Translating testimonial claims into evidence for category-based induction

Tamar Kushnir (tk397@cornell.edu)
Cornell University, Department of Human Development, Martha Van Rensselaer Hall
Ithaca, NY 14853 USA

Susan A. Gelman (gelman@umich.edu)
University of Michigan, Department of Psychology, 530 Church Street
Ann Arbor, MI 48109 USA

Abstract

Inductive generalizations about the properties of *kinds* are based on evidence. But evidence can come either from our observations, or from the testimony of knowledgeable informants. The current study explores how we combine information from these two sources to make inductive inferences. Participants learned about a novel object category, and observed the property occur with some frequency in a sample of category members. Different groups of participants also heard an informant making either Generic, Quantified, or Specific claims about the prevalence of the property. Participants who heard generic claims were more resistant to a straightforward use of statistical evidence in their generalizations. Moreover, participants who rated the informant as more knowledgeable (across conditions) gave higher prevalence estimates. The results suggest two pathways through which testimony translates into evidence for category learning, and raise questions on how to best combine evidence from these different sources into a common representational form.

Keywords: category-based induction; probabilistic reasoning; generics; rational models; testimony; epistemic trust

Introduction

It's nearly impossible to learn something new without in some way generalizing what you learned. This core feature of cognition – that our experiences teach us about *kinds* of things and not just individual instances – supports the inferences, predictions, and explanations that help us make sense of the world around us. Much empirical research – guided by formal models based on Bayesian inference – suggests that these generalizations are principled and rational (Gopnik & Wellman, 2012; Oaksford & Chater, 2007; Schulz, 2012; Tenenbaum & Griffiths, 2001); that is, they are both based on our current conceptual knowledge and appropriately responsive to new evidence. This process of rational (i.e. Bayesian) updating is how we are able to get so much inductive power from so little new information.

The rational models approach has empirical support from concept learning studies in infants (Denison, Trikutam & Xu, 2014; Teglas et al, 2011), children (Kushnir & Gopnik, 2007; Schulz, Bonawitz & Griffiths, 2001), and adults (Griffiths et al, 2011; Griffiths & Tenenbaum, 2009). In all of these studies, evidence presented to participants is statistical (associations that are observed to occur with some frequency) and as such can be represented as probability distributions. But new information comes to us in many forms, and, as human learners, the process of translating evidence into probabilistic representations may not always be as straightforward as these studies would imply.

Consider the following example. My best friend just visited an exotic island and brings back with her two Ylang-Ylang seeds. One seed grows into fragrant flowers, and the other grows into flowers with no smell at all. A reasonable generalization, based on my observations and no other information, might be to assume that these plants (in their natural habitat) are fragrant about half of the time.

Consider, instead, what would happen if she brought home stories from her travels instead of seeds, and told me about the Ylang-Ylang flowers she saw with an amazing fragrance. What can I conclude from these stories? Based on her general description, and the fact that I trust my friend is telling the truth, I'd probably assume that all (or most) of the plants had a lovely smell.

The example illustrates two very different ways of translating information about a new category and its properties into probabilities: One is by sampling from the category and observing the prevalence of that property in the sample (i.e., the statistical likelihood of a member of the sample having said property) then treating the sample as representative of the population. Both adults and children do this (Gweon, Tenenbaum & Schulz, 2010; Xu & Tenenbaum, 2007).[1]

Another is by believing the testimony of trusted sources. Especially when we are young, but also when we are older, we rely on others to impart generalities by making kind-based claims (Cimpian & Markman, 2009; Gelman et al, 1998; Koenig et al., 2015). Arguably, much of our conceptual knowledge is acquired in this way, rather than through direct sampling, because many of the deep, non-obvious, and essential properties of categories are not directly observable (Gelman, 2003). Thus, it remains a

[1] We leave aside for now, specifics on the way the sample was generated. In our example, it could have been a random sample (she happened to pick some flowers while on vacation), a sample chosen intentionally (she liked those flowers specifically, so she picked them) or a sample chosen pedagogically (she wanted to share information about their fragrant properties, so she picked an representative set to show me). For details on how rational models account for the sample-generating process in inductive generalization, see Shafto, Goodman & Frank, 2012; Shafto, Goodman & Griffiths, 2013.

critical question: does it make sense to straightforwardly translate new information about kinds garnered through testimony into probability estimates?

