

**Generics about Categories and Generics about Individuals:
Same Phenomenon or Different?**

Lin Bian¹ and Andrei Cimpian²

¹Department of Psychology, University of Chicago

²Department of Psychology, New York University

In press, *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Acknowledgments

We are grateful to the participants in our studies and to April Bailey, Mark Bowker, Cindy Fisher, Emily Foster-Hanson, Jess Gladstone, James Hampton, Sunny Khemlani, Joshua Knobe, Jillian Lauer, Vivian Liu, Meredith Meyer, Melis Muradoglu, Greg Murphy, George Newman, Arber Tasimi, MH Tessler, Andrea Vial, and the members of the Cognitive Development Lab at the University of Illinois for helpful discussion.

Abstract

Language can be used to express broad, unquantified generalizations about both categories (e.g., “Dogs bark”) and individuals (e.g., “Daisy barks”). Although these two classes of statements are commonly assumed to arise from the same linguistic phenomenon—*genericity*—the literature to date has not offered a direct experimental comparison of the conditions under which they are endorsed (i.e., their truth conditions). Here, we provided such a comparison by testing whether endorsement of generics about categories and individuals is affected in similar ways by the conceptual content of the properties being generalized. Consistent with this possibility, six experiments ($N = 1,265$) and an internal meta-analysis revealed that endorsement of generics about both categories and individuals is facilitated when the properties being generalized are distinctive or dangerous, suggesting systematic similarities in the truth conditions of generics about categories and individuals. The experiments also suggested that these facilitative effects do not extend to statements with overt quantification (*all/always* and *some/sometimes*). Thus, the present studies provide the first empirical evidence indicating that generics about categories and generics about individuals represent a unified phenomenon. These findings contribute to theories of genericity and inform our understanding of how language shapes social interactions.

Keywords: generics, generic statements, generalization, quantifiers

Generics about Categories and Generics about Individuals:

Same Phenomenon or Different?

Languages all over the world provide the means to convey that a property generalizes across the members of a category, as in “Cats drink milk,” or across time-slices of an individual, as in “Fluffy drinks milk” (e.g., Behrens, 2005; Carlson & Pelletier, 1995; Leslie & Lerner, 2016). Although these two classes of statements clearly differ in scope, prominent theoretical accounts suggest that the conditions under which they are endorsed (that is, their truth conditions) are nevertheless similar in key respects (e.g., Krifka et al., 1995; Leslie & Lerner, 2016; Tessler & Goodman, 2019). This is also why they are labeled with the same term: *generic statements* or, simply, *generics*. To date, however, the claim of similarity in truth conditions has been made exclusively on the basis of formal semantic analyses or computational models (e.g., Krifka et al., 1995; Tessler & Goodman, 2019), and a direct experimental comparison of the circumstances under which people endorse generics about categories and individuals is lacking. Thus, performing a systematic comparison of the truth conditions of these statements was our goal here.

Genericity is a complex linguistic phenomenon. Current accounts of this phenomenon suggest that there are two distinct types of linguistic structures that fall under its scope (e.g., Krifka et al., 1995; Leslie & Lerner, 2016). The first—which is not our focus here—consists of sentences in which a predicate is used to describe a category as a whole. Consider, for instance, the sentence “The T-Rex is extinct.” Here, *being extinct* is applied to the entire *Tyrannosaurus Rex* species. These sentences do not express a generalization per se but rather simply a fact about a category considered as a whole, in the same way that saying “Barney the T-Rex is dead” expresses a fact about an individual considered as a whole.

Our focus here is on the second type of linguistic structure that is typically classified as generic. These sentences do express generalizations, either across individual members of a category or across situations in an individual's life. In their landmark treatment of genericity, Krifka et al. (1995) describe these generic sentences as “propositions which do not express specific episodes or isolated facts, but instead report a kind of *general property*, that is, report a regularity which summarizes groups of particular episodes or facts” (p. 2). For example, according to Krifka and colleagues, “Italians drink wine with their dinner” and “Luigi drinks wine with his dinner” are both generic sentences (p. 12) because they express a regularity or generalization that holds across Italians (what we term here a *category generic*) and across events in a particular Italian individual's life (what we term here an *individual generic*), respectively. Many other theoretical accounts include category and individual generics under the same rubric (e.g., Dahl, 1975; Papafragou, 1996; Schubert & Pelletier, 1989). Reflecting this seeming consensus, in their entry summarizing current research on genericity for the *Stanford Encyclopedia of Philosophy*, Leslie and Lerner (2016) report that the term *generic* is used in the literature to refer to both sentences such as “Cats lick themselves” and sentences such as “Mary smokes after work,” even though the former expresses a generalization about a category and the latter does not.

The claim that category and individual generics are part of the same linguistic phenomenon is based on an argument that the meaning of both types of statements can be formalized using the same (unpronounced) generic operator **GEN**, where **GEN** stands for “generic” (e.g., Krifka et al., 1995). **GEN** is hypothesized to function like a quantificational adverb such as *usually* or *generally*, but with a much more flexible and context-sensitive quantificational scope. Critically, however, the conclusion that the meaning of category and

individual generics is similar because it can be formalized with the same semantic operator is based almost exclusively on the intuitions of the formal semanticists who formulated it. Other than these intuitions, there is currently no empirical evidence that bears on the claim of similarity between the truth conditions of category and individual generics. Thus, in the present research we treated this claim as a *hypothesis* to be tested empirically. This hypothesis is worthy of investigation not just because it is a cornerstone of the ever-growing literature on genericity in linguistics and philosophy but also because generics about categories and individuals are probably the most common linguistic means of expressing general beliefs about the world (e.g., Cimpian, 2016; Krifka et al., 1995; Leslie, 2008), with implications for how people navigate their social environments (e.g., Rhodes et al., 2012). In addition, insofar as generics about categories and individuals express generalizations, the claim that the truth conditions of these two types of statements are similar has implications for the cognitive mechanisms that produce the generalizations being expressed. Specifically, this claim might suggest that the cognitive mechanisms that output generalizations about categories and individuals are one and the same or at least share key operating parameters. Thus, a systematic comparison of the truth conditions of category and individual generics—the circumstances under which people endorse or reject them—is likely to make a meaningful contribution to both (psycho)linguistics and cognitive psychology.

The closest thing to empirical evidence on this topic can be found in recent computational modeling work on generic meaning. Tessler and Goodman (2016, 2019) elicited participants' endorsement of a wide range of category generics (e.g., "Cardinals are red") and, separately, individual generics (e.g., "Bill goes to the ballet") and then fit a computational model to these data. Their model provided a good fit to the human data on both category and individual

generics, a result that is consistent with the hypothesized similarity in the truth conditions of these statements. At the same time, closer inspection of their methodology suggests that the comparison they performed was rather indirect. First, the category and individual generics used as stimuli in Tessler and Goodman's studies differed in multiple ways, beyond just their category vs. individual scope. For instance, the category generics were about familiar categories of non-human animals (e.g., cardinals), whereas the individual generics were about unfamiliar human individuals (e.g., Bill). Second, participants' endorsement of the category generics was measured in the absence of information about how frequent the relevant feature was (e.g., how many cardinals are red)—in part because these were all familiar facts—whereas participants' endorsement of individual generics was measured after they were told how frequent the relevant behavior was (e.g., how often Bill goes to the ballet). Third, Tessler and Goodman fit their computational model separately on the two types of generics rather than simultaneously. These aspects of Tessler and Goodman's methodology highlight the value of a more closely matched, direct comparison between the truth conditions of category and individual generics.

Comparing the Truth Conditions of Generics about Categories and Individuals

To our knowledge, no studies have systematically compared the truth conditions of category and individual generics in an experimental setup. In what follows, we detail our proposed experimental test, highlighting specific predictions. We then summarize prior evidence bearing on these predictions.

The Overall Logic. To test whether category and individual generics are meaningfully similar in their truth conditions, we first need to articulate what features we might expect their truth conditions to share. For this purpose, we rely on the so-called *content-based hypothesis* about the semantics of **GEN** (for a review, see Leslie & Lerner, 2016; Lerner & Leslie, 2016).

According to this hypothesis, **GEN** is sensitive to the content of the properties being generalized, all else being equal. So far, tests of this hypothesis have focused specifically on category generics and have suggested that category generics about properties that are *distinctive* or *dangerous* are more easily endorsed than ones about properties that are otherwise similar but are neither distinctive nor dangerous (e.g., Cimpian et al., 2010; Leslie, 2007, 2008). For example, a category generic about a property such as *having red feathers* is more readily endorsed if the property is present only among members of that category (i.e., when the property is distinctive) than when members of other categories display it as well. Similarly, people more readily endorse a category generic about a property that poses a danger to humans (e.g., *having poisonous red feathers*) than about a neutral property such as *having red feathers* (see also Zhu & Murphy, 2013). Notably, the endorsement advantage for distinctive and dangerous properties is stronger when the relevant features are relatively rare (e.g., Cimpian et al., 2010). Category generics about properties that are common within a category are typically judged to be true regardless of the conceptual content of these properties (Prasada et al., 2013).¹

The evidence that endorsement of category generics is sensitive to property content motivates our key question: Will property content matter in the same way for *individual* generics as well? That is, will distinctive or dangerous behaviors also lead to higher endorsement of the corresponding individual generics compared to similar behaviors that are not distinctive or dangerous? Thus, the present experiments investigate the prediction that endorsement of category and individual generics will be affected similarly by the content of the properties they describe. This sensitivity to property content is the critical aspect of **GEN**'s semantics that

¹ There are exceptions, of course. Even though prevalence is an important factor driving endorsement of category generics, high prevalence is neither necessary nor sufficient for endorsement (e.g., Cimpian et al., 2010). For instance, "Mosquitoes don't carry malaria" seems false even though most mosquitoes don't in fact carry malaria.

should be observed with respect to both types of generics.

Prior Evidence: Distinctiveness. There is some prior evidence suggesting that endorsement of individual generics, like that of category generics (e.g., Cimpian et al., 2010), may be influenced by distinctiveness. Much of this evidence comes from the literature in social psychology concerning how people infer traits from behaviors. For example, Jones and Davis (1965) argued that people are more likely to draw broader conclusions from a specific behavior if this behavior departs from the typical behavior expected in that situation (i.e., if it is distinctive): If Carlos asks for a beer when visiting a winery, then the generic “Carlos drinks beer” seems particularly appropriate (e.g., Ajzen & Fishbein, 1975; Higgins & Bryant, 1982; Kelley, 1967, 1973; McArthur, 1972; Orvis et al., 1975; Reeder & Brewer, 1979; Skowronski & Carlston, 1987). However, in none of these studies were participants also asked to make judgments about category generics based on analogous evidence, suggesting again the need for a direct comparison.

Prior Evidence: Danger. The prior evidence on how endorsement of individual generics is influenced by danger connotations is more mixed. For instance, some studies—also from the literature on trait inferences—have suggested that people are particularly likely to draw broad conclusions from negatively valenced (including dangerous and immoral) behaviors (e.g., Aloise, 1993; Gidron et al., 1993; Rothbart & Park, 1986; see also Martijn et al., 1992; Skowronski & Carlston, 1987). Similarly, Carlson (1977) pointed out that less evidence seems needed to endorse the truth of a generic sentence such as “John beats small children” than of a sentence such as “John repairs cars.” Other studies, however, have found that generalizations about individuals based on negative behaviors are sometimes *less* likely, although we note that this result emerges almost exclusively in studies of children (e.g., Heyman & Giles, 2004;

Lockhart et al., 2002; Rholes & Ruble, 1984). For instance, Lockhart and colleagues (2002) found that 5- to 6-year-old children believed dangerous behaviors (e.g., hitting others) were less stable, less generalizable, and easier to change compared to positive ones. Further complicating the overall picture, none of the studies cited here assessed participants' judgments about category generics based on analogous evidence, so their relation to the question at hand is indirect at best.

