Memory & Cognition
2009, 37 (5), 596-607
doi:10.3758/MC.37.5.596

Sample selection and inductive generalization

Chris A. Lawson
Carnegie Mellon University, Pittsburgh, Pennsylvania

and

Charles W. Kalish
University of Wisconsin, Madison, Wisconsin

In two experiments with adults ($N = 126$), we examined the influence of sampling procedure on inductive generalization. In predicate sampling, participants learned the category identity of individuals known to possess some property. In subject sampling, individuals selected for category identity were discovered to possess a novel property. In both experiments, sampling procedure influenced induction. Predicate sampling resulted in very narrow generalization, whereas subject sampling yielded a fairly high and constant rate of projection. Differences in confidence of generalizations were also observed. Conditions in which evidence was described as randomly sampled from a collection of animals yielded a consistent decrease in projections as predicted by similarity-based models. The results are presented as support for an evidence-based view of induction.

Just how people make predictions about new objects and events is a core question in the study of inductive reasoning. What processes govern the inference that a basketball game will last about 2 h, that a bird will fly, or that a piece of chocolate will taste good? Most accounts of inductive inference depend on the relation between known exemplars and novel exemplars. We expect that a bird will fly because most birds encountered in the past have flown. The basis of that expectation, the function that maps from known to novel exemplars, remains a source of debate within the field of psychology. In the present article, we consider two general approaches to inductive inference, characterized as similarity based and evidence based. A key distinction between these two approaches is the significance of exemplar selection, or *sampling*.

Suppose that one is told that all vertebrate animals have one of two variants of the hemoglobin molecule in their blood: plaxium or drotium. One then encounters several robins with plaxium blood (base exemplars). A long tradition of research leads us to expect that inferences from this experience will follow a similarity-based gradient (Rips, 1975; Sloman, 1993; Sloutsky & Fisher, 2004). The likelihood of projecting a property from one individual to another decreases as their similarity decreases. Animals very similar to robins, animals dissimilar to robins, and animals of intermediate similarity will be predicted to have plaxium at rates proportional to their similarity to robins. Two features of this projection function warrant comment. First, projection is a matter of degree, rather than all or none. Second, learning that some things have plaxium convinces people that other things lack it; learning that robins have plaxium makes it seem less likely that mice have it. Psychological accounts of inductive inference must explain these two features. The central question of the present study is whether exemplar selection influences patterns of projection. Would the same pattern of graded predictions result when the base exemplars were selected on the basis of their category membership and then discovered to have plaxium as when they were selected on the basis of having plaxium and then discovered to be members of the same category?

A large number of theories present inductive projection as governed by a similarity-based process (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975; Sloutsky & Fisher, 2004; Smith, 1989). Recent evidence suggests that dissimilarity may also play a role in induction (Stewart & Brown, 2005; Stewart & Morin, 2007). Similarity naturally accommodates the graded pattern of projection described above. As an illustration, we consider similarity-based models based on feature overlap (without meaning to single out feature overlap as more or less adequate than other similarity models; Heit, 1998; Sloman, 1993). When deciding whether a novel exemplar (the target) will share a property with a known exemplar (the base), one compares the number of known features common to the two exemplars with the number of known features not shared by the two exemplars. The ratio of common to distinctive features determines the response. With a base of robins with plaxium, people are likely to predict that a sparrow has plaxium (because robins and sparrows are known to have many common features and few distinctive ones) and to predict that a mouse lacks plaxium (because robins and mice have many distinctive and few common features).

C. A. Lawson, clawson@andrew.cmu.edu

Similarity relations between base and target exemplars determine the shape (slope) and position (intercept) of the projection function. Such models are sensitive to the range of exemplars encountered (which may determine the sensitivity of the similarity computation) and the property being generalized (which may determine the weighting of different features). The critical constraint on such models is that similarity depends on features of the actual exemplars. More general or abstract representations of categories are not involved. For example, the fact that a robin and a penguin are both members of the category BIRD does not affect their similarity (or inferences from one to the other) above and beyond the features of the two individuals (potentially including common labels).

Alternative accounts characterize the projection from known to unknown exemplars in terms of evidence rather than of similarity. The common claim is that known (base) exemplars contribute support for hypotheses about unknown (target) exemplars. Exact characterizations of the support relation range from a kind of similarity metric (Osherson et al.'s, 1990, "coverage") to formal constructs involving informativeness and Bayesian belief revision (e.g., Tenenbaum, Griffiths, & Kemp, 2006; Xu & Tenenbaum, 2007). These approaches are distinguished from similarity-based accounts in that base exemplars are understood to provide information about a more general category or population. The category representation (updated with information about the base exemplars) then determines inferences about the targets. Encountering a set of robins with plaxium blood provides some evidence about the frequency of plaxium within the category of BIRD. The category representation then guides inferences about novel targets, such as penguins.

Evidence-based accounts do not directly predict a similarity gradient. Indeed, one of the attractions of similarity-based models is that such a gradient is a natural consequence. In evidence-based accounts, the characterization of projection in terms of features of the base and target exemplars underspecifies the problem. The shape of the projection function cannot be determined without specifying the priors: the hypothesis space or the set of categories available. Indeed, a major debate between similarity- and evidence-based accounts concerns the nature of such priors, the ability to specify the priors in some principled way, and the distinction (if any) between priors and the contextual sensitivity required of similarity-based theories (see Tenenbaum et al., 2006). However, priors are only part of the missing information.

In general, the significance of a piece of evidence cannot be specified independent of information about how one came to encounter that evidence (Eells, 1982). Information about sampling (which is part of the likelihood in Bayesian formulations) is critical. For example, the implications of encountering a set of robins with plaxium blood for projection to other animals depends on the circumstances that led to the encounters. If the exemplars were selected from a population of robins, what we will refer to as *subject sampling*, then discovering that all of them have plaxium suggests that all robins have plaxium (see Fiedler, Brinkmann, Betsch, & Wild, 2000, on "predictor" sampling). The experience is informative about the population of robins and provides some basis for estimating the probability of plaxium conditional on belonging to that population. In contrast, under *predicate sampling* (also called *strong sampling*, Tenenbaum & Griffiths, 2001, or *criterion sampling*, Fiedler et al., 2000), individuals selected for having plaxium blood are discovered to be robins. This experience conveys information about the population of things with plaxium blood and provides some basis for estimating the probability of being a robin conditional on belonging to that population. There are infinitely many other sampling strategies, notably a random or "weak" (Tenenbaum & Griffiths, 2001) strategy of selecting animals at random and investigating their blood type and species identity.