There is some evidence that generic claims ("Ylang-Ylang flowers have a lovely smell") can be used as the basis for probability estimates. Novel generic claims lead people to make high estimates of the prevalence of a property in the category, and to generalize broadly to new category members (Brandone, Gelman & Hedglen, 2015, Cimpian, Brandone & Gelman, 2010; Rhodes, Leslie & Tworek, 2010). However, there are plenty of cases in which generic claims could lead to estimates of *low* probability (as in the claim "Mosquitoes carry West Nile virus"). Morover, both children (Brandone, Cimpian, Leslie, & Gelman, 2012) and adults (Cimpian, Gelman, & Brandone, 2010; Prasada, Khemlani, Leslie & Glucksberg, 2013) view generic claims as distinct from claims about quantities. Thus, with the addition of even a little prior knowledge (such as the fact that West Nile is a rare disease) it is easy to see how a generic claim could be recast as a probability estimate.

Are more precise, quantified ("some/most/all Ylang-Ylang flowers have a lovely smell") statements about kinds easier to translate into probabilities? Perhaps. But logically, absolute claims ("all Ylang-Ylang flowers have a lovely smell") fail to be true after even *one* exception, and, a discovery of that exception may call into question the credibility of the person making the claims. This would present its own complication: how do we create probability estimates from *unreliable* testimony?

The current study represents a first attempt at addressing these less-than-straightforward cases. We use a simple design to teach participants about a novel object category and a property (a causal property, discoverable, but not immediately obvious) of a sample of category members. Participants observed the property occur with some frequency in the sample. Two groups also heard an informant making category-based claims – either Generic or Quantified (i.e. "all") – about the property. We compared these two groups with a group of participants who heard the informant make a Specific (and accurate) claim about the property of only one of the objects in the sample.

Our central aim was to investigate how participants would integrate the statistical data with category-based testimonial claims into estimates of property prevalence. Most straightforwardly, participants should take the observed frequency of the property in the sample as representative of the prevalence of the property in the category as a whole. However, based on prior work, we might expect two differences unique to generic testimonial claims. On one hand, given the flexibility of generic claims (that they allow for exceptions, and can be true for low-prevalence properties), participants who hear generic claims may not rely as much on observed frequency in their prevalence estimates. On the other hand, given that generic claims are often taken as indicating high prevalence, participants who hear generic claims may over-estimate the prevalence of the property across the range of observed frequencies.

Our second aim was to investigate how knowledge attributions are influenced by the nature of the testimonial claims made, and in turn how these attributions influence inferences about the category and its properties. In our study, observed frequencies afford participants the opportunity to *verify* testimonial claims. Thus, most straightforwardly, specific (and verifiably accurate) claims made about a single object should lead to high knowledge attributions. On the other hand, category-based claims that are not well-matched by observed frequencies (e.g. a general claim about the property, but a low observed frequency in the sample) may lead to low knowledge attributions. This might be particularly true when the observations logically contradict the claim (as in the absolute, quantified "all" claims). But, it may also be true of generic claims (as they suggest high prevalence). On the other hand, given the flexibility of generic claims to varying interpretations, knowledge attributions following such claims may remain high. These knowledge attributions may be an additional pathway through which category-based claims influence prevalence estimates; perhaps leading people to assign greater weight to testimony that is seen as coming from a knowledgeable source.

Methods

Participants

Nine-hundred-and-thirty-three adults (512 female, 412 male, age range 16yrs – 86yrs, mean age = 37.37, SD = 12.16) participated in the study through Amazon Mechanical Turk in exchange for monetary compensation. Data was collected on US participants only. Participants were majority white (79%) and non-Latino/a (92.9%). A majority had attended college (35% with a 4-year degree, 12.9% with a 2-year degree, 27.5% some college).

