NIH Public Access
Author Manuscript
Cogn Sci. Author manuscript; available in PMC 2011 November 1.

Published in final edited form as:
Cogn Sci. 2010 November 1; 34(8): 1452–1482. doi:10.1111/j.1551-6709.2010.01126.x.

Generic statements require little evidence for acceptance but have powerful implications

Andrei Cimpian,
Department of Psychology, University of Illinois at Urbana-Champaign

Amanda C. Brandone, and
Department of Psychology, University of Michigan, Ann Arbor

Susan A. Gelman
Department of Psychology, University of Michigan, Ann Arbor

Abstract

Generic statements (e.g., "Birds lay eggs") express generalizations about categories. In this paper, we hypothesized that there is a paradoxical asymmetry at the core of generic meaning, such that these sentences have extremely strong implications but require little evidence to be judged true. Four experiments confirmed the hypothesized asymmetry: Participants interpreted novel generics such as "Lorches have purple feathers" as referring to nearly all lorches, but they judged the same novel generics to be true given a wide range of prevalence levels (e.g., even when only 10% or 30% of lorches had purple feathers). A second hypothesis, also confirmed by the results, was that novel generic sentences about dangerous or distinctive properties would be more acceptable than generic sentences that were similar but did not have these connotations. In addition to clarifying important aspects of generics' meaning, these findings are applicable to a range of real-world processes such as stereotyping and political discourse.

Keywords

generic language; concepts; truth conditions; prevalence implications; quantifiers; semantics

1. Introduction

A statement is generic if it expresses a generalization about the members of a kind, as in "Mosquitoes carry the West Nile virus" or "Birds lay eggs" (e.g., Carlson, 1977; Carlson & Pelletier, 1995; Leslie, 2008). Such generalizations are commonplace in everyday conversation and child-directed speech (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998; Gelman, Taylor, & Nguyen, 2004; Gelman, Goetz, Sarnecka, & Flukes, 2008), and are likely to foster the growth of children's conceptual knowledge (Cimpian & Markman, 2009; Gelman, 2004, 2009). Here, however, we explore the semantics of generic sentences —and, in particular, the relationship between generic meaning and the statistical prevalence of the relevant properties (e.g., what proportion of birds lay eggs).

Consider, first, generics' *truth conditions*: Generic sentences are often judged true despite weak statistical evidence. Few people would dispute the truth of "Mosquitoes carry the West Nile virus", yet only about 1% of mosquitoes are actually carriers (Cox, 2004). Similarly,

Contact information, Correspondence concerning this article should be addressed to Andrei Cimpian, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel St., Champaign, IL 61820. Phone: (217) 333-0852. acimpian@psych.illinois.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

only a minority of birds lays eggs (the healthy, mature females), but "Birds lay eggs" is uncontroversial. This loose, almost negligible relationship between the prevalence of a property within a category and the acceptance of the corresponding generic sentence has long puzzled linguists and philosophers, and has led to many attempts to describe the truth conditions of generic statements (for reviews, see Carlson, 1995; Leslie, 2008).

Though generics' truth conditions may be unrelated to property prevalence (cf. Prasada & Dillingham, 2006), the same cannot be said about the *implications* of generic statements. When provided with a novel generic sentence, one often has the impression that the property talked about is widespread. For example, if we were unfamiliar with the West Nile virus and were told (generically) that mosquitoes carry it, it would not be unreasonable to assume that *all*, or at least a majority of, mosquitoes are carriers (Gelman, Star, & Flukes, 2002).

It is this paradoxical combination of flexible, almost prevalence-independent truth conditions, on the one hand, and widespread prevalence implications, on the other, that is the main focus of this article. We will attempt to demonstrate empirically that the prevalence level that is sufficient to judge a generic sentence as true is indeed significantly lower than the prevalence level implied by that very same sentence. If told that, say, "Lorches have purple feathers," people might expect almost all lorches to have these feathers (illustrating generics' *high implied prevalence*), but they may still agree that the sentence is true even if the actual prevalence of purple feathers among lorches turned out to be much lower (illustrating generics' *flexible truth conditions*). Additionally, we propose that this asymmetry is peculiar to generic statements and does not extend to sentences with quantified noun phrases as subjects. That is, the prevalence implied by a sentence such as "Most lorches have purple feathers" may be more closely aligned with the prevalence that would be needed to judge it as true.

Before describing our studies, we provide a brief overview of previous research on the truth conditions and the prevalence implications of generic statements.

1.1. Generics' truth conditions

Some of the first experimental evidence for the idea that the truth of a generic statement does not depend on the underlying statistics was provided by Gilson and Abelson (1965; Abelson & Kanouse, 1966) in their studies of "the psychology of audience reaction" to "persuasive communication" in the form of generic assertions (Abelson & Kanouse, 1966, p. 171). Participants were presented with novel items such as the following:

> Altogether there are three kinds of tribes—Southern, Northern, Central. Southern tribes *have* sports magazines. Northern tribes *do not have* sports magazines. Central tribes *do not have* sports magazines. *Do tribes have sports magazines?*

All items had the same critical feature: only one third of the target category possessed the relevant property. Despite the low prevalence, participants answered "yes" approximately 70% of the time to "Do tribes have sports magazines?" and other generic questions similar to it. Thus, people's acceptance of the generics did not seem contingent on strong statistical evidence, leaving the door open for persuasion, and perhaps manipulation, by ill-intentioned communicators.

A similar conclusion about the relationship between statistical prevalence and generics' truth conditions emerged from the linguistics literature on this topic (e.g., Carlson, 1977; Carlson & Pelletier, 1995; Dahl, 1975; Declerck, 1986, 1991; Lawler, 1973). For example, Carlson (1977) writes that "there are many cases where […] less than half of the individuals under consideration have some certain property, yet we still can truly predicate that property of the appropriate bare plural" (p. 67), as is the case with "Birds lay eggs" and "Mosquitoes carry

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

the West Nile virus" but also with "Lions have manes" (only males do), "Cardinals are red" (only males are), and others. He points out, moreover, that there are many properties that, although present in a majority of a kind, nevertheless *cannot* be predicated truthfully of that kind (e.g., more than 50% of books are paperbacks but "Books are paperbacks" is false). Thus, acceptance of a generic sentence is doubly dissociated from the prevalence of the property it refers to—not only can true generics refer to low-prevalence properties, but high-prevalence properties are also not guaranteed to be true in generic form.

A recent resurgence of interest in generics within psychology (e.g., Cimpian, Gelman, & Brandone, 2010; Gelman, 2004; Gelman & Bloom, 2007; Prasada, 2000) and philosophy (e.g., Leslie, 2008) brought further confirmation of this dissociation, as well as a new and compelling theoretical framework to explain it.[1] Two basic assumptions are at the core of this account: First, generic sentences are a linguistic outlet for our knowledge about kinds and are thus intimately bound up with how we *represent* kinds. Second, an adequate description of how we represent kinds cannot be based solely on statistical information about the features that co-occur with category membership (e.g., how often *laying eggs* co-occurs with *being a bird*) but rather needs to incorporate the vast causal knowledge that people have about the world (e.g., about the process of sexual reproduction, about the reproductive biology of different species, etc.; for reviews, see Murphy, 2004; Murphy & Medin, 1985).

Thus, whether a property is true of a kind is not just a function of its prevalence (though prevalence does play a role; e.g., Prasada & Dillingham, 2006) but also of the other causal-explanatory knowledge that binds that feature to the category. As an illustration of this general idea, Cimpian et al. (2010) demonstrated that adults' interpretation of generic sentences is influenced by a particular causal belief, namely that some features that characterize a kind emerge only with development. To take a familiar example, baby swans are not beautiful, yet because the adults of the species are, the generic "Swans are beautiful" is true. Similarly, participants in Cimpian et al.'s study were more likely to map novel generic sentences such as "Dontrets have long tails" onto kinds where the key feature (e.g., long tails) was absent in all the babies but present in all the adults than onto kinds where the key feature was *more prevalent* overall but randomly distributed among the babies and the adults. People are also sensitive to the *origins* of the properties referred to in generic sentences, in that they endorse generics about inborn features more often than generics about acquired, but equally prevalent, properties (Gelman & Bloom, 2007).

1.2. Generics' prevalence implications

The implications of generic sentences have received much less attention. Consistent with the argument we make here, Abelson and Kanouse (1966) claimed that generic statements, "once accepted psychologically, […] appear to be commonly taken in a rather strong sense, as though the qualifier *always* had implicitly crept into their interpretation" (p. 172). Also implicit in their interpretation may be an assumption of *widespread prevalence*, as the results of a study by Gelman et al. (2002) seem to suggest. Participants were given generic statements about unfamiliar properties of familiar kinds (e.g., "Bears like to eat ants") and asked to estimate the percentage of category members that displayed the relevant features (e.g., the percentage of bears that like to eat ants). The average estimate was indeed very high (84%), confirming the idea that generics have powerful prevalence implications.

[1]However, this was by no means the first attempt to explain this dissociation. Carlson (1995) and Leslie (2008) provide excellent reviews of prior theoretical accounts.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

However, because some of the items used in Gelman et al.'s study were probably familiar to the adult participants (e.g., bears like to eat ants, cats have eyes that glow in the dark), their results may provide only a rough index of the prevalence implications of generic sentences. Familiarity with the items is an issue here because participants could have used their previous knowledge that these facts are widespread rather than the generic form of the statements to generate their estimates. While no measure of generics' implications may be *completely* free from the influence of prior knowledge (in the form of abstract overhypotheses, theory-based plausibility estimates, etc.), using novel categories might avoid some of these problems. This is the approach we adopted in the current studies.

Familiarity is also an issue when considering generics' truth conditions. Intuitions about the acceptability of sentences such as "Birds lay eggs" or "Lions have manes" may be artificially inflated by previous exposure to these generic predications. In contrast, "Books are paperbacks" is not often heard, which may add to the impression that it just sounds false. Of course, this observation cannot account for why speakers produce some generics but not others. There must be some antecedent reason why "Lions have manes" is acceptable and thus commonly said. Even so, familiarity probably makes these sentences more likely to be judged true than they would be otherwise. Our use of novel categories will eliminate this source of potential bias.