The Present Research

We report six experiments that tested whether endorsement of category and individual generics is similarly affected by property content (specifically, distinctiveness and danger). Experiment 1, along with a preregistered replication (Experiment 1R; R stands for "replication"), and Experiment 3 directly compared the contribution of distinctiveness and danger, respectively, to participants' endorsement of category and individual generics. Because sensitivity to property content is a feature of the truth conditions of generic sentences but not of sentences expressing quantified generalizations, whose endorsement depends solely on the prevalence of the properties being described (e.g., Brandone et al., 2015; Cimpian et al., 2010; Krifka et al., 1995), in Experiments 2 and 4 we tested the prediction that distinctive and dangerous behaviors would *not* increase endorsement of quantified statements about categories (some/all *Xs* do *Y*) and individuals (*X* sometimes/always does *Y*). Finally, Experiment 5 investigated whether generics about dangerous behaviors are more readily endorsed simply because these behaviors are also more distinctive.

Experiment 1: Similar Effects of Distinctiveness on Category and Individual Generics

In Experiment 1, we investigated the influence of distinctiveness on endorsement of category and individual generics. Participants were presented with information about a novel target (a category or an individual, depending on the condition) and were then asked whether that

information licenses a generic statement. In the *category-generics* condition, participants were provided with information about the prevalence of a behavior among the members of an unfamiliar category (e.g., zorbs) and were then asked whether the corresponding category generic is true. Half of the stimuli described behaviors that were distinctive (i.e., uncommon among other categories), and half described behaviors that were non-distinctive. In the *individual-generics* condition, participants were provided with information about the prevalence of a behavior in the life of an unfamiliar individual (e.g., Zorb) and were then asked whether the corresponding individual generic is true. As before, half of the stimulus behaviors were said to be distinctive, whereas the other half were said to be non-distinctive.

We predicted that the generic statements about both categories and individuals would be endorsed more readily about behaviors said to be distinctive. Considering the previous results on category generics (e.g., Cimpian et al., 2010; Tasimi et al., 2017), we also expected the effects of distinctiveness to emerge only—or more strongly—at low levels of prevalence.

Method

Participants. Participants ($N = 194$; 69 men, 123 women; 1 participant did not report gender and another reported “male and female”) were recruited from two sources: a university undergraduate subject pool and Amazon’s Mechanical Turk service. Participants received course credit or \$0.75, respectively, for their participation. An additional 16 subjects were tested but excluded from the final sample either because their IP addresses were from outside the US ($n = 2$) or because they failed an attention check (see below; $n = 14$).

Across the five studies, we generally aimed to recruit between 150 and 300 participants. Although we did not determine these sample sizes based on a priori power analyses, they are considerably larger than those of older studies on this topic (e.g., Cimpian et al., 2010) and on

par with the sample sizes in more recent work (e.g., Tessler & Goodman, 2019). To examine the statistical power of these samples to detect the effect of interest, we conducted a series of sensitivity analyses (Faul et al., 2007). Sensitivity analyses answer the following question: “What is the minimum effect size to which your test is sufficiently sensitive?” Given that our critical question is whether endorsement of category and individual generics is similarly affected by property content, we calculated the minimum effect size for the interaction between type of generic (category vs. individual; between subjects) and property content (distinctive or dangerous vs. not; within subject) that our samples would allow us to detect in the context of an analysis of variance (ANOVA). Power calculations for mixed-design ANOVAs, which include both between- and within-subjects factors, are mathematically complex (e.g., Westfall et al., 2014) and are often only tractable via simulation. We used a recent simulation-based application created for this purpose (*Superpower*; Lakens & Caldwell, 2021) and found that samples of 150 to 300 participants afford 80% power to detect medium ($\eta_p^2 = .057$) to small-to-medium ($\eta_p^2 = .029$) interaction effects between type of generic and property content, respectively.² To foreshadow, we will also conduct a meta-analysis of our studies, which will further boost our power to detect small effects.

Instructions. The procedure for this and following experiments was approved by the [blinded for review] Institutional Review Board. The instructions were identical for all participants, regardless of condition: “In this study, we will tell you about some animals that live on a planet in a remote galaxy. For each question, you will be given some evidence and then you will be asked to judge if a certain conclusion follows from that evidence.”

Items and Measures. On each trial in the category-generics condition, participants were

² We adopt Cohen’s (1988)’s conventional thresholds for η_p^2 : .01 is a small effect and .06 is a medium effect.

told about a category (e.g., zorbs) and were then provided with evidence about the prevalence of a behavior among its members (e.g., 30% of zorbs climb trees; see Table 1). On each trial in the individual-generics condition, participants were told about an individual (e.g., Zorb) and were then provided with evidence about the prevalence of a behavior in this individual's life (e.g., Zorb has climbed trees in 30% of the situations where he had the opportunity; see Table 1). The qualifier "where he had the opportunity" was added to avoid ambiguity. Without this qualifier, the meaning of the percentage is ambiguous because it is unclear what is included in the denominator (e.g., are we counting situations in which Zorb is sleeping? situations in which there are no trees around?).

Each participant heard about four behaviors: climbing trees, digging holes in the ground, making nests out of grass, and washing oneself with water. We manipulated distinctiveness by including additional information on each trial that suggested the relevant behavior was shared either by very few other animals and was thus distinctive, or by many other animals and was thus non-distinctive (see Table 1). Each participant judged generics about two distinctive and two non-distinctive behaviors. Across participants, each behavior was presented approximately equally as distinctive and non-distinctive. This aspect of the design was intended to avoid confounding the distinctiveness manipulation with particular sets of properties.

Each of the four behaviors was presented at six levels of prevalence (1%, 5%, 10%, 30%, 50%, and 70%). The order of the 24 trials (4 behaviors [2 distinctive + 2 non-distinctive] \times 6 levels of prevalence), and thus of the different prevalence levels, was randomized individually for each participant. We oversampled low prevalence levels (rather than spacing the prevalence levels evenly between 0 and 100%) because, as mentioned above, we expected that the effects of distinctiveness would be greater at lower prevalence levels (Cimpian et al., 2010).

A different novel name was used on each trial (e.g., blins/Blin, ludinos/Ludino). Each name was assigned to a distinctive trial for half of the subjects and to a non-distinctive trial for the other half. This counterbalancing was designed to avoid confounding the distinctiveness manipulation with particular sets of names.

To assess whether the evidence on each trial licensed endorsement of the corresponding generic statements, we asked participants to judge whether the generic statement was true or false: for example, “zorbs climb trees” (category-generics condition) or “Zorb climbs trees” (individual-generics condition). It is noteworthy that this is the most stringent test of the effect of property distinctiveness to date. In Cimpian et al. (2010), participants judged the truth of different generic statements for distinctive properties (e.g., “Lorches have *distinctive* purple feathers”; emphasis ours) and non-distinctive properties (e.g., “Lorches have purple feathers”). In contrast, the test sentences in the present study were identical regardless of whether the property was distinctive or non-distinctive.

To summarize, the design of the study was as follows: 2 (type of generic: category vs. individual; between subjects) \times 2 (distinctiveness: distinctive vs. non-distinctive; within subject) \times 6 (level of prevalence: 1%, 5%, 10%, 30%, 50%, and 70%; within subject).

Attention Check, Demographics, and Debriefing. At the end of the task, we asked a question designed to check whether participants had paid attention: “Please think back to the task you just completed. Did the questions ask about single individuals or entire species?” Fourteen participants failed this attention check and were thus excluded from the sample. Participants then completed a demographics questionnaire and were debriefed about the goals of the study.

Analytic Strategy. Following Cimpian and colleagues’ (2010) analytic strategy, we submitted the proportion of “true” responses to a 2 (type of generic) \times 2 (distinctiveness) \times 6

(prevalence level) ANOVA. To explore the robustness of our results, we conducted three additional sets of analyses. First, we conducted ANOVAs on proportions that had been arcsine-transformed (Howell, 2013; but see Warton & Hui, 2011). These additional analyses, which were conducted for Experiments 2–4 as well, replicated the results reported below. Second, all follow-up comparisons between endorsement of generics about distinctive (or, in later studies, dangerous) and control properties at specific prevalence levels were also conducted with non-parametric Wilcoxon Signed-Ranks Tests, which revealed the same patterns of results as we report here. Third, we conducted mixed-effects logistic regressions with crossed random intercepts for participants and items on the true/false responses on each trial. The results of these mixed-effects models are reported in Supplementary Text A and Tables S1 and S2 in the Supplementary Online Materials (SOM).

Open Data. The data for this and all other studies can be found on the Open Science Framework (OSF): https://osf.io/vga63/?view_only=d83c43e4dadb419c9351c03d4845bae8.

Results

We predicted that category and individual generics will be endorsed more often for distinctive behaviors. In the context of the ANOVA, this predicted difference should translate into a significant main effect of distinctiveness or—more likely given prior research on category generics (e.g., Cimpian et al., 2010)—a significant two-way interaction between distinctiveness and prevalence, with greater endorsement of generics about distinctive (vs. non-distinctive) behaviors at lower prevalence levels but perhaps not at the higher ones, where we may see ceiling effects. The main effect of distinctiveness was not significant, $F(1, 192) = 2.10, p = .15, \eta_p^2 = .011$, but there was indeed a significant interaction between distinctiveness and prevalence, $F(5, 960) = 2.57, p = .025, \eta_p^2 = .013$. As illustrated in Figure 1, the effect of distinctiveness was

larger at (some of the) low prevalence levels (5% level: $p = .089$; 10% level: $p = .023$) than when the behaviors were more prevalent (30% level: $p = .18$; 50% level: $p = .45$; 70% level: $p = .62$). The effect of distinctiveness was absent at the very lowest prevalence level, 1% ($p = .82$), an unexpected result that we discuss further below.

Most relevant to our argument here, the facilitative effect of distinctiveness did not differ for generics about categories vs. individuals, consistent with the claim that the truth conditions of these sentences are similar in this respect: Neither the two-way interaction between type of generic (category vs. individual) and distinctiveness, $F(1, 192) = 0.23$, $p = .63$, $\eta_p^2 = .001$, nor the three-way interaction between these two factors and prevalence, $F(5, 960) = 0.32$, $p = .90$, $\eta_p^2 = .002$, was significant.

Rather than relying on these non-significant interaction effects as evidence for the claim of similar sensitivity to property distinctiveness for category and individual generics, we calculated a Bayes Factor (BF_{01}) to quantify the amount of support for the null hypothesis (H_0 : the effect of distinctiveness is the same for category and individual generics) over the alternative hypothesis (H_1 : the effect of distinctiveness is different for category and individual generics). BF_{01} was calculated as the ratio of two likelihoods: the likelihood of the data given H_0 and the likelihood of the data given H_1 . Values of BF_{01} greater than 3 would indicate that there is substantially more evidence for H_0 than H_1 (e.g., Dienes, 2014).³

To calculate BF_{01} , we first calculated a difference score for each participant by taking the difference between the proportion of “true” responses to distinctive items and the proportion of “true” responses to non-distinctive items. This difference score provides an overall index of the

³ We note that Bayes Factors provide information that is independent of the power of the sample to detect a null result (e.g., Dienes, 2014; Lakens et al., 2020). For example, Lakens et al. (2020) write, “Bayes factors can show that ... a high-powered nonsignificant result provides no evidence for H_0 relative to H_1 ” (p. 53).

facilitative effect of distinctiveness. We then conducted a Bayesian independent-samples t test that compared these difference scores for category vs. individual generics using JASP (JASP Team, 2020). BF_{01} was 5.74, indicating substantial support for H_0 over H_1 . We also calculated the BFs specifically for the 5% and 10% prevalence levels, where the distinctiveness effects were largest (see Figure 1). These BF_{01} values were 6.41 and 6.28, respectively, suggesting again support for H_0 over H_1 . Robustness checks indicated that these conclusions held across a wide range of prior assumptions about the magnitude of the difference in the distinctiveness effect for category vs. individual generics (see OSF).

