Prior expectations in pedagogical situations

Patrick Shafto[1], Noah D. Goodman[2], Ben Gerstle[1], & Francy Ladusaw[1]
[1] Department of Psychological and Brain Sciences, University of Louisville
[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Abstract

Much of human learning occurs in social situations, and among these, pedagogical situations may afford the most powerful learning. In pedagogical situations, a teacher chooses the concept that they are going to teach and the examples that they use to teach the concept. If learners know that a teacher is helpful and understands the implications, this could support strong inferences. In previous work, Shafto and Goodman (2008) proposed and tested a model of pedagogical data selection. We integrate special-purpose pedagogical expectations in this framework, and derive a task that allows independent assessment of pedagogical expectations. Two experiments contrast people's expectations about pedagogical and communicative situations. The results show that people's expectations differ in these situations, and that in pedagogical situations people expect teachers to present generalizable and semantically coherent knowledge. We discuss the implications for modeling learning in pedagogical settings, as well as for understanding human learning more broadly.

Keywords: Pedagogy; Learning; Bayesian Model

Much of human learning occurs in social contexts. We learn from siblings, parents, friends, and teachers by observing, imitating, and teaching. Among these social learning settings, pedagogical situations stand out as potentially the most important. Pedagogical situations are situations in which one person, a teacher, chooses information for the purpose of helping another person, a learner, arrive at some belief. Pedagogical situations might provide uniquely powerful learning situations, especially if learners are privy to, and understand the implications of, teachers' intentions to help.

Indeed, recent theories argue that an intuitive understanding of pedagogical situations may be what sets us apart from other animals (Csibra, 2007). Under this proposal, learners' intuitive understanding of pedagogical situations consists of two components: inferences how teachers choose examples to teach a concept, and expectations about what kinds of concepts teachers are more likely to teach.

The issue of how teachers choose information and learners' understanding of these situations have been investigated in detail (for a review, see Csibra and Gergely, 2009). Recently, Shafto and Goodman (2008) have proposed a computational model of reasoning in pedagogical situations. This account provides a formal explanation of why and how teachers decide which examples to choose, and how learners can capitalize on the teacher's intent to make stronger inferences.

Researchers have also argued that young children come prepared with expectations about what kinds of knowledge to expect in pedagogical situations. Specifically, Csibra and Gergely (2009) argue that very young children expect that knowledge provided in pedagogical contexts is semantically generalizable. For instance, Topal et al. (2008) show that children make A-not-B errors in pedagogical contexts, but not in neutral contexts. They argue that the perseverative errors are a consequence of children misinterpreting initial pedagogical demonstrations as indicating that the A box is where the ball belongs. While these results are quite compelling, they contain influences of both the learner's inference about the teachers' choice of data, and the learners' expectations about what kinds of properties are likely to be taught.

In the current paper, we investigate the hypothesis that people expect semantically generalizable knowledge in teaching situations. We begin by discussing the role of prior knowledge in pedagogical reasoning, and how this can be integrated with Shafto and Goodman's (2008) model of pedagogical reasoning. We then use this framework to develop a method for separating the role of pedagogical priors from pedagogical data selection. Two experiments use this method to investigate whether adults expect generalizable knowledge (Experiment 1) and whether adults expect semantically coherent knowledge (Experiment 2). In each case, we contrast pedagogical situations with communicative situations to address whether these prior expectations are specific to pedagogical contexts. We conclude by discussing implications for modeling human learning and understanding reasoning in social situations.

