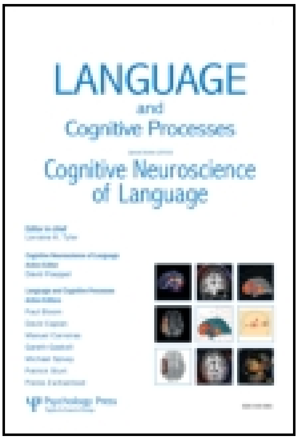

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Inferences about members of kinds: The generics hypothesis

Sangeet Khemlani [a], Sarah-Jane Leslie [b] & Sam Glucksberg [a]
[a] Department of Psychology, Princeton University, Princeton, NJ, USA
[b] Department of Philosophy, Princeton University, Princeton, NJ, USA
Version of record first published: 25 Oct 2011.

To cite this article: Sangeet Khemlani, Sarah-Jane Leslie & Sam Glucksberg (2012): Inferences about members of kinds: The generics hypothesis, Language and Cognitive Processes, 27:6, 887-900

To link to this article: http://dx.doi.org/10.1080/01690965.2011.601900

LANGUAGE AND COGNITIVE PROCESSES
2012, 27 (6), 887–900

Psychology Press
Taylor & Francis Group

Inferences about members of kinds: The generics hypothesis

Sangeet Khemlani[1], Sarah-Jane Leslie[2], and Sam Glucksberg[1]

[1]Department of Psychology, Princeton University, Princeton, NJ, USA
[2]Department of Philosophy, Princeton University, Princeton, NJ, USA

People routinely make inferences based on kind membership. For example, if you were told that a particular kind of animal is a tiger, then you would likely infer that it has stripes. Under what conditions are people willing to infer that a member of a given kind has a property? Two hypotheses were examined. The base rate or *prevalence hypothesis* holds that people rely only on their knowledge of the statistical frequency of a property among its kind to infer whether a member has that property. An alternative is the *generics hypothesis*, which states that people are influenced by their belief that the relevant generic generalization is true. In other words, if people agree to the generalization, "ducks lay eggs", then they should be willing to make the inference that an arbitrary individual duck lays eggs, despite their knowledge that the majority of ducks do not lay eggs (i.e., juveniles, males, and infertile females). We present data that support the second hypothesis. Rather than being driven solely by beliefs about prevalence, agreement to the relevant generic predicted performance on an inference task beyond estimated prevalence or cue validity. These findings suggest that models of categorization that are based solely on statistical or simple probabilistic principles are incomplete. They also provide support for the idea that generics articulate core conceptual beliefs that guide our interactions with the world.

Keywords: Generics; Generics hypothesis; Default inferences; Non-monotonic reasoning; Concepts.

Consider an arbitrary dog. How confident are you that the dog has four legs? The inference is compelling, and it is a type of inference known as a *default* inference because, absent any information to the contrary, it leads one to conclude that an arbitrary dog has four legs. Now consider an arbitrary Canadian. How confident are you that he or she is right-handed? We might be less inclined to draw the default conclusion that the Canadian is right-handed than that the dog has four legs.

What might account for the difference between the two inferences? It may be that a very high percentage of dogs have four legs, while a somewhat lower percentage of

---

Correspondence should be addressed to Sangeet Khemlani, Department of Psychology, Princeton University, 3-C-6 Green Hall, Princeton, NJ 08540, USA. E-mail: skhemlani@gmail.com

This work was supported in part by a National Science Foundation Graduate Research Fellowship awarded to Sangeet Khemlani. We thank Jeremy Boyd, Adele Goldberg, Phil Johnson-Laird, Mark Johnston, Olivia Kang, Niklas Kunze, Max Lotstein, Adam Moore, and Lance Rips for their many helpful discussions and suggestions.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2011.601900

Canadians are right-handed. Alternatively, if our conceptual knowledge concerning kinds contains structure richer than bare information about base rates or prevalence, then that structure might be reflected in our inferences about members of kinds. This paper presents an experiment designed to test whether such conceptual structure can be detected in our default reasoning. If so, then the view that categories and categorical induction can be adequately characterised in terms of the encoding and representation of statistical relationships, including prototype and exemplar theories, must, at best, be incomplete (e.g., Maddox & Ashby, 1993; Rosch, 1978; for a recent and comprehensive review see Murphy, 2002). Similar concerns would also apply to much of the recent work on Bayesian approaches to categorization however, sophisticated Bayesian approaches would likely be able to model this conceptual structure (Griffiths, Kemp, & Tenenbaum, 2008; Tenenbaum & Griffiths, 2001).