Returning to the results of the ANOVA, we observed several other main effects and interactions. First, there was a significant main effect of prevalence, $F(5, 960) = 112.82, p < .001, \eta_p^2 = .370$, indicating that participants were more likely to endorse generics at higher prevalence levels. Second, individual generics were endorsed more often than category generics, $F(1, 192) = 19.36, p < .001, \eta_p^2 = .092$, and this difference was larger at lower prevalence levels, $F(5, 960) = 6.63, p < .001, \eta_p^2 = .033$ (compare top and bottom panels of Figure 1).

Discussion

The results of Experiment 1 suggest that category and individual generics are both endorsed more easily—at (some) low prevalence levels—when the properties being considered are distinctive. Moreover, the effect of distinctiveness was similar in magnitude for the two types of generics. Overall, these results are consistent with our core prediction that endorsement of category and individual generics is similarly affected by property content—in this case, a property's distinctiveness. However, there are a few caveats to consider as well.

Modest Effects of Distinctiveness. First, the effect of property distinctiveness was more subtle here than in prior research on category generics (e.g., Experiment 4 in Cimpian et al.,

2010). One reason for this difference may just be sampling variability. That is, perhaps the modest effect of distinctiveness observed here is simply a reflection of the variability that is to be expected when administering a similar task to different—and differently sized—subsets of a population. To further explore this possibility, we will recruit a new, ~50% larger sample from a different online platform and conduct a direct replication of this first study (see Experiment 1R below).

A second potential reason for the modest distinctiveness effects in Experiment 1 was foreshadowed above: The present test of the relation between property distinctiveness and generic endorsement was conservative because the test sentences were *identical* regardless of whether the property was distinctive or non-distinctive. To show an effect of distinctiveness, participants needed to notice and remember the distinctiveness-relevant information provided to them and then factor it spontaneously (i.e., without any prompting) into their truth judgments. In contrast, Cimpian and colleagues (2010) used the word “distinctive” in their test sentences (e.g., “Lorches have distinctive purple feathers”), which likely made it easier for participants to show sensitivity to property content. In subsequent experiments in this paper (namely, Experiments 3 and 5), we will embed the relevant information (namely, that the properties are dangerous) directly in the generic test sentences, thereby facilitating participants’ use of this information.⁴

A third possible reason for the modest effects of distinctiveness in the present study is the lack of granularity in our dependent measure: Participants were unable to indicate a *degree* of endorsement—they had to choose either “true” or “false.” In Experiment 5, we will elicit participants’ endorsement on a continuous 0–100 scale, which is better suited to capture the

⁴ Even if we observe larger effects in subsequent studies, we will, of course, not be able to pinpoint this methodological change as the sole reason. For example, it may also be the case that danger-related information is more potent than distinctiveness-related information in participants’ reasoning about generic statements.

nuances in participants' reasoning.

Null Effect of Distinctiveness at the Lowest Prevalence Level. Second, although we had predicted that the effects of distinctiveness would be present (and/or larger) at lower prevalence levels, the lowest level (1%) showed no such effect. Although unexpected from the standpoint of our argument, this pattern of results has been observed before in the literature. For example, Experiment 4 in Cimpian et al. (2010), which investigated the effects of distinctiveness and danger information on endorsement of category generics, showed a nearly identical pattern: The effects of property content were strongest at the second and third lowest prevalence levels and entirely absent at the lowest level (see their Figure 5 on p. 1470). However, this pattern of results is observed somewhat inconsistently in prior work. For instance, in Cimpian et al.'s (2010) Experiment 1, the lowest prevalence level showed the *strongest* effect of property content. The next experiment reported here, Experiment 1R (the direct replication of the present study), will provide further evidence on whether this finding is robust and worth interpreting.

Higher Overall Endorsement for Individual (vs. Category) Generics. Third, we observed that participants were more likely to endorse individual generics than category generics. This result suggests that the truth conditions of category and individual generics (at least as assessed here) do differ in some respects. To reiterate, however, the prediction we set out to test is a relatively narrow one: that of similarity between the truth conditions of these two classes of statements in a particular respect that is key to **GEN**'s semantics—namely, its sensitivity to property content. Thus, we will proceed with this more-specific focus and return to the overall endorsement difference between individual and category generics in the General Discussion.

Experiment 1R: A Direct, Preregistered Replication of Experiment 1

To examine the robustness of the effect of distinctiveness, we conducted a direct replication of Experiment 1 with a ~50% larger sample. We recruited this sample through a different online platform (Prolific rather than Mechanical Turk), in part because some have suggested an advantage for Prolific participants in terms of comprehension and attention (Peer et al., 2021; but see Litman et al., 2021). This replication was preregistered:

https://aspredicted.org/blind.php?x=/D9S_CSS (blinded).

Method

Participants. Participants ($N = 302$; 56 men, 235 women, 9 non-binary, 1 transgender [female to male], 1 did not report gender) were recruited from Prolific (www.prolific.co) and paid \$1.20 for their participation. An additional 42 subjects were tested but excluded from the final sample based on a preregistered criterion: namely, because they failed an attention check (same as in Study 1). We did not exclude any participants based on our second preregistered criterion (having an IP address from outside the US) because Prolific allows researchers to select the location of their participants, so we restricted access to this study to US participants only.

We note that we had preregistered a sample of 300 participants post-exclusions, not 302 participants. The discrepancy occurred because the anticipated exclusion rate we factored into the sample size posted on Prolific ended up being different from the actual exclusion rate.

Materials, Procedure, and Analytic Strategy. This study was a direct replication of Experiment 1. All materials and procedures were the same as in Experiment 1, as was our analytic strategy.

Results

Participants were more likely to endorse the generic statements about distinctive properties than the ones about non-distinctive properties, $F(1, 300) = 32.25, p < .001, \eta_p^2 = .097$.

This effect was considerably larger than in Experiment 1. The interaction between distinctiveness and prevalence was significant as well, $F(5, 1500) = 5.11, p < .001, \eta_p^2 = .017$. As illustrated in Figure 1R, the effect of distinctiveness was again most noticeable at the low prevalence levels (1% through 30% levels: $ps < .001$) and smaller or absent when the behaviors were more prevalent (50% level: $p = .060$; 70% level: $p = .74$).

As in Experiment 1, the effect of distinctiveness did not differ for generics about categories vs. individuals, as suggested by a non-significant two-way interaction between type of generic (category vs. individual) and distinctiveness, $F(1, 300) = 2.37, p = .13, \eta_p^2 = .008$, and a non-significant three-way interaction between these two factors and prevalence, $F(5, 1500) = 0.42, p = .84, \eta_p^2 = .001$.

To avoid relying on null interaction effects, we again computed a Bayes Factor to compare the likelihood of the data under two hypotheses: that the endorsement advantage for distinctive behaviors was the same (H_0) vs. different (H_1) for category and individual generics. The BF_{01} value for the overall difference between category and individual generics in the magnitude of the distinctiveness effect was 2.55, indicating some support for H_0 (see OSF for robustness checks). The BF_{01} values at the prevalence levels where we observed significant distinctiveness effects (1%, 5%, 10%, and 30%) were 7.49, 7.55, 7.24, and 1.76, respectively. Thus, while there was some variability in the values of BF_{01} , the preponderance of the evidence supported H_0 over H_1 .

The ANOVA revealed several other significant effects that mirrored those observed in Experiment 1. Participants were more likely to endorse generics at higher prevalence levels, $F(5, 1500) = 336.45, p < .001, \eta_p^2 = .529$. In addition, individual generics were endorsed more often than category generics, $F(1, 300) = 33.06, p < .001, \eta_p^2 = .099$, particularly at the lower

prevalence levels, $F(5, 1500) = 13.49, p < .001, \eta_p^2 = .043$ (see Figure 1R).

Discussion

The results of the present replication experiment support our main prediction that endorsement of category and individual generics is similarly boosted by the distinctiveness of the properties being described. In addition, the distinctiveness effect in Experiment 1R was noticeably larger than that in Experiment 1 and was present at the lowest prevalence level as well (see Figure 1R). Thus, two of the more surprising results of the preceding experiment (namely, the small size of the distinctiveness effect and its absence at the 1% level) were not observed in this direct, higher-powered replication. At this point, it is unclear what accounts for the differences between these two studies. While it is possible that Prolific participants are more attentive than Mechanical Turk and subject pool participants (Experiment 1), the exclusion rate due to missed attention checks was actually higher in the present study than in Experiment 1, so differences in attentiveness per se may not explain the difference in the magnitude of the effects. Regardless of the reasons for this difference, the key take-away is that both experiments supported the claim of similarity in the truth conditions of category and individual generics.

Experiment 2: No Effect of Distinctiveness on Quantified Statements

In Experiment 2, we tested the prediction that the truth conditions of quantified statements about categories (e.g., “Some/All zorbs climb trees”) and individuals (e.g., “Zorb sometimes/always climbs trees”) would not show sensitivity to property distinctiveness, which is theorized to be a unique feature of generalizations expressed via **GEN** (e.g., Leslie, 2008; Lerner & Leslie, 2016). Instead, endorsement of quantified statements about categories and individuals should be just a function of how prevalent the properties in question are.

Method

Participants. Participants ($N = 213$; 88 men, 118 women, 1 not reporting gender; demographic information for the remaining 6 subjects was not recorded due to a programming error) were recruited from two sources: a university undergraduate subject pool and Amazon’s Mechanical Turk service. Participants received course credit or \$0.75, respectively, for their participation. An additional 20 subjects were excluded because they failed an attention check (same as in Experiment 1).

Materials, Procedure, and Design. The method of this experiment was identical to that of Experiment 1, with one key exception: Participants were asked whether they endorse quantified statements about categories and individuals. Specifically, participants were asked to judge statements expressing either existential generalizations (e.g., “*Some zorbs climb trees,*” “*Zorb sometimes climbs trees*”) or universal generalizations (e.g., “*All zorbs climb trees,*” “*Zorb always climbs trees*”). Thus, the design of this study was as follows: 2 (quantification type: existential [some/sometimes] vs. universal [all/always]; between subjects) \times 2 (statement type: category [some/all] vs. individual [sometimes/always]; between subjects) \times 2 (distinctiveness: distinctive vs. non-distinctive; within subject) \times 6 (prevalence level: 1%, 5%, 10%, 30%, 50%, and 70%; within subject).

Results

The data revealed no evidence that endorsement of quantified statements was influenced by distinctiveness (see Figure 2). In fact, none of the interactions involving distinctiveness even approached significance, neither did its main effect, $F_s < 0.54$, $p_s > .46$, $\eta_p^2 < .005$. A Bayes Factor comparing the likelihood of the data given the hypothesis that the distinctiveness effect is absent (H_0) and the hypothesis that it is present (H_1) strongly favored the absence of a

distinctiveness effect, $BF_{01} = 9.66$ (see OSF for robustness checks).⁵

The only significant results involved prevalence and quantification type: Participants were more likely to endorse statements with existential quantifiers than universal quantifiers, $F(1, 209) = 1523.69$, $p < .001$, $\eta_p^2 = .879$, particularly at the lower prevalence levels, $F(5, 1045) = 12.60$, $p < .001$, $\eta_p^2 = .057$.⁶ Endorsement also increased with prevalence, $F(5, 1045) = 29.24$, $p < .001$, $\eta_p^2 = .123$. An unexpected result was that participants accepted significantly more existentially-quantified statements about categories (e.g., “Some zorbs climb trees”) than about individuals (e.g., “Zorb sometimes climbs trees”) at the lowest two prevalence levels (1% and 5%), $ps < .038$. Because similar differences were absent at higher prevalence levels, as well as for the universally-quantified statements, the ANOVA revealed a significant three-way interaction between quantification type, statement type, and prevalence, $F(5, 1045) = 2.84$, $p = .015$, $\eta_p^2 = .013$.⁷ For our purposes here, however, the important take-away is that distinctiveness was not involved in any of these effects.

Discussion

The results of Experiment 2 suggest that endorsement of quantified statements is not influenced by the distinctiveness of the behaviors being described. In tandem with the results of Experiment 1, this finding suggests that sensitivity to property distinctiveness is a unique feature of generic statements about categories and individuals.

⁵ Comparing Experiments 1 and 2, there were more participants who showed a distinctiveness effect (i.e., higher overall endorsement of statements about distinctive than non-distinctive behaviors) in Experiment 1 (generics) than in Experiment 2 (quantified statements), $\chi^2(1, N = 407) = 20.90$, $p < .001$. The same was true when we compared Experiments 1R and 2, $\chi^2(1, N = 515) = 55.12$, $p < .001$

⁶ We note that *always* and *sometimes* are not quantifiers per se but rather adverbs of quantification (or quantificational adverbs; Lewis, 1975). We use the term *quantifier* loosely here.

⁷ This three-way interaction subsumed two lower-order interactions: that between statement type (category vs. individual) and prevalence, $F(5, 1045) = 4.82$, $p < .001$, $\eta_p^2 = .023$, and that between statement type and quantification type, $F(1, 209) = 5.27$, $p = .023$, $\eta_p^2 = .025$.

Experiment 3: The Effect of Danger on Category and Individual Generics

In Experiment 3, we investigated the effect of danger information on endorsement of category and individual generics. We predicted that the endorsement of both types of generics will be higher for dangerous (vs. non-dangerous) behaviors, and to a similar extent.

Method

Participants. Participants ($N = 198$; 86 men, 112 women) were recruited from Amazon's Mechanical Turk service. They received \$0.75 for participation. An additional 36 participants were tested but excluded from the final sample because their IP addresses were from outside the United States ($n = 7$) or because they failed the attention check (same as in Experiment 1; $n = 29$).

Materials, Procedure, and Design. The method of this experiment was identical to that of Experiment 1, except for the items. Here, we used two items that had strong danger connotations (chopping people's heads off and ripping out people's guts), and two items that did not (climbing trees and digging holes in the ground). Thus, the design of this study was as follows: 2 (type of generic: category vs. individual; between subjects) \times 2 (danger: dangerous vs. non-dangerous; within subject) \times 6 (prevalence level: 1%, 5%, 10%, 30%, 50%, and 70%; within subject).

Results

We predicted that the category and individual generics about dangerous properties would be endorsed more often than the corresponding generics about non-dangerous properties, especially at the lower levels of prevalence. Consistent with this prediction, we found overall higher endorsement of the generics about dangerous behaviors, $F(1, 196) = 7.16, p = .008, \eta_p^2 = .035$, which was qualified by an interaction with prevalence, $F(5, 980) = 2.85, p = .015, \eta_p^2 =$

.014. As illustrated in Figure 3, the endorsement difference between generics about dangerous and non-dangerous properties was most apparent at lower prevalence levels (1% level: $p = .004$; 5% level: $p = .039$). In contrast, higher levels of prevalence licensed almost uniform endorsement of the generic statements, with no differences by whether the properties were dangerous or not.

As in Experiments 1 and 1R, there was no evidence that the effect of property content (in this experiment, dangerousness) differed for category vs. individual generics: Neither the two-way, type of generic (category vs. individual) \times danger interaction, $F(1, 196) = 0.04, p = .85, \eta_p^2 < .001$, nor the three-way, type of generic \times danger \times prevalence interaction, $F(5, 980) = 0.93, p = .46, \eta_p^2 = .005$, was significant.

BF_{01} was 6.36, indicating support for H_0 (i.e., that the endorsement advantage for dangerous behaviors was the same for category and individual generics; see OSF for additional robustness checks). Similar levels of support for H_0 were observed at the 1% and 5% prevalence levels, where danger information had the strongest effects on endorsement, $BF_{S01} = 4.56$ and 6.21, respectively.

Finally, the ANOVA revealed several other main effects and interactions. Endorsement rose with prevalence, $F(5, 980) = 84.66, p < .001, \eta_p^2 = .302$ (see Figure 3). We again saw overall higher endorsement for individual than category generics, $F(1, 196) = 39.24, p < .001, \eta_p^2 = .167$, particularly at low prevalence levels, $F(5, 980) = 11.12, p < .001, \eta_p^2 = .054$.

Discussion

The results of Experiment 3 suggest that category and individual generics are endorsed more readily when the properties being described are dangerous. Importantly, the facilitative effect of danger information was equivalent in magnitude for the two types of generics,

providing additional support for our main prediction that the truth conditions of these two types of sentences are similarly sensitive to property content.

It is also noteworthy that the effects of property content on generic endorsement were larger in the present study compared to Experiment 1, which was conducted with a similar participant sample (unlike Experiment 1R). A potential reason for this difference could be that the danger information was embedded directly in the test sentences in the present experiment. Another contrast with Experiment 1 is that the effects of property content in the present study were largest at the lowest prevalence level, similar to Experiment 1R. It is unclear what explains this inconsistency, both here and in prior work (e.g., compare Experiments 1 and 4 in Cimpian et al., 2010). We suspect, however, that the cause is some yet-to-be-identified aspect of the experimental setup or the participant samples rather than a meaningful feature of the phenomenon under investigation.

Experiment 4: No Effect of Danger on Quantified Statements

In Experiment 4, we investigated whether the danger information—conveyed with the same stimuli as in Experiment 3—increases endorsement of quantified statements. We predicted that it would not, just as the distinctiveness information did not affect endorsement of quantified statements in Experiment 2.

Method

Participants. Participants ($N = 221$; 67 men, 109 women; demographic information for the remaining 45 subjects was not recorded due to a programming error) were recruited from two sources: a university undergraduate subject pool and Amazon's Mechanical Turk service. Participants received partial course credit or \$0.75, respectively, for their participation. An additional 27 subjects were tested but excluded from the final sample either because their IP

addresses were from outside the US ($n = 6$) or because they failed the attention check at the end of the study (same as in Experiment 1; $n = 21$).

Materials, Procedure, and Design. The method of this study was identical to that of Experiment 2, which also used quantified statements, except that the items were taken from Experiment 3. Thus, this experiment had the following design: 2 (quantification type: existential [some/sometimes] vs. universal [all/always]; between subjects) \times 2 (statement type: category [some/all] vs. individual [sometimes/always]; between subjects) \times 2 (danger: dangerous vs. non-dangerous; within subject) \times 6 (level of evidence: 1%, 5%, 10%, 30%, 50%, and 70%; within subject).

Results

The ANOVA revealed no evidence that the danger information increased endorsement of quantified statements (see Figure 4). There was no overall difference between participants' endorsement of quantified statements about dangerous and non-dangerous behaviors, $F(1, 217) = 0.18, p = .67, \eta_p^2 = .001$, nor was there an interaction between danger and prevalence, $F(5, 1085) = 1.50, p = .19, \eta_p^2 = .007$. A Bayes Factor comparing the likelihood of the data given the hypothesis that the danger effect is absent (H_0) and the hypothesis that it is present (H_1) strongly favored the absence of a danger effect, $BF_{01} = 11.38$ (see OSF for robustness checks).⁸

The only significant effect involving danger was the four-way interaction between all factors in the ANOVA, $F(5, 1085) = 2.67, p = .021, \eta_p^2 = .012$. Importantly, however, an inspection of the relevant means and follow-up comparisons uncovered no evidence of an endorsement advantage for the dangerous items. The four-way interaction seemed to be driven

⁸ Comparing across Experiments 3 and 4, there were more participants who showed a danger effect (i.e., higher overall endorsement of statements about dangerous than non-dangerous behaviors) in Experiment 3 (generics) than in Experiment 4 (quantified statements), $\chi^2(1, N = 419) = 9.22, p = .002$.

by an unusual pattern of differences in participants' endorsement of the existentially-quantified statements about individuals (X sometimes does Y): At the 1% evidence level, participants endorsed the non-dangerous items more than the dangerous ones ($p = .005$), whereas at the 10% evidence level they did the opposite ($p = .003$). Analogous differences were not found for the existentially-quantified statements about categories (some X s do Y), or for either type of universally-quantified statement; these asymmetries in turn led to the significant four-way interaction. While this interaction is not easy to interpret, it also provides little support for the claim that danger information increases endorsement of quantified statements.

The other significant effects uncovered by the ANOVA were as follows: Existentially-quantified statements were endorsed more often than universally-quantified ones, $F(1, 217) = 1353.17, p < .001, \eta_p^2 = .862$, particularly at the lower prevalence levels, $F(5, 1085) = 12.84, p < .001, \eta_p^2 = .056$. Also, as expected, endorsement increased with prevalence, $F(5, 1085) = 17.69, p < .001, \eta_p^2 = .075$.

Discussion

The results of Experiment 4 indicate that danger information, provided in exactly the same way as in Experiment 3, does not increase endorsement of quantified statements. Together with the results of Experiment 3, this finding suggests that sensitivity to the dangerousness of the property being generalized is a unique feature of generic statements about categories and individuals.

Experiment 5: The Effect of Danger on Category and Individual Generics

While Holding Distinctiveness Constant

A skeptical perspective on Experiment 3 (generics, danger) might suggest that it is redundant with Experiments 1 and 1R (generics, distinctiveness): Perhaps participants assumed

that the dangerous behaviors (e.g., chopping heads) were also more distinctive than the non-dangerous behaviors (e.g., climbing trees; Tessler & Goodman, 2019). To test this alternative, in Experiment 5 we explicitly equated the distinctiveness of the dangerous and non-dangerous behaviors. We expected that the endorsement advantage for category and individual generics about dangerous behaviors would replicate even when distinctiveness is held constant.

A secondary goal of Experiment 5 was to sharpen our understanding of the effects of danger information. Is it danger per se that increases endorsement of generics? Or is the effect of danger explained by (1) the negative valence of the dangerous behaviors, or (2) the fact that they are relevant to people (since these behaviors targeted people, whereas the non-dangerous behaviors did not)? To explore these possibilities, we asked participants to rate the valence of each stimulus behavior, its relevance to people, as well as how dangerous it was. We then tested whether the relationship between participants' danger ratings and their generic endorsement remained significant when adjusting for their ratings of the behaviors' valence and relevance to people.

Method

Participants. Participants ($N = 137$; 50 men, 87 women) were recruited from Amazon's Mechanical Turk service. They received \$0.75 for participation. An additional 32 participants were tested but excluded from the final sample either because their IP addresses were from outside the United States ($n = 1$) or because they failed one of our attention checks or catch trials (see below; $n = 31$).

Materials, Procedure, and Design. The method of this experiment was identical to that of Experiment 3, with several exceptions. First, we equated the distinctiveness of all the items by explicitly stating that "on this planet, this behavior is rare" on every trial, regardless of whether

the behavior was dangerous or non-dangerous (see Table 5).⁹ Second, to avoid ambiguity, we explicitly stated that the dangerous behaviors are dangerous and that the non-dangerous behaviors are not (see Table 5).

Third, for a more precise estimate of endorsement, we asked participants to indicate whether they agree with each statement on a 0–100 scale (rather than requesting just a true/false response, as in previous studies). Responses on this scale have the additional advantage of being less prone to heteroskedasticity, and thus less likely to violate the assumptions underlying ANOVA, than proportions typically are (Howell, 2013).

Fourth, we added two catch trials after the 24 trials of the main task to identify participants who were not paying attention. On these trials, we provided no evidence for the relevant generics (i.e., the behavior was present in 0% of category members or situations). Participants who gave non-zero responses on *both* of these catch trials were excluded from the sample ($n = 16$).

Fifth, after completing the main task, participants were asked to recall the distinctiveness information provided for each item (“How frequent was this behavior said to be?”). Participants chose their answer from among three options: an option indicating the behavior was rare, an option indicating it was common, and an “I don’t know” option. For each participant, we computed the proportion of correct responses (“rare”) for the dangerous and non-dangerous items. We used these responses to check whether we successfully conveyed to participants that both the dangerous and non-dangerous behaviors were distinctive.

Finally, participants were presented with each item again and were asked to rate (1) its

⁹ We chose to portray all properties as distinctive rather than non-distinctive because we reasoned that participants might find it implausible if we said that the dangerous behaviors (e.g., chopping people’s heads off) were non-distinctive.

valence (on a sliding scale from -100 to $+100$; this question was preceded by instructions on what we mean by “valence”), (2) its relevance to people (“To what extent is [animal’s behavior] relevant to human beings?” and “If you encountered [animal], would his behavior matter to you?” on 1–9 scales), and (3) its dangerousness (“How dangerous is [animal’s behavior]?” and “How likely is it that [animal’s behavior] will kill you?” on 1–9 scales). Responses to the two questions that assessed each behavior’s relevance to people were averaged, as were responses to the two questions that assessed its dangerousness.

Results

The results replicated those of Experiment 3. First, we found greater endorsement for category and individual generics about dangerous than non-dangerous behaviors (see Figure 5), $F(1, 134) = 25.45, p < .001, \eta_p^2 = .160$. This difference was again larger at lower prevalence levels, $F(5, 670) = 8.68, p < .001, \eta_p^2 = .061$. Specifically, endorsement was higher for the dangerous than the non-dangerous items at the 1% through 50% prevalence levels ($ps \leq .002$) but not at the 70% level ($p = .59$; see Figure 5).

Importantly, these endorsement differences were observed despite the fact that participants correctly recalled both the dangerous and non-dangerous behaviors as being rare ($M_s = 70.1\%$ and 67.2% “rare” responses, respectively), $t(136) = 0.73, p = .47$. Moreover, whenever the behaviors were misremembered as being common, the rates of such errors did not differ between the dangerous and non-dangerous items ($M_s = 25.9\%$ and 24.5% , respectively), $t(136) = 0.38, p = .71$. These results suggest that we were successful in holding distinctiveness constant while manipulating the dangerousness of the stimulus behaviors.

Returning to the results of the ANOVA, the facilitative effect of danger information did not differ for category vs. individual generics: Neither the two-way interaction between type of

generic (category vs. individual) and danger, $F(1, 134) = 1.83, p = .18, \eta_p^2 = .013$, nor the three-way interaction between these two factors and prevalence, $F(5, 670) = 1.52, p = .18, \eta_p^2 = .011$, was significant. As in previous studies, we computed a Bayes Factor to compare the likelihood of the data under two hypotheses: that the endorsement advantage for dangerous behaviors was the same (H_0) vs. different (H_1) for category and individual generics. BF_{01} was 2.01, indicating some, albeit weak, support for H_0 over H_1 (see OSF for additional robustness checks). Stronger support for H_0 was observed at the 1%, 10%, and 50% prevalence levels ($BF_{s01} > 3.06$). At the 5% and 30% levels, the evidence was ambiguous ($BF_{s01} = 1.12$ and 1.52 , respectively).

The ANOVA revealed several other main effects and interactions. A main effect of prevalence indicated that, as before, people were more likely to endorse generics when the behaviors were more prevalent, $F(5, 670) = 321.52, p < .001, \eta_p^2 = .706$. Additionally, we again found that endorsement was higher for individual generics than category generics, $F(1, 134) = 12.32, p = .001, \eta_p^2 = .084$, especially at low prevalence levels, $F(5, 670) = 3.08, p = .009, \eta_p^2 = .022$.

Finally, we examined whether the effect of danger on generic endorsement was explained by the two other dimensions on which the dangerous and non-dangerous behaviors differed: valence and relevance to people. To test this explanation, we added each participant's valence and relevance ratings for each item as covariates in a mixed-effects linear regression predicting participants' endorsement. The model also included each participant's danger rating for each item (continuous), the item's prevalence level (1% to 70%; continuous), and the type of generic being endorsed (category vs. individual; categorical), as well as all interactions among these variables. Crossed random intercepts for participants and items were also included, allowing each participant's and item's intercept to vary randomly. In this mixed-effects model, computed

with the *mixed* command in Stata 16.1 (StataCorp, 2019), neither valence, $b = .01$, $SE = .01$, $p = 0.35$, nor relevance to people, $b = .04$, $SE = .28$, $p = 0.88$, emerged as significant predictors of endorsement. In contrast, danger did significantly predict endorsement, $b = .89$, $SE = .29$, $p = .003$, suggesting that its effect on endorsement does not boil down to an effect of valence and relevance to people.

Discussion

The results of Experiment 5 demonstrate that danger information influences participants' endorsement of category and individual generics even when distinctiveness is held constant. Moreover, the effect of danger information on endorsement of category and individual generics was again similar, providing additional support for our main prediction (namely, that property content affects endorsement of category and individual generics to a similar degree).

It is also notable that the effects of property content on generic endorsement in the present experiment were larger than in Experiments 1, 1R, and 3, the latter of which used the same test sentences. The larger effects are likely due to a combination of methodological improvements, such as the use of a continuous scale for assessing truth judgments, which is more sensitive to nuances in participants' reasoning, and the addition of an explicit statement that the dangerous behaviors were in fact dangerous, which may have facilitated participants' attention to and use of the danger information (relative to Experiment 3).

Finally, Experiment 5 revealed that the effect of danger information on participants' endorsement of generic statements is not parasitic on other semantic features, such as negative valence or relevance to humans. The extent to which a behavior was deemed dangerous strongly predicted participants' endorsement of the corresponding generic even when statistically adjusting for these other semantic features (i.e., negative valence and relevance to people). Of

course, these results do not address the deeper question of why endorsement of generics is sensitive to danger information in the first place, a question to which we return in the General Discussion.

Internal Meta-Analysis of Experiments 1, 3, and 5

Experiments 1, 1R, 3, and 5 all tested our core prediction that endorsement of category and individual generics is similarly affected by property content. For a more powerful test of this prediction, we meta-analyzed the data across these four experiments (e.g., Goh et al., 2016). If any differences between the effect of property content on endorsement of category vs. individual generics are present, a meta-analysis is well positioned to bring them to light. In addition, the practice of conducting internal meta-analyses of the studies within a paper is aligned with current best-practice recommendations for a cumulative psychological science (e.g., Cumming, 2014).

For purposes of this analysis, we first calculated a Cohen's d for the difference between category and individual generics in their sensitivity to property content.¹⁰ Positive values of this metric indicate that endorsement of individual generics is more sensitive to property content; negative values indicate that endorsement of category generics is more sensitive. If our prediction is correct, we should see values of d close to 0. Separate d s were computed for each of the four experiments. Within each experiment, we computed a d for the overall difference between category and individual generics in their sensitivity to property content (aggregating across prevalence levels), as well as separate d s for each of the six prevalence levels (1%, 5%, 10%, 30%, 50%, and 70%). These seven sets of values (one for the overall comparison + six for the comparisons at each prevalence level) were then submitted to seven Bayesian random-effects meta-analyses, computed in JASP (JASP Team, 2020). Each of these meta-analyses produced a

¹⁰ The standard errors of these Cohen's d s, which are needed for computing the meta-analytic estimates, were calculated via bootstrapping (1,000 resamples).

meta-analytic effect size, d_+ , and its 95% credible interval,¹¹ as well as a Bayes Factor (BF_{01}) that quantifies the amount of support for the null hypothesis (H_0 : the effect of property content is the same for category and individual generics) over the alternative hypothesis (H_1 : the effect of property content is different for category and individual generics).

These meta-analyses provided strong support for the prediction that endorsement of category and individual generics is similarly affected by property content (see below and OSF for detailed results): All BF_{01} were greater than 3, indicating substantially more evidence for H_0 than H_1 (e.g., Dienes, 2014), and all meta-analytic d_+ estimates were considerably below the conventional threshold for a small effect ($d = |0.20|$; Cohen, 1988). The overall estimate of the difference between category and individual generics in their sensitivity to property content was $d_+ = -0.03 [-0.24, 0.20]$, $BF_{01} = 8.56$. The estimates at specific prevalence levels were as follows:

- 1% prevalence level: $d_+ = 0.03 [-0.17, 0.23]$, $BF_{01} = 9.34$
- 5% prevalence level: $d_+ = 0.06 [-0.15, 0.27]$, $BF_{01} = 7.74$
- 10% prevalence level: $d_+ = 0.02 [-0.18, 0.21]$, $BF_{01} = 9.90$
- 30% prevalence level: $d_+ = -0.10 [-0.33, 0.16]$, $BF_{01} = 4.73$
- 50% prevalence level: $d_+ = -0.02 [-0.26, 0.18]$, $BF_{01} = 9.46$
- 70% prevalence level: $d_+ = -0.13 [-0.32, 0.12]$, $BF_{01} = 3.04$

General Discussion

Sentences expressing broad, unquantified generalizations about categories (e.g., “Dogs bark”) and individuals (e.g., “Daisy barks”) are common in daily conversation and a primary

¹¹ A sensitivity analysis revealed that (a frequentist version of) this meta-analysis would have 80% power to detect a minimum between-subjects difference of $d = 0.19$, which is a small effect by Cohen’s (1988) standards.

means of expressing and acquiring general beliefs about the world (e.g., Gelman et al., 1998; Gelman et al., 2008; Gelman et al., 2004). Here, we provided the first direct experimental comparison of the circumstances under which people endorse these two classes of so-called generic statements. Experiments 1, 1R, 3 and 5 suggested that endorsement of category and individual generics was similarly facilitated by distinctive and dangerous behaviors. Of note, the stimuli used to elicit endorsement for these two types of generics were—in all relevant respects—identical (see Tables 1, 3, and 5); this feature of the design enables us to draw strong conclusions from these comparisons. Moreover, Experiments 2 and 4 established that the distinctive and dangerous stimuli did not facilitate endorsement of existentially- and universally-quantified statements about categories and individuals, consistent with theoretical arguments that sensitivity to property content is a distinguishing property of the truth conditions of generic sentences (e.g., Lerner & Leslie, 2016; Leslie, 2008).

Contributions to Theory on Generic Language

These studies contribute to theory on the semantics of generic statements in two critical respects. First, they provide the first systematic empirical evidence for the claim that generics about categories and generics about individuals represent a unified phenomenon. Although theoretical work in formal semantics had proposed that the meaning of these sentences can be analyzed in the same way—with the flexible and context-sensitive generic operator **GEN** (e.g., Krifka et al., 1995; Leslie & Lerner, 2016)—to date the empirical evidence for this claim was restricted to armchair intuitions. However, insofar as generics are an aspect of everyday language, any theory that seeks to describe their truth conditions must also be borne out by the judgments of language users who have not had training in formal semantics. Thus, the present evidence that laypeople’s endorsement of category and individual generics is similarly sensitive

to property content is an important step forward for this literature.

Second, the present studies contribute to ongoing debates about the semantics of the **GEN** operator. One of the key points of disagreement on this topic is whether **GEN** can be adequately described without taking into account the content of the properties being generalized (for a summary of the opposing views, see Lerner & Leslie, 2016; Leslie & Lerner, 2016). So far, the debate between content-neutral (e.g., Asher & Morreau, 1995; Cohen, 2004) and content-based (e.g., Leslie, 2007, 2008) perspectives on the semantics of **GEN** has largely played out in the context of generics about categories; we know of no arguments that have compared these perspectives in the context of generics about individuals, even though the **GEN** formalism is assumed to describe their meaning as well. From this point of view, the present research marks a new frontier in this debate. Our finding that the truth conditions of generics about individuals are also sensitive to property content, similar to the truth conditions of generics about categories, provides additional support for the content-based hypothesis about the semantics of **GEN**.

Implications for Theory on the Cognitive Mechanisms Underlying Generalization

The present findings indicate that generics about categories and generics about individuals give voice to structurally analogous sorts of generalizations, suggesting perhaps that the semantics of **GEN** are supplied, or at least constrained, by the same cognitive mechanism. Articulating an account of this shared mechanism is beyond the scope of this article, but we offer some speculative suggestions below.

This putative mechanism that computes both generic generalizations about categories (across specific members) and generic generalizations about individuals (across specific events, episodes, behaviors, etc.) is likely to possess at least four key characteristics. Two of these characteristics map directly onto the present experiments—namely, the sensitivity to

distinctiveness and danger—so we will not dwell on them here, except to say that there is a reasonable argument to be made for each of these characteristics having some adaptive value. For instance, the generalization mechanism’s sensitivity to distinctiveness may be explained by the high informational value of distinctive features and behaviors. A distinctive feature (e.g., being red) differentiates a category (e.g., cardinals) from others and is thus particularly informative; the same can be said about an individual’s distinctive behavior. It seems adaptive (and rational) for a generalization mechanism to privilege features and behaviors with high informational value. Similarly, threat-relevant information of the sort we investigated here has obvious consequences for survival and is privileged in other cognitive–perceptual systems as well (for reviews, see Baumeister et al., 2001; Vaish et al., 2008). Even young children show faster orienting responses and enhanced memory for threatening stimuli (e.g., Baltazar et al., 2012; Hamlin et al., 2010; Kinzler & Shutts, 2008; LoBue & DeLoache, 2010). The other two putative characteristics of the cognitive mechanism that computes generic generalizations go beyond the present studies and suggest fruitful avenues for future research.

First, generic generalizations about categories and individuals may be intimately linked with people’s *explanations* for the evidence at hand. A fundamental aspect of human psychology is the drive to make sense of the world via explanations (e.g., Cimpian & Salomon, 2014a, 2014b; Gopnik, 1998; Horne et al., 2019; Keil, 2006; Lombrozo, 2006; Murphy & Medin, 1985; Ross, 1977; Weiner, 1985). The explanation generated for a certain observation imposes strong constraints on the breadth of the inferences that can be drawn from it. For example, if a feature of an unfamiliar animal (e.g., patches of hairless skin) is explained as an accident (e.g., the animal was in a fight), then it is unlikely that this feature will be inferred to characterize that category of animals (e.g., Cimpian et al., 2010; Cimpian & Markman, 2008; Gelman, 1988). Similarly, if an

instance of behavior (e.g., a nervous laughter) is explained in terms of the particular circumstances in which the behavior occurred (e.g., a first date), then broader inferences about personal dispositions are unlikely.

The key role of explanation in formulating generic generalizations has been independently recognized in several relevant literatures. For instance, a currently dominant view of categorization has been termed the “theory-based view” to highlight the crucial role it assigns to causal-explanatory understandings in people’s reasoning about categories (e.g., Ahn et al., 2000; Carey, 1985; Gelman, 2003; Murphy & Medin, 1985; Rehder, 2003). Similarly, social psychologists have conceptualized the process of generalizing from individuals’ behaviors to their (generic) traits as an instance of causal attribution (which is a form of explanation) since their first systematic explorations of this subject matter (e.g., Heider, 1958; Jones & Davis, 1965). In light of these considerations, it seems plausible to claim that the mechanism by which generic generalizations about both categories and individuals are formulated is influenced by the explanations people generate to make sense of the world.¹²

Second, borrowing an idea from Leslie (2007, 2008) and Gelman (2010; Hollander et al., 2002), we also speculate that the cognitive mechanism underlying generic generalizations—and thus generic language—is the *default* means available to our species for drawing general conclusions from limited evidence. By *default*, we mean two things here. First, this mechanism may be developmentally primitive—the first-emerging and most basic means of going beyond the evidence to derive expectations about new exemplars or situations (Cimpian, 2016).

Consistent with this possibility, (seemingly generic) generalizations about both categories and individuals seem to be available to infants even by their first birthdays (e.g., Baldwin et al.,

¹² This may also be the reason why both types of generic generalizations (i.e., about categories and individuals) are shaped by cultural differences in causal-explanatory frameworks (e.g., Atran & Medin, 2010; Morris & Peng, 1994).

1993; Dewar & Xu, 2009; Graham et al., 2004; Hamlin et al., 2007; Keates & Graham, 2008; Kuhlmeier et al., 2003; Luo & Baillargeon, 2005; Song et al., 2005; Woodward, 1998, 1999).¹³

There is also evidence that the ability to draw these generic generalizations *precedes* the ability to draw quantified (e.g., existential, universal) generalizations, which further strengthens claims of developmental primacy (Hollander et al., 2002; Mannheim et al., 2011; Tardif et al., 2012). It should be noted, however, that the latter studies pertain exclusively to generalizations over category members (not situations). In future work, it will be important to test whether this conclusion also holds with respect to generalizations over situations.

The mechanism for drawing generic generalizations may be default in a second sense as well: It may be the mechanism that people call upon spontaneously and without conscious awareness in order to extract general conclusions from particular samples of evidence. The idea that generic generalizations about individuals are computed ubiquitously, spontaneously, and outside of conscious awareness is commonly accepted in social psychology, even if the terminology used in this literature is different (e.g., “trait inference,” “person perception”; for a review, see Uleman et al., 2008). Recent evidence suggests this conclusion extends to generic generalizations about categories as well. Sutherland, Cimpian, Leslie, and Gelman (2015) found that participants drew generic generalizations about categories in a task in which such generalizations detracted from correct performance (which speaks to their ubiquity and spontaneity), and even when participants were under a cognitive load. In fact, generic generalizations about categories and individuals may be default in this second sense of

¹³ Several studies seem to suggest that, contrary to this claim, children cannot formulate inferences about the (generic) traits of individuals until they are 6 or 7 years old (e.g., Boseovski & Lee, 2006; Rholes & Ruble, 1984). Our view is that these studies underestimate children’s ability to reason about traits, in large part because they use tasks with high information-processing demands. Tasks that avoid some of these demands—and that are similar in this respect to the tasks investigating category generalizations in infancy (e.g., Graham et al., 2004)—generally reveal precocious dispositional inferences (for a review, see Cimpian, 2017).

“ubiquitously running in the background” in part because they require few cognitive resources to be computed (e.g., Hampton, 2012; Leslie & Gelman, 2012; Leslie et al., 2011; Todorov & Uleman, 2003).

Why Were Individual Generics Endorsed More Easily Than Category Generics?

Although the truth conditions of category and individual generics were found to be similar in their sensitivity to property content, they differed in other respects. Specifically, individual generics were endorsed more often than category generics, especially at low prevalence levels. Our studies leave open the question of why this result was observed, but we offer some speculative suggestions here.

To begin, we note that the lower endorsement threshold for individual generics than for category generics does not in and of itself provide strong reason to doubt that these sentences are both instances of genericity or that they are the output of the same generalization mechanism. Consider, for example, that different types of *category* generics also differ in their overall likelihood of being endorsed: Category generics with indefinite singular noun phrases (e.g., “A fluorescent light bulb is energy-efficient”) are generally understood to express stronger generalizations than category generics with bare plural noun phrases (e.g., “Fluorescent light bulbs are energy-efficient”)—the former are more definitional and law-like, and as a result also differ from bare plural generics in their truth conditions (e.g., Krifka et al., 1995).

Even so, what explains the finding that individual generics were more easily endorsed than category generics in our studies? One possibility is that this result was an artifact of our methodology. For instance, perhaps a low-level cue such as the addition of “where he had the opportunity” to the prevalence information for individual generics made them seem more acceptable. A second, perhaps more interesting, possibility is suggested by prior evidence that

two arbitrary time-slices of an individual are usually assumed to be more similar to each other than are two arbitrary members of a category (Lawson & Kalish, 2006). As a result, people are more likely to generalize a property that is true of an individual in a particular situation to other situations in that individual's life than they are to generalize a property that is true of a particular member of a category to other members of that category. In other words, people seem to operate with a baseline assumption of homogeneity across the time-slices in an individual's life—an assumption that likely makes it easier to endorse individual generics in our task regardless of the prevalence information provided. Notably, this is not a blanket assumption. As discussed above, the putative mechanism that computes generic generalizations about categories and individuals is guided by the explanations generated for the relevant property or behavior (e.g., Heider, 1958; Jones & Davis, 1965; Murphy & Medin, 1985); these explanations inform the expected likelihood that the property or behavior will recur in other members of the same category or other time-slices of the same individual. The connection to the broader causal-explanatory frameworks with which people make sense of the world ensures that the output of the shared generalization mechanism is appropriately sensitive to the specifics of the properties being considered and the circumstances under which they are observed. Thus, a more precise version of claim here is that, holding the property and the explanations generated for it constant, people are on average more likely to assume that other time-slices of an individual will display it than that other members of a category will.

A third possibility, related to the second, appeals to a difference in the coherence or entitativity of individuals vs. categories as targets of inference. This claim has two premises: First, it is easier to draw generic generalizations about targets that are more coherent or entitative (e.g., birds vs. animals in North America; e.g., Gelman, 2003; Markman, 1989). Second,

individuals are more easily perceived as coherent units than categories are (e.g., Hamilton & Sherman, 1996; Kashima et al., 2005; Susskind et al., 1999). With respect to the second premise, Hamilton and Sherman (1996) reviewed a body of evidence that led them to conclude that “perceivers expect less entitativity—less unity, consistency, organization, and coherence—in group targets than they do in individual targets” (p. 351). Although their conclusion pertains specifically to social targets (i.e., human individuals and groups), it can be extrapolated to a more general contrast between individuals and categories as targets of generalization: While categories vary in entitativity, as do individuals, on average individuals are more likely to be perceived as coherent units, which may in turn make it easier to formulate generic generalizations (and endorse generic statements) about them.

Limitations and Future Directions

Related to the preceding section, a fruitful goal for future research would be to provide a more comprehensive account of the ways in which the semantics of category and individual generics are similar vs. different. For example, it would be informative to assess whether their endorsement is affected to similar degrees by manipulations of entitativity, of property content beyond distinctiveness and danger, and of the domain to which the categories and individuals belong (e.g., non-human animals, humans, artifacts). With respect to the last point, Tasimi et al. (2017) found that the facilitative effect of danger information on endorsement of category generics is eliminated when the generics describe *human* categories, perhaps in part because of an assumption that, deep down, human beings are good rather than threatening or dangerous (e.g., Newman et al., 2014, 2015). It is an open question whether the effect of danger information is reduced or eliminated for generics about human *individuals* as well (for some reasons to expect it might not be, see Aloise, 1993; Gidron et al., 1993; Rothbart & Park, 1986).

It would also be worthwhile to investigate whether the conceptual consequences of exposure to generics about individuals parallel those of exposure to generics about categories. For instance, hearing a series of generic statements about an unfamiliar social or animal category prompts people to “essentialize” it—to see it as a coherent, informative grouping that embodies a rich cluster of nonobvious properties emerging from an inherent causal source (e.g., Gelman et al., 2010; Leshin et al., 2021; Rhodes et al., 2012). By analogy, it is possible that hearing a series of generics about an individual would likewise serve to essentialize this individual—to create the impression that the individual embodies a range of other yet-to-be-discovered characteristics that emerge from an inherent causal source. Although individuals are generally perceived as coherent units even in the absence of any additional information (e.g., Hamilton & Sherman, 1996), generic language could still convey a sense of inferential depth and inherent causal sources that are not present in default expectations about individuals. This line of argument leads to an interesting prediction with respect to self-concepts as well: When repeatedly endorsed or produced about oneself, individual generics may lead people to assume that the source of their behaviors and traits resides in an underlying, essence-like “true self” (Christy et al., 2019).

Finally, it will also be important to explore the claim of similar truth conditions for category and individual generics from a developmental perspective. If these statements express the output of a single, developmentally primitive generalization mechanism, as we speculated above, we should observe (similar degrees of) sensitivity to property content even in young children’s endorsement and production of category and individual generics. Many other interesting predictions follow as well, including that the developmental primacy of generic relative to quantified statements about categories (e.g., Leslie & Gelman, 2012) should also be observed for generic (vs. quantified) statements about individuals.

Conclusion

The present findings suggest that the truth conditions of generics about categories and individuals are analogous with respect to their sensitivity to property content: Endorsement of both classes of statements is facilitated when the properties being generalized are distinctive or dangerous. This work contributes to theory on a topic—genericity—that has drawn sustained interest from linguists, philosophers, and cognitive and developmental psychologists over the last several decades. In addition, because generics are a primary vehicle for conveying general beliefs about the social world, this work can inform our understanding of how language shapes interpersonal and intergroup attitudes. We hope that this work will inspire others to investigate category and individual generics jointly, in closely matched experimental designs, further refining theory on how these aspects of language relate to the underlying cognitive mechanisms that give them meaning.

References

- Ahn, W.-k., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, *41*, 361-416.
- Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, *82*(2), 261-277.
- Aloise, P. A. (1993). Trait confirmation and disconfirmation: The development of attribution biases. *Journal of Experimental Child Psychology*, *55*(2), 177-193.
- Asher, N., & Morreau, M. (1995). What some generic sentences mean. In G. N. Carlson, & F. J. Pelletier (Eds.), *The Generic Book* (pp. 300-338). Chicago, IL: The University of Chicago Press.
- Atran, S. & Medin, D. L. (2010). *The native mind and the cultural construction of nature*. Cambridge, MA: MIT Press.
- Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, *64*(3), 711-728.
- Baltazar, N. C., Shutts, K., & Kinzler, K. D. (2012). Children show heightened memory for threatening social actions. *Journal of Experimental Social Psychology*, *112*, 102-110.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323-370.
- Behrens, L. (2005). Genericity from a cross-linguistic perspective. *Linguistics*, *43*(2), 275-344.
- Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, *42*(3), 500-513.
- Brandone, A. C., Gelman, S. A., & Hedglen, J. (2015). Young children's intuitions about the

- truth conditions and implications of novel generics. *Cognitive Science*, 39, 711-738.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Carlson, G. N. (1977). *Reference to kinds in English*. Doctoral dissertation, University of Massachusetts, Amherst.
- Christy, A. G., Schlegel, R. J., & Cimpian, A. (2019). Why do people believe in a “true self”? The role of essentialist reasoning about personal identity and the self. *Journal of Personality and Social Psychology*, 117(2), 386-416.
- Cimpian, A. (2016). The privileged status of category representations in early development. *Child Development Perspectives*, 10(2), 99-104.
- Cimpian, A. (2017). Early reasoning about competence is not irrationally optimistic, nor does it stem from inadequate cognitive representations. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of Competence and Motivation (2nd Edition): Theory and Application* (pp. 387-407). New York: Guilford Press.
- Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, 34(8), 1452-1482.
- Cimpian, A., Gelman, S. A., & Brandone, A. C. (2010). Theory-based considerations influence the interpretation of generic sentences. *Language and Cognitive Processes*, 25(2), 261-276.
- Cimpian, A., & Markman, E. M. (2008). Preschool children’s use of cues to generic meaning. *Cognition*, 107(1), 19-53.
- Cimpian, A., & Salomon, E. (2014a). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, 37(5), 461-480.

- Cimpian, A., & Salomon, E. (2014b). Refining and expanding the proposal of an inference heuristic in human understanding. *Behavioral and Brain Sciences*, 37(5), 506-527.
- Cohen, A. (2004). Generics and mental representations. *Linguistics and Philosophy*, 27, 529-556.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29.
- Dahl, Ö. (1975). On generics. In E. L. Keenan (Ed.), *Formal semantics of natural language: Papers from a colloquium sponsored by the King's College Research Centre, Cambridge* (pp. 99-111). Cambridge, England: Cambridge University Press.
- Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, 20(2), 252-257.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, 20, 65-95.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.
- Gelman, S. A. (2010). Generics as a window onto young children's concepts. In F. J. Pelletier (Ed.), *Kinds, things, and stuff: Mass terms and generics* (pp. 100-120). New York: Oxford University Press.

- Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A., & Keil, F. C. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, *i*-157.
- Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development*, *4*(1), 1-31.
- Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs: I. Introduction. *Monographs of the Society for Research in Child Development*, *69*(1), 1-14.
- Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*(3), 273-301.
- Gidron, D., Koehler, D. J., & Tversky, A. (1993). Implicit quantification of personality traits. *Personality and Social Psychology Bulletin*, *19*, 594-604.
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta- analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*(10), 535-549.
- Gopnik, A. (1998) Explanation as orgasm. *Minds and Machines* *8*, 101-118.
- Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*(2), 409-427.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336-355.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923-929.
- Hampton, J. A. (2012). Thinking intuitively: The rich (and at times illogical) world of concepts.

Current Directions in Psychological Science, 21(6), 398-402.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons.

Heyman, G. D., & Giles, J. W. (2004). Valence effects in reasoning about evaluative traits.

Merrill-Palmer Quarterly, 50(1), 86-109.

Higgins, E. T., & Bryant, S. L. (1982). Consensus information and fundamental attribution error:

The role of development and in-group versus out-group knowledge. *Journal of Personality and Social Psychology*, 43(5), 889-900.

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

Horne, Z., Muradoglu, M., & Cimpian, A. (2019). Explanation as a cognitive process. *Trends in Cognitive Sciences*, 23(3), 187-199.

Howell, D. C. (2013). *Statistical methods for psychology (8th ed.)*. Belmont, CA: Wadsworth.

JASP Team (2020). JASP (Version 0.14.1) [Computer software].

Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219-266). New York: Academic Press.

Kashima, Y., Kashima, E., Chiu, C. Y., Farsides, T., Gelfand, M., Hong, Y. Y., ... & Yzerbyt, V. (2005). Culture, essentialism, and agency: Are individuals universally believed to be more real entities than groups? *European Journal of Social Psychology*, 35(2), 147-169.

Keates, J., & Graham, S. A. (2008). Category markers or attributes: Why do labels guide infants' inductive inferences? *Psychological Science*, 19, 1287-1293.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, 57, 227-254.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska*

- Symposium on Motivation* (Vol. 15, pp. 192-240). Lincoln, NE: University of Nebraska Press.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28(2), 107-128.
- Kinzler, K. D., & Shutts, K. (2008). Memory for “mean” over “nice”: The influence of threat on children’s face memory. *Cognition*, 107, 775-783.
- Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., & Chierchia, G. (1995). Genericity: An introduction. In G. N. Carlson, & F. J. Pelletier (Eds.), *The generic book* (pp. 1-124). Chicago, IL: The University of Chicago Press.
- Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, 14(5), 402-408.
- Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 1-14.
- Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology: Series B*, 75(1), 45-57.
- Lawson, C. A., & Kalish, C. W. (2006). Inductive inferences across time and identity: Are category members more alike than single individuals? *Journal of Cognition and Development*, 7(2), 233-252.
- Lerner, A., & Leslie, S. J. (2016). Generics and experimental philosophy. In J. Sytsma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 404-416). Hoboken, NJ: Wiley.
- Leshin, R. A., Leslie, S. J., & Rhodes, M. (2020). Does it matter how we speak about social

- kinds? A large, preregistered, online experimental study of how language shapes the development of essentialist beliefs. *Child Development*, 92(4), e531-e547.
- Leslie, S. J. (2007). Generics and the structure of the mind. *Philosophical Perspectives*, 21(1), 375-403.
- Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, 117(1), 1-47.
- Leslie, S. J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, 64(3), 186-214.
- Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, 65(1), 15-31.
- Leslie, S. J., & Lerner, A. (2016). Generic generalizations. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Available at <https://plato.stanford.edu/archives/win2016/entries/generics/>
- Lewis, D. (1975). Adverbs of quantification. In E. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 178-188). Cambridge, UK: Cambridge University Press.
- Liebesman, D. (2011). Simple generics. *Noûs*, 45(3), 409-442.
- Litman, L., Moss, A., Rosenzweig, C., & Robinson, J. (2021). Reply to “MTurk, Prolific or panels? Choosing the right audience for online research.” Available at <http://dx.doi.org/10.2139/ssrn.3775075>
- LoBue, V., & DeLoache, J. S. (2010). Superior detection of threat-relevant stimuli in infancy. *Developmental Science*, 13(1), 221-228.
- Lockhart, K. L., Chang, B., & Story, T. (2002). Young children’s beliefs about the stability of traits: Protective optimism? *Child Development*, 73(5), 1408-1430.
- Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*,

10(10), 464-470.

- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16*(8), 601-608.
- Mannheim, B., Gelman, S. A., Escalante, C., Huayhua, M., & Puma, R. (2011). A developmental analysis of generic nouns in Southern Peruvian Quechua. *Language Learning and Development, 7*(1), 1-23.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
- Martijn, C., Spears, R., Van der Pligt, J., & Jakobs, E. (1992). Negativity and positivity effects in person perception and inference: Ability versus morality. *European Journal of Social Psychology, 22*(5), 453-463.
- McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology, 22*(2), 171-193.
- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*(6), 949-971.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289-316.
- Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin, 40*(2), 203-216.
- Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science, 39*(1), 96-125.
- Orvis, B. R., Cunningham, J. D., & Kelley, H. H. (1975). A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of*

Personality and Social Psychology, 32(4), 605-616.

- Papafragou, A. (1996). On generics. *UCL Working Papers in Linguistics*, 8, 165-198.
- Peer, E., Rothschild, D. M., Evernden, Z., Gordon, A., & Damer, E. (2021). *Data quality of platforms and panels for online behavioral research*. Available at <http://dx.doi.org/10.2139/ssrn.3765448>
- Prasada, S., Khemlani, S., Leslie, S. J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition*, 126(3), 405-422.
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, 86(1), 61-79.
- Rehder, B. (2003). Causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1141-1159.
- Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, 109(34), 13526-13531.
- Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, 55(2), 550-560.
- Ross, L. D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173-220). New York: Academic Press.
- Rothbart, M., & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology*, 50(1), 131-142.
- Schubert, L. K., & Pelletier, F. J. (1989). Generically speaking, or, using discourse representation theory to interpret generics. In G. Chierchia, B. H. Partee, & R. Turner (Eds.), *Properties, types and meaning* (pp. 193-268). Dordrecht: Kluwer.

- Song, H., Baillargeon, R., & Fisher, C. (2005). Can infants attribute to an agent a disposition to perform a particular action? *Cognition*, *98*, B45-B55.
- StataCorp. (2019b). *Stata Statistical Software: Release 16.1*. College Station, TX: StataCorp LLC.
- Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*(4), 689-699.
- Sutherland, S. L., Cimpian, A., Leslie, S. J., & Gelman, S. A. (2015). Memory errors reveal a bias to spontaneously generalize to categories. *Cognitive Science*, *39*(5), 1021-1046.
- Susskind, J., Maurer, K., Thakkar, V., Hamilton, D. L., & Sherman, J. W. (1999). Perceiving individuals and groups: Expectancies, dispositional inferences, and causal attributions. *Journal of Personality and Social Psychology*, *76*(2), 181-191.
- Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2012). Acquisition of generic noun phrases in Chinese: Learning about lions without an ‘-s’. *Journal of Child Language*, *39*(1), 130-161.
- Tasimi, A., Gelman, S. A., Cimpian, A., & Knobe, J. (2017). Differences in the evaluation of generic statements about human and non- human categories. *Cognitive Science*, *41*(7), 1934-1957.
- Tessler, M. H., & Goodman, N. D. (2016). Communicating generalizations about events. In A., Papafragou, D., Grodner, D., Mirman, & J. C., Trueswell. (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review*, *126*(3), 395-436.
- Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to

- actors' faces. *Journal of Experimental Social Psychology*, 39(6), 549-562.
- Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59, 329-360.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, 134(3), 383-403.
- Warton, D. I., & Hui, F. K. C. (2011). The arcsine is asinine: The analysis of proportions in ecology. *Ecology*, 92, 3-10.
- Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, 97(1), 74-84.
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020-2045.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1-34.
- Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, 22(2), 145-160.
- Zhu, J., & Murphy, G. L. (2013). Influence of emotionally charged information on category-based induction. *PLOS ONE*, 8(1), e54286.

Table 1. *Sample item from Experiments 1 and 1R*

Property Content	Condition	
	Category	Individual
Distinctive	Background information: One animal is called morseths. They live up to 50 years. ^a	Background information: One animal is called Morseth. He is 50 years old. ^a
	Evidence: (1) $x\%$ of morseths climb trees. (2) On this planet, very few other animals perform this behavior. This behavior is extraordinary.	Evidence: (1) Morseth has climbed trees in $x\%$ of the situations where he had the opportunity. (2) On this planet, very few other animals have ever performed this behavior when they had the opportunity. This behavior is extraordinary.
	Conclusion: Morseths climb trees. Is this conclusion true or false?	Conclusion: Morseth climbs trees. Is this conclusion true or false?
Non-distinctive	Background information: One animal is called zorbs. They live up to 50 years. ^a	Background information: One animal is called Zorb. He is 50 years old. ^a
	Evidence: (1) $x\%$ of zorbs wash themselves with water. (2) On this planet, many other animals also perform this behavior. This behavior is unremarkable.	Evidence: (1) Zorb has washed himself with water in $x\%$ of the situations where he had the opportunity. (2) On this planet, many other animals have also performed this behavior when they had the opportunity. This behavior is unremarkable.
	Conclusion: Zorbs wash themselves with water. Is this conclusion true or false?	Conclusion: Zorb washes himself with water. Is this conclusion true or false?

^a This second sentence served to disambiguate the referent of “one animal” in the preceding sentence (and of “some animals” in the instructions to participants, which were the same across conditions). “They live up to 50 years” signaled that the item is about a category, whereas “He is 50 years old” signaled that the item is about a specific individual. These sentences were the same across all trials in a condition.

Table 2. *Sample item from Experiment 2*

Property Content	Condition	
	Quantified Statements over Individuals	Quantified Statements over Situations
Distinctive	<p>Background information: One animal is called morseths. They live up to 50 years.</p> <p>Evidence: (1) $x\%$ of morseths climb trees. (2) On this planet, very few other animals perform this behavior. This behavior is extraordinary.</p> <p>Conclusion: [Some/All]^a morseths climb trees. Is this conclusion true or false?</p>	<p>Background information: One animal is called Morseth. He is 50 years old.</p> <p>Evidence: (1) Morseth has climbed trees in $x\%$ of the situations where he had the opportunity. (2) On this planet, very few other animals have ever performed this behavior when they had the opportunity. This behavior is extraordinary.</p> <p>Conclusion: Morseth [sometimes/always]^a climbs trees. Is this conclusion true or false?</p>
	Non-distinctive	<p>Background information: One animal is called zorbs. They live up to 50 years.</p> <p>Evidence: (1) $x\%$ of zorbs wash themselves with water. (2) On this planet, many other animals also perform this behavior. This behavior is unremarkable.</p> <p>Conclusion: [Some/All]^a zorbs wash themselves with water. Is this conclusion true or false?</p>

^a “Some” and “sometimes” were used in existentially-quantified statements. “All” and “always” were used in universally-quantified statements.

Table 3. *Sample item from Experiment 3*

Property Content	Condition	
	Category	Individual
Dangerous	Background information: One animal is called morseths. They live up to 50 years.	Background information: One animal is called Morseth. He is 50 years old. ^a
	Evidence: $x\%$ of morseths chop off people's heads.	Evidence: Morseth has chopped off people's heads in $x\%$ of the situations where he had the opportunity.
	Conclusion: Morseths chop off people's heads.	Conclusion: Morseth chops off people's heads.
	Is this conclusion true or false?	Is this conclusion true or false?
Non-dangerous	Background information: One animal is called zorbs. They live up to 50 years. ^a	Background information: One animal is called Zorb. He is 50 years old. ^a
	Evidence: $x\%$ of zorbs climb trees.	Evidence: Zorb has climbed trees in $x\%$ of the situations where he had the opportunity.
	Conclusion: Zorbs climb trees.	Conclusion: Zorb climbs trees.
	Is this conclusion true or false?	Is this conclusion true or false?

^a This second sentence served to disambiguate the referent of “one animal” in the preceding sentence (and of “some animals” in the instructions to participants, which were the same across conditions). “They live up to 50 years” signaled that the item is about a category, whereas “He is 50 years old” signaled that the item is about a specific individual. These sentences were the same across all trials in a condition.

Table 4. *Sample item from Experiment 4*

Property Content	Condition	
	Quantified Statements over Individuals	Quantified Statements over Situations
Dangerous	Background information: One animal is called morseths. They live up to 50 years.	Background information: One animal is called Morseth. He is 50 years old.
	Evidence: $x\%$ of morseths chop off people's heads.	Evidence: Morseth chopped off people's heads in $x\%$ of the situations where he had the opportunity.
	Conclusion: [Some/All] ^a morseths chop off people's heads. Is this conclusion true or false?	Conclusion: Morseth [sometimes/always] ^a chops off people's heads. Is this conclusion true or false?
Non-dangerous	Background information: One animal is called zorbs. They live up to 50 years.	Background information: One animal is called Zorb. He is 50 years old.
	Evidence: $x\%$ of zorbs climb trees.	Evidence: Zorb climbed trees in $x\%$ of the situations where he had the opportunity.
	Conclusion: [Some/All] ^a zorbs climb trees. Is this conclusion true or false?	Conclusion: Zorb [sometimes/always] ^a climbs trees. Is this conclusion true or false?

^a “Some” and “sometimes” was used in existentially-quantified statements. “All” and “always” was used in universally-quantified statements.

Table 5. Sample item from Experiment 5

Property Content	Condition	
	Category	Individual
Dangerous	Background information: One animal is called morseths. They live up to 50 years.	Background information: One animal is called Morseth. He is 50 years old.
	Evidence: $x\%$ of morseths chop off people's heads.	Evidence: Morseth has chopped off people's heads in $x\%$ of the situations where he had the opportunity.
	On this planet, this behavior is rare and extremely dangerous.	On this planet, this behavior is rare and extremely dangerous.
	Conclusion: Morseths chop off people's heads.	Conclusion: Morseth chops off people's heads.
	How much do you agree with this conclusion?	How much do you agree with this conclusion?
Non-dangerous	Background information: One animal is called zorbs. They live up to 50 years.	Background information: One animal is called Zorb. He is 50 years old.
	Evidence: $x\%$ of zorbs climb trees.	Evidence: Zorb has climbed trees in $x\%$ of the situations where he had the opportunity.
	On this planet, this behavior is rare and not dangerous at all.	On this planet, this behavior is rare and not dangerous at all.
	Conclusion: Zorbs climb trees.	Conclusion: Zorb climbs trees.
	How much do you agree with this conclusion?	How much do you agree with this conclusion?

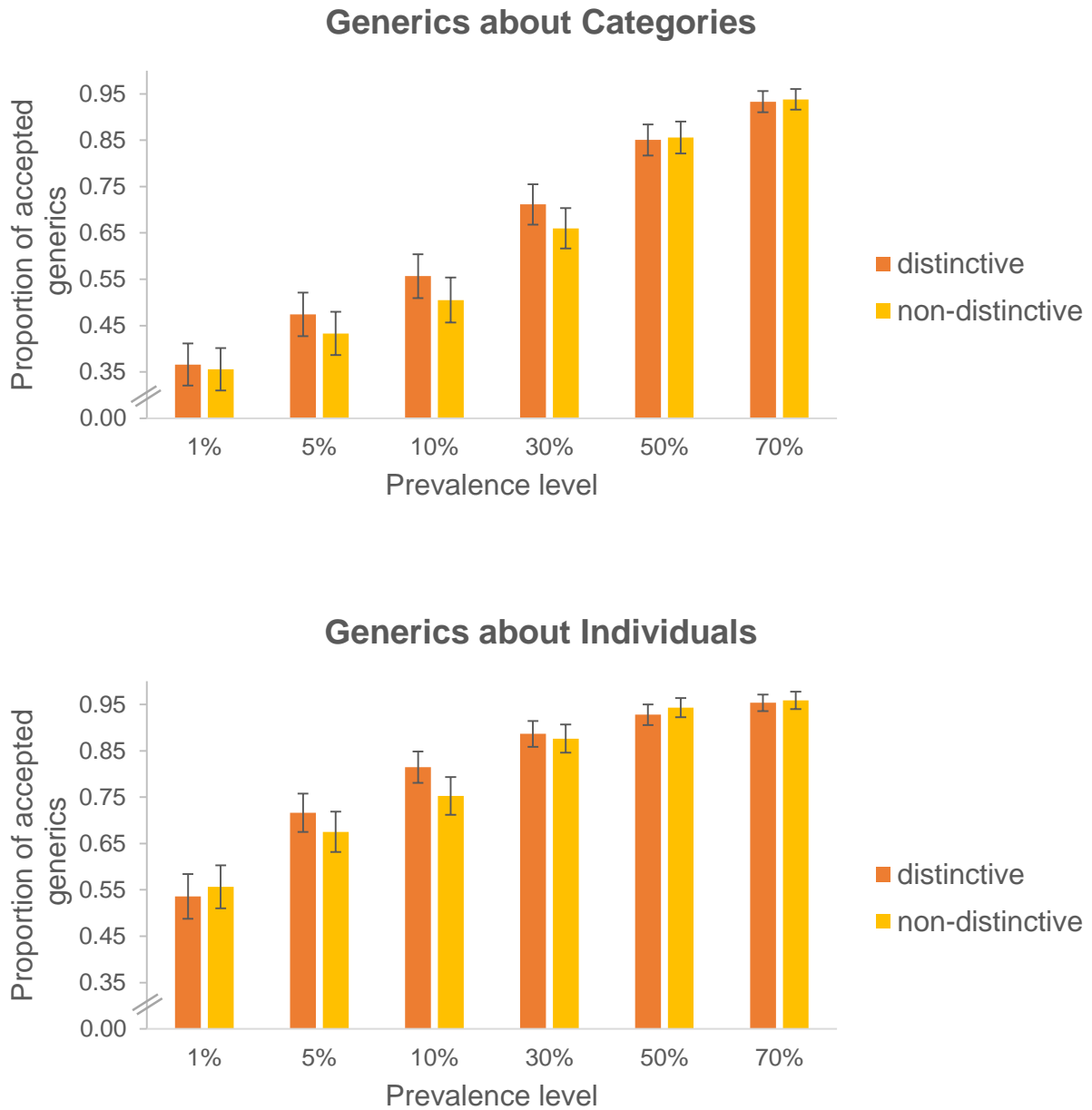


Figure 1. Average proportion of “true” responses in Experiment 1, by type of generic (category vs. individual), distinctiveness, and prevalence level. Error bars represent $\pