of naïve similarity to guide their categorizations and generalizations. On the other hand, experts' projections of novel properties were well predicted by similarity, whereas similarity was a weaker predictor of projections of novel diseases and was augmented by beliefs about causal relations among marine creatures.

General Discussion

We have sought to extend past findings by comparing experts and novices from the same general population, systematically varying the properties being inferred, and using a more sensitive induction task. Our aim in this study was to examine the degree to which taxonomic similarity and causal/ecological knowledge influenced how experts and novices categorized and reasoned about the natural world, thereby characterizing the acquisition of expertise. Our results clearly show that the shift from novice to expert does not involve abandoning similarity as a guide to reasoning. Rather, experts augment taxonomic similarity with other kinds of relations, when relevant.

Novices and experts produced remarkably similar sorts, despite providing different rationales for those sorts. This makes perfect sense, as perceptual similarity is often a very good indicator of other relations. Shape is indicative of speed; long slender fish move fast. Color is indicative of surroundings; brown fish live on or near the bottom. Novices are able to sort items in ways that are much like experts, by calling on less experienced yet very effective intuitive notions of similarity. Names were also important to this consensus: both groups found whales and sharks to be salient groupings. Groups are named in ways that convey important information. Whales are mammals, different in kind from sharks because they breathe air, although many appear similar.

Despite similarities in sorting, experts' sorts showed clear influence of ecological knowledge lacking in novices. Novice categorization was largely driven by taxonomic similarity. Novices represented the basic level categories sharks and whales in their sorts, tended to lump all fish together at some point, made no systematic distinctions within the fish category, and justified their groups largely on the basis of appearance and taxonomic affinity. In contrast, experts rarely placed all fish into a single class. Instead, they reliably subdivided fish in pelagic and ground fish groups, and justified these groups ecologically. Experts were more likely than novices to justify sorts using environmental, commercial or behavioral factors, and less likely to justify sorts using appearance. These differences reflect the impact of specific ecological knowledge about experts' classification of marine creatures. This finding supports those of López et al. (1997), who found that when sorting mammals, the folk-biologically sophisticated Itzaj Maya used ecological considerations to a much greater degree than relatively naïve U.S. students.

Like their categorizations, novices' inductions were guided by their notions of taxonomic similarity, as shown in previous work (e.g., López et al., 1997; Osherson et al., 1990). Experts' inferences—though guided in part by taxonomic similarity—also showed clear influence of theory-based causal beliefs and ecological relations. Like novices, experts' projections of blank properties were well explained by their similarity ratings. Granted, experts' similarity showed more sensitivity to ecological concerns (e.g., in the ecological division of fish), and therefore experts' similarity-based inductions may be seen as more ecological than novices' similarity-based inductions. Nonetheless, the relationship between similarity and reasoning for experts when reasoning about blank properties was the same as that for novices. In contrast, experts were more willing to project novel diseases than blank properties, and although similarity partially predicted projection of novel diseases, these projections were also guided by specific causal beliefs about disease transmission.

[1] We thank an anonymous reviewer for this suggestion.

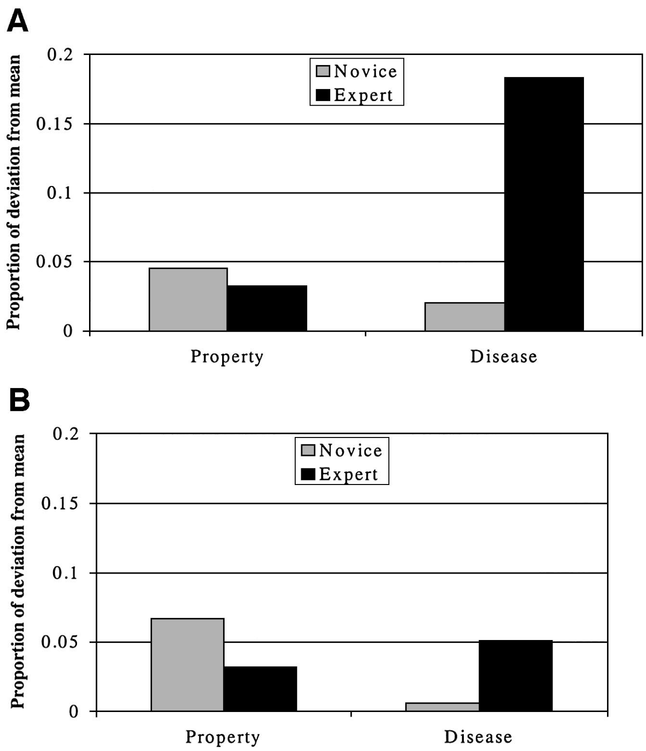

Figure 4. Mean deviation of creatures higher (A) and lower (B) in the food chain.

Previous findings have shown that experts perform at or below chance when compared with predictions derived from similarity-based models of induction; further analysis of justifications suggests ecological reasoning (López et al., 1997; Proffitt et al., 2000).[2] In contrast, our finding that experts projected novel diseases disproportionately from prey to predators represents more direct evidence of ecological relations guiding inductive reasoning. Reasoning about disease leads experts to search for potential causal mechanisms of disease transmission, and to use knowledge of food chain relations to guide inductions. Moreover, a specific causal bias—reasoning with the salient causal direction or reasoning downstream—led to unidirectional causal inferences in these experts (see Medin et al., in press, for further discussion of causal asymmetry in induction).

Moreover, our finding that experts' causal reasoning is contingent on the property being inferred suggests that previous findings may exaggerate differences between experts and novices. Clearly, causal and ecological relations play a role for experts that is absent in novice reasoning. However, rather than characterizing expertise as a general shift away from reasoning on the basis of taxonomic similarity, our results support a more nuanced view. Specifically, experts appear to reason more flexibly than novices. Experts are sensitive to the property being generalized in a way that novices are not. They have access to causal/ecological relations that are unknown to novices; however, rather than relying on these causal/ecological relations instead of taxonomic similarity, experts use these relations to guide inferences only when they are deemed relevant. In our case, they were deemed relevant when reasoning about disease, but not when reasoning about blank properties. Therefore, one component of expert reasoning is having knowledge of a greater variety of relations among categories in the domain of expertise, but another component is knowing which relations are relevant in which contexts.

Certainly questions of how experts and novices organize the world and deal with the uncertainty therein are bound to be, at minimum, vast and complex. We argue that the road from novice to expert is a continuous one. Novices' notions of taxonomic similarity offer them a fundamental bootstrap by which to understand the natural world. Expertise represents finding deeper reasons underlying that similarity, learning new causal and ecological relations that are potentially orthogonal to taxonomic relations, and learning when to rely on one type of relation versus another. However, ultimately, expertise seems to be an augmentation of naïve beliefs about taxonomic similarity with deeper reasons and additional relations, rather than a complete reconceptualization of a domain of experience. Perhaps this is what should be expected; after all, the road from novice to expert should be a tractable one.

[2] In both of these studies, experts were reasoning about novel disease. The only previous studies examining property effects on experts' generalizations (Bailenson et al., 2002; Medin et al., 1997) found no differences in experts' projections of disease *x*, enzyme *x*, and property *x*. However, both used a forced-choice inference task and manipulated property within subjects. It is possible that the task was not sufficiently sensitive to pick up property differences or that experts settled on one response strategy and stuck with it despite variations in property.

References

Bailenson, J. B., Shum, M. S., Atran, S., Medin, D., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84*(1), 1–53.

Boster, J. S., & Johnson, J. C. (1989). Form or function: A comparison of expert and novice judgments of similarity among fish. *American Anthropologist, 88,* 866–889.

Goodman, N. (1972) Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–447). New York: Bobbs-Merrill.

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

Hubert, L., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology, 29,* 190–241.

López, A., Atran, S. A., 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.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (in press). A relevance theory of induction. *Psychonomic Bulletin and Review.*

Medin, D. L., Goldstone R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

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.

Nakao, K., & Romney, A. K. (1984). A method for testing alternative theories: An example from English kinship. *American Anthropologist, 88,* 318–338.

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

Paxton, J. R., & Eschmeyer, W. N. (1995). *The encyclopedia of fishes.* Boston: Academic Press.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 811–828.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and information accuracy. *American Anthropologist, 88,* 313–338.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

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

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

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Weller, S. C. (1987). Shared knowledge, intracultural variation, and knowledge aggregation. *American Behavioral Scientist, 31,* 178–193.

Whitefield, P. (1998). *The Simon and Schuster encyclopedia of animals.* New York: Simon & Schuster.

Appendix A

Marine Creatures Used as Stimulus Items in Experiments 1 and 2

Pelagic fish: bluefish, herring, mackerel, salmon, shad, striped bass, swordfish, tuna
Ground fish: cod, haddock, halibut, flounder, whiting, wolfish, sole*
Whales: blue whale, dolphin, humpback, killer whale
Sharks: blue shark, mako shark, dogfish*, porbeagle*
Other marine creatures: clam, krill, lobster, mussel, octopus, skate, stingray, weakfish, oyster*

Note. An asterisk indicates that a creature was used as a potential inference target in Experiment 2 but not as a sorting stimulus in Experiment 1.

Appendix B

Premise Pairs Used in Experiment 2

Halibut and flounder, Herring and shad, Striped bass and salmon, Mako shark and blue shark, Wolfish and mackerel, Cod and swordfish, Bluefin tuna and halibut, Mussel and dolphin, Lobster and mako shark, Whiting and blue shark, Lobster and wolfish, Bluefin tuna and dolphin

Received April 2, 2002
Revision received November 18, 2002
Accepted December 18, 2002 ■