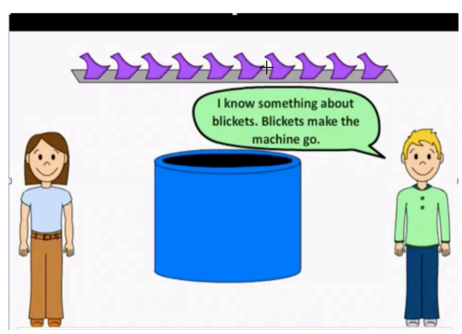

Figure 1: A sample of the video seen by participants on trial 1 in the *Generic Language* condition. In this condition, the cartoon figure on the right, a character named "Zorg," makes a claim about the objects before they enter the "special music machine."

Procedure

A random number generator assigned each participant to one of 30 conditions, for a total of 30-33 participants per condition. Each condition combined one of 3 types of testimonial claims (Specific, Generic, Quantified) and one of 10 observed frequencies of the novel causal property (ranging from 1/10 to 10/10 inclusive).

The procedure consisted of two trials, each with the same characters, observed frequency, and testimony, but with different objects (Trial 1: “blickets” and Trial 2: “midos”) and a differently-colored machine. On each trial, participants first watched a short video then answered a set of questions. A screen shot of the first video is shown in Figure 1. The video began with a girl who introduced the “special music machine,” her friend Zorg, and an array of objects on a shelf that she labeled (e.g. “Here are some blickets” in Trial 1). Zorg then made one of the following testimonial claims, depending on condition:

Specific claim: *“I know something about this blicket. This blicket makes the machine go.”*[2]

Generic claim: *“I know something about blickets. Blickets make the machine go.”*

Quantified (“All”) claim: *“I know something about all blickets. All blickets make the machine go.”*

After hearing Zorg’s claim, the video showed the 10 objects passing through the machine (they went into the top and came out of the bottom) one at a time. When the machine “activated” it changed color and made a loud trumpeting sound. The first object always had the property of activating machine insuring that, in the Specific condition, Zorg’s claim was accurate. After the first object, a random subset of the remaining objects activated the machine depending on the observed frequency assigned.

Following the video, participants made two knowledge attribution ratings: A **Category Knowledge Attribution** (*How much do you think Zorg knows about blickets?)* and a **Machine Knowledge Attribution** (*How much do you think Zorg know about the machine?).* Answer choices for both were on a three-point scale: 0 – Nothing, 1 – A little bit, 2 - A lot. Participants also answered a **Property Prevalence Estimate**: *“Imagine there were more blickets here. What percentage of these blickets would make the machine go?* Participants were allowed to enter numbers ranging from from 0-100. Following the questions for Trial 1, participants began Trial two by watching the second video and answering the same set of questions about “midos.”

Results

Property Prevalence Estimates: We examined the influence of Trial (within participants), Observed Frequency (between participants), Testimony (between participants), and the interaction of the two manipulated variables on Prevalence Estimates using a linear mixed effects model.[3] Parameter estimates for the model are shown in Table 1. There was no significant effect of Trial ($F(1,1836)=.02$, *ns*). There was a significant main effect of Observed Frequency ($F(1,1836)=2034.6$, $p<.001$) such that estimates increased as frequency of activation increased, regardless of the type of testimony heard. There was also a main effect of Testimony ($F(2,1836)=96.4$, $p<.001$). Prevalence estimates were on average highest after hearing Quantified (i.e. “All”) claims (Mean=58%, SD=31%), and on average 3.3% above than the observed frequencies in each condition (SE=1.16, $t(297) = 2.84$, $p<.01$). Estimates were next highest after hearing Generic claims (Mean=54%, SD=30%) and were not different on average from observed frequencies in each condition (M=-.12, SE=1.33, $t(311) = -0.09$, *ns*). Estimates were lowest after hearing Specific claims (Mean=52%, SD=30%) and were on average 4.18% lower than the observed frequencies in each condition (SE=1.06, $t(302) = -3.94$, $p<.001$).