1.3. The present studies

We hypothesized an asymmetry at the core of generic meaning: Despite often being judged true at low prevalence levels, generics nevertheless imply high prevalence. To test this asymmetry, we provided participants with facts about novel categories (animals from "a remote island") and collected two measures. First, we determined what the average prevalence level is that leads to acceptance of these facts in generic form. For example, would participants judge the generic "Lorches have purple feathers" to be true if they were told that 10% of lorches have purple feathers? What if 30% of lorches have them, or 50%, or 70%, or 90%? Second, we determined what average prevalence level is *implied* by a generic statement. For example, when provided with the generic "Lorches have purple feathers," what percentage of lorches would participants estimate to have this feature? If the hypothesized asymmetry holds, then the average prevalence level at which generics are judged true should be significantly lower than the average prevalence level implied by the very same generics. For example, participants might expect over 90% of lorches to have purple feathers when provided with the generic, but they might judge the generic true not only when told that 90% of lorches have purple feathers but also at much lower prevalence levels.

We also hypothesized that sentences expressing quantified generalizations would not show this asymmetry. Consider, for instance, universally quantified sentences: The prevalence needed to assent to "All lorches have purple feathers" and the prevalence implied by this statement are identical: 100% (though see Khemlani, Leslie, Glucksberg, & Fernandez, 2007). In this case, the quantifier "all" explicitly fixes the size of the referent set, preventing any slippage between truth conditions and implications. In our studies, however, we will use as controls sentences quantified with "most." "Most" seems to be the English quantifier that comes closest to capturing generic meaning (Carlson, 1977; Cimpian et al., 2010), thus allowing for the strongest test of our hypothesis. If generic sentences show the predicted asymmetry but their closest quantified analogues do not, that would provide solid evidence for the claim that generics are unique in this respect.

Another goal of this series of studies is to further test the claim that generics' truth conditions are bound up with how we represent kinds. If generics are a direct linguistic instantiation of our kind representations, then their truth conditions should be sensitive to the

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

biases inherent in these representations. On this view, generic predications of features that are privileged in our concepts may be more acceptable than generic predications of features that are not, all other things being equal. In these studies, we tested whether features that are *distinctive* or *dangerous* are more likely to be thought true in generic form than equally prevalent properties that do not have the same connotations (Leslie, 2008). A feature that is distinctive (that is, possessed by few members of comparable categories) or that poses a threat to humans has high informational value and may thus be featured more prominently in the representation of the kind, even when its statistical association with the kind is not particularly strong. This prominence, in turn, may make the feature acceptable in a generic statement—that is, a bias in conceptual structure may be reflected in a bias in the acceptability of the corresponding generic statement. For example, the fact that very few birds have red feathers makes male cardinals' plumage noteworthy, which in turn may make "Cardinals are red" acceptable. (Since female cardinals' brown feathers are not similarly distinctive, "Cardinals are brown" is thought to be false.) Analogously, West Nile encephalitis is a potentially fatal disease, which makes information about its path of transmission noteworthy and lends truth to the generic "Mosquitoes carry the West Nile virus," despite the extremely low prevalence of actual virus carriers.

Leslie (2008) argued convincingly for the role of this information in determining generics' truth conditions and concluded that "the more striking, appalling, or otherwise gripping we find the property predicated in the generic, the more tolerant the generic is to exceptions" (p. 15). However, her argument was based entirely on *familiar* generic sentences such as those above. A test involving *novel* categories would provide much stronger evidence both for Leslie's specific claim and for the more general point that generics' truth conditions cannot be separated from the structure of the underlying kind representations. To perform this test, we simply added the danger/distinctiveness information to our novel items: For instance, after telling participants that a certain percentage of lorches have purple feathers, we sometimes added that no other animals on the island have such plumage (i.e., it is distinctive) and that these feathers are lethal to humans because they are sharp as needles (i.e., they are dangerous). We then measured whether, as predicted, the combined danger/distinctiveness information increased the likelihood that the property is accepted in generic form. (The separate effects of danger and distinctiveness were also assessed in a later experiment.)

We conducted four studies. Experiment 1 tested both the asymmetry between generics' truth conditions and their prevalence implications, and the contribution of danger/distinctiveness information to their acceptability. Experiments 2 and 3 were designed to investigate alternative explanations of the initial results. Experiment 3 also investigated the limits of this asymmetry by varying the content of the properties used. Finally, in Experiment 4 we separated the danger information and the distinctiveness information, which had been combined in the previous experiments, and asked whether they are individually sufficient to increase participants' willingness to accept the relevant generics.

2. Experiment 1

In this study, participants were randomly assigned to one of two tasks: the *truth conditions* task or the *implied prevalence* task. Both tasks involved reasoning about *novel* animal categories, but they differed in the types of judgments participants had to make. In the truth conditions task, participants were told that a certain percentage of category members have a feature (e.g., 30% of lorches have purple feathers) and were then asked if the corresponding generic (e.g., "Lorches have purple feathers") is true. The implied prevalence task was the converse: Participants were told that the generic statement is true and were then asked what percentage of the animal kind possesses the feature (e.g., what percentage of lorches have purple feathers). Comparing the average percentage that led to a “true” response in the truth conditions task to the average percentage generated in the implied prevalence task afforded a simple test of the hypothesized asymmetry.

The study also included a separate group of subjects who were tested on “most”-quantified items (e.g., “Most lorches have purple feathers”) instead of generics. We predicted that these participants would not show the asymmetry—that is, that their responses on the truth conditions and implied prevalence tasks would be equivalent.

To test for the effect of the danger/distinctiveness information on generics’ truth conditions, we compared three types of items: (1) *plain* items, which included only information about the category (e.g., lorches) and the property (e.g., purple feathers); (2) *dangerous/distinctive* items, which specified that the property is dangerous to humans and unique; and (3) *non-distinctive control* items, which contained a few extraneous facts about the property (e.g., that the feathers are wide and smooth to the touch) and specified that it is not unique. This last item type was included for two reasons. First, we wanted to control for the possibility that simply providing more information relative to the plain items, regardless of what the information is, would influence participants’ judgments. Second, the contrast with the distinctiveness information provided for the dangerous/distinctive items was meant to highlight this information and increase the chances that subjects will try to use it in their judgments. If the dangerous/distinctive items are accepted more frequently than both the plain and the non-distinctive control items, then it is likely that the conceptual information about the dangerous and distinctive nature of the properties was in fact responsible for this increase.

2.1. Method

2.1.1. Participants—Seventy-one undergraduates (33 males, 38 females) from two large public universities participated in the study either for course credit or for $5.

2.1.2. Design—The design is illustrated in Fig. 1. The participants were randomly assigned to either the generic condition or the “most” condition. Within each of these conditions, participants completed either the truth conditions task or the implied frequency task. Each of these tasks consisted of 30 items: 10 plain, 10 dangerous/distinctive, and 10 non-distinctive. In addition, for the truth conditions task only, each item type was presented twice at each of 5 prevalence levels: 10%, 30%, 50%, 70%, and 90%. This design can be summarized as follows: 2 (item wording: generic vs. “most”; between subjects) × 2 (task: truth conditions vs. implied prevalence; between subjects) × 3 (item type: plain vs. dangerous/distinctive vs. non-distinctive; within subject), with an additional five-level prevalence factor (within subject) for the truth conditions task only.

2.1.3. Procedure and materials—The 30 items were randomized, printed in booklet form, and handed to the participants, who were tested in groups. The first page of the booklets contained the instructions:

> In this study, we will tell you about some animals that live on a remote island. This island is very large and has many different animals on it. For each item, you will be given some information and asked a question. Please try to answer our questions to the best of your ability.

The items were divided into 3 subsets of 10. Within each subset, the 10 properties referred to 10 different animal body parts (feathers, scales, tails, shells, etc.). However, all 3 subsets used these same 10 body parts, varying only their color and the name of the category they were associated with (see Appendix A). For example, subset 1 contained a property about

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

morseths' silver fur, subset 2 contained a property about ollers' yellow fur, and subset 3 contained a property about kweps' copper fur. Participants saw one of these subsets in plain form, one in dangerous/distinctive form, and one in non-distinctive control form. Each subset was rotated through these three possible forms across subjects.

To be able to present each subset in all three different forms, we constructed a plain, dangerous/distinctive, and non-distinctive control version of each of the 30 items. Table 1 contains an example of an item in all three forms as it could appear to participants. (Note, though, that a single participant would only see one of these possible forms.)

Finally, the materials and procedure in the "most" condition were identical to those in the generic condition, the only difference being that the generic noun phrases were replaced with quantified noun phrases (e.g., "*Most morseths* have silver fur" instead of "*Morseths* have silver fur").

2.1.4. Data analysis strategy—Participants' responses were submitted to analyses of variance (ANOVAs), followed up with tests of simple effects and Bonferroni-adjusted post-hoc tests. Although our data violated some of the assumptions underlying parametric tests such as ANOVA, we nevertheless chose to report these statistics for several reasons. In addition to being widely used and thus familiar to a variety of audiences, they are in fact robust to moderate violations of their assumptions, especially when the sample sizes are similar across groups (see, e.g., Glass, Peckham, & Sanders, 1972; Howell, 2009). Moreover, all the crucial pairwise comparisons in this and subsequent experiments were also significant—at α levels of .05 or, rarely, .10—when analyzed with non-parametric Mann-Whitney U and Wilcoxon Matched-Pairs Signed-Ranks tests for between- and within-subject comparisons, respectively.

We also performed a separate set of ANOVAs with items, rather than subjects, as a random variable. In the vast majority of cases, these analyses replicated the results of the by-subjects ANOVAs. The few F ratios that were significant by subjects but not by items are indicated below.

2.2. Results and discussion

2.2.1. Is there an asymmetry between generics' truth conditions and their prevalence implications?—The first analysis we performed asked (a) whether there is an asymmetry between the prevalence implied by generic statements and the prevalence required for judging them as true, and (b) whether this asymmetry is specific to generics or extends to "most"-quantified sentences as well. In effect, this analysis required us to compare subjects' average percentage estimates from the implied prevalence task with their responses in the truth conditions task. However, since the responses in the truth conditions task were true/false judgments, they had to be converted to a metric that could be compared directly to the percentages obtained from the implied prevalence task. We thus used the true/false responses in the truth conditions task to compute, for each subject, the average prevalence level that led to "true" responses. For example, if a subject circled "true" whenever the prevalence was 70% or 90% and "false" for anything else, then that person's average prevalence that led to "true" responses would be 80%. The only participants for whom this average could not be computed were those who said "false" to every single item, regardless of prevalence. Since these were participants who did not deem even prevalence levels of 90% sufficient to accept the relevant statement, we decided to assign them a score of 100%. This decision was in keeping with what the pattern of their responses seemed to imply—that they needed even more than 90% prevalence to accept the statements—and was also conservative. The three subjects who said "false" to everything were all in the generic condition, so by assuming they would have needed 100% to accept the generic statements,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

we stacked the odds against finding the predicted result that the prevalence required for accepting a generic is *lower* than the prevalence implied by it.