Treating base exemplars in a projection task as evidence provides a set of clear hypotheses about the effects of sampling strategies on the shape of the projection function. In contrast, similarity-based theories do not provide predictions about the effects of exemplar selection. In the remainder of this article, we develop the evidence-based predictions about subject and predicate sampling and report two experiments in which we tested these hypotheses.

In the experiments in the present article, we held base exemplars constant but varied the sampling strategy that generated the exemplars. For example, participants encountered 20 robins with plaxium blood. The strategy that generated those 20 exemplars is described as *subject*, *predicate*, or *random* sampling. Evidence was presented as a collection of single individuals to heighten the sense that the instances were drawn from a sample and to provide multiple opportunities to encode the strategy that generated the evidence. Following the encounters with the base exemplars, the participants were asked to project the property of having plaxium blood to a series of targets varying in similarity to the base exemplars. In addition to indicating whether targets have the property, the participants also indicated their confidence in their judgments. The predictions for the effects of sampling strategy are as follows (see Figure 1):

Subject sampling. Base exemplars suggest that all robins have plaxium blood. Rates of projection will be high to targets that are highly similar to the base. Rates of projection to medium- and low-similarity targets will also be relatively high. The base exemplars provide evidence consistent with hypotheses that the property is widely shared (e.g., all birds have plaxium, all animals have plaxium) and provide no evidence against a particular target having the property. Confidence in projections will be high for highly similar targets but low for others. The sampling strategy is only informative about robins and provides a poor basis for decisions about nonrobins.

Predicate sampling. Base exemplars suggest that only robins have plaxium blood. This condition suggests that nonrobins lack plaxium. Rates of projection for nonrobin targets will be very low; projections will drop sharply with decreasing base–target similarity. The participants will

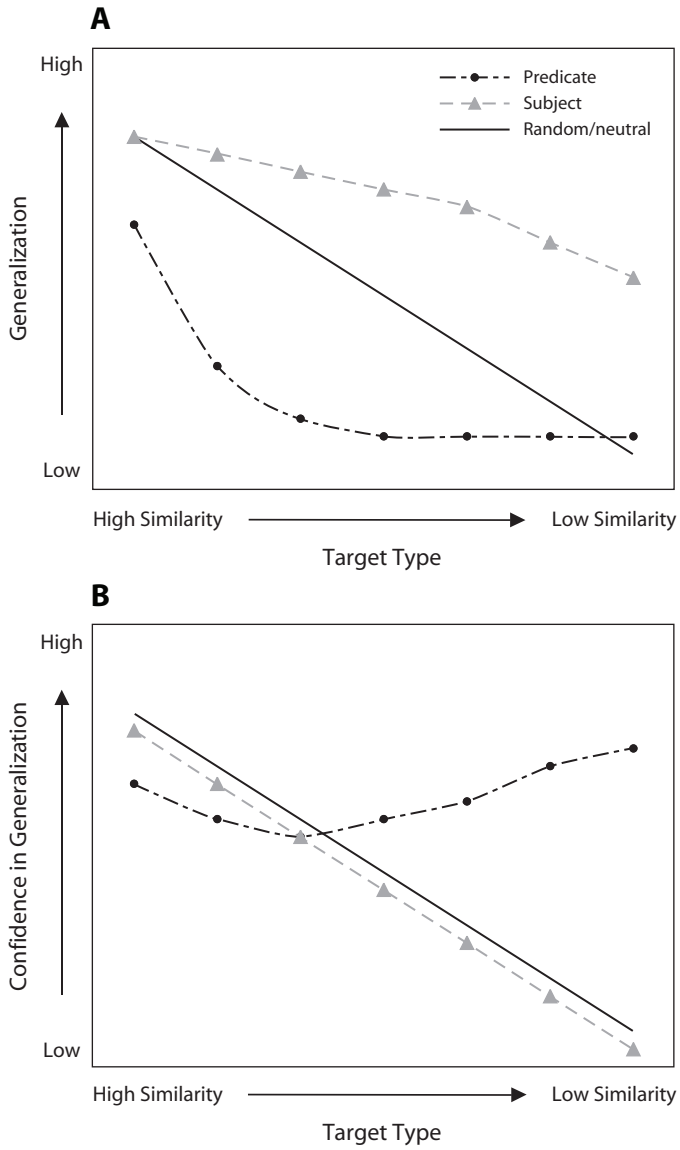

Figure 1. Predicted patterns of generalization (A) and confidence in generalization (B) as a function of similarity between base and target exemplars and sampling strategy.

also be confident in predictions that low-similarity targets do not share the property with base exemplars. Predictions about high-similarity targets (e.g., novel robins) are slightly more complex. On their own, the base exemplars do not provide a strong warrant for conclusions about the population of robins. However a plausible prior, or *overhypothesis* (Goodman, 1955; Kemp, Perfors, & Tenenbaum, 2007; Shipley, 1993), is that blood type is constant across animals of the same species. Thus, evidence that one robin has plaxium blood indicates that all robins do. Nonetheless, the warrant for this conclusion seems weaker than that under subject sampling. The participants may make fewer and less confident projections to high-similarity targets for predicate than for subject sampling.

Random sampling. A natural hypothesis is that random sampling is the default assumption that participants use when sampling strategy is not indicated (at least for adults; see Kalish & Lawson, 2007). Thus, projections and confidence under random sampling may produce a constant decrease in projections with decreasing similarity, similar to the patterns of projections reported in the literature (Sloutsky & Fisher, 2004). Of course, this is only one possible similarity function (see Shepard, 1987, for alternatives). Critically, this condition is intended to provide a context in which sampling procedures hold no specific implications that a certain range of targets are privileged, thus causing participants to rely on base–target relations. An alternative interpretation of this pattern is that it represents something of an average between predicate and subject sampling.

Existing literature has provided ambiguous evidence regarding people's sensitivity to sampling strategy. Previ-

ous research characterizes people as poor intuitive statisticians (e.g., Nisbett, Krantz, Jepson, & Kunda, 1983). Fiedler and colleagues (Fiedler, 2000, 2008; Fiedler et al., 2000), in particular, argued that people misunderstand the implications of sampling. Fiedler's (2000, 2008) basic claim is that people do not appreciate limitations of different sampling strategies. For example, from a disease-based sample that yields unbiased information only about $p(\text{symptom} \mid \text{disease})$, people will estimate $p(\text{disease} \mid \text{symptom})$ without correcting for—or perhaps even noticing—the potential bias introduced by the sampling strategy (Fiedler et al., 2000).

Evidence from word learning and inductive projection provides a somewhat more optimistic picture of people's understanding of sampling. People draw samples differently when asked to assess whether all Xs have Y versus whether only Xs have Y (Kincannon & Spellman, 2003). This finding stands in contrast to the biasing information search observed in more classic decision-making tasks (Fiedler et al., 2000; Wason & Johnson-Laird, 1972; but see Oaksford & Chater, 1994, for a more positive assessment of information search). A recent study by Xu and Tenenbaum (2007) indicated that even young children take sampling into account when learning the extensions of novel words. As they encounter more exemplars, people match the extension of the term to the range of exemplars encountered. After seeing one beagle called a *blicket*, people may extend the term to other breeds of dog. After seeing three beagles called *blicket*, people restrict the term to beagles. Xu and Tenenbaum argued that this pattern reflects the assumption that the exemplars were generated via predicate sampling. Critically, when this assumption is violated, by having someone ignorant of the true distribution of the property doing the selecting (producing something like subject sampling), the extension is not narrowed; people do not assume that the examples reflect the population (Tenenbaum & Xu, 2000).

The predictions from previous research for the present experiments are not entirely clear. Decision-making tasks have tended to use complex probabilities with substantial memory demands. Word-learning tasks have not explicitly manipulated sampling of the information provided to participants. Nonetheless, participants do seem to be sensitive to information relevant to sampling. Thus, we predict that participants in the present experiments will adjust their inductive generalizations according to the sampling strategy producing the base of evidence that they have to work from. A plausible alternative hypothesis is that inferences will be guided by base–target similarity alone. This alternative predicts no differences between sampling conditions but, rather, predicts a common graded pattern of decreasing projections with decreasing similarity in all cases.

EXPERIMENT 1

Method

Participants. Seventy-two adult university students at a large midwestern university participated in Experiment 1 for course credit. There were approximately equal numbers of male and female participants.

Design. An equal number of participants were randomly assigned to one of four conditions representing the sampling procedures: random, subject, predicate, and neutral. Information about sampling procedure was manipulated with a cover story that described a group of explorers learning about the animals on a newly discovered island. The participants learned that they would be shown some examples and then asked to make some predictions. In the *subject* condition, examples were a collection of robins that happened to be in the lab. In the *predicate* condition, examples were animals known to have plaxium. In the *random* condition, the examples were chosen at random from animals on the island. In the *neutral* condition, we provided no information about sampling. Most existing projection research has used a procedure similar to the neutral condition. The exact wording for each cover story is presented in Appendix A.

The experimental session involved a training phase and a projection phase. In the training phase, participants learned about 20 individual exemplars. The same exemplars were used in all four conditions. Exemplars were four highly similar subordinate-level items (American robins), each presented five times in random order. Items were presented in different orientations, so that no single instance was identical. The presentation of exemplars varied slightly across the four conditions. In the random and subject conditions, the participants saw an image of an exemplar and then clicked on a box to test the blood type of the animal. Upon clicking, the participants discovered that the animal had plaxium blood. The procedure was repeated for all 20 exemplars. In the predicate condition, the participants were instructed to click on a box in order to see the animal that had tested positive on the plaxium test. This procedure was repeated for all 20 items. In the neutral condition, the animal and its blood type were presented simultaneously. Thus, participants in all four conditions encountered 20 robins with plaxium blood.

In the projection phase, the participants made estimates of the frequency of plaxium blood within a group of 10 target exemplars. For each target set, the participants were told, "Now your assistants have brought in some more animals for testing. This time, they went out and collected 10 instances of one particular species. You have before you 10 <targets>. We would like you to estimate how many of these <targets> have plaxium blood." Responses were recorded on a slider ranging from 0 to 10. There were six targets, designed to vary in similarity to the base exemplars: One of the targets was highly similar to those used in the base (a novel set of robins), three were drawn from the same basic level as the base exemplar (birds: pigeons, owls, and ostriches), and two were members of different basic-level categories (nonbirds: lizards and mice). After estimating the frequency of plaxium, the participants rated their confidence in their estimate on a scale of 0 (*not at all confident*) to 10 (*very confident*).

Procedure. All stimuli and responses were presented as text and images on a computer screen. The participants worked at individual computers in a classroom containing 12 workstations. The projection phase followed the training phase with no delay. Target sets in the training phase appeared in random order. Together, the training and projection phases lasted approximately 15 min.

Similarity measure. A separate group of 12 undergraduates participated in a similarity-rating task. The task asked for similarity judgments for all target–conclusion pairs on a scale of 1 (*not very similar*) to 9 (*very similar*). These ratings revealed that the stimuli were ordered in terms of decreasing similarity to robins: pigeon, owl, ostrich, lizard, mouse. Ratings for owl and ostrich and those for lizard and mouse were not significantly different from each other in this small sample (each pair member was equally similar to robin).

Results

Figure 2A shows the mean number of targets predicted to have the property (projection scores) for each target

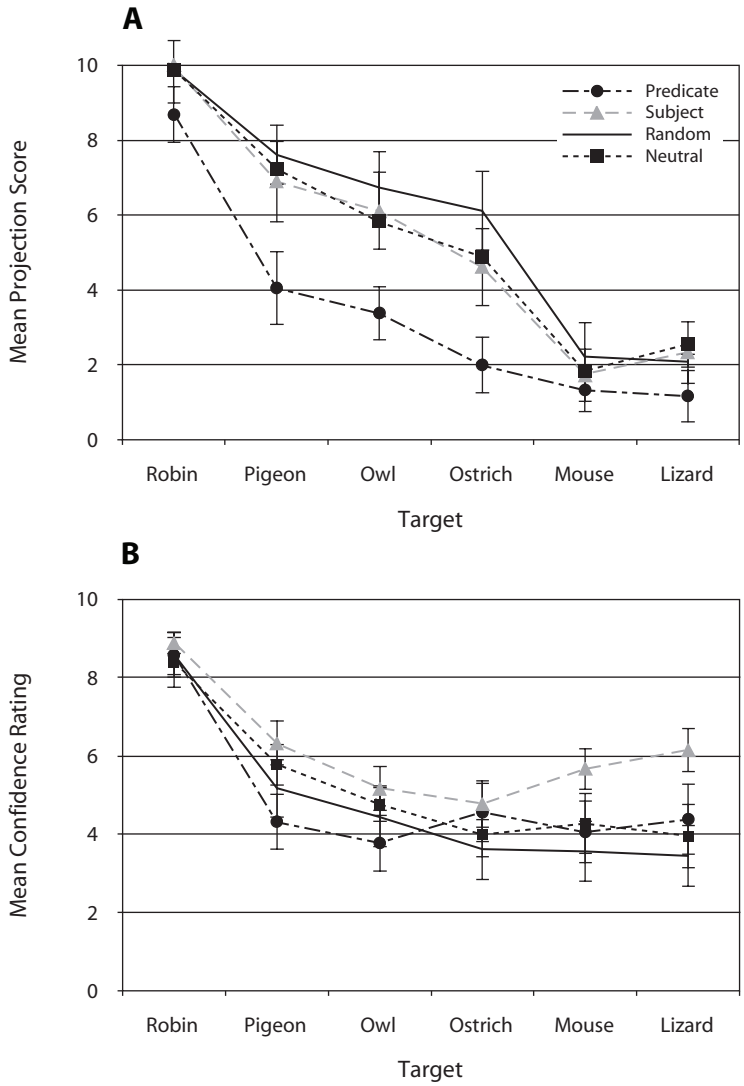

Figure 2. Patterns of generalization (A) and confidence in generalization (B) for all targets in all three sampling conditions in Experiment 1. Bars indicate ±1 *SE*.

across the four conditions. Projection scores showed a bimodal distribution: Participants tended to reply either that all 10 exemplars would have the property or that none would. Given the nonnormal distribution, we analyzed the data using a series of planned nonparametric (Wilcoxon signed-rank) comparisons. For each set of comparisons, p values were adjusted using Holm's procedure. In the first analysis, we tested the general prediction that projections would be higher in the subject than in the predicate condition by comparing projections to each of the targets. There were no differences in projections between the subject and predicate conditions for the high- (robin) or low- (mouse, lizard) similarity targets (all Wilcoxon signed-rank tests, $p > .40$). One of the medium-similarity items (other birds) received higher ratings in the subject than in the predicate condition (owl, $T = 2.73$, $p < .001$); the other two medium-similarity items received higher scores under subject sampling but failed to reach significance after controlling for familywise error (pigeon, $T = 2.44$, $p = .02$; ostrich, $T = 2.37$, $p = .01$).

A central evidence-based prediction is that projections should drop off sharply under predicate but not under subject sampling. One test of this prediction compares projection scores for high- and medium-similarity targets. All medium-similarity targets received significantly lower projection scores than did the high-similarity target under predicate sampling (robin vs. pigeon, $T = 3.06$, $p < .01$; robin vs. owl, $T = 3.52$, $p < .001$; robin vs. ostrich, $T = 3.30$, $p < .001$). Under subject sampling, two medium-similarity targets received projection scores that did not differ significantly from the high-similarity target (pigeon, $T = 1.4$, $p > .15$; owl, $T = 1.6$, $p > .10$), though the difference between one pair approached significance (robin vs. ostrich, $T = 2.23$, $p = .026$). Rates of projection to each of the different birds were significantly lower than to the robin both in the random condition (robin vs.

pigeon, $T = 2.37$, $p < .01$; robin vs. owl, $T = 2.80$, $p < .01$; robin vs. ostrich, $T = 3.18$, $p < .01$) and in the neutral condition (robin vs. pigeon, $T = 3.18$, $p < .01$; robin vs. owl, $T = 3.52$, $p < .001$; robin vs. ostrich, $T = 3.62$, $p < .001$). These results are consistent with the prediction that predicate but not subject sampling would lead to a drop-off in projections.

A similar analysis follows for differences between medium- and low-similarity items. If predicate sampling indicates that only robins have plaxium, projections should be equal (low) to all nonrobins. In the predicate condition, projection scores for two of the three medium-similarity items (owl and ostrich) were not significantly different from the low-similarity items (lizard and mouse). The exception was the pigeon item, which was rated higher than either low-similarity item (lizard, $T = 2.66$, $p < .01$; mouse, $T = 2.66$, $p < .01$). In contrast, projection scores for all of the medium-similarity items under subject, random, and neutral sampling projections were higher than projections to each of the low-similarity targets (all ps $< .05$), with one exception: Under random sampling, the difference between projections to ostrich and lizard did not reach significance ($T = 1.5$, n.s.). Thus, the prediction of a sharper decline in projections under predicate than under subject sampling was largely supported in both the high- to medium-similarity and the medium- to low-similarity comparisons.

The next analysis involved comparisons among mean confidence ratings for targets at all three levels. In order to reduce the number of comparisons, this analysis tested differences among high-, medium- (average of all bird items), and low- (average of all nonbirds) similarity items. Subject sampling was expected to result in high confidence in predictions for high-similarity items (the participants should be confident that other robins have the property) but low confidence for low-similarity items (the participants should be unsure about things unlike robins). Predicate sampling should lead to the opposite pattern: The participants should be confident that nonrobins lack the property, but unsure whether animals similar to robins will have the property. The analysis failed to support this prediction. Confidence in projections to robins was higher than confidence in projections to all other targets in the four sampling conditions (see Figure 2B). Additionally, comparisons across the four conditions revealed that confidence in projections to low-similarity items was actually higher in the subject condition than in any of the other three conditions (Wilcoxon, two-tailed comparisons, all ps $< .02$).

Finally, we considered individual patterns of responses. The predicate sampling should show a sharp decrease in projection scores from high- to medium-similarity targets, whereas subject sampling is predicted to yield a constant level of projection. The participants were considered to show *constant* projection if their projection scores for two or more of the three medium-similarity targets (other birds) were no less than their projection scores for high-similarity targets (robins). Eleven participants in the subject condition made constant projections, whereas only 3 did so in the predicate condition ($p < .01$, Fisher's exact test). Eight participants showed constant projection in the random condition. In contrast, only 1 participant showed constant projection in the neutral condition. Confidence was expected to show a positive relation to similarity under subject sampling but a negative relation under predicate sampling. The participants were considered to show a *positive* relation if their mean confidence increased from low- to medium-similarity targets and from medium- to high-similarity targets. Four participants in the subject condition showed a positive relation, as did 5 in the predicate condition and 8 in the random condition. These frequencies were not significantly different from each other. Only 2 participants showed decreasing confidence (from low- to medium- to high-similarity targets): 1 in the subject condition and 1 in the predicate condition. Nineteen participants overall showed a quadratic pattern of greater confidence for high- and low-similarity targets than for medium-similarity targets. The frequency of this quadratic pattern did not vary by condition.

Discussion

The results from Experiment 1 support our general hypotheses about the influence of predicate and subject sampling. The prediction was that predicate sampling would lead to a narrow range of projections and that subject sampling would lead to a broad range of projections. Two results confirmed these predictions. First, responses in the subject condition followed a constant pattern of high projections for the high- and medium-similarity targets. Second, predicate sampling led to a pattern of projecting properties at an equal and low rate to most of the target exemplars; participants in this condition were more likely to restrict projections to high-similarity targets. However, we failed to support the prediction that projections to exemplars most similar to the base would be higher in the subject than in the predicate condition. We also found no support for our prediction of high confidence for low-similarity items in the predicate condition, and low confidence for such items in the subject condition.

As expected, projections under random sampling followed a similarity-based pattern—high- $>$ medium- $>$ low-similarity—basically matching the pattern in the neutral condition. However, results from the random condition did resemble the patterns from the other conditions, suggesting that each of the sampling cases hold some of the same implications for how evidence generalizes. However, it is possible that participants did not believe that sampling in the random condition was truly random. It seems unlikely that random sampling would generate 20 animals of the same species. Also, in the random condition, the cover story may have been interpreted as nonrandom. For instance, sampling from the first animals encountered on the island may have lead participants to expect that the evidence was about those animals that are more visible or more abundant (on islands). More generally, the differences in sampling strategies across conditions may not have been very salient for the participants. Differences were conveyed largely through textual descriptions. Projections might have been more distinct if the sampling strategies were made more apparent. In Experiment 2, we address this possibility.

EXPERIMENT 2

In the training phase of Experiment 1, participants either discovered a novel property of a known species of animal or the species type of an animal known to have a novel property. Although this procedure established that different conditional probabilities were involved in the two cases, it did not highlight the idea of a sample being drawn from a population. One interpretation of the projections in Experiment 1 is that the participants were reasoning as if they had the entire population of evidence. To make sampling a more salient feature of the evidence, in Experiment 2, we presented the participants with a large number of exemplars, only some of which could be examined. The participants selected a sample to examine drawn from some larger population (although the population was really just a larger sample selected on a particular feature). For example, in the predicate condition of Experiment 2, the participants saw 50 files representing animals known to have plaxium blood. Each participant was allowed to select 20 of those files to examine. Thus, a participant had the experience of discovering that the 20 plaxium cases that he or she selected all turned out to be birds of a certain species. In the subject condition, the 50 cases were selected for (and labeled by) their species identity. In the random condition, the 50 cases were selected for being animals.

A second concern with Experiment 1 is that the set of base exemplars was extremely homogeneous. Especially in the random condition, the participants may have doubted the stated sampling strategy. It is also possible that participants in the subject condition were somewhat unwilling to draw conclusions from such a narrow sample. Learning about the distribution of a property by studying a single species is a poor strategy. We might expect broader and more constant generalizations under subject sampling with a more diverse evidential base. For these reasons, the base exemplars in Experiment 2 consisted of a set of songbirds of various species.

Predictions remain the same as those in Experiment 1. Predicate sampling should result in a sharp drop in projections to nonsongbird targets, even to those similar to the base exemplars. Subject sampling should produce consistently high rates of projection, even to targets moderately dissimilar to the base exemplars. Similarly, we expected that random sampling would lead to a stepwise decrease in projections as in Experiment 1, with responses following a pattern of high- $>$ medium- $>$ low-similarity.

Method

Participants. Fifty-four participants were recruited from undergraduate courses at a large public university in a medium-sized midwestern city and received course credit for participation. The population was predominately white and of middle income. There were approximately equal numbers of male and female participants.

Design and Procedure. An equal number of participants were randomly assigned to the random, subject, and predicate conditions. Each condition was established through a cover story about scientists studying the distribution of two variants of the hemoglobin molecule: plaxium and drotium. The cover stories were modified from those in Experiment 1 in order to accommodate the selection format in the training phase. In the subject condition, the participants saw 50 rectangles representing 50 songbirds available for testing for plaxium blood. The rectangles were labeled *Songbirds 1–50*. The random condition was quite similar, although the rectangles represented 50 animals chosen at random and were labeled *Animals 1–50*. In the predicate condition, we presented 50 rectangles representing animals known to have plaxium blood, labeled *Plaxium 1–50*. Appendix B provides the exact wording for each condition.

In the training phase, 50 rectangles appeared on the computer screen, and the participants were told they were to select 20 of the items (one at a time). In the subject and random conditions, selecting a rectangle produced a picture of an animal (in each case, one of four songbirds). Clicking on the picture revealed the blood type—plaxium in each case. In the predicate condition, selecting a rectangle produced the written text "Animal Number X with plaxium blood." Clicking on the text then revealed a picture of the animal—a songbird in each case. At the end of the training phase, all of the participants had seen 20 songbirds with plaxium blood.

In the projection phase, we used the same wording and exemplars as in Experiment 1. After completion of the projection phase, the participants received a manipulation check to ensure that they remembered the sampling procedure. The participants were shown descriptions of the three options (animals were randomly selected, the animals were a select group of songbirds, or the animals represented a collection of items testing positive on a plaxium test) and asked, "Do you remember how the original examples you saw were chosen? How were the first examples of animals you saw actually selected? Please click on the text that best describes the sampling process."

Similarity measure. A separate group of 12 undergraduates were asked to rate the pairwise similarity between items. The first test compared similarity ratings between the four base exemplars (songbirds). A one-way ANOVA revealed no differences in ratings between all possible similarity comparisons for the base exemplars [$F(5,56) < 1$]. Thus, analyses were conducted on the averaged similarity ratings across the base exemplars for each target exemplar. A one-way ANOVA with each evidence–conclusion pair within subjects revealed a main effect [$F(5,55) = 180.9, p < .0001$]. Scheffé's tests revealed no significant similarity difference between the base exemplars and one of the bird targets (pigeon). The remaining targets were ordered (in decreasing similarity to songbirds: owl $>$ ostrich $>$ mouse $>$ lizard).

Results

The first analysis looked at responses to the manipulation check in which participants were asked to recall the method by which the evidence was sampled. All but one of the participants chose the correct sampling method (53 out of 54)—a pattern significantly greater than chance (binomial theorem, .33, $p < .0001$). This result indicates that the participants remembered the condition under which the evidence was sampled.

Our analytic strategy in Experiment 2 was the same as that in Experiment 1. The first set of analyses tested the prediction that projections would be higher under subject than under predicate sampling. The results shown in Figure 3 revealed no difference in projections to high-similarity items (songbirds, identical to the training exemplars) in the subject and predicate conditions ($T = .07$, n.s.). The analyses revealed higher rates of projections to both of the low-similarity items in the subject than in the predicate condition (mouse, $T = 2.84$, $p < .01$; lizard, $T = 2.51$, $p < .01$). One of the medium-similarity items received higher ratings under subject sampling (owl, $T = 2.43$, $p < .01$), whereas differences for the other two medium-similarity items were significant before control-

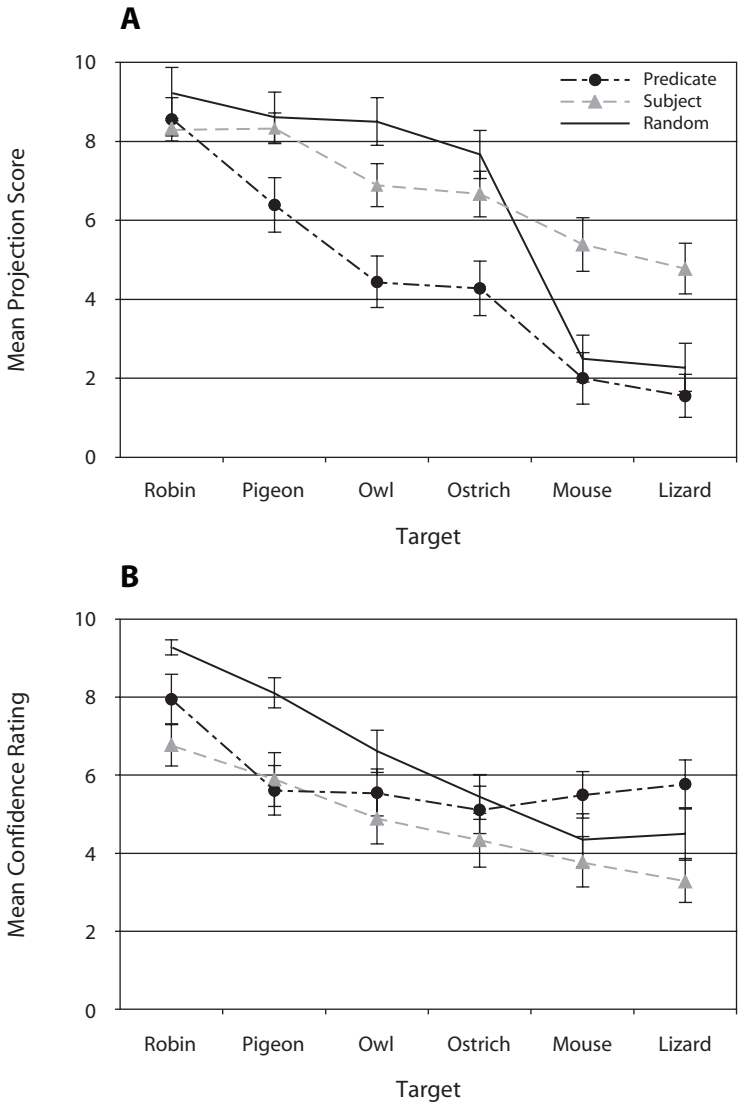

Figure 3. Patterns of generalization (A) and confidence in generalization (B) for all targets in all three sampling conditions in Experiment 2. Bars indicate ±1 *SE*.

ling for familywise error (pigeon, $T = 2.00$, $p < .05$; ostrich, $T = 1.92$, $p = .05$). Thus, these results are consistent with the general prediction that subject sampling would lead to higher rates of projections than would predicate sampling.

The next set of analyses examined differences in projections to the high- and medium-similarity items. In the predicate condition, there was a clear drop-off in responses with higher rates of projections to high-similarity items than to each of the medium-similarity items (songbird vs. pigeon, $T = 2.58$, $p < .01$; songbird vs. owl, $T = 2.79$, $p < .01$; songbird vs. ostrich, $T = 2.79$, $p < .01$). In contrast, subject and random sampling led to a constant rate of projection: In no case were projections to the high-similarity target greater than projections to medium-similarity targets. These results confirm one of the predicted sampling effects and are consistent with the results observed in Experiment 1. However, the results run counter to the prediction of a graded decrease under random sampling.

Differences between medium- and low-similarity items were also examined to test the prediction that predicate sampling would lead to an equal rate of projections to nonsongbirds. The analysis revealed significant differences between all medium-similarity (pigeon, owl, ostrich) and low-similarity (lizard, mouse) item targets in all three conditions (all ps $< .02$). These results suggest that there was a graded decrease in projections under all three sampling conditions, counter to the predictions for the subject and predicate conditions.

A second set of analyses tested predictions about confidence: Would confidence show a direct relation to target similarity for subject sampling but an inverse

relation for predicate sampling? This analysis involved comparisons among the high-, medium- (all birds), and low- (all nonbird) similarity items. The prediction was that, under subject sampling, confidence in projections would be lower for nonrobins than for robins and that predicate sampling would yield the opposite relationship. All conditions supported higher confidence ratings for high- than for low-similarity items (all ps $< .01$). However, comparisons across the three conditions revealed that confidence in projections to low-similarity items were higher under predicate than under subject sampling ($T = 2.70, p < .01$) or under random sampling ($T = 2.30$, $p < .03$). These results do not support the prediction that predicate sampling would cause participants to be more confident in making projections to nonsampled items than to sampled items. However, the predicate sample did lead to higher confidence ratings than did subject and random sampling.

With a final set of analyses, we considered individual patterns. The evidence-based hypothesis is that participants will be more likely to project at constant rates using subject than using predicate sampling. For this analysis, *constant projection* was defined as the score given to the high target, being no more than 1 point greater than the score for the other target. A participant was considered to have shown the constant pattern if he or she gave constant projection scores for at least two of the three medium-similarity items (and for both of the low-similarity items). Only 5 participants showed the constant pattern for medium-similarity items under predicate sampling, whereas 12 and 11 did so under random and subject sampling, respectively (predicate $<$ random or subject, both ps $< .05$, Fisher's exact test). Five participants in the subject sampling condition responded constantly for low-similarity items as well (1 participant did so in each of the other conditions). As in Experiment 1, three patterns of confidence may be defined: directly, inversely, or quadratically related to similarity. More participants in the random than in the predicate conditions gave consistently decreasing confidence scores (7 vs. 15, $p < .05$, Fisher's exact test; 10 in the subject condition). No other patterns showed significant condition differences. Two participants did show the predicted inverse relation between confidence and similarity in the predicate condition, whereas no participants displayed this pattern in the other conditions.

Discussion

The results confirm and extend the findings from Experiment 1. Under conditions of predicate sampling, projections were restricted to songbirds. Furthermore, the prediction of a constant rate of projections for subject sampling received strong support: Projections to high- and medium-similarity targets were equivalent, and projections to all nonsongbird items were high, relative to other sampling conditions. The results also provided some support for the prediction that confidence would increase as targets became more dissimilar to the evidence in the predicate case but it would decrease in the subject case. Overall, the results are consistent with our interpretation that predicate and subject sampling support different inductive generalizations. Finally, the results again failed to support the prediction of lower projections for the high-similarity targets under the predicate than under other sampling conditions.

GENERAL DISCUSSION

In the present study, we explored the influence of sampling procedures on inductive generalizations. Participants projected properties to novel targets differently, depending on the sampling strategy that generated the base exemplars during training. Specifically, subject-based sampling led to a relatively high and constant rate of property projection across a range of targets. Predicate-based sampling produced a sharp drop-off in projections: Only highly similar targets were expected to share the base property. Both of these patterns contrasted with the familiar similarity-based gradient observed when the participants were not provided with information about the sampling strategy. The participants' confidence in their projections also varied by sampling strategy. At least in Experiment 2, when strategy differences were highlighted, the participants expressed low confidence in predictions about dissimilar items under subject sampling but higher confidence under predicate sampling.

Overall, the results support the evidence-based predictions for the effects of sampling strategy. The basis for these predictions is the assumption that base exemplars in a projection task represent a sample drawn from some population. The sample observed provides evidence about the population. Different ways of drawing samples will provide different kinds of evidence. In the present study, predicate sampling supported a hypothesis about the population of things with plaxium blood: Only certain kinds of animals were members of this population. Subject sampling supported a hypothesis about the population of animals of a specific type: All members of the population had plaxium blood. Participants receiving the two different kinds of evidence were predicted to make different generalizations to novel targets. Specifically, properties would be generalized more broadly given subject than given predicate sampling. Participants given predicate sampling should be confident that dissimilar exemplars lacked the property, whereas subject sampling does not provide a strong basis for inferences about dissimilar exemplars. Of course, these predictions, and the specific patterns of responses observed, also depend on other beliefs about the properties and exemplars involved (e.g., animals of the same species share blood type; plaxium blood is not particularly rare). However, it is plausible that these other beliefs were constant across the sampling manipulation. Thus, the different patterns observed in Experiments 1 and 2 may be taken to reflect participants' sensitivity to sampling strategy.

The results of the two experiments reported above are predicted by an evidence-based account of inductive projection but not by a similarity-based account and, therefore, support the former approach over the latter. The similarity relations among the exemplars did not vary across sampling conditions as the same base and target exemplars

appeared. This should yield a prediction of no condition difference. That said, all viable similarity models admit contextual influences. Such models can accommodate the present results if subject sampling results in an emphasis on shared features of base and target, whereas predicate sampling results in an emphasis on distinctive features. For example, under subject sampling, the features that robins and other birds have in common contribute more to the similarity computation, whereas the nonshared features receive lower weights. The result is that robins and other birds are predicted to share novel properties. Under predicate sampling, the nonshared features receive more weight; robins and other birds are seen as dissimilar and unlikely to share properties. Although similarity-based theories can accommodate the results, it is not clear that they predict them. Why should subject sampling result in higher weights for shared features and predicate sampling higher weights for distinctive features? We suspect that incorporating such predictions will involve something like a representation of a category or population, reducing the differences between similarity- and evidence-based accounts.

One feature of the sampling manipulation that may be relevant for similarity-based accounts is that the direction of inference differed across conditions. In some cases, participants saw the animal and then learned its property; in other cases, they saw the property and then learned about the features of the animal possessing the property. Temporal relations and order have been shown to have strong effects on the inferences people are willing to endorse (Baumann & Krems, 2002; Dawes, 1993; Lagnado & Sloman, 2004).

The suggestion that people's inductive projections are sensitive to sampling strategy raises the question of how to interpret prior studies of inductive inference that have not, by and large, informed participants of the origins of the base exemplars. A natural hypothesis is that participants assume random sampling in the absence of information to the contrary. At least in the present study, this hypothesis received support: Patterns of projection were quite similar in the random and neutral sampling conditions of Experiment 1. Alternatively, in the absence of sampling information, the participants may have realized that they lacked the necessary information to assess the evidential value of the base exemplars and instead used some more basic heuristics, such as assessments of similarity.

People did take sampling into account when using evidence to make inferences, and they did so in ways consistent with normative theory. Thus, the results of the two experiments described above seem inconsistent with claims that people are poor intuitive statisticians who fail to consider biases introduced by sampling strategy (Fiedler, 2008; Fiedler et al., 2000). Before discussing these implications, it is worth considering some respects in which the participants did seem to make unwarranted inferences from samples. Given subject sampling, which suggested that all robins have the property, people tended to conclude that nonbirds lacked the property—a conclusion warranted by predicate sampling. Similarly, given predicate sampling, which suggested that only robins have the property, people tended to conclude that all robins have the property—a conclusion warranted by subject sampling. We suggest that such inferences reflected additional beliefs: overhypotheses, such as *all animals of the same species have the same kind of blood*, or priors about base-rates (see Kemp et al., 2007; Xu & Tenenbaum, 2007). Just how, or if, people's interpretation and use of sampling information will respond to changes in these background conditions/assumptions (e.g., unequal prior probabilities) is an open empirical question. Recent research suggests that inference strategies are sensitive to such conditions (McKenzie & Mikkelsen, 2007; Oaksford & Chater, 1994) but that interpretations of sampling strategies have not been a direct focus.

There are also several methodological differences between the present study and prior research on sampling. First, in the present study, we used very simple structures of evidence; the training exemplars were always consistent (e.g., 100% had the property). It may be that people would be less attentive to sampling in more complex information environments with less consistent evidence, as is characteristic of decision-making research. Second, memory demands were quite low. Past research can be taken to demonstrate that long-term representations of evidence do not include information about sampling (Fiedler, 2008). It may be that sampling strategy is part of episodic memory and, in a process akin to source amnesia, such detail is lost as memories are consolidated into long-term representations. In contexts such as the present study in which sampling procedures were explicitly stated and samples were notably small, there are limited processing demands, perhaps making it easier for people to encode information about sampling. Ultimately, the most significant questions will be when and how people respond to sampling strategy when assessing evidence, not whether they do so.

The present results converge with other accounts in suggesting that reasoning is sensitive to the origins of evidence. A full understanding of induction requires considering reasoners' attention to the implications of example selection. The present study demonstrates that people do, in fact, modify their inductive projections in light of information about sample selection. Moreover, the observed sensitivity is consistent with the evidential value, or *informativeness*, of the observed exemplars for hypotheses about populations. It may be possible to accommodate this sensitivity into similarity-based or associative models of inference, but for now, the results of the present study seem most consistent with theories that treat inference as a process of evidence evaluation.

AUTHOR NOTE

This research was supported by NIH Postdoctoral Training Grant T32 MH019102, awarded to the first author, and by National Science Foundation-DLS Grant 0745423, awarded to the second author. We thank Charles Kemp, Brian Ross, and two anonymous reviewers for valuable comments on earlier versions of the manuscript. Correspondence concerning this article should be sent to C. Lawson, Department of Psychology, 5000 Forbes Avenue, Carnegie Mellon University, Pittsburgh, PA 15213 (clawson@andrew.cmu.edu).

REFERENCES

BAUMANN, M., & KREMS, J. F. (2002). Frequency learning and order effects in belief updating. In P. Sedlmeier & T. Betsch (Eds.), *Etc.: Frequency processing and cognition* (pp. 221-237). New York: Oxford University Press.

DAWES, R. M. (1993). Prediction of the future versus an understanding of the past: A basic asymmetry. *American Journal of Psychology*, **106**, 1-24.

EELLS, E. (1982). *Rational decision and causality*. New York: Cambridge University Press.

FIEDLER, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review*, **107**, 659-676.

FIEDLER, K. (2008). The ultimate sampling dilemma in experience-based decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 186-203.

FIEDLER, K., BRINKMANN, B., BETSCH, T., & WILD, B. (2000). A sampling approach to biases in conditional probability judgments: Beyond base-rate neglect and statistical format. *Journal of Experimental Psychology: General*, **129**, 399-418.

GOODMAN, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University 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.

KALISH, C. W., & LAWSON, C. A. (2007). Negative evidence and inductive generalization. *Thinking & Reasoning*, **13**, 394-425.

KEMP, C., PERFORS, A., & TENENBAUM, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, **10**, 307-321.

KINCANNON, A., & SPELLMAN, B. A. (2003). The use of category and similarity information in limiting hypotheses. *Memory & Cognition*, **31**, 114-132.

LAGNADO, D. A., & SLOMAN, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 856-876.

MCKENZIE, C. R. M., & MIKKELSEN, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, **54**, 33-61.

NISBETT, R. E., KRANTZ, D. H., JEPSON, C., & KUNDA, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, **90**, 339-363.

OAKSFORD, M., & CHATER, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608-631.

OSHERSON, D. N., SMITH, E. E., WILKIE, O., LOPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.

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

SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

SHIPLEY, E. (1993). Categories, hierarchies, and induction. *Psychology of Learning & Motivation*, **30**, 265-301.

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

SLOUTSKY, V. M., & FISHER, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, **133**, 166-188.

SMITH, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, **96**, 125-144.

STEWART, N., & BROWN, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, **49**, 403-409.

STEWART, N., & MORIN, C. (2007). Dissimilarity is used as evidence of category membership in multidimensional perceptual categorization: A test of the similarity–dissimilarity generalized context model. *Quarterly Journal of Experimental Psychology*, **60**, 1337-1346.

TENENBAUM, J. B., & GRIFFITHS, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral & Brain Sciences*, **24**, 629-640.

TENENBAUM, J. B., GRIFFITHS, T. L., & KEMP, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, **10**, 309-318.

TENENBAUM, J. B., & XU, F. (2000). Word learning as Bayesian inference. In L. Gleitman & A. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 517-522). Hillsdale, NJ: Erlbaum.

WASON, P. C., & JOHNSON-LAIRD, P. N. (1972). *Psychology of reasoning*. Cambridge, MA: Harvard University Press.

XU, F., & TENENBAUM, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, **10**, 288-297.

APPENDIX A
Cover Stories Used in Experiment 2

Random

Your assistants have spread over the island and collected a random sample of animals. Your assistants each went out and came back with a couple animals to test. They didn't do a systematic search: Each just caught the first couple of animals they happened to encounter. We will show you the 20 animals they brought back, one at a time. In each case you will test whether the animal has Plaxium blood. In this way you can learn about what kinds of animals have Plaxium blood.

Subject

It happens that the explorers have a collection of robins that they have found on the island. These robins were collected for some other project. The robins are already in the lab, so you decide to do some blood tests on the animals. We will show you 20 of the robins, one at a time. In each case you will test whether the bird has Plaxium blood. In this way you can learn about what kinds of animals have Plaxium blood.

Property

Your assistants have gone around the island giving a Plaxium test to all the animals they encounter. They have brought back all the animals that have registered positive on this test. Now, in your lab you can examine these animals. We will show you the 20 animals that had a positive Plaxium test, one at a time. In this way you can learn about what kinds of animals have Plaxium blood.

APPENDIX B
Cover Stories Used in Experiment 3

Random

Your assistants have spread out over the island and collected a random sample of 50 animals. Your assistants each went out and came back with a couple animals to test. They didn't do a systematic search. Each just caught the first couple animals they happened to encounter. You can examine 20 of the animals they brought back, one at a time by clicking the buttons (An_1 thru An_50). In each case you will test whether the animal has plaxium or drotium blood.

Subject

It happens that the explorers have a collection of 50 small songbirds that they have found on the island. These animals were collected for some other project. These small songbirds are already in the lab, so you decide to do some blood tests on the animals. You can examine 20 of the 50 songbirds, one at a time by clicking on the buttons (Sb_1 thru Sb_50). In each case you will test whether the bird has plaxium or drotium blood.

Property

It turns out there is a simple test for Plaxium blood. Your assistants have gone around the island giving the test to the animals they encounter. They have brought back all 50 animals that have registered positive on the plaxium test. You can examine 20 of the animals that had the positive reading by clicking on the buttons (Px_1 thru Px_50). In each case you will see what kind of animal had plaxium blood.

(Manuscript received September 4, 2008;
revision accepted for publication January 19, 2009.)