How might generics affect categorization and default inferences beyond consideration of statistical relationships such as prevalence? People often use generic statements, such as "tigers are striped", "barns are red", "ducks lay eggs", or "ticks carry Lyme disease" to express generalizations about kinds (Brandone, Cimpian, Leslie, & Gelman, in press; Carlson & Pelletier, 1995; Gelman, 2003; Krifka et al., 1995). Unlike quantified statements—for example, "all tigers are striped", "most barns are red"—generics do not communicate information about *how many* members of the kind have the property in question. For example, one cannot answer the question, "How many tigers are striped?" by replying that "tigers are striped" (Brandone et al., in press; Carlson, 1977; Leslie, 2007). Correspondingly, there is no direct relation between the prevalence of a property among members of a kind and the acceptability of the relevant generic. Indeed, "ducks lay eggs" is accepted even though only mature fertile females lay eggs, but "ducks are female" is rejected (Brandone et al., in press; Leslie, Khemlani, & Glucksberg, 2011). Further, "ticks carry Lyme disease" is accepted even though only 2% of ticks carry the disease, but "books are paperbacks" is rejected, despite the fact that over 80% of books are paperbacks (Khemlani, Leslie, Glucksberg & Rubio-Fernandez, 2007; Leslie, 2007, 2008).

Leslie (2007, 2008) hypothesised that these acceptance patterns arise because generics are sensitive to four main factors:

1. If a generic predicates a *characteristic* property of a kind, it is accepted, regardless of prevalence (e.g., "ducks lay eggs", and "lions have manes"). Examples of a kind's characteristic properties include (for animal kinds) its method of locomotion, diet, salient physical characteristics, method of reproduction, method of nurturing its young; function (for artifact kinds); and societal role (for professional kinds).
2. If a generic predicates a striking or dangerous property of a kind, it is accepted if just some members of the kind have the property (e.g., "ticks carry Lyme disease", "sharks attack bathers"; see Cimpian, Brandone, & Gelman, 2010).
3. If the property is neither dangerous nor characteristic, then the majority of the kind must have the property for the generic to be accepted (e.g., "barns are red", "cars have radios"; see also Prasada & Dillingham, 2006, 2009), with one caveat.
4. Generics are *rejected*—even if they predicate highly prevalent properties—if the members of the kind that lack the property have an equally salient alternative property instead, for example, being a hardcover book instead of a paperback ("books are paperbacks"); being male instead of female ("ducks are female"); being left-handed instead of right-handed (e.g., "Canadians are right-handed").

Recent findings suggest that generics may articulate our core conceptual knowledge about kinds (e.g., Gelman, 2010; Hollander, Gelman, & Star, 2002; Leslie, 2007, 2008; Leslie et al., 2011; Leslie & Gelman, 2011; Prasada & Dillingham, 2006, 2009; Prasada, Khemlani, Leslie, & Glucksberg, 2011). If this hypothesis is correct, then our conceptual knowledge is not based solely on quantitative information such as prevalence but rather is sensitive to the semantic factors outlined above.

To date, most studies that have tested this hypothesis have been developmental (e.g., Brandone et al., in press). In this paper, we take a different approach and address the question of how conceptual knowledge is structured by examining systematic patterns of default inference among adults. In particular, we ask: what factors promote default inferences? If generics express core conceptual knowledge, then one would predict that default inferences would be influenced by the same factors that determine our agreement to a generic, and not be based solely on beliefs about prevalence.

How do adults and children reason with generics? A fair number of studies have addressed this question (Chambers, Graham, & Turner, 2008; Cimpian, Brandone, & Gelman, 2010; Connolly, Fodor, Gleitman, & Gleitman, 2007; Elio & Pelletier, 1993, 1996; Gelman & Bloom, 2007; Graham, Nayer, & Gelman, in press; Gelman & Raman, 2003; Gelman, Star, & Flukes, 2002; Hampton, 2009; Pelletier & Elio, 2003). For both adults and children, belief in the generic generalization seems to dispose people to judge that a given member of the kind will have the relevant property, a view we call the *generics hypothesis.* However, previous work has not systematically considered "low prevalence" accepted generics (e.g., "ticks carry Lyme disease"; for one exception see Cimpian, Brandone, & Gelman, 2010) or high prevalence rejected generics (e.g., "books are paperbacks"). As a result, we do not know whether the tendency to make strong inferences from generics is a result of generic belief per se, or whether these inferences are driven by beliefs about the relative frequency with which the property occurs. For example, if people learn that Tweety is a bird, they may use their knowledge that *most* birds fly to infer by default that Tweety flies (Gerla, 1994). We refer to this account as the *base rate* or *prevalence* hypothesis.

The experiment reported here examined whether default inferences are guided by the relevant generic generalization (the generics hypothesis) or by beliefs about the prevalence of the relevant property for the kind (the prevalence hypothesis). To test this, we chose a variety of items for which generic agreement is dissociated from estimates of prevalence. Table 1 includes the taxonomy of these types of generics along with descriptions of the operationalisations of the categories to which they belong.