Finally, there was a significant interaction between Testimony and Frequency ($F(2,1836)=4.8$, $p<.01$). An illustration of this interaction can be seen in the graph of the model predicted values in Figure 2. In the Generic conditions, the slope of prevalence estimates was significantly flatter than in the other two conditions (see parameter estimate in bold in Table 1): they were on average higher for low-frequency properties than in the Specific condition and lower for high-frequency properties than in the Quantified condition.

Table 1: Parameter Estimates of the mixed model predicting Prevalence Estimates by condition

Parameter	Estimate	SE	*t*
Intercept	5.01	1.91	2.62**
Slope of Specific condition	.84	.03	27.38***
Intercept difference between Generic and Specific	10.13	2.67	3.79***
Intercept difference between Quantified and Specific	7.08	2.70	2.62**
Slope difference between Generic and Specific	**-.12**	**.04**	**-2.69****
Slope difference between Quantified and Specific	.00	.04	.02

*p<.05; **p<.01, ***p<.001

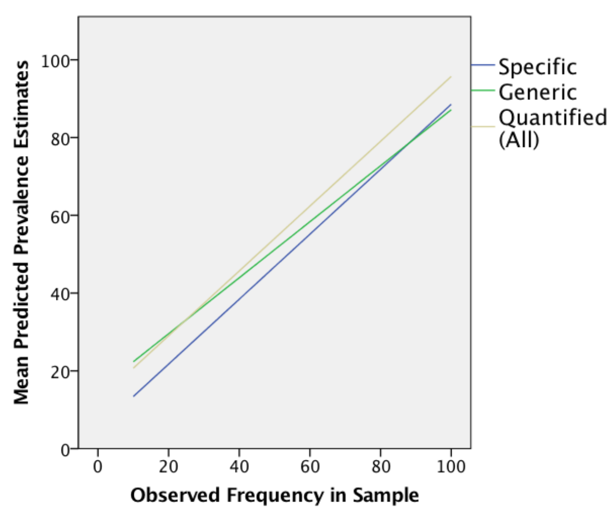

Figure 2: An illustration of the main effects of Frequency, Testimony, and interaction between the two on participants’ estimates of property prevalence.

[2] In the Specific claim condition, the blicket Zorg refers to is always the one closest to him, and as he speaks he “points” to it with a dashed line connecting his hand to the object.

[3] A Q-Q plot of the prevalence estimates showed that they were approximately normally distributed.

Object Knowledge Attributions: We then examined the influence of Trial, Observed Frequency, Testimony, and their interaction on attributions of Category Knowledge (e.g. "How much does Zorg know about blickets?") using an ordinal logistic GEE. There was no significant effect of Trial (Wald χ^2 (1) = 2.5, *ns*). There was a significant main effect of Observed Frequency (Wald χ^2 (1) = 59.4, p<.001), such that attributions of knowledge increased as the frequency in the sample increased. There was also significant main effect of Testimony (Wald χ^2 (2) = 38.2, p<.001): attributions of knowledge were highest in the Specific condition (Mean = 1.34, SD= .55), next highest in the Generic condition (Mean=1.25 SD=.54) and lowest in the Quantified condition (Mean=1.09 SD=.58).

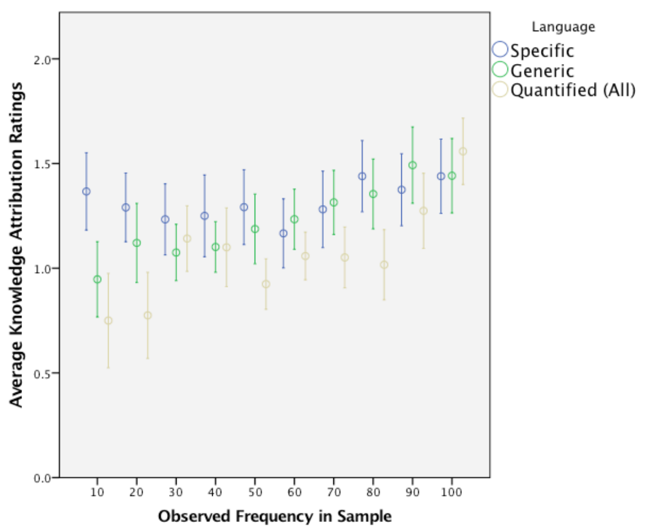

Figure 3: Average knowledge attributions (Category and Machine Knowledge Questions, Trials 1 and 2) by Observed Frequency and type of Testimony. Error bars represent 95% confidence intervals of the mean.

Critically, there was a significant interaction between Observed Frequency and Testimony (Wald χ^2 (2) = 17.2, *p*<.001). Parameter estimates reveal the nature of the interaction: while the slope in the Specific condition was not significantly different from 0 (Wald χ^2 (1) = 2.1, *ns*) the slopes in the Generic and Quantified conditions were significantly different from the Specific condition (Generic vs Specific: Wald χ^2 (1) = 15.3, *p*<.001; Quantified vs Specific: Wald χ^2 (1) = 9.6, *p*<.01) and significantly positively correlated with Observed Frequency.

Object Knowledge Attributions: We ran the same analysis on attributions of Machine Knowledge ("How much does Zorg know about the Machine?"). There was no significant effect of Trial (Wald χ^2 (1) = .2, *ns*). There was again significant main effect of Observed Frequency (Wald χ^2 (1) = 46.1, p<.001), a significant main effect of Testimony (Wald χ^2 (2) = 26.8, p<.001), and a significant interaction (Wald χ^2 (2) = 12.2, p<.01). The interaction followed the same pattern as above: the slope in the Specific condition was not significantly different from 0 (Wald χ^2 (1) = 1.8, *ns*) the slopes in the Generic and Quantified conditions were significantly different from the Specific condition (Generic vs Specific: Wald χ^2 (1) = 9.3, *p*<.01; Quantified vs Specific: Wald χ^2 (1) = 8.8, *p*<.01) and significantly positively correlated with Observed Frequency.

Together these results suggest that, when Zorg made claims about a specific object, participants' knowledge attributions did not depend on observed frequency. However, when Zorg made category-based claims about the properties of the objects, participants' knowledge attribution ratings increased as observed frequency of the property in the sample increased. This interaction is illustrated in a graph of the average knowledge attributions across both types of questions (Figure 3).

Knowledge Attribution Influences on Prevalence Estimates: Our final question concerned the influence of knowledge attributions on inductive generalizations. We addressed this question by adding knowledge attributions to our mixed effects model from Table 1 to see if either type of knowledge attribution (Category Knowledge or Machine Knowledge) predicted prevalence estimates above and beyond the main effects and interactions shown. Only the main effect of Category Knowledge attributions was significant (F(2,1832)= 23.16, p<.001). An illustration of the knowledge attribution effect is shown in Figure 4. Participants who rated Zorg as less knowledgeable about blickets/midos gave lower than average prevalence estimates. Participants who rated Zorg as more knowledgeable about blickets/midos gave higher than average prevalence estimates. Partial correlations, controlling for observed frequencies, show that the relationship was strongest in the Quantified (r=.45, p<.001) and Generics conditions (r=.40, p<.001), and still significantly positive but less strong in the Specific condition (r=.14, p<.05).

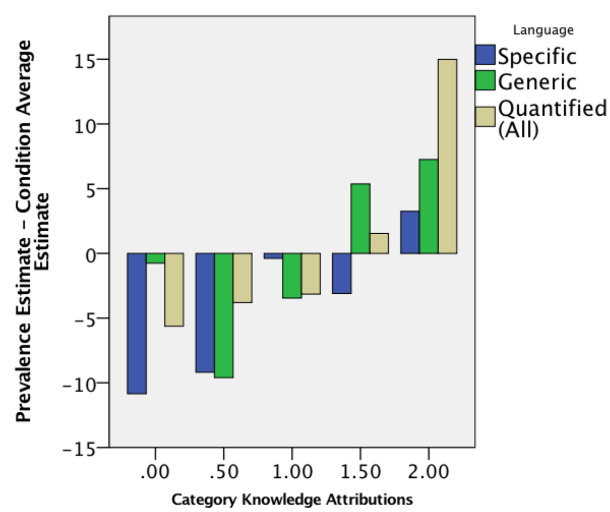

Figure 4: An illustration of the effect of Category Knowledge Attributions on prevalence estimates. Y axis shows the difference between actual estimates and the condition average (to account for condition differences found in prior analyses).

General Discussion

New information comes to us in different ways; we may make our own observations, or we can believe what others tell us. How do we integrate evidence from these different sources into a common representational form? The present study begins to address this broad question. We designed a category learning study in which participants heard a testimonial claim about a category and one its properties. They then observed the property with some frequency in a sample of category members. Participants were asked to translate this evidence into an estimate of the prevalence of the property in the category as a whole, and also asked to attribute knowledge to the informant who made the testimonial claim.

To begin with, it should be noted that participants in our study responded rationally to the statistical evidence from the sample they observed. Low observed frequencies led to low prevalence estimates, and high observed frequencies led to high prevalence estimates. However, testimony influenced this basic and well-replicated result.

Generic claims ("*blickets* make the machine go") had a unique influence on prevalence estimates compared to the other types of testimony. After hearing generic claims about the category and its novel property, participants' estimates of low prevalence properties were higher, and their estimates of high prevalence properties were lower. This explains why the estimates were, on average, well-matched to the observed frequencies. But it also suggests that, at the extremes, these claims would be less well matched. Thus, we offer one intriguing interpretation of this "flattened" relationship: that generics have the effect of making people *resistant* to straightforward use of statistical evidence in their inductive inferences about categories.

Generic claims also led to some degree of skepticism about the knowledge of the informant, in particular when the observed frequency of the property was low. This is consistent with prior work suggesting that generic claims are often interpreted as indicating high prevalence (Cimpian et al., 2010). So, observing high frequencies helps verify the credibility of these claims, and observing low frequencies calls their credibility into question.

Quantified, absolute claims ("*all blickets* make the machine go") led to prevalence estimates that were on average higher than straightforwardly predicted by the sample. Absolute claims also led to the greatest degree of skepticism about the knowledge of the informant. But this skepticism did not seem to be based on a straightforward logical relation between the sample and the claim; if it were, then any frequency of less than 100% would have overwhelmingly led to responses that Zorg knows "nothing" or "little" about the category (which it did not). Rather, the relation between knowledge attributions and frequency paralleled the results in the Generic condition, suggesting that perhaps a documented tendency to interpret quantified claims as generic (Leslie & Gelman, 2012) played a role. This possibility, and other potential influences on how we use frequency evidence to verify the reliability of logical claims, would be interesting to explore in future work.

Non category-based claims about a specific object ("*this blicket* makes the machine go") also influenced prevalence estimates; they led to estimates lower than straightforwardly predicted by the observed frequency. Knowledge attributions in this condition tended to be high, and were independent of the observed frequencies. This suggests that participants based their knowledge attributions on the accuracy of the specific claim, disregarding evidence (i.e. the properties of the remaining objects) irrelevant to verifying them.

The final result was that, across all types of testimony and all observed frequencies, knowledge attributions positively influenced prevalence estimates. Thus, all else equal, participants who were skeptical gave lower prevalence estimates relative to others in the same condition, and participants who were convinced gave higher relative prevalence estimates.

All together, the results suggest two pathways through which testimony translates into evidence for category learning. Claims about categories and their properties lead to adjustments to the way observed frequencies translate into general category knowledge. And, the credibility of claims about categories and their properties themselves can be verified through observation, and subsequent skepticism or belief in these claims leads to further adjustments to inductive generalizations.

The influence of generic testimony in our study has particularly important implications for a principled, rational account of category learning, inductive inference, and belief revision. Generics may, as stated, make us resistant to straightforward interpretations of probabilistic data. They may also, by extension, lead to a greater resistance to counter-evidence in belief revision. Generic *speakers* also have the advantage of being judged as knowledgeable despite counter-evidence. In support of this idea, other work has shown that generic claims lead to high estimates of knowledge even when they are *not* verifiable (Koenig et al., 2016). Thus, what may make generics so influential in category learning is that they are resistant to statistical evidence by both pathways.

Do these findings pose a problem for the rational view? Not necessarily. Our preliminary findings suggest that evidence from testimony doesn't completely discount our own observations, but rather can be integrated with them. The details of the integration may depend in part on the type of testimony (Generic, Specific, and Quantified among others). It may also be expected to interact with domain-specific knowledge, and additional types of evidence. Many of these details could be worked out in future empirical and computational work.

Our data suggests that perhaps not all of the interesting nuances in human learning and inductive inference are explained by differences in prior knowledge, or by assumptions made about the evidence-generating mechanism (e.g. strong vs. weak sampling, teacher vs.

learner driven, intervention vs. observation). Instead we suggest that evidence from *what people say* is not straightforwardly reducible to these distinctions. When we relax our assumption that evidence comes in only one form, we can come closer to understanding the potential of human learning.

Acknowledgments

We'd like to thank Alexandra Was and Natalie Davidson for help creating stimuli and with data collection. Thanks to three anonymous reviewers, and to Felix Thommes for helpful comments on the revised version. This research was supported by funding from NSF (DLS #1023179) to T.K. and NICHD grant HD-36043 to S.G.

References

Brandone, A. C., Cimpian, A., Leslie, S., & Gelman, S. A. (2012). Do lions have manes? For children, generics are about kinds rather than quantities. *Child Development,* 83(2), 423-433.

Brandone, A. C., Gelman, S. A., & Hedglen, J. (2015). Children's developing intuitions about the truth conditions and implications of novel generics versus quantified statements. *Cognitive Science*, *39*(4), 711-738.

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science,* 34(8), 1452-1482.

Cimpian, A., Gelman, S. A., & Brandone, A. C. (2010). Theory-based considerations influence the interpretation of generic sentences. *Language and Cognitive Processes,* 25(2), 261-276.

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*(1), 14-25.

Denison, S., Trikutam, P., & Xu, F. (2014) Probability versus representativeness in infancy: Can infants use naïve physics to adjust population base rates in probabilistic inference? *Developmental Psychology.* 50(8):2009-19.

Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A., & Keil, F. C. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, i-157.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.

Gopnik A, & Wellman H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085-1108.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science, 35*, 1407-1455.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661-716.

Gweon, H., Tenenbaum, J., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107*(20), 9066-9071.

Koenig, M. A., Cole, C. A., Meyer, M., Ridge, K. E., Kushnir, T., & Gelman, S. A. (2015). Reasoning about knowledge: Children's evaluations of generality and verifiability. *Cognitive Psychology, 83*, 22-39.

Leslie, S., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, *64*(3), 186-214.

Oaksford, M., & Chater, N. (2007) *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Prasada, S., Khemlani, S., Leslie, S., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition,* 126(3), 405-422.

Rhodes, M., Leslie, S., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, 109*(34), 13526-13531.

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*, 382-389.

Schulz, L.E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared make your tummyache? Naive theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology, 43*(5), 1124-1139.

Shafto, P., Goodman, N. D., & Griffiths, T. L (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71*, 55-89.

Teglas, E., Girotto, V., Gonzalez, M., & Bonatti, L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences, 104*, 19156–19159.

Xu, F., & Tenenbaum, J.B. (2007) Word learning as Bayesian inference. *Psychological Review, 114*, 245-272.