We compared participants' scores in the implied prevalence and truth conditions tasks with an ANOVA that also included item wording (generic vs. "most"; between subjects) and item type (plain vs. distinctive/dangerous vs. non-distinctive control; within subject) as factors. The ANOVA revealed a main effect of task, $F(1, 67) = 23.03$, $p < .001$, $\eta_p^2 = .26$, qualified by a task × item wording interaction, $F(1, 67) = 19.47$, $p < .001$, $\eta_p^2 = .23$ (see Fig. 2). We used tests of simple effects to check for an effect of task within each of the two wording conditions. As predicted, in the generic condition, the average prevalence that led participants to accept the generics ($M_{truth\ conditions}$ = 69.1%) was significantly lower than the average prevalence implied by them ($M_{implied\ prevalence}$ = 95.8%), $F(1, 67) = 42.00$, $p < .001$, $\eta_p^2 = .39$.[2] Also as predicted, no such difference was found in the "most" condition, $F(1, 67) = 0.08$, $p = .785$, $\eta_p^2 = .00$. The prevalence that led participants, on average, to assent to "most"-quantified generalizations ($M_{truth\ conditions}$ = 76.9%) was almost identical to the prevalence implied by these generalizations ($M_{implied\ prevalence}$ = 78.0%).

2.2.2. The impact of the danger/distinctiveness information on generics' truth conditions—

We hypothesized, following Leslie (2008), that generic statements about properties that are distinctive and dangerous should be more acceptable than generic statements about properties that, although broadly similar, do not contain this information. To test this hypothesis, we analyzed participants' responses in the truth conditions task, expecting to find a higher proportion of "true" responses for the dangerous/distinctive items than for the plain and the non-distinctive control items. The "most" condition was included as a control; participants' acceptance of "most"-quantified statements, which are statements about sets of a particular size and not about kinds per se, should not be influenced by this conceptual information.

We performed a 2 (item wording: generic vs. "most"; between subjects) × 3 (item type: plain vs. dangerous/distinctive vs. non-distinctive; within subject) × 5 (prevalence level: 10%, 30%, 50%, 70%, and 90%; within subject) ANOVA on the proportion of "true" responses in the truth conditions task. This analysis uncovered a main effect of item type, $F(2, 78) = 5.99$, $p = .004$, $\eta_p^2 = .13$, which is best interpreted in light of an interaction between item type and item wording, $F(2, 78) = 4.58$, $p = .013$, $\eta_p^2 = .11$. Bonferroni-adjusted post-hoc tests revealed the predicted pattern of results (see Fig. 3, top two panels): In the generic condition, participants were more likely to accept the dangerous/distinctive items, for which the mean proportion of "true" responses was .68, than either the plain items (M = .55), p = .019, or the non-distinctive control items (M = .48), p = .001. In the "most" condition, however, participants were insensitive to this information: The proportions of "true" responses for the plain, dangerous/distinctive, and non-distinctive control items were .46, .47, and .45, respectively, all ps = 1.00.

[2]One might argue that this difference was due, at least in part, to the fact that participants' *implied prevalence* responses could go up to 100%, whereas the maximum score in the *truth conditions* task was 90% (except for the rare cases in which participants were assigned a score of 100% because they circled "false" for every item). To test whether this difference in range could have been responsible for the significant difference between the two tasks, we performed an additional analysis where participants' implied prevalence estimates were capped at 90% (i.e., any responses greater than 90% were converted to 90%). The main result was replicated: The (capped) prevalence implied by the generic statements was still significantly higher than the prevalence that led to their acceptance, $F(1, 67) = 20.13$, $p < .001$, $\eta_p^2 = .23$. Similar analyses were performed in all subsequent experiments that tested this asymmetry, and they all replicated the results reported in the main text.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The effect of the conceptual information was particularly strong *at the lower percentage levels* in the generic condition, as can be clearly seen in Fig. 3 (top panel) and was also indicated by an interaction between item type and prevalence level, $F(8, 312) = 2.66$, $p = .008$, $\eta_p^2=.06$, and by a three-way interaction among item type, prevalence level, and item wording, $F(8, 312) = 2.22$, $p = .026$, $\eta_p^2=.05$. (These two interactions were not significant in the ANOVA with items as a random variable.) Bonferroni-adjusted post-hoc tests confirmed that, in the generic condition only, the dangerous/distinctive items were more likely to be accepted than either of the other two item types at the 10% prevalence level (*p*s < .001) and at the 30% prevalence level (*p*s < .035), and more likely to be accepted than the non-distinctive control items at the 50% prevalence level ($p = .040$). The differences between the item types became smaller as the prevalence level increased. Since high-prevalence features are often judged true in generic form regardless of whether or not they are privileged in our concepts (Prasada & Dillingham, 2006), the danger/distinctiveness information could not add much to their acceptability. However, it did have a powerful impact on the acceptability of the lower-prevalence generics.

The ANOVA also revealed several other main effects and interactions. These included, first, a main effect of prevalence level, $F(4, 156) = 93.94$, $p < .001$, $\eta_p^2=.71$, suggesting that the proportion of "true" responses increased with prevalence level (see Fig. 3); and, second, a significant interaction between prevalence level and item wording, $F(4, 156) = 17.53$, $p < .001$, $\eta_p^2=.31$, which indicated that the effect of prevalence was more pronounced in the "most" condition than in the generic condition (see Fig. 3).

2.2.3. Conclusion—The results of this experiment provide support for both of our hypotheses. First, they confirm the hypothesized asymmetry between generics' implied prevalence and their truth conditions: When provided with generic statements, participants assumed that they apply to *nearly all* members of the relevant categories. In stark contrast, the same generic statements were often judged to be true even when the prevalence of the properties was very low. Second, the results confirm that generics' truth conditions are relaxed by making the properties distinctive and dangerous.[3] The fact that we obtained this result with *novel* categories and properties is particularly compelling: Participants' acceptance of the dangerous/distinctive generic items could not have been inflated by their prior familiarity with these generics and was thus based solely on their judgment that a feature that is dangerous and distinctive is more characteristic of a kind and therefore more acceptable in a generic predication.

3. Experiment 2

One of the most striking results from the first study was participants' willingness to accept generic statements at low prevalence levels. Although this result was predicted by the "asymmetry" hypothesis, one possible alternative explanation is that participants treated our bare plural statements as *existential* or *indefinite* rather than generic—that is, they may have treated sentences such as "Lorches have purple feathers" as if they were synonymous with "There exist some lorches that have purple feathers" or, simply, "Some lorches have purple feathers." This interpretation would indeed lead to a high proportion of "true" responses even at very low prevalence levels. (Note, however, that it would not account for the high percentage estimates in the implied prevalence task, nor would it account for the distinction

[3]An additional 40 participants (20 generic and 20 "most") were tested on a version of the truth conditions task that included, in addition to a true/false question, an 11-point scale on which participants could rate the extent to which they agreed with the statements provided (e.g., "Lorches/Most lorches have purple feathers"). This second question was added to provide a more fine-grained measure of participants' truth value judgment. Participants' answers on both questions replicated the results reported for Experiment 1 above.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between dangerous/distinctive properties and other properties.) To test this alternative hypothesis, we ran an additional set of participants on a "some" version of our truth conditions task (e.g., "Some lorches have purple feathers"). If the high proportion of "true" responses to the bare plural statements in Experiment 1 was driven by an implicit indefinite interpretation, then participants should be equally likely to accept the *explicitly* indefinite sentences in Experiment 2. If, on the other hand, participants interpreted the bare plural statements in the first study as generic, the proportion of "true" responses should be considerably higher in Experiment 2 because, arguably, indefinite statements have looser truth conditions than generics.

3.1. Method

3.1.1. Participants—Twenty-two undergraduates (5 males, 17 females) from a large public university participated in the study for course credit.

3.1.2. Design, procedure, and materials—The design, procedure, and materials were the same as in the truth conditions task of Experiment 1, with only one exception: The generic sentences were replaced with "some"-quantified indefinite sentences.

3.2. Results and discussion

To compare participants' willingness to accept the bare plural and "some"-quantified statements, we submitted their responses to an ANOVA with item wording (bare plural [from Experiment 1] vs. "some" [from Experiment 2]; between subjects), item type (plain vs. dangerous/distinctive vs. non-distinctive; within subject) and prevalence level (10%, 30%, 50%, 70%, and 90%; within subject) as factors. This ANOVA revealed a number of significant main effects and interactions. For brevity, we report only those involving item wording, the focus of this study. The average proportion of "true" responses was significantly higher for "some"-quantified statements (M = .90) than for bare plural statements (M = .57), $F(1, 40) = 20.54$, $p < .001$, $\eta_p^2 = .34$. An interaction between item wording and item type, $F(2, 80) = 6.04$, $p = .004$, $\eta_p^2 = .13$, suggested that, although this difference was significant for all three item types (ps < .019), it was particularly pronounced for the plain and non-distinctive items. The item wording × prevalence level interaction was also significant, $F(4, 160) = 15.34$, $p < .001$, $\eta_p^2 = .28$. The differences between the "some"-quantified and bare plural statements were significant at the 10%, 30%, and 50% prevalence levels (ps < .001) but not at 70% and 90% (ps > .52; see Fig. 3). Interestingly, part of the reason for this significant interaction may be that the "some"-quantified sentences were accepted *less often* at the higher than at the lower prevalence levels (see Fig. 3), probably due to scalar implicature (e.g., Levinson, 1983). Finally, the three-way interaction among item wording, item type, and prevalence level was also significant (but only in the by-subjects analysis), $F(8, 320) = 2.04$, $p = .041$, $\eta_p^2 = .05$. Although the "some" vs. bare plural differences at the lower prevalence levels held up for all three item types (ps < .003), they were somewhat larger for the plain and non-distinctive control items.

These findings suggest that participants' interpretation of bare plural sentences such as "Lorches have purple feathers" is not indefinite or existential. Had the preferred interpretation been indefinite, participants' responses should have been similar across these two experiments. The clear differences we found support the argument that the bare plural statements in Experiment 1 were in fact interpreted as *generic*. Thus, we are justified in concluding that *generics* exhibit an asymmetry between their truth conditions and their implied prevalence.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

4. Experiment 3

The results so far provide compelling evidence for the "asymmetry" hypothesis—that generic statements require little evidence to be judged true but have implications that go far beyond what is needed to accept them. In this experiment, we explored the *limits* of this asymmetry. Since the meaning of generic sentences is inextricably linked to people's causal beliefs about kinds (Cimpian et al., 2010; Gelman & Bloom, 2007), it is possible that the asymmetry between generics' truth conditions and their prevalence implications would be sensitive to these beliefs as well. Note that the properties we used so far were all plausibly construed as biological, stable, and "essential" (Gelman & Wellman, 1991; Gelman, 2003). This type of content may be particularly compatible with the inference that the properties apply to nearly all category members (in the implied prevalence task) and that they are true of the kind despite low prevalence (in the truth conditions task). If the properties were more superficial, though, perhaps the asymmetry between the two tasks would be reduced or even eliminated. For example, if participants in the truth conditions task were told that 10% of lorches have *muddy* feathers or that 30% of ludinos have *broken* legs, they may be reluctant to endorse the corresponding generics (e.g., "Lorches have muddy feathers," "Ludinos have broken legs"). Intuitively, the fact that a subset of these animals displays such temporary and accidental-sounding properties does not seem sufficient to accept them as characterizing the entire kind. Along the same lines, if provided with a generic such as "Ludinos have broken legs" in the implied prevalence task, participants' prevalence estimates might be lower than before, given what an implausible coincidence would be needed for, say, a majority of ludinos to have broken legs at any one time. To test these predictions, in Experiment 3 we ran a version of the truth conditions and implied prevalence tasks on a set of generic items that referred to *temporary and accidental states* (e.g., broken legs, muddy feathers, wet fur; see Appendix B).[4]

In addition, this study can serve as a test of another potential alternative explanation: One might argue that the asymmetry obtained in Experiment 1 was driven by the demand characteristics of our tasks. In the truth conditions task, for instance, participants had to make a true/false judgment based on very little information. Uncertain about how to respond, participants may have assented to the generic statements simply as a coping strategy. Responses on the implied prevalence task were similarly underdetermined, so it is possible that the high percentage estimates also reflected, at least in part, some task demand. The influence of these demands may have been weaker in the "most" condition because the explicit quantifier established clearer guidelines for responding; in contrast, the bare plurals in our generic statements were considerably more ambiguous, perhaps prompting participants to search for response strategies that were not anchored in the meaning of the sentences. If task demands were indeed responsible for the asymmetry obtained in the generic condition of Experiment 1, then this result should obtain despite changes in property content—as long as the structure of the tasks remains the same. However, if participants' responses were genuinely based on a consideration of the kind concepts to which generics refer, then they should be sensitive to a change in the content of the generic statements. That is, given the accidental nature of the properties in this experiment, the asymmetry between the two tasks should become smaller or disappear altogether.

---

[4]In principle, temporary features *may* be compatible with a more kind-relevant, essential interpretation under certain circumstances. For example, if told that lorches intentionally splash mud on their feathers as part of a mating ritual, participants may treat this feature (muddy feathers) similarly to the biological ones in Experiment 1. It is temporary properties with *accidental* origins that may be particularly likely to reduce asymmetry between generics' truth conditions and their implied prevalence. Although we were not specific about the origin of the properties in Experiment 3, it is likely that they were typically construed as accidental.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

4.1. Method

4.1.1. Participants—Thirty-five undergraduates (20 males, 15 females) from two large public universities participated in the study either for course credit or for $5.

4.1.2. Design, procedure, and materials—Participants were randomly assigned to either a truth conditions task (n = 20) or an implied prevalence task (n = 15). Each task consisted of 20 items, all of which were *generic* and referred to temporary, accidental, or disease states (see Appendix B for full list). Thus, in the truth conditions task, participants were told that a certain percentage of animals have a certain feature (e.g., 10% of lorches have muddy feathers) and then asked whether the corresponding generic statement (e.g., "Lorches have muddy feathers") is true or false. In the implied prevalence task, participants were provided with the generic statements (e.g., "Lorches have muddy feathers") and then asked to estimate what percentage of category members display the relevant features (e.g., what percentage of lorches have muddy feathers).

4.2. Results and discussion

To test whether accidental/superficial properties show less of an asymmetry between truth conditions and implied prevalence, we compared responses from this experiment to responses from Experiment 1, averaged across the plain, dangerous/distinctive, and non-distinctive item types. Thus, we performed an ANOVA with property content (accidental [from Experiment 3] vs. biological [from Experiment 1]; between subjects) and task (truth conditions vs. implied prevalence; between subjects) as factors. This analysis revealed a single significant effect—the interaction between property content and task, $F(1, 66) = 11.26$, $p = .001$, $\eta_p^2 = .15$. As shown in Fig. 4, this interaction was due to the absence of an asymmetry for the accidental properties (muddy feathers, broken legs, etc.). For these items, the average prevalence that led to a "true" response ($M_{truth\ conditions}$ = 78.3%) was, in fact, *higher* than the prevalence implied by them ($M_{implied\ prevalence}$ = 69.5%), although not significantly so, $F(1, 66) = 1.38$, $p = .244$, $\eta_p^2 = .02$. (Recall that the biological properties showed a strong asymmetry, $M_{truth\ conditions}$ = 69.1% vs. $M_{implied\ prevalence}$ = 95.8%, $F[1, 66] = 12.74$, $p = .001$, $\eta_p^2 = .16$.)

These results suggest that the asymmetry between generics' truth conditions and their implied prevalence is—just as is the meaning of generic sentences itself—sensitive to theory-based causal considerations. Relative to subjects in the first experiment, subjects who were told about muddy feathers, broken legs, and other such accidental properties were (a) less likely to assume that all category members share these features and (b) less likely to agree with generic predications of these features.[5] This pattern of responses led to the disappearance of the asymmetry between truth conditions and implied prevalence (see Fig. 4). Finally, these results also argue against the possibility that the original results were due to task demands. The structure of the tasks was the same as before, yet the participants modulated their responses based on the content of the information provided.

[5]The low proportion of "true" responses in this experiment also speaks against the possibility that participants have a default tendency to interpret bare plural sentences as existential or indefinite rather than generic. Had the indefinite been their preferred interpretation, subjects' acceptance of these statements should have been uniformly high *regardless of property content*. That is, even if only 10% of lorches have *muddy* feathers, it is still true that there exist some lorches that have muddy feathers. The fact that participants were relatively reluctant to endorse these bare plural statements when the properties were accidental (e.g., the proportion of "true" responses at the 10% prevalence level was .11 for accidental properties vs. .34 for biological properties, $t[38] = 2.14$, $p = .039$) is another piece of evidence against an existential interpretation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

5. Experiment 4

This experiment was designed to clarify the role of the danger and distinctiveness information in reasoning about the truth of generic statements. We hypothesized that the meaning of these statements reflects the biases of the underlying conceptual representations, such that features that are privileged in the representation of a concept may be more easily accepted in generic form than other, equally prevalent, features. Moreover, we argued that features that are *either* dangerous *or* distinctive have high informational value and may thus be weighed more heavily in their respective concepts, which may in turn lead to an increase in the acceptability of the generic statements that describe them. So far, we have shown that the *combination* of danger and distinctiveness information leads to an increase in the proportion of "true" responses (see Fig. 3). However, given our original argument, we predict that either of these pieces of information should be *independently* sufficient to relax generics' truth conditions. Testing this prediction would clarify which aspects of our conceptual representations are reflected in the meaning of generic statements.

For this study, then, we separated the dangerous/distinctive items from Experiment 1 into *dangerous* items and *distinctive* items and tested whether each of these item types would draw a higher proportion of "true" responses in the truth conditions task than the plain items. We also ran the implied prevalence task with these items to check whether the asymmetry between generics' truth conditions and implied prevalence would replicate. Since these properties are inherent and biological rather than accidental, we predict it should.

5.1. Method

5.1.1. Participants—Thirty-five undergraduates (20 males, 15 females) from two large public universities participated in the study either for course credit or for $5.

5.1.2. Design, procedure, and materials—Participants were randomly assigned to complete either an implied prevalence task (n = 15) or a truth conditions task (n = 20), each consisting of 30 items. There were three types of items: plain, dangerous, and distinctive. The *plain* items were the same as in Experiment 1. The *dangerous* items were generated by retaining only the danger information from the dangerous/distinctive items used in Experiment 1. For example, a subject in the truth conditions task would be provided with items such as,

> 30% of lorches have dangerous purple feathers. These feathers are as sharp as needles and can easily get lodged in you, causing massive bleeding.

The subject would then be asked whether the generic sentence, "Lorches have dangerous purple feathers," is true or false. In the implied prevalence task, the generic sentence and the statement containing the danger information were provided, and a percentage estimate was required. The *distinctive* items were generated by retaining only the distinctiveness information from the dangerous/distinctive items used in Experiment 1. Recall, however, that the sentence that carried this information in Experiment 1 was identical across items ("No other animals on this island have this kind of [body part]"). Since the danger information varied from item to item, we decided it may be best to vary the distinctiveness information as well. We thus supplemented it with some of the facts from the non-distinctive control items used in Experiment 1 (e.g., that the feathers are wide and smooth; see Appendix A). To illustrate, a subject in the truth conditions task might read,

> 30% of lorches have distinctive purple feathers. No other animals on this island have wide, smooth feathers like these.

The subject would then have to decide whether the generic sentence, “Lorches have distinctive purple feathers,” is true or false. In the implied prevalence task, the generic sentence and the statement about the feature’s distinctiveness were provided, and a prevalence estimate was required.

We used only *generic* versions of the items in this study. Since participants’ acceptance of “most”-quantified sentences was not influenced even by the combined danger-plus-distinctiveness information in Experiment 1, it is unlikely it would have shown any differences here.

5.2. Results and discussion

5.2.1. The impact of the danger and distinctiveness information on generics’ truth conditions—An ANOVA was performed on the proportion of “true” responses provided in the truth conditions task. The factors were item type (plain vs. dangerous vs. distinctive; within subject) and prevalence level (10%, 30%, 50%, 70%, and 90%; within subject). Most importantly, this analysis revealed a significant main effect of item type, $F(2, 38) = 3.66$, $p = .035$, $\eta_p^2 = .16$. As predicted, the proportion of the “true” responses was higher for the dangerous ($M = .68$) and the distinctive ($M = .66$) items than for the plain items ($M = .55$), *p*s = .028 and .048, respectively.[6] In sum, it appears that both danger and distinctiveness have *independent* effects on generics’ truth conditions, making these statements overall more likely to be judged true.