Participants were told that an arbitrary individual was a member of a particular kind, and they were then asked how confident they were that the individual had a given property. For example, they were told that *Jumpy is a tick*, and then asked to report their confidence in the claim that *Jumpy carries Lyme disease*. No generics were presented here—participants evaluated "Jumpy carries Lyme disease" without seeing the generic "ticks carry Lyme disease". That is because the presentation of an explicit generic might lead participants to rely more on the generic than they otherwise would. Instead, participants could draw on whatever knowledge or information they deemed relevant. The generics hypothesis predicts that participants' inferences would be sensitive to the same factors as generic generalizations, and thus agreement to a generic would be correlated with the tendency to make a default inference, over and above beliefs about prevalence.

TABLE 1
Various types of generic generalizations and operational definitions of the categories to which they belong

Item type	*Example*	*Truth value of the generic*	*Definition*
Definitional	Triangles have three sides	True	Property must be universally true of all the members of the kind; no exceptions.
Majority characteristic	Tigers are striped	True	Property must be prevalent though not universal among members of the kind; some exceptional members (e.g., albino tigers) fail to possess it. The property must be central, principled or essential (Gelman, 2003; Medin & Ortony, 1989; Prasada & Dillingham, 2006, 2009). In our study, we included only items that passed linguistic tests as outlined by Prasada and Dillingham (2006, 2009).
Minority characteristic	Lions have manes	True	Property must be held by only a minority of the kind, but must be central, principled or essential (Gelman, 2003; Medin & Ortony, 1989), and so pass Prasada and Dillingham's tests (Prasada & Dillingham, 2006, 2009). For our purposes we restricted these items to methods of gestation, methods of nourishing the very young, and characteristic physical traits held by only one gender.
Majority statistical	Cars have radios	True	Property must be prevalent among members of the kind, and must not be an essential or principled property.
Striking	Pit bulls maul children	True	Property must hold for only a small minority of the kind, and must signify something dangerous and to be avoided.
Majority FGs	Canadians are right-handed	False	Property must be prevalent among members of the kind, and there must exist a sufficiently salient alternative property (e.g., being left-handed), such that the generic form of the predication sounds false or mistaken.
Minority FGs	Rooms are round	False	Property must be held by very few members of the kind, and must not signify something dangerous. The generic form of the predication must appear false or mistaken.
False	Sharks have wings	False	No members of the kind may hold the property.

Note: FGs = False generalizations.

METHOD

Norming study

A norming study was conducted to provide data on prevalence estimates, cue validity ratings, and generic agreement for the items used in the experiment (see Table 2). Sixty-four online volunteers were randomly assigned to perform one of three tasks:

1. *Prevalence estimation.* Fifteen participants estimated the prevalence of a property within a category, for example, they were asked "What percentage of mosquitoes carry malaria?" and responded on a 0–100 scale.
2. *Cue validity rating*. Twenty-seven participants evaluated the cue validity (Beach, 1964) of each item, as cue validity could potentially mediate prevalence estimation and generic agreement. That is, people may determine that a given mosquito has malaria because the probability of having malaria and *not* being a mosquito is quite low. Thus, participants in the norming study were told, "Suppose x carries malaria" and were asked, "How likely is it that x is a mosquito?" They responded on a 7-point Likert scale in which +3 = "very likely", 0 = "not sure", and –3 = "very unlikely".
3. *Generic agreement.* Twenty-two participants agreed or disagreed with generic statements, for example, "mosquitoes carry malaria", by responding on a 7-point Likert scale in which +3 = "definitely agree", 0 = "can't tell", and –3 = "definitely disagree".

Participants

Twenty-nine volunteers completed the study for monetary compensation through Mechanical Turk, an online platform hosted through Amazon.com (for an analysis of the validity of results obtained through this platform, see Paolacci, Chandler, & Ipeirotis, 2010). None of the participants reported that they had received any training

TABLE 2
Mean ratings of confidence to the given conclusion (+3 = "I'm confident it's true" and −3 = "I'm confident it's false" and prevalence estimates (0–100), cue validities (+3 = "very likely", 0 = "not sure", –3 = "very unlikely"), and generic agreement (+3 = "definitely agree", 0 = "can't tell", –3 = "definitely disagree") for the corresponding generic generalization as a function of item type

	Experimental data	*Norming data*		
Item type	*Mean confidence rating*	*Prevalence estimate*	*Cue validity*	*Generic agreement*
Definitional	2.5	89	1.0	2.6
Majority characteristic	2.6	89	1.1	2.7
Minority characteristic	1.7	65	0.6	2.1
Majority statistical	1.5	67	0.5	1.2
Striking	0.7	24	0.9	1.3
Majority FG	0.6	60	0.6	0.0
Minority FG	−0.3	18	−0.5	−0.9
False	−2.5	–	–	–

Note: Norming data were not obtained for the false items. FG = False generalization.

in logic, and none had participated in experiments involving generics before. The participants completed the experiment online using an interface written in Javascript, HTML, and PHP.

Materials

Five types of items that are generally accepted in generic form were used (see Table 1): definitional, e.g., "triangles have three sides"; majority characteristic, e.g., "tigers are striped"; minority characteristic, e.g., "ducks lay eggs"; majority statistical, e.g., "cars have radios"; and striking item types, e.g., "mosquitoes carry malaria". Three item types that are generally rejected as generics were also used: majority false generalizations (FGs), e.g., "Canadians are right-handed"; minority FGs, e.g., "Rooms are round"; and false items, e.g., "sharks have wings". The participants were never presented with generic statements, but rather the generics were used to construct two statements: one was a premise concerning the category membership of an individual (e.g., "Buzzy is a mosquito") and a second was a conclusion about a property that the individual might possess (e.g., "Buzzy carries malaria"). The names of the individuals were chosen so as not to imply anything relevant about the truth of the property statement. For example, for items for which gender was relevant, only gender-neutral names were used. The materials are provided in Appendix 1.

Design and procedure

Participants carried out a default inference task in which they provided their level of confidence that a conclusion (e.g., "Buzzy carries malaria") was true by selecting from a 7-point Likert scale where $+3$ meant "I'm confident it's true" and –3 meant "I'm confident it's false". The eight different item types were distributed such that participants received 20 false items and 10 each of the other item types. Thus, they were given 90 items, in which 50 of the corresponding generics were accepted and 40 were rejected. The majority FGs (e.g., "books are paperbacks", "Canadians are right-handed") served to provide a comparison for the majority statistical items (e.g., "cars have radios"). The generics hypothesis predicted that people would be more willing to make default inferences from the majority statistical items (e.g., cars have radios) than from the majority FG items (e.g., Canadians are right-handed). The items that corresponded to rejected generic types—especially minority FGs and false statements—also served to provide opportunities for low-confidence ratings, and thereby balanced the Likert scale responses. Each participant received a different randomised order of the experimental items, and the participants were familiarised with the response scales before performing the default inference task.[1]

RESULTS AND DISCUSSION

The mean confidence ratings for default inferences and the norming data for each generic type are provided in Table 2.

[1] We also ran a separate, counterbalanced block of the generic agreement task on the participants in the main experiment. For brevity, these data were not used as predictors in any of the analyses in the paper, though they produced effects comparable to the generic agreement data collected in the norming study (as analyzed in Khemlani, Leslie, & Glucksberg, 2009). The participants' default inferences were not affected by whether the generic agreement task or the inference task came first, and so the two samples were collapsed in all subsequent analyses.

TABLE 3
Correlation matrix among the three predictors taken from the norming study ($N = 64$)

Variable	*1*	*2*	*3*
1. Prevalence estimates	–	–	–
2. Cue validity ratings	0.67	–	–
3. Generic agreements	0.77	0.83	–

To test whether the strength of default inferences is accounted for by prevalence or cue validity (but not generic agreement), the prevalence estimates, cue validity ratings, and generic agreement data (collapsing over item type) were regressed against mean confidence ratings as implemented in the lme4 package (linear mixed-effects models using S4 classes; Bates, Maechler, & Dai, 2008) in the statistical software R (version 2.10; R Development Core Team, 2008). The data were subjected to a hierarchical linear regression analysis that compared three models: Model 1 used prevalence alone as a predictor, Model 2 added cue validity as a predictor, and Model 3 added generic agreement data as a predictor. Table 3 provides a correlation matrix for the three predictors used in the models, and Table 4 gives the results of the analysis.

In all three models, prevalence estimates were significant positive predictors of confidence ratings. The model that considered generic agreement as a predictor accounted for significantly more variance than either of the other two models ($R^2_{\text{Model 3}} = .65$ vs. $R^2_{\text{Model 2}} = .58$, $p < .0001$). Likewise, prevalence and cue validity (Model 2) accounted for less variance than the model that considered generic agreement (Model 3). The analysis suggests that a strictly prevalence-based account of default reasoning (Model 1) is incomplete. Generic agreement and prevalence accounted for more variance than the account based on prevalence alone.

Analysis of item types by default inference confidence ratings and prevalence estimates

For an analysis of participants' performance for specific item types, mean confidence ratings and prevalence estimates were subjected to Friedman nonparametric analyses of variance, which revealed that confidence ratings varied significantly as a function of

TABLE 4
Hierarchical regression analysis for three predictors of confidence ratings: prevalence estimates, cue validity ratings, and generic agreement data

Model	*Predictor*	*B*	*SE (b)*	β
1	Intercept	−0.66	0.14	
	Prevalence estimates	0.03	0.00	0.74 ***
2	Intercept	−0.57	0.13	
	Prevalence estimates	0.03	0.00	0.58 ***
	Cue validity ratings	0.56	0.14	0.23 ***
3	Intercept	−0.22	0.13	
	Prevalence estimates	0.01	0.00	0.36 ***
	Cue validity ratings	−0.19	0.18	−0.07 ns
	Generic agreement	0.54	0.08	0.55 ***

Note: $R^2 = .55$ for Model 1; $R^2 = .58$ for Model 2; $R^2 = .65$ for Model 3 (all ps $< .001$).
***$p < .001$.

item type, $\chi^2(7, 28) = 170.29$, $p < .0001$, as did prevalence estimates, $\chi^2(6, 28) = 79.49$, $p < .0001$. Additional planned comparisons tested whether confidence ratings and prevalence estimates differed significantly across the various item types and are provided in Table 5.

For all comparisons, we computed the Bonferroni correction with a family-wise alpha rate of .05, and we include values of Cliff's δ, a nonparametric effect size indicator whose value ranges from –1 to 1 (see Cliff, 1993). We found that default inferences and prevalence estimates often coincide. For instance, definitional (e.g., "triangles have three sides") and majority characteristic (e.g., "tigers have stripes") items did not differ reliably on their mean confidence ratings or their corresponding prevalence estimates of 89% and 89%, respectively. Similarly, the mean confidence ratings and prevalence estimates for minority characteristic (e.g., "ducks lay eggs") and majority statistical (e.g., "cars have radios") were not significantly different. These comparisons suggest a strong relationship between mean confidence ratings and estimated prevalence in line with the prevalence hypothesis.

However, the theoretically instructive comparisons are those in which both estimated prevalence and agreement to the generic differed. Striking (e.g., "ticks carry Lyme disease") and majority FG items (e.g., "books are paperbacks") provide clear evidence that prevalence was not the only factor driving default inference. Consider first the striking item types. If participants relied primarily on prevalence information, the mean confidence ratings for striking items should have been lower than the ratings for majority FGs because the respective prevalence estimates for these two item types are reliably different (24% vs. 60%). Despite this difference, the mean confidence rating for striking generics is comparable to those for majority FGs (0.70 vs. 0.60).

The opposite effect holds for the mean confidence ratings of majority FG items (e.g., "books are paperbacks"), which were lower than one would expect if participants relied exclusively on prevalence. The estimated prevalence of 60% for majority FGs is comparable to the estimated prevalence of 65% for minority characteristic items, yet the mean confidence ratings for these two item types are

TABLE 5
Planned comparisons between different item types used in the experiment and the norming study. For each comparison, we provide a pairwise analysis of mean confidence ratings obtained for the main experiment as well as mean prevalence estimates obtained from the norming study

Planned comparison	*Wilcoxon test, z-value*	*p-value*	*Cliff's δ*
Definitional vs. majority characteristic			
Mean confidence rating: 2.5 vs. 2.6	0.88	> .30	.10
Mean prevalence estimate: 89% vs. 89%	0.51	> .30	.17
Minority characteristic vs. majority statistical			
Mean confidence rating: 1.7 vs. 1.5	0.83	> .30	.12
Mean prevalence estimate: 65% vs. 67%	0.97	> .30	.09
Striking vs. majority FG			
Mean confidence rating: 0.70 vs. 0.60	1.17	> .30	.16
Mean prevalence estimate: 24% vs. 60%	3.35	< .005	.82
Minority characteristic vs. majority FG			
Mean confidence rating: 1.7 vs. 0.60	4.12	< .0005	.75
Mean prevalence estimate: 65% vs. 60%	1.65	> .30	.27

reliably different (0.6 vs. 1.7, respectively). Also note that agreement for majority FG and minority characteristic items is 0.0 and 2.1 respectively, that is, generic agreement corresponds with the confidence ratings while prevalence estimates do not. Since the only difference between these two item types is whether or not they are agreed to as generics, it is strong evidence that participants relied on information other than prevalence, in this case, whether or not the item is agreed to as a generic. These two cases can be construed as instances of base rate neglect (Kahneman & Tversky, 1973; Tversky & Kahneman, 1982).

The above analyses show that striking, majority statistical, and majority FG items yield default inferences that a prevalence-based model cannot account for. These results corroborate the generics hypothesis, namely that default inferences are guided in part by belief in the underlying generic generalization, and thus that default inferences are sensitive to the same factors as generics are (e.g., whether the property is strikingly dangerous or not).

GENERAL DISCUSSION

What factors guide default reasoning about arbitrary members of kinds? We examined two alternative hypotheses. The *prevalence* hypothesis posits that confidence in a default inference is proportional to a reasoner's belief about the statistical properties of a kind, for example, that *most* dogs have four legs. The *generics* hypothesis proposes that inferences concerning the properties of individual members of a kind are driven by belief in the corresponding generic generalization, for example, the belief that *dogs have four legs*. If it is indeed the case that generic generalizations are conceptually fundamental (e.g., Gelman, 2010; Leslie, 2007, 2008), then one would expect that the generics hypothesis would be correct, and that default inferences would be sensitive to the same factors that drive acceptance of generic generalizations. Our findings suggest that this is indeed the case.

We tested these hypotheses by providing participants with category information such as, "Speedy is a car", and then asking them to rate their confidence in a default inference, for example, "Speedy has a radio". We compared confidence ratings with prevalence estimates, cue validity ratings, and generic agreement from a separate group of participants, and found that generic agreement and estimated prevalence is a better predictor of performance than estimated prevalence alone. For instance, prevalence estimates for striking (e.g., "mosquitoes carry malaria") and majority FG items (e.g., "Canadians are right-handed") differed reliably (24% and 60%, respectively), yet ratings of confidence about default inferences were comparable (0.7 and 0.6, respectively). We propose that the explanation for this is that, despite the fact that striking items were associated with lower prevalence estimates than majority FG items, striking items were more likely to be accepted in generic form than majority FGs. Since people's default inferences are guided by both their beliefs about prevalence as well as their belief in the relevant generic generalization, this led to the striking items being rated comparably to the majority FGs in the default inferences task, despite their disparate associated prevalence estimates. In other words, the fact that people accept striking generic generalizations led them to be more confident of the default inference than their beliefs about prevalence alone would license and conversely for majority FGs. Likewise, the prevalence estimates for majority FG items like "Canadians are right-handed" were comparable to those for majority statistical items (e.g., "cars have radios") and minority characteristic items (e.g., "ducks lay eggs''), yet people were much less likely to agree that the arbitrary Canadian is right-handed than that the arbitrary car has a radio or that the arbitrary duck lays eggs. People know, of course, that only female ducks lay eggs (Leslie et al., 2011), and so the tendency to agree that an arbitrary duck, Quacky, lays eggs is high even though people know that immature ducks and male ducks cannot lay eggs.

Cimpian, Brandone, and Gelman (2010) report that people may also discount information about prevalence when considering generic assertions that concern novel kinds. Their participants were willing to agree to striking and characteristic generics even when a property's prevalence was very low (e.g., accepting *lorches have dangerous feathers* even if only 30% of lorches have dangerous feathers). However, their participants expected that, if a striking or characteristic generic is true, the property would be prevalent (e.g., upon being told that *lorches have dangerous feathers*, they estimated that over 90% of lorches would have dangerous feathers). Thus they found, as we did, that inferences concerning the properties of individuals were influenced by agreement to the generic.

The study presented here indicates that default inferences are sensitive to the same factors that generics are sensitive to, and thus may well be guided by belief in the generic generalization. The findings presented here provide additional evidence for the hypothesis that our core conceptual information is richly structured and sensitive to the nature of the property being generalised, for example, whether it is characteristic of the kind or strikingly dangerous, rather than being driven solely by considerations such as how many members of the kind have the property. Any model of concepts that is built solely on statistical measures such as prevalence and cue validity (e.g., Rosch, 1978 and successors) will not be adequate to model human cognition.

In recent years there has been a surge of interest in Bayesian models of cognition, with theorists arguing that generalization, categorization, and induction can be unified under a Bayesian framework (Griffiths et al., 2008; Tenenbaum & Griffiths, 2001). Despite successful applications of these models, critics have countered that human cognition is far too complex to be explained by such a framework (Boroditsky & Ramscar, 2001; Jones & Love, in press; Love, 2001). Though the global adequacy of Bayesian models is beyond the scope of our analysis, we note that our findings suggest that inferences about particular members of kinds cannot be modelled by a simple or unstructured application of the probabilistic calculus. That is, people's judgments about whether a given member of a kind K has a property P does not simply reduce to their beliefs about the conditional probability of *having* P given that something is a K (nor to their beliefs about the conditional probability that something is a K given that it has P; nor to the conjunction of the two sets of beliefs). Our results indicate that a far more complex and sophisticated model—whether Bayesian or otherwise—will be needed to account for how we reason about members of kinds. The model must reflect the rich conceptual structure we find in generic generalizations.

Manuscript received 15 November 2010
Revised manuscript received 22 June 2011
First published online 19 October 2011

REFERENCES

Bates, D. M., Maechler, M., & Dai, B. (2008). *lme4: Linear mixed-effects models using S4 classes* [R package Version 0.999375-24]. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Beach, L. R. (1964). Cue probabilism and inference behavior. *Psychological Monographs: General and Applied*, *78*, 1–20.

Boroditsky, L., & Ramscar, M. (2001). "First, we assume a spherical cow…". *Behavioral and Brain Sciences*, *24*, 656–657.

Brandone, A., Cimpian, A., Leslie, S. J., & Gelman, S. A. (in press). Do lions have manes? For children, generics are about kinds rather than quantities. *Child Development*.

Carlson, G. (1997). *Reference to kinds in English*. Doctoral dissertation, University of Massachusetts, Amherst, MA, USA.

Carlson, G., & Pelletier, F. J. (1995). *The generic book*. Chicago, University of Chicago Press.

Chambers, C. G., Graham, S. A., & Turner, J. N. (2008). When hearsay trumps evidence: How generic language guides preschoolers' inferences about unfamiliar things. *Language and Cognitive Processes*, *23*, 749–766.

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

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, *114*, 494–509.

Connolly, A., Fodor, J., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition*, *103*, 1–22.

Elio, R., & Pelletier, F. J. (1993). Human benchmarks on AI's benchmark problems. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 406–411). Hillsdale, NJ: Lawerence.

Elio, R., & Pelletier, F. J. (1996). On reasoning with default rules and exceptions. *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 131–136). Hillsdale, NJ: Lawrence Erlbaum.

Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press.

Gelman, S. A. (2010). Generics as a window onto young childrens concepts. In F. J. Pelletier (Ed.), *Kinds, Things, and Stuff* (pp. 100–123). New York: Oxford University Press.

Gelman, S. A., & Bloom, P. (2007). Developmental changes in the understanding of generics. *Cognition*, *105*(1), 166–183.

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, *74*(1), 308–325.

Gelman, S. A., Star, J. R., & Flukes, J. E. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development*, *3*(2), 179–199.

Gerla, G. (1994). The probability that Tweety is able to fly. *International Journal of Intelligent Systems*, *9*, 403–409.

Graham, S. A., Nayer, S. L., & Gelman, S. A. (2011). Two-year-olds use the generic/nongeneric distinction to guide their inferences about novel kinds. *Child Development*, *82*(2), 493–507.

Griffiths, T., Kemp, C., & Tenenbaum, J. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology*. Cambridge: UK Cambridge University Press.

Hampton, J. (2009). Stability in concepts and evaluating the truth of generic statements. In F. J. Pelletier (Ed.), *Kinds, things, and stuff* (pp. 80–100). New York: Oxford University Press.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, *38*(6), 883–894.

Jones, M., & Love, B. (in press). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (2009). Generics, prevalence, and default inferences. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S., Leslie, S. J., Glucksberg, S., & Rubio-Fernandez, P. (2007). Do ducks lay eggs? How humans interpret generic assertions. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 395–400). Nashville, TN: Cognitive Science Society.

Krifka, M., Pelletier, F. J., Carlson, G., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: An introduction. In G. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 1–125). Chicago, IL: Chicago University Press.

Leslie, S. J. (2007). Generics and the structure of the mind. *Philosophical Perspectives*, *21*(1), 375–405.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *The Philosophical Review*, *117*(1), 1–49.

Leslie, S. J., & Gelman, S. A. (2011). *Quantified statements are recalled as generics: Evidence from preschool children and adults*. Manuscript under review.

Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, *65*, 15–31.

Love, B. C. (2001). Three deadly sins of category learning modelers. *Behavioral and Brain Sciences*, *24*, 687–688.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, *53*, 49–70.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York, NY: Cambridge University Press.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411–419.

Pelletier, F. J., & Elio, R. (2003). Logic and cognition: human performance in default reasoning. In P. Gärdenfors, J. Wolenski, & K. Kijania-Placet (Eds.), *In the scope of logic, methodology, and philosophy of science, Vol. I* (pp. 137–154). Dordrecht: Kluwer Academic.

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*, 73–112.

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense perception. *Cognitive Science*, *33*, 401–448.

Prasada, S., Khemlani, S., Leslie, S. J., & Glucksberg, S. (2011). *Conceptual distinctions amongst generics.* Manuscript under review.

R Development Core Team (2008). R: A language and environment for statistical computing (Version 2.10). Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Earlbaum Associates.

Tenenbaum, J., & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–640.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 153–160). Cambridge, UK: Cambridge University Press.

APPENDIX 1

Materials used in the experiment. Participants were given a category membership premise and were asked to rate their confidence in the predication

Item type	*Category membership*	*Predication*
Definitional	Barky is a dog	Barky is a mammal
Definitional	Breezy is an elm	Breezy is a tree
Definitional	Fluffy is a poodle	Fluffy is a dog
Definitional	John is a bachelor	John is unmarried
Definitional	Righty is a rectangle	Righty has four sides
Definitional	Snoozy is a cat	Snoozy is an animal
Definitional	Snuffly is a sow	Snuffly is a female pig
Definitional	Triggy is a triangle	Triggy has three angles
Definitional	Vivi is a vixen	Vivi is a female fox
Definitional	x is an even number	x is divisible by 2
Majority characteristic	Tweety is a bird	Tweety has wings
Majority characteristic	Cheety is a cheetah	Cheety runs fast
Majority characteristic	Grazer is a cow	Grazer moos
Majority characteristic	Fido is a dog	Fido has a tail
Majority characteristic	Aussie is a kangaroo	Aussie hops
Majority characteristic	Regal is a lion	Regal roars
Majority characteristic	Hempel is a raven	Hempel is black
Majority characteristic	Squiggy is a squirrel	Squiggy eats nuts
Majority characteristic	Prowly is a tiger	Prowly has stripes
Majority characteristic	Spotty is a cat	Spotty has four legs

Appendix (*Continued*)

Item type	*Category membership*	*Predication*
Majority characteristic	Hammy is a hamster	Hammy has nipples
Minority characteristic	Bleaty is a sheep	Bleaty has an udder
Minority characteristic	Browny is a moose	Browny has antlers
Minority characteristic	Fuzzy is a goat	Fuzzy has horns
Minority characteristic	Galloper is a horse	Galloper gives live birth
Minority characteristic	Oinky is a pig	Oinky suckles its young
Minority characteristic	Quacky is a duck	Quacky lays eggs
Minority characteristic	Roary is a lion	Roary has a mane
Minority characteristic	Sparky is a mammal	Sparky produces milk
Minority characteristic	Tweety is a cardinal	Tweety is red
Majority statistical	Boozey is a bar	Boozey is noisy
Majority statistical	Brassy is a trumpet	Brassy is loud
Majority statistical	Cozy is a jacket	Cozy has a zipper
Majority statistical	Footsy is a shoe	Footsy has laces
Majority statistical	Luigi is an Italian	Luigi eats spaghetti
Majority statistical	Metroline is a subway	Metroline is crowded
Majority statistical	Nike Relax is a shirt	Nike Relax has a collar
Majority statistical	Old Rickety is a barn	Old Rickety is red
Majority statistical	Speedy is a car	Speedy has a radio
Majority statistical	Ticky is a clock	Ticky is round
Striking	Bob had a stroke	Bob's stroke caused paralysis
Striking	Buzzy is a mosquito	Buzzy carries malaria
Striking	Fido is a Rottweiler	Fido mauls children
Striking	Fins is a shark	Fins attacks swimmers
Striking	Jumpy is a tick	Jumpy carries Lyme disease
Striking	Leo is a lion	Leo eats people
Striking	Marsha is a hurricane	Marsha damaged buildings
Striking	Squeaky is a rat	Squeaky carries disease
Striking	Stripey is a tiger	Stripey attacks people
Striking	Wingy is a bird	Wingy carries avian flu
Majority FG	Happy Times is a book	Happy Times is a paperback
Majority FG	Joe is a Canadian	Joe is right-handed
Majority FG	X-Rig is a computer	X-Rig is a PC
Majority FG	Beaky is a duck	Beaky is female
Majority FG	Dr. Jones is an engineer	Dr. Jones is male
Majority FG	Jane is an American	Jane is brunette
Majority FG	Pat Brown is a school teacher	Pat Brown is female
Majority FG	Plucky is a lion	Plucky is male
Majority FG	Southern cross is a tree	Southern cross is a deciduous tree
Majority FG	Viv is an athlete	Viv is a student
Minority FG	Fluffy is a dog	Fluffy is blind
Minority FG	June's is a restaurant	June's is a Chinese restaurant
Minority FG	Kate is an American	Kate is dyslexic
Minority FG	Kitty is a cat	Kitty is white
Minority FG	Nurse Jones is a nurse	Nurse Jones is a man
Minority FG	Pablo is a Spaniard	Pablo is Jewish
Minority FG	Pouncer is a tiger	Pouncer is albino
Minority FG	Table 371 is a table	Table 371 is 10ft long
Minority FG	The Thornberry Suite is a room	The Thornberry Suite is round
Minority FG	Waggy is a dog	Waggy has three legs
False	Ally is an alligator	Ally has fur
False	Bushy is a fox	Bushy lays eggs
False	Cheeky is a hamster	Cheeky has stripes
False	Father O'Reilly is a Catholic priest	Father O'Reilly is married
False	Fawny is a deer	Fawny eats meat
False	Feathers is a bird	Feathers suckles its' young
False	Freddy is a frog	Freddy can fly

Appendix (*Continued*)

Item type	*Category membership*	*Predication*
False	Gills is a fish	Gills is a mammal
False	Grizzly is a wolf	Grizzly says' meow'
False	Hexy is a hexagon	Hexy has four sides
False	Hissy is a snake	Hissy has legs
False	Kanga is a kangaroo	Kanga has fins
False	Larry is a leaf	Larry is blue
False	n is an odd number	n is divisible by two
False	Robby is a robin	Robby is four-legged
False	Sammy is a shark	Sammy has wings
False	Savannah is a lion	Savannah is black
False	Sharpy is a television	Sharpy makes coffee
False	Whiskers is a tiger	Whiskers has horns
False	Zebediah is a zebra	Zebediah has spots