The role of priors in pedagogical reasoning

The proposal that learners expect generalizable information can be integrated naturally into a Bayesian reasoning framework. From this perspective the problem of learning is one of inferring the probability of different hypotheses, h, given observed data, d. Bayes' theorem provides a way of updating our posterior beliefs about hypotheses, $P(h|d)$, given prior beliefs, $P(h)$, and as-

[1]Please address correspondence to Patrick Shafto, p.shafto@louisville.edu

sumptions about how data are sampled, $P(d|h)$,

$$P(h|d) \propto P(d|h)P(h). \tag{1}$$

In standard Bayesian learning, it is typically assumed that the prior, $P(h)$, is determined by some stochastic generative process, and data are sampled from a hypothesis, $P(d|h)$, randomly; however, these standard assumptions are not appropriate for pedagogical situations. In pedagogical situations, a teacher chooses both the hypothesis that they are going to teach, and the data that they use to teach the hypothesis. It is, therefore, reasonable for the leaner to expect that teachers' choices are not random, but are instead purposeful.

Recent work has formalized pedagogical sampling – how teachers may choose data for the purpose of teaching a hypothesis (Shafto and Goodman, 2008). This model suggests that learners may use the knowledge that the teacher is choosing data for the purpose of teaching them a hypothesis, to replace the random sampling with an assumption that the teacher will choose data that tend to make the learner believe the correct hypothesis,

$$P(d|h) \propto P(h|d). \tag{2}$$

Their results show that this pedagogical sampling assumption allows for stronger inferences than random sampling. They also provided evidence that people's inferences differ in pedagogical and ostensibly random contexts, and the pedagogical model accurately predicted people's intuitive pedagogical inferences.

Here, the question is whether people have prior expectations about which kinds of hypotheses should be taught. That is, are there a specific set of prior probabilities, $P(h)$, that apply to teaching situations. Intuitively, the question is whether learners expect teachers to choose particular hypotheses that are important or worthy of teaching. For example, more general hypotheses might be expected because the knowledge is more likely to be applicable in future situations. Formally, we capture these expectations as a utility function which defines how hypotheses are weighted in pedagogical situations, $U(h; pedagogy)$. Integrating this prior into Equation 1 would allow us to capture how prior expectations, specific to pedagogical situations, affect inferences.

Investigating pedagogical priors

In this paper, we attempt to answer two questions: (1) what are the characteristics of people's pedagogical prior expectations, and (2) are these expectations specific to pedagogical situations? To address the second question, we need to choose appropriate (non-pedagogical) control conditions. To address the first question, we can ask people to make judgments about which of two teachers they would rather have teaching them, while varying the particular hypothesis that each is teaching. However, given our above analysis, this requires that we separate the contribution of pedagogical data sampling from the contributions the pedagogical prior expectations.

Judgments about the teachers can be formalized as an inference about whether the teacher chooses hypotheses

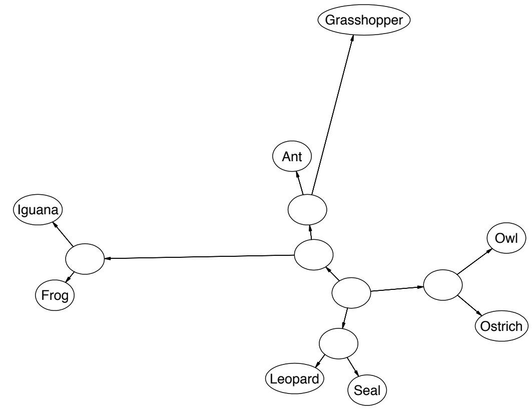


Figure 1: A tree representing the intuitive taxonomic relationships among 8 animals.

that the learner expects to be taught about. More formally, we assume a parameter, θ, which specifies how systematically a teacher chooses her examples. The probability of choosing hypotheses depends on how systematic the teacher is,

$$P(h|\theta) \propto e^{-\theta U(h)}, \tag{3}$$

where systematic teachers tend to choose hypotheses that have higher utility (Luce, 1959).

The learner can then infer how systematic a teacher is, given some data,

$$P(\theta|d) \propto \sum_h P(d|h)P(h|\theta)P(\theta) \tag{4}$$

where $P(\theta)$ is a prior distribution specifying whether people tend to be systematic or not. This equation states that teachers are considered systematic to the degree that they choose hypotheses that agree with the learner's prior expectations, as specified by Equation 3. However, in this inference, the influence of $P(h|\theta)$ is not isolated.

To see how we could isolate the effects of $P(h|\theta)$, consider hypotheses about properties of the the animals in Figure 1. The set of possible hypotheses can be defined extensionally, by enumerating all possible combinations of animals that have or do not have the property. For example, one hypothesis about a property is {ostrich=true, owl=false, grasshopper=false, ant=false, iguana=false, frog=false, leopard=false, seal=false}. Our goal is to eliminate the contribution of the sampling assumption, $P(d|h)$. Assuming that we want to teach the learner the hypothesis that all of the animals have a particular property, how would we choose which animal or animals to provide as examples? By presenting all of the data – each animal labeled as having or not having the property – we essentially choose one hypothesis. Thus, the contribution of the sampling of data is to simply indicate a particular hypothesis. Formally, the $P(d|h) = 1$ for the true hypothesis, and zero for all others. Equation 4 reduces to,

$$P(\theta|d) \propto P(h|\theta)P(\theta). \quad (5)$$

Given the fully labeled data, the learner's judgments about the teacher's systematicity depend on whether the learner expects that hypothesis to be chosen, and their prior expectations about systematicity.

To isolate the influence of learners' prior expectations about hypotheses $P(h|\theta)$, we can ask learners to choose between two teachers. Because each teacher is equally likely to be systematic *a priori*, judgments about which of two teachers is preferred isolate the effects of a learner's prior expectations. Formally, the judgment becomes a ratio of two inferences, each individually specified by Equation 5,

$$\frac{P(\theta_1|d_1)}{P(\theta_2|d_2)} \propto \frac{P(h_1|\theta)P(\theta)}{P(h_2|\theta)P(\theta)} = \frac{P(h_1|\theta)}{P(h_2|\theta)}. \quad (6)$$

In the following, we present two experiments in which people make judgments about which of two teachers they want to have teaching them in the future (presumably the one that chooses a hypothesis that is more consistent with their expectations). In our investigations, we have two goals: (1) identifying the prior expectations that people bring to pedagogical situations, and (2) establishing whether these expectations are unique to pedagogical situations. The experiments test two claims related to prior expectations about pedagogical situations: that learners expect more generalizable information, and that learners expect semantically coherent information.

Experiment 1: Testing the bias toward generalizability

Experiment 1 investigated whether people have an expectation that teachers would teach generalizable information. To investigate this question, we choose a domain for which we have a good understanding of the possible hypotheses, the domain of animals. Figure 1 shows the animals, and the intuitive taxonomic relations among these animals.[2] We operationalize generalizable concepts here as a concept that is true of a broader class of animals.

To investigate whether people expected generalizable knowledge, we presented participants with scenarios in which pairs of teachers taught concepts of different levels of generality. The generalizable teacher taught a property that was consistent with the tree structure and was true of a greater number of exemplars. For instance, the generalizable teacher might teach a property that was true of all 8 animals, while the less generalizable teacher might teach a property that was true of only ostriches and none of the other animals. If people expect teachers to teach generalizable information, we expect to find that people choose the teacher who teaches properties that were true of broader sets of examples.

[2]The tree was derived using the tree learning algorithm and a subset of the animals used in Kemp and Tenenbaum (2008).

Methods:

Participants: Twenty-four university undergraduates participated in this experiment in exchange for course credit. Participants were randomly assigned to the pedagogical or the communication scenarios.

Procedure: In the pedagogical situation, people were presented with a series of questions asking them to decide which of two teachers they would like to learn from in the future. Each teacher was presented as teaching about a novel enzyme, e.g. "Teacher 1 is teaching about enzyme P23T." The names of the enyzmes were random combinations of letters and numbers. This was followed by lists indicating which of the eight animals had the enzyme and which did not. Each question contrasted two teachers, where teachers differed in the generality of the properties taught. For instance, one teacher might teach a property that was true of owls, ostriches, leopards, and seals, but not of grasshoppers, ants, iguanas, and frogs, while the other was teaching a property that was true of all eight animals. Paired teachers always taught concepts where one was a subset of the other, so the more generalizable concept included all of the positive examples of the property in the less generalizable concept, with additional positive examples (e.g. ostrich versus ostrich and owl). Participants indicated which teacher they would rather have teaching them about new enzymes using a Likert scale ranging from -10 to 10, where the extremes indicated the teacher on the left or the right and zero indicated indifference. Participants rated all possible pairings of teachers, resulting in a total of 34 questions. Order of the questions, as well as the side (right or left) of the more general concept, were randomized.

The communication condition was identical to the pedagogical condition, with the exception of some of the wording. Instead of teaching about enzymes, the situations described people who were talking about enzymes. For example, "Person 1 is talking about enzyme P23T." Additionally, participants were asked to provide ratings about which one they would rather talk to in the future. Otherwise, the questions and response sheets were identical.

Results & Discussion

We coded people's judgments as positive if they were in the direction of the more generalizable teacher and negative if they were in the direction of the teacher with the less general property. To test whether people expected more general properties, we compared the average ratings to chance (zero). In the pedagogical condition, people chose the teacher with the more general information, $mean = 0.66, t(407) = 2.06, p < 0.05$. In contrast, in the communication condition, people choose the less general information, $mean = -0.56, t(407) = -2.09, p < 0.05$. The difference between the two conditions was significant, $t(814) = 2.84, p < 0.01$. These results suggest that people expect that more general properties will be taught in pedagogical situations, in contrast with communicative settings, where people expect less general properties.

To follow up on these results, we investigated the pat-

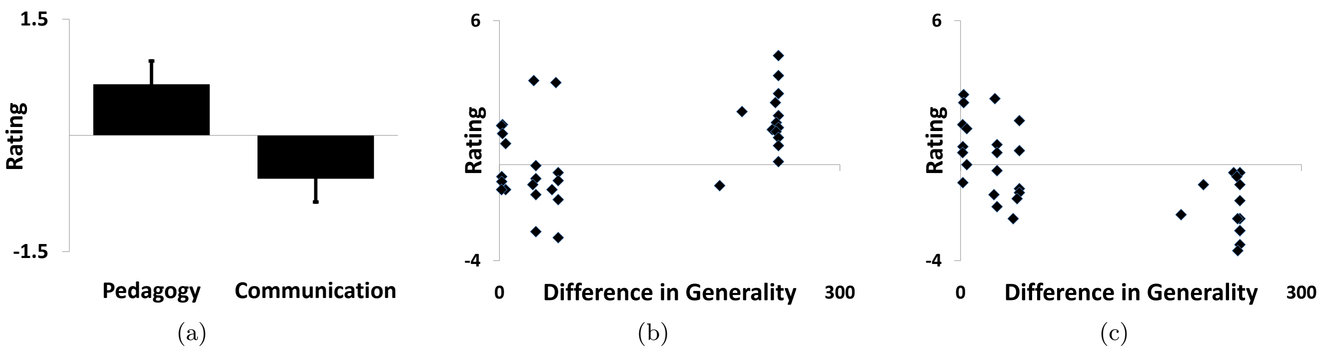


Figure 2: Experiment 1 results: (a) Average human ratings in the pedagogy and communication conditions. Positive ratings indicate the more generalizable teacher. (b) Scatterplot showing the relationship between the difference in generalizability for pairs of teachers (x axis) and people's ratings (y axis) for the pedagogy condition. The strength of people's ratings increases with an increasing difference in generalizability, $r = 0.51$, indicating that they expect more generalizable concepts in pedagogical settings. (c) Scatterplot showing the relationship between difference in generalizability and people's ratings for the communication condition. The strength of ratings decreases with an increasing generalizability, $r = -0.66$, indicating that people expect less generalizable concepts in communicative settings.

tern of ratings for individual items. If people choose more generalizable concepts, then pairs for which there was the greatest gap between the more and less generalizable teacher should have the strongest ratings. To investigate this question, we needed to quantify how general each hypothesis was. We consider two possible measures of generality: the number of positive examples, and the sum of the distances among items in the tree. To test whether ratings indicated an expectation that properties would be generalizable, we collapsed individual judgments into a single average rating for each question, resulting in 34 ratings. To investigate which measure of generality best predicted people's judgments, we conducted a stepwise regression with item averages as the dependent variable. The independent variables included the number of positive examples in more general concept, the number of positive examples in the less general concept, the difference in number of positive examples, as well as the summed distances for the more and less general concepts, and the difference in the summed distance. The two difference scores allowed us to test whether people's judgments take into account both teachers, or just a single teacher when making their judgments. Stepwise regression greedily selects the variable that accounts for the greatest variance, and iterates until no variables account for significant variance. Analysis of the pedagogy condition showed that the difference in summed distances accounted for the greatest variance, $r = 0.51$, $F(1, 32) = 11.49, p < 0.01$, and that no other variables accounted for significant residual variance. The correlation indicates that the bigger the difference in generalizability was, the stronger people's ratings were toward the more generalizable teacher. In contrast, regression analyses on the communication condition showed that while the difference in summed distance was a significant predictor of ratings, the relationship was negative, $r = -0.66$, $F(1, 32) = 24.52, p < 0.001$. This suggests that in communicative settings, people's expectations about generalizability are the opposite of their expectations in pedagogical settings.

The number of positive examples, while a straightforward measure of generality for this task, is undesirable for two reasons. First, if this leads to an accurate characterization of people's inferences, then one might wonder to what degree the results are a consequence of task demands (given that people were answering questions about lists of animals). Second, the number of positive examples is not a very good measure of generality because it bears no necessary relationship with actual semantic generalizability. As can be seen in Figure 1, many possible sets with the same number of positive examples differ markedly in their coverage of the tree. Instead, we prefer to measure the generalizability of a concept in terms of the sum of distances between all pairs of positive examples. This provides a measure that is not subject to task demands, and is related to the semantic generality of the concept. Our analyses show that distance in the tree provides a better description of people's behavior, providing evidence that people's judgments do not simply reduce to task demands, and that their judgments are based on semantic generalizability.

It appears that people have strong prior expectations that they bring specifically to pedagogical situations. In pedagogical situations, learners expect that teachers will choose to teach generalizable information. In contrast, when in communicative situations, people expect that speakers are likely to talk about specific information. Our analyses showed that people's judgments are better predicted by distance in a semantic tree, consistent with a bias toward semantically generalizable information.

Experiment 2: Testing the bias toward semantic coherence

Experiment 2 investigated whether people have an expectation that teachers will choose semantically coherent concepts. To investigate this, we presented participants with scenarios in which two teachers each taught concepts with two positive exemplars. The semantically coherent teacher taught a property that was true of two tree-consistent exemplars, such as owl and ostrich. They were contrasted with a semantically incoherent teacher who taught a property that was true of two tree-inconsistent exemplars, such as ostrich and leopard. If people expect teachers to teach semantically coherent concepts, we expect to find that people choose the teacher who teaches tree-consistent properties.

Methods:

Participants: Twenty university undergraduates participated in this experiment in exchange for course credit.

Procedure: The procedure was identical to that used in Experiment 1 with the exception of the questions used. Each scenario provided information taught by two teachers. All properties were true of two animals, but were absent in the other six. In each scenario, one teacher taught a property that was semantically coherent – it was consistent with the structure of the tree – and the other taught a property that was semantically incoherent – it was inconsistent with the structure of the tree. For instance, a semantically coherent property might be true of owls and ostriches, but no other animals. Contrarily, a semantically incoherent property might be true of owls and leopards but no other animals. Questions were designed such that semantically coherent pairs were contrasted with all minimally different semantically incoherent pairs that overlapped one animal. For example, owls and ostriches were contrasted with owls and leopards, owls and seals, ostrich and leopards, and ostrich and seals. This resulted in a total of 16 questions. Order of the questions, as well as the side of the semantically coherent pair (left or right), were randomized.

Results & Discussion

Do people expect teachers to teach semantically coherent concepts? To address this question, we coded people's ratings as positive numbers if they were in the direction of the semantically coherent teacher, and negative numbers if they were not. We then ran separate t-tests comparing the means in the pedagogical and communicative conditions to zero. In the pedagogical condition, people tended to choose teachers of semantically coherent concepts, $mean = 0.97, t(159) = 2.04, p < 0.05$, one-tailed. In the communication condition, people also chose teachers of semantically coherent concepts, $mean = 2.21, t(159) = 6.76, p < 0.001$. The difference between the two conditions was also significant, $t(308) = 2.16, p < 0.05$.

To further investigate the role of semantic coherence, we computed the distance between all of the positive examples in each scenario (see Figure 1). If people expect semantically coherent concepts, then more semantically coherent pairs – those with shorter distances – should have the strongest ratings. We ran a stepwise regression with people's ratings as the dependent variable, and independent variables including distance between the positive examples in the more and less coherent sets, and the difference in the distances. For the pedagogical condition, the distance between positive examples in the coherent concept was the only predictor selected, $r = -0.70, F(1, 14) = 13.24, p < 0.01$. Of the coherent hypotheses, the teachers teaching the more coherent concepts were rated more strongly. For the communication condition, regression analyses showed that distance between positive examples of coherent pairs did not strongly predict people's ratings, $r = 0.23, F(1, 14) = 0.80, p > 0.3$. [3]

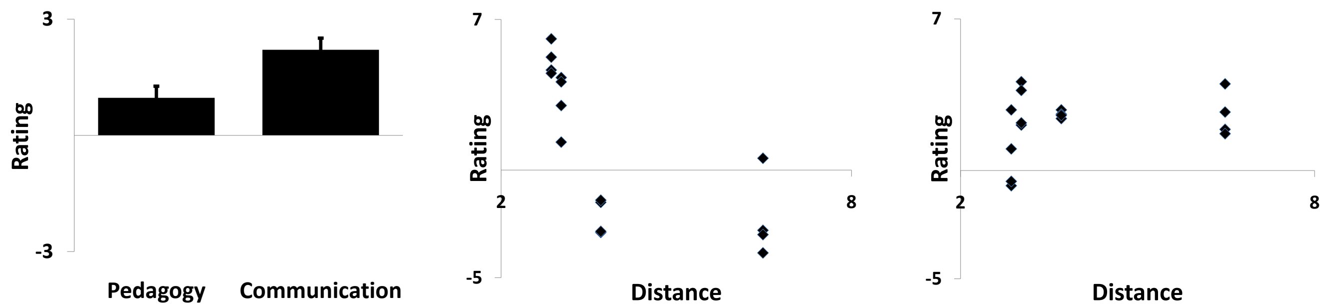
Interestingly, unlike in Experiment 1, people's judgments in Experiment 2 were best predicted by the coherence of the more coherent hypothesis alone (as opposed to the difference in coherence). This suggests that the semantically incoherent hypotheses did not play a large role in people's judgments. This may reflect an explicit judgment that these cases are so unexpected that they, in effect, have zero weight.

The evidence suggests that people expect teachers to teach semantically coherent concepts: overall, people chose teachers of more semantically coherent concepts, and the strength of people's ratings decreased as the strength of coherence decreased. The evidence also suggests that people's expectation of coherence may apply across more than just pedagogical situations. Results from the communication condition showed that people tended to choose the more coherent speaker, but the strength of their ratings was not related to the degree of coherence. These results suggest that people's expectation of semantic coherence may not be limited to pedagogical situations.

Discussion

Pedagogical situations play a central role in human learning. In pedagogical situations, teachers choose which concepts to teach and which examples to use to teach the concept. We have presented an extension of Shafto and Goodman's (2008) model of pedagogical data selection that incorporates specific expectations about pedagogical situations. Using this framework, we have derived a method for isolating the effects of prior expectations about pedagogical situations. The results of Experiment 1 showed that people expect teachers to provide generalizable knowledge, and that this expectation does not apply in more general communicative settings. The results of Experiment 2 showed that people expect teachers to provide semantically coherent information, although this appears not to be specific to pedagogical situations. Taken together, these results provide evidence that people have specific expectations—intuitive

[3] A separate stepwise regression showed that none of the independent variables accounted for significant variance in people's judgments.



(a) (b) (c)

Figure 3: Experiment 2 results: (a) Average human ratings in the Pedagogy and Communication conditions. Positive ratings indicate the more semantically coherent teacher. (b) Scatterplot showing the relationship between distance among positive examples of coherent hypotheses (x axis) and people's ratings (y axis) in the pedagogy condition. People's ratings increase with decreasing distance, $r = -0.70$, suggesting that people expect coherent hypotheses in pedagogical settings. (c) Scatterplot showing the relationship between distance among positive examples of coherent hypotheses (x axis) and people's ratings (y axis) in the communication condition. People's ratings are only weakly related to distance, $r = 0.23$.

theories of pedagogical situations.

Our results provide additional evidence in support of Csibra and Gergely's (2009) claim that people expect generalizable information in pedagogical contexts. Where previous results focused on young children, our results suggest that this expectation continues into adulthood. Our results also provide evidence that semantic coherence, while expected in pedagogical situations, is not specific to these contexts. Rather, the expectation of semantically coherent concepts extends to communicative, as well as pedagogical situations.

Here we have focused on learners' expectations, but for these pedagogical expectations to be reasonable, it is important that teachers meet their expectations. Specifically, do people choose to teach concepts that are more generalizable and more coherent? If so, what are the implications of these matching (or mismatching expectations) in terms of the kinds of concepts that can be learned, the speed at which they are acquired, and the robustness of knowledge transmission? Future research will aim to answer these questions.

Our experiments have provided information about people's prior expectations in pedagogical situations, but it is also important to explain why people have these biases. There is work to be done in formalizing computational models that explain why certain hypotheses would be more or less likely to be taught. This may not turn out to be entirely straightforward because while there is a reasonable motivation for teaching generalizable concepts, there are also motivations for teaching other kinds of concepts. For instance, one might also want to teach sparse concepts because they may be difficult to discover on one's own. Further empirical research may help narrow down the possibilities and provide guidance for more explanatory models.

More generally, previous approaches to modeling human learning have focused on a single unitary set of prior expectations that apply generically across situations (but see Shafto et al., 2006). However, this approach seems obviously too simple. We all intuitively understand that we have different expectations that apply when, for example, we talk to children as opposed to adults. Pedagogical situations are but one case of a more general problem. Understanding how social situations affect learning will require understanding how different contexts affect both learners' prior expectations and learners' assumptions about how information is selected.

Acknowledgments

Thanks to Russell Warner and Carissa Shafto for helpful comments and suggestions during the writing process.

References

Csibra, G. (2007). Teachers in the wild. *Trends in Cognitive Sciences*, 11:95–96.

Csibra, G. and Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 14:148–153.

Kemp, C. and Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, 105:10687–10692.

Luce, R. D. (1959). *Individual choice behavior*. John Wiley, New York.

Shafto, P. and Goodman, N. D. (2008). Teaching games: Statistical sampling assumptions for pedagogical situations. In *Proceedings of the 30th annual conference of the Cognitive Science Society*.

Shafto, P., Kemp, C., Mansinghka, V., Gordon, M., and Tenenbaum, J. B. (2006). Learning cross-cutting systems of categories. In *Proceedings of the 28th annual conference of the Cognitive Science Society*.

Topal, J., Gergely, G., Miklosi, A., Erdohegyi, A., and Csibra, G. (2008). Infants perseverative search errors are induced by pragmatic misinterpretation. *Science*, 321:1831–1834.