The ANOVA also uncovered an interaction between item type and prevalence level, $F(8, 152) = 2.99$, $p = .004$, $\eta_p^2 = .14$. (This interaction was not significant in the by-items ANOVA.) As illustrated in Fig. 5, the danger and distinctiveness information were most effective in boosting generics’ acceptability at some of the lower prevalence levels, particularly 30%. Bonferroni-adjusted post-hoc tests revealed that the 30% prevalence level was the only place where the advantage for the dangerous and distinctive items was significant (*p*s < .025).

Finally, the main effect of prevalence level was also significant, $F(4, 76) = 25.84$, $p < .001$, $\eta_p^2 = .58$, indicating that participants were more likely to accept the generic statements at higher prevalence levels (see Fig. 5).

5.2.2. Is there an asymmetry between generics’ truth conditions and their prevalence implications?—We asked whether, for this set of properties as well, the prevalence needed to accept a generic statement is significantly lower than the prevalence implied by it. To answer this question, we performed an ANOVA on participants’ prevalence scores, with task (truth conditions vs. implied prevalence; between subjects) and item type (plain vs. dangerous vs. distinctive; within subject) as factors. The only significant result was a main effect of task, $F(1, 33) = 21.40$, $p < .001$, $\eta_p^2 = .39$. As expected, the average prevalence that led to a “true” response in the truth conditions task ($M = 67.4\%$) was significantly lower than the average prevalence estimate generated in the implied prevalence task ($M = 90.4\%$). This asymmetry replicates closely the one obtained for the generic items in Experiment 1 (see Fig. 2).

[6]Note that the Bonferroni adjustment was not applied to these post-hoc tests. In the special case when one is following up on a significant main effect involving three levels, unadjusted post-hoc tests keep the familywise Type I error rate at $\alpha = .05$, so no correction is needed (Levin, Serlin, & Seaman, 1994; see especially p. 155).

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

5.2.3. Conclusions—These results confirm our original prediction that properties that are either distinctive or dangerous to humans would be more acceptable in generic form than properties that are otherwise similar but do not have these semantic elements. The impact of this information was particularly noticeable at some of the lower prevalence levels, which agrees well with Leslie's (2008) observation that many of the low-prevalence generics we judge to be true of *familiar* categories have connotations of danger or distinctiveness (e.g., "Mosquitoes carry the West Nile virus," "Cardinals are red").

In addition, the comparison between participants' responses on the truth conditions and implied prevalence tasks revealed, once again, that the prevalence needed to accept generic statements is lower than the prevalence implied by them.

6. General Discussion

These four studies demonstrate that generics embody a paradoxical combination of flexible truth conditions and near-universal prevalence implications. Generic statements were often judged true given prevalence levels of 50% or lower (see Figs. 3 and 5), and yet they suggested to participants that the relevant properties occur in over 90% of category members. In contrast, sentences quantified with "most" showed no such differences, suggesting that this asymmetry may be specific to generics. Our studies also provide evidence for the idea that the truth conditions of generic sentences reflect the internal structure of kind representations. Specifically, we hypothesized that properties that are either distinctive or dangerous to humans are likely to be highlighted, or given more weight, in the representation of a kind; this may, in turn, make them more likely to be seen as characterizing the entire kind and thus more likely to be judged true in generic form (see also Leslie, 2008). The results supported this claim, as generic predications of properties that were described as dangerous, distinctive, or both were accepted more often than generic predications of other, similar, properties. This pattern was especially strong when the prevalence of the properties was low (see Figs. 3 and 5). We now go on to discuss each of these contributions in more detail.

6.1. The asymmetry

6.1.1. Evidence and limitations—These findings suggest that the asymmetry between generics' implications and their truth conditions is a robust, general aspect of the meaning of generic statements. In addition, the contrast with participants' responses to the "most" items —arguably the closest quantified analog of generics (e.g., Carlson, 1977)—reinforces the conclusion that this asymmetry is a feature peculiar to *generic* meaning per se.

Though the truth-conditions-vs.-implied-prevalence asymmetry may be a robust feature of generics' meaning, it is not an obligatory one. In Experiment 3, for instance, participants were relatively reluctant to accept generic statements about temporary or accidental states (e.g., "Ludinos have broken legs"), and they also did not expect these generic statements to apply to quite as many category instances. Note, however, that properties of this type do not lend themselves very well to generic predication (Cimpian & Markman, 2008; Gelman, 1988), so generics about broken legs, itchy skin, etc. are infrequent outside the laboratory. Thus, the limitation they reveal is mostly a theoretical one.

Even in cases where there is an asymmetry, however, there are probably many factors that modulate its magnitude. For example, verbs differ systematically in their "generalizing power"—that is, in their ability to express acceptable generalizations based on sparse evidence (Abelson & Kanouse, 1966; Gilson & Abelson, 1965). To illustrate, the verb "have" appears to have more generalizing power than "like." Given the same low prevalence, participants in Gilson and Abelson's study were more likely to judge as true

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

generics involving "have" (e.g., "Do tribes *have* sports magazines?") than ones involving "like" (e.g., "Do tribes *like* sports magazines?"). Variability on this dimension would certainly affect the size of the asymmetry: The stricter the truth conditions, the smaller the discrepancy with implied prevalence.

Another factor to consider here is the prototypicality of the sample on which the true/false judgments are based. Rips (1975), among others, has shown that more prototypical inductive bases lead to wider generalizations. For example, people generalize a novel feature more broadly within the *bird* category if they are told that sparrows possess it than if they are told that ostriches do (see also Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rhodes, Brickman, & Gelman, 2008). Similarly, people's willingness to accept generic statements such as "Lorches have purple feathers" may depend on their assumptions about the prototypicality of the sample provided as evidence. If the purple-feathered lorches, however few they may be, are assumed to be in other ways prototypical, the generic should be judged true more often than if this prototypicality assumption was blocked (e.g., if the purple-feathered lorches were said to be very old or sick). In sum, the more prototypical the evidence is assumed to be, the higher the likelihood that the generic statement is accepted, and thus the greater the magnitude of the asymmetry between its truth conditions and implied prevalence.

The ontological domain of the referred-to entities might also modulate the size of the asymmetry. The generics we used in these studies were about natural kinds, and natural kinds are particularly cohesive, in that their members are alike on a variety of dimensions (e.g., biology, diet, habitat; see Gelman, 1988, 2003; Keil, 1989). The more homogenous the category, the more likely it is that (a) a generic about the category applies very widely and (b) a property observed in a subset is true of the category as a whole. Thus, the asymmetry may be stronger in domains whose categories are cohesive or homogeneous (such as natural kinds) than in domains whose categories are more loosely bound together (such as artifacts).

6.1.2. Implications—The discrepancy between generics' truth conditions and implied prevalence is not just an arcane bit of experimental data. On the contrary, there are several real-world phenomena to which this feature of generics' meaning is relevant. Take, for example, political discourse. As Abelson and Kanouse (1966) pointed out, generics are a powerful means of manipulating public opinion. Since these generalizations are legitimized even by scant evidence, their truth is rarely questioned. Yet, after they become part of accepted discourse, they take on a life of their own, turning what may have originally been a nuanced, contextualized fact into a definitive pronouncement: A few cases of successful school voucher programs morph into "School vouchers work"; a few salient incidents at nuclear power plants become "Nuclear power plants are dangerous"; and so on.

This asymmetry has immediate relevance to stereotyping as well. Generics are a common and natural means of expressing stereotypes (Gelman et al., 2004). Consider, for example, a sentence such as "Boys are good at math." Due to generics' flexible truth conditions, this statement may be accepted based on very little evidence—perhaps even no evidence, especially if it comes from a trustworthy source (e.g., Jaswal & Neely, 2006; Koenig, Clement, & Harris, 2004). Once believed to be true, though, this statement may imply that being good at math is a normative, near-universal fact about boys, which may in turn have a powerful effect on one's perceptions and behavior (e.g., Cimpian, in press; Gelman et al., 2004).

This asymmetry also raises broader questions about the process of learning facts about the world through generic sentences. When children first hear, say, that "Ducks lay eggs", do they infer (erroneously) that all ducks lay eggs? It is possible that they do, especially given

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

our current results and children's presumed lack of knowledge about reproductive biology. If so, what happens when they eventually learn that only female ducks can lay eggs? Are children able to revise downward their expectations about the prevalence of this property, or do the near-universal prevalence expectations somehow co-exist with the knowledge that only less than half of ducks lay eggs? Although this last possibility seems far-fetched, it is in fact supported by a recent study. Khemlani, Leslie, and Glucksberg (2009) asked adult participants to imagine an arbitrary member of a category (e.g., "Suppose you are told: Quacky is a duck") and then judge whether it displays a certain property (e.g., "What do you think of the following statement: Quacky lays eggs"). On a scale from −3 (false) to +3 (true), participants' rating of the truth of "Quacky lays eggs" and of other similar items was at 1.7; translated into percentages, this rating corresponds to about 78.3% confidence in the truth of these statements. This confidence level is surprisingly high when compared with the actual probability that a randomly selected duck would be able to lay eggs. Khemlani et al. argue that this pattern of responses "is explicable only if one posits that people understand 'ducks lay eggs' to be an unrestricted generalization across the entire kind *duck*" (p. 447)—or, from our perspective, if people fail to revise generics' near-universal prevalence implications on the basis of other known facts (about, for example, sexual reproduction).

6.1.3. Potential causes of the asymmetry—What might explain the asymmetry between generics' truth conditions and their prevalence implications? Though we do not have a definitive answer to this question, a few ideas are provided below. First, why do generics have such strong prevalence implications? A sentence such as "Lorches have purple feathers" establishes a mapping between the property *having purple feathers* and the abstract kind *lorches*. However, in and of itself, this mapping leaves open the question of *how many* lorches display this feature. The fact that a property is considered true of a kind does not automatically imply that it is true of every member of that kind. On the contrary, people are willing to say that a property is true of a kind given quite a range of prevalence levels (see our experiments; Cimpian et al., 2010). Something extra is needed, then, to explain the near-universal prevalence estimates. One proposal is that generics about high-prevalence features (e.g., "Dogs have four legs") are most *prototypical*, which might lead to a default, but defeasible, assumption that any new generic one hears is about a high-prevalence feature as well. A second option would be to invoke a Gricean (Grice, 1975) principle, following Declerck's (1991) lead:

> a statement about the set as a whole will be interpreted as a statement about all the members of the set [because] if a statement is applicable only to a subset of a set, then it is misleading to make it in connection with the set as a whole (p. 84).

If the speaker had meant to pick out a subset, this should have been indicated by the use of another description (e.g., "some lorches"). Using an expression that is compatible with a stronger interpretation when one intends to refer to something weaker violates the norms of cooperative communication (in particular, Grice's maxim of quantity).[7] Of course, both this pragmatic principle and the default assumption described under the first proposal would interact with many other sources of information, such as context and general knowledge (see Declerck, 1991). For example, when the properties are accidental (e.g., broken legs, as in Experiment 3), the tendency to interpret the generic noun phrases as referring to nearly all members of the category would probably be curbed by the realization that it would be quite improbable for an accidental feature to be so prevalent.

[7]This pragmatic principle could be construed as the inverse of scalar implicature (e.g., Levinson, 1983). The idea behind scalar implicature is that using an expression which is compatible with a *weaker* interpretation (e.g., "some") when one really intends to refer to something *stronger* (e.g., all) violates the norms of cooperative communication. Thus, when a speaker says "I solved *some* of the problems," the listener is justified in inferring that *not all* of the problems were solved.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Evaluating the truth of generic statements might require a different set of computations: Instead of simply trying to reconstruct the speaker's intended meaning, the listener has to (1) map out the range of circumstances under which the sentence is true and (2) determine whether the circumstances provided fall within that range. The fact that generics lack an explicit quantifier leaves quite undetermined the size of their referent sets, so there is often considerable ambiguity in their interpretation (Abelson & Kanouse, 1966; Declerck, 1986). (Note the contrast with "most"-quantified sentences.) This ambiguity in scope may broaden the range obtained in step (1), resulting in "true" responses even at relatively low prevalence levels. Also, we should not lose sight of the fact that generic sentences refer to kinds. Thus, in judging whether a generic is true or not, participants may not be merely mapping out what prevalence levels fall under the ambiguous quantificational scope of the generic NP (as implied above), but also deciding whether there is sufficient evidence to attribute the property to the kind as an abstract entity. Importantly, the evidence needed for these kind attributions is not limited to prevalence but rather includes naive-theoretical considerations as well, which can sometimes compensate for low prevalence (e.g., the present studies; Cimpian et al., 2010). Thus, the wide range of prevalence levels at which generic statements are accepted may also be explained by the kind-reference aspect of their meaning.

6.1.4. Relationship to the literature on inductive inferences—The processes involved in reasoning about generics' truth conditions and prevalence implications are clearly inductive rather than deductive. As such, our findings add to the rich literature on inductive inference and generalization (e.g., Carey, 1985; Gelman & E. M. Markman, 1986; Osherson et al., 1990; Rips, 1975; Sloman, 1993; Tversky & Kahneman, 1983; Yamauchi, 2005; Yamauchi & A. B. Markman, 2000). Typically, previous work on this topic has investigated the factors that affect people's inductive generalizations from one category to another (e.g., if dogs have golgi inside, do bees have them as well?), from one category member to another (e.g., if this bird feeds its babies mashed-up food, does this other bird do the same?), or from one feature to another (e.g., if Linda is a feminist, is she also a philosophy major?). The inductive judgments required of participants in our studies are of a different sort. They involve alternating between two different modes of thinking about the *same* category-feature pairs: one that operates over the *extension* of the relevant concept (e.g., the frequency of purple feathers in the set of all *lorches*) and one that operates over the *intension* of the concept (e.g., purple feathers as a feature of the abstract concept *lorches*). In the truth conditions task, the inductive inference consists of deciding whether an extensional fact licenses the corresponding intensional conclusion—that is, whether the feature can be legitimately considered part of the meaning of the concept given that a certain percentage of entities in the concept's extension possess it. The implied prevalence task requires an inductive inference in the opposite direction, translating an intensional fact into its extensional equivalent.

This description applies specifically to reasoning about *generic* statements. "Most"-quantified noun phrases do not refer to kinds and therefore do not call up the intensional mode. Rather, participants have to translate between two types of extensional, or frequency-based, considerations—the prevalence of the relevant property and the meaning of the quantifier "most." The greater compatibility between the base and the target of the inductive inferences in the "most" condition may facilitate the translation between the two, which may in turn explain the symmetry between people's responses in the truth conditions and implied prevalence tasks.

In contrast, the asymmetry obtained in the generic condition may suggest that extensional and intensional facts are not as commensurable. This idea is supported not only by our own results but also by a number of induction studies in which intensional and extensional reasoning produce discrepant outcomes (e.g., Jönsson & Hampton, 2006; Sloman, 1993;

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Tversky & Kahneman, 1983). For example, if a property is true of a set (e.g., "All sofas have back rests"), then it should be true of all subsets as well (e.g., "All uncomfortable handmade sofas have back rests"). An extensional reasoning process based on the relevant class inclusion relationships would arrive at this (normatively correct) conclusion. However, people in these studies often agree with the general statements and disagree with the more specific ones (Jönsson & Hampton, 2006) because they rely on the *intension* of the concepts involved. For example, because *having a backrest* is not a strong feature of our concept of *uncomfortable handmade sofas*, the statement that pairs the two is often deemed false. From this perspective, the asymmetry obtained in our studies is another compelling demonstration of the non-equivalence between the intensional and extensional modes of reasoning.

6.2. The impact of the danger and distinctiveness information

Across our experiments, participants accepted novel generics about distinctive or dangerous properties more often than generics about properties that were overall similar but lacked these semantic elements. These results support the argument that generic statements are best thought of as a direct linguistic outlet for our conceptual representations. As such, the meaning of these statements should reflect the biases of our conceptual representations. Since properties that are either distinctive or dangerous to humans are likely to be weighed more heavily or otherwise privileged in the representation of a concept, it makes sense that they would also be more easily accepted as being true of the kind as a whole even when they are not very prevalent.

However, what reason is there to accept the assumption that dangerous and distinctive properties occupy a privileged place among the features represented in a concept? Besides its intuitive appeal, there is in fact independent empirical evidence for this claim. For example, when provided with a concept (e.g., *beaver*), people are faster to verify features that are distinctive (e.g., *builds dams*) than features that are not (e.g., *swims*), even when the two types of features are equally strongly associated with the respective concept (Cree, McNorgan, & McRae, 2006). Several demonstrations that danger- and threat-related properties are also privileged can be found in the perception literature. For example, even infants orient faster to pictures of snakes than to pictures of perceptually similar, but non-threatening, stimuli (LoBue & DeLoache, 2010); analogous processing advantages for angry faces, which typically signal threat, can be found at all ages as well (e.g., LoBue, 2009; Öhman, Lundqvist, & Esteves, 2001).

6.3. Conclusion

Generic sentences are a common and powerful means of conveying knowledge about the natural and social worlds. As a result, they have been argued to play a significant role in children's conceptual development (Cimpian & Markman, 2009; Gelman, 2009) and in stereotyping and other social psychological processes (Abelson & Kanouse, 1966; Cimpian, in press; Gelman et al., 2004). However, generic language remains in many respects a puzzle. By demonstrating the influence of danger and distinctiveness information on generics' truth conditions, the present studies support the view that these sentences give linguistic expression to our conceptual representations. More importantly, our paper identifies a core feature of generics' meaning: Generic statements are often judged true based on weak evidence but have implications that go far beyond what is needed to accept them.

Acknowledgments

This research was supported in part by research funds from the University of Illinois to Cimpian, an NSF Graduate Research Fellowship to Brandone, and NICHD Grant HD-36043 to Gelman. We would like to thank Sarah-Jane Leslie and Brian Ross for helpful discussion; Joe Robinson, Cindy Fisher, Vikram Jaswal, Gedeon Deák, and two

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

anonymous reviewers for their comments on previous drafts of the manuscript; and Jenna Hedglen, Ben Boldt, and the UIUC Cognitive Development Lab team for their assistance in collecting and entering the data.

References

Abelson, RP.; Kanouse, DE. Subjective acceptance of verbal generalizations. In: Feldman, S., editor. Cognitive consistency: Motivational antecedents and behavioral consequents. New York: Academic Press; 1966. p. 171-197.

Carey, S. Conceptual change in childhood. Cambridge, MA: MIT Press; 1985.

Carlson, GN. Doctoral dissertation. Amherst: University of Massachussetts; 1977. Reference to kinds in English.

Carlson, GN. The truth conditions of generic sentences: Two contrasting views. In: Carlson, GN.; Pelletier, FJ., editors. The generic book. Chicago, IL: The University of Chicago Press; 1995. p. 224-237.

Carlson, GN.; Pelletier, FJ. The generic book. Chicago, IL: The University of Chicago Press; 1995.

Cimpian A. The impact of generic language about ability on children’s achievement motivation. Developmental Psychology. (in press).

Cimpian A, Gelman SA, Brandone AC. Theory-based considerations influence the interpretation of generic sentences. Language and Cognitive Processes. 2010; 25(2):261–276. [PubMed: 20352078]

Cimpian A, Markman EM. Preschool children’s use of cues to generic meaning. Cognition. 2008; 107(1):19–53. [PubMed: 17765216]

Cimpian A, Markman EM. Information learned from generic language becomes central to children’s biological concepts: Evidence from their open-ended explanations. Cognition. 2009; 113(1):14–25. [PubMed: 19674739]

Cox, A. U.S. poised for epidemic West Nile year. 2004. Retrieved from http://edition.cnn.com/2004/HEALTH/05/03/wnv.outlook/

Cree GS, McNorgan C, McRae K. Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32(4):643–658.

Dahl, Ö. On generics. In: Keenan, EL., editor. Formal semantics of natural language: Papers from a colloquium sponsored by the King’s College Research Centre, Cambridge. Cambridge: Cambridge University Press; 1975. p. 99-111.

Declerck R. The manifold interpretations of generic sentences. Lingua. 1986; 68:149–188.

Declerck R. The origins of genericity. Linguistics. 1991; 29:79–102.

Gelman SA. The development of induction within natural kind and artifact categories. Cognitive Psychology. 1988; 20:65–95. [PubMed: 3338268]

Gelman, SA. The essential child: Origins of essentialism in everyday thought. London: Oxford University Press; 2003.

Gelman, SA. Learning words for kinds: Generic noun phrases in acquisition. In: Hall, DG.; Waxman, SR., editors. Weaving a lexicon. Cambridge, MA: MIT Press; 2004. p. 445-484.

Gelman SA. Learning from others: Children’s construction of concepts. Annual Review of Psychology. 2009; 60:115–140.

Gelman SA, Bloom P. Developmental changes in the understanding of generics. Cognition. 2007; 105(1):166–183. [PubMed: 17094957]

Gelman SA, Coley JD, Rosengren KS, Hartman E, Pappas A. Beyond labeling: The role of parental input in the acquisition of richly-structured categories. Monographs of the Society for Research in Child Development. 1998; 63 Serial No. 253.

Gelman SA, Goetz PJ, Sarnecka BW, Flukes J. Generic language in parent-child conversations. Language Learning and Development. 2008; 4:1–31.

Gelman SA, Markman EM. Categories and induction in young children. Cognition. 1986; 23:183–209. [PubMed: 3791915]

Gelman SA, Star JR, Flukes JE. Children’s use of generics in inductive inferences. Journal of Cognition and Development. 2002; 3:179–199.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Gelman SA, Taylor MG, Nguyen SP. Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. Monographs of the Society for Research in Child Development. 2004; 69 Serial No. 275.

Gelman SA, Wellman HM. Insides and essences: Early understandings of the non-obvious. Cognition. 1991; 38:213–244. [PubMed: 2060270]

Gilson C, Abelson RP. The subjective use of inductive evidence. Journal of Personality and Social Psychology. 1965; 2(3):301–310. [PubMed: 14333301]

Glass GV, Peckham PD, Sanders JR. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research. 1972; 42(3): 237–288.

Grice, HP. Logic and conversation. In: Cole, P.; Morgan, JL., editors. Syntax and semantics. Vol. vol. 3. New York: Academic Press; 1975. p. 41-58.

Howell, DC. Statistical methods for psychology. 7th ed.. Belmont, CA: Wadsworth; 2009.

Jaswal VK, Neely LA. Adults don’t always know best: Preschoolers use past reliability over age when learning new words. Psychological Science. 2006; 17(9):757–758. [PubMed: 16984291]

Jönsson ML, Hampton JA. The inverse conjunction fallacy. Journal of Memory and Language. 2006; 55:317–334.

Keil, FC. Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press; 1989.

Khemlani, S.; Leslie, SJ.; Glucksberg, S. In: Taatgen, N.; van Rijn, H., editors. Generics, prevalence, and default inferences; Proceedings of the 31st Annual Conference of the Cognitive Science Society; Amsterdam, Netherlands: Cognitive Science Society; 2009. p. 443-448.

Khemlani, S.; Leslie, SJ.; Glucksberg, S.; Fernandez, PR. Do ducks lay eggs? How people interpret generic assertions. In: McNamara, DS.; Trafton, JG., editors. Proceedings of the 29th Annual Cognitive Science Society. Austin, TX: Cognitive Science Society; 2007. p. 395-400.

Koenig MA, Clement F, Harris PL. Trust in testimony: Children’s use of true and false statements. Psychological Science. 2004; 15:694–698. [PubMed: 15447641]

Lawler JM. Studies in English generics. University of Michigan Papers in Linguistics. 1973; 1(1):1–184.

Leslie SJ. Generics: Cognition and acquisition. Philosophical Review. 2008; 117(1):1–47.

Levin JR, Serlin RC, Seaman MA. A controlled, powerful multiple-comparison strategy for several situations. Psychological Bulletin. 1994; 115(1):153–159.

Levinson, SC. Pragmatics. Cambridge, England: Cambridge University Press; 1983.

LoBue V. More than just another face in the crowd: Superior detection of threatening facial expressions in children and adults. Developmental Science. 2009; 12(2):305–313. [PubMed: 19143803]

LoBue V, DeLoache JS. Superior detection of threat-relevant stimuli in infancy. Developmental Science. 2010; 13(1):221–228. [PubMed: 20121878]

Murphy, GL. The big book of concepts. Cambridge, MA: MIT Press; 2004.

Murphy GL, Medin DL. The role of theories in conceptual coherence. Psychological Review. 1985; 92:289–316. [PubMed: 4023146]

Öhman A, Lundqvist D, Esteves F. The face in the crowd revisited: A threat advantage with schematic stimuli. Journal of Personality and Social Psychology. 2001; 80(3):381–396. [PubMed: 11300573]

Osherson DN, Smith EE, Wilkie O, Lopez A, Shafir E. Category-based induction. Psychological Review. 1990; 97(2):185–200.

Prasada S. Acquiring generic knowledge. Trends in Cognitive Sciences. 2000; 4:66–72. [PubMed: 10652524]

Prasada S, Dillingham EM. Principled and statistical connections in common sense conception. Cognition. 2006; 99:73–112. [PubMed: 16443448]

Rhodes M, Brickman D, Gelman SA. Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. Cognition. 2008; 108:543–556. [PubMed: 18436200]

Rips LJ. Inductive judgments about natural categories. Journal of Verbal Learning and Verbal Behavior. 1975; 14:665–681.

Sloman SA. Feature-based induction. Cognitive Psychology. 1993; 25:231–280.

Tversky A, Kahneman D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review. 1983; 90(4):293–315.

Yamauchi T, Markman AB. Inference using categories. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000; 26(3):776–795.

Yamauchi T. Labeling bias and categorical induction: Generative aspects of category information. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31(3):538–553.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

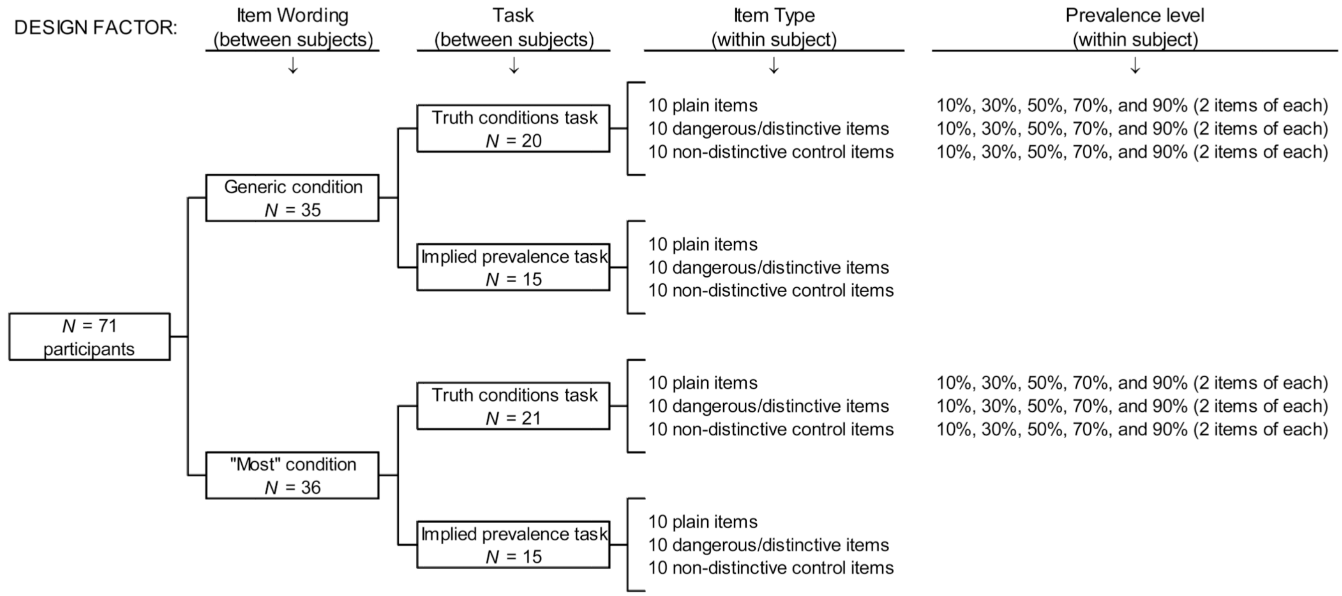

Fig. 1.
The design for Experiment 1.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

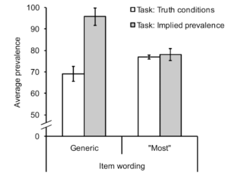

Fig. 2.
The average prevalence that led participants in Experiment 1 to accept statements (the truth conditions task, white bars) vs. the average prevalence implied by the same statements (the implied prevalence task, gray bars). The error bars represent ± 1 standard error of the mean.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

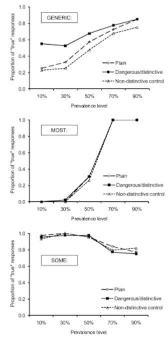

Fig. 3.
Average proportion of "true" responses in Experiments 1 and 2 (truth conditions task), by item wording, item type, and prevalence level.

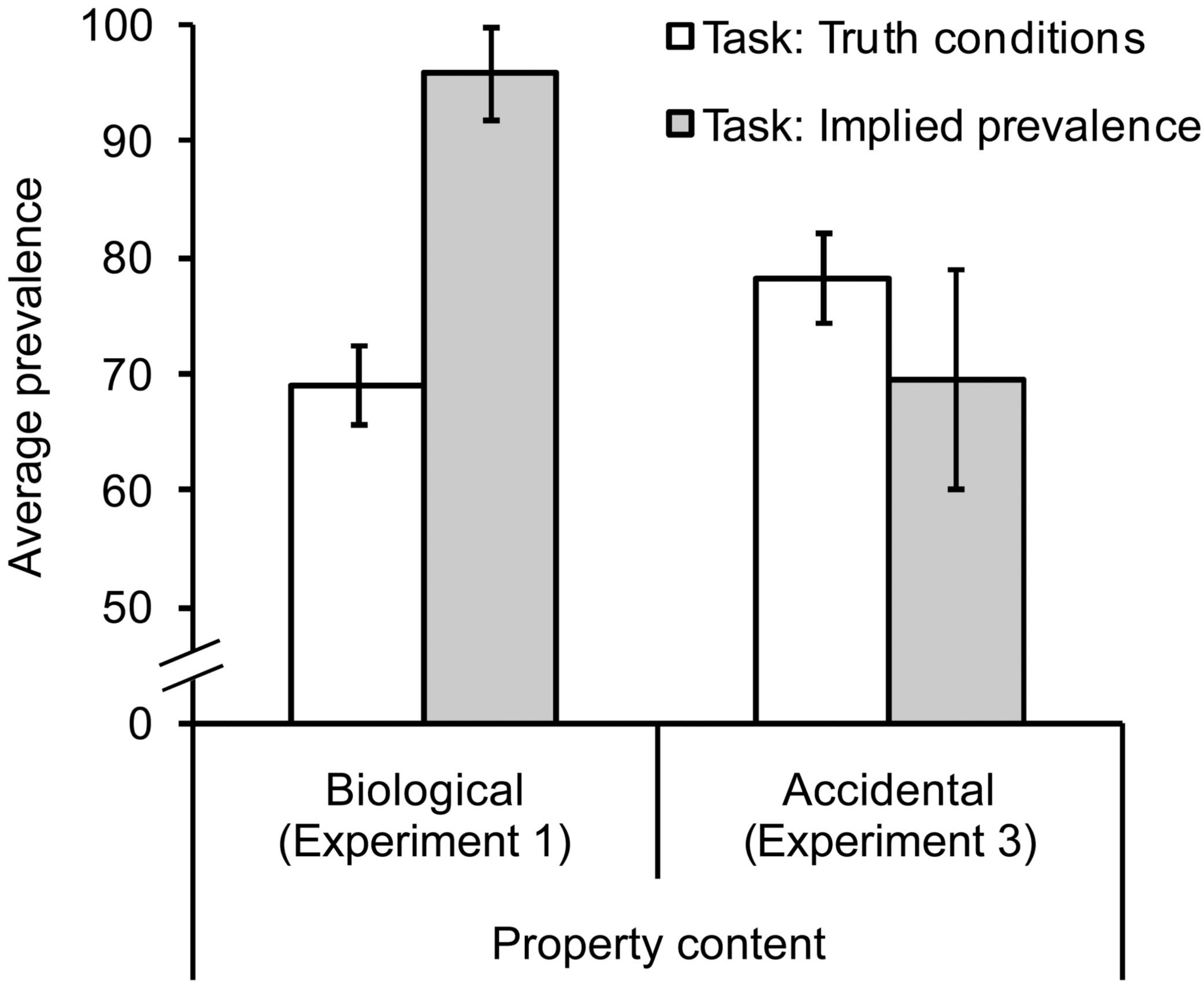

Fig. 4.

The average prevalence that led participants to accept generic statements (the truth conditions task, white bars) vs. the average prevalence implied by the same generic statements (the implied prevalence task, gray bars), by property content. The error bars represent ± 1 standard error of the mean.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

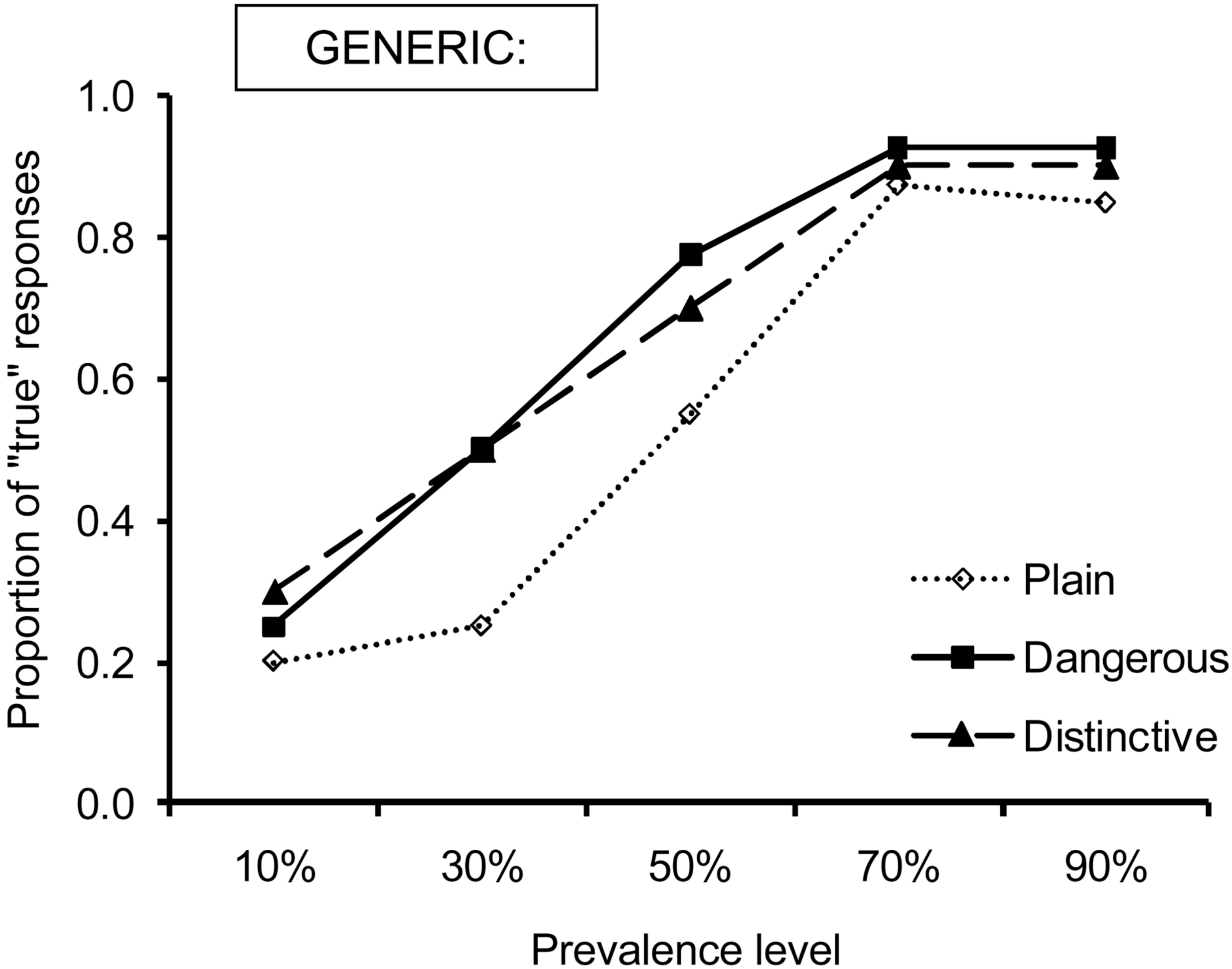

Fig. 5.
Average proportion of “true” responses in Experiment 4 (truth conditions task), by item type and prevalence level.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Sample item from Experiment 1

PLAIN	DANGEROUS/DISTINCTIVE	NON-DISTINCTIVE CONTROL
IMPLIED PREVALENCE TASK:		
Information: [**Morseths/Most morseths**][a] have silver fur.	Information: [**Morseths/Most morseths**][a] have dangerous silver fur. This fur sheds particles that get lodged in your lungs and make it impossible to breathe. No other animals on this island have this kind of fur.[b]	Information: [**Morseths/Most morseths**][a] have curly silver fur. This fur is very curly and rough to the touch. Other animals on this island also have this kind of fur.[c]
Question: What percentage of morseths do you think have silver fur?	Question: What percentage of morseths do you think have dangerous silver fur?	Question: What percentage of morseths do you think have curly silver fur?
TRUTH CONDITIONS TASK:		
Information: xx% of morseths have silver fur.	Information: xx% of morseths have dangerous silver fur. This fur sheds particles that get lodged in your lungs and make it impossible to breathe. No other animals on this island have this kind of fur.[b]	Information: xx% of morseths have curly silver fur. This fur is very curly and rough to the touch. Other animals on this island also have this kind of fur.[c]
Question: Is the following sentence true or false? [**Morseths/Most morseths**][a] have silver fur.	Question: Is the following sentence true or false? [**Morseths/Most morseths**][a] have dangerous silver fur.	Question: Is the following sentence true or false? [**Morseths/Most morseths**][a] have curly silver fur.

[a]"Morseths" was used in the generic condition, "most morseths" in the "most" condition.

[b]The sentence about the property's distinctiveness was the same across all dangerous/distinctive items, but the sentence describing the danger varied.

[c]The sentence implying that the feature is not distinctive was the same across all non-distinctive control items, but the sentence describing the irrelevant information varied.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

APPENDIX A

Complete set of items used in Experiments 1 and 2

Body part	Item subset – Category name – Color	Danger description[a]	Irrelevant fact description[b]
fur	1 – morseths – silver 2 – ollers – yellow 3 – kweps – copper	This fur sheds particles that get lodged in your lungs and make it impossible to breathe.	This fur is very curly and rough to the touch.
scales	1 – blins – red 2 – reesles – blue 3 – dorbs – yellow	These scales secrete a strong venom that kills you on the spot.	These scales are soft, flexible, and very shiny.
tail	1 – zorbs – orange 2 – taifels – purple 3 – trufts – green	These tails are so long and muscular that they can suffocate you in a matter of minutes.	These tails are very long and usually get curled up in a ball.
stripes	1 – daiths – gold 2 – mooks – copper 3 – frams – silver	These stripes deliver a powerful electric shock that’s deadly to anyone within a few feet.	These stripes are very thin and closely spaced.
shell	1 – moxes – green 2 – luzaks – orange 3 – javs – purple	These shells are so very heavy that they would immediately crush your bones.	These shells have an octagonal shape and are very light.
legs	1 – ludinos – yellow 2 – ackles – silver 3 – feps – pink	These legs are so powerful that a single blow could kill you.	These legs are very long and covered with bumpy skin.
ears	1 – cheebas – blue 2 – elleps – pink 3 – kazzes – orange	These ears are the home to dangerous parasites that can make you go deaf.	These ears are small and round.
feathers	1 – lorches – purple 2 – plovs – gold 3 – noobs – red	These feathers are as sharp as needles and can easily get lodged in you, causing massive bleeding.	These feathers are wide and very smooth to the touch.
spots	1 – glippets – copper 2 – sapers – red 3 – stups – gold	These spots are the home to a contagious fungus that is deadly to anyone who becomes infected with it.	These spots are very sensitive and cover most of their bodies.
teeth	1 – krivels – pink 2 – zoovs – green 3 – thups – blue	These teeth are razor-sharp and so powerful that a single bite can be lethal.	These teeth are long and narrow.

[a] All the dangerous/distinctive items also contained the following sentence: “No other animals on this island have this kind of [body part].”

[b] All the non-distinctive control items also contained the following sentence: “Other animals on this island also have this kind of [body part].”

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

APPENDIX B

Complete set of items used in Experiment 3

Category name	Property
ludinos	broken legs
zorbs	broken tails
glippets	cracked claws
moxes	cracked shells
sapers	fungus-covered claws
blins	fungus-covered scales
elleps	infected ears
reesles	infected scales
mooks	itchy skin
taifels	itchy tails
lorches	muddy feathers
morseths	muddy fur
luzaks	rotting shells
krivels	rotting teeth
ackles	sore legs
zoovs	sore teeth
cheebas	swollen ears
daiths	swollen skin
plovs	wet feathers
ollers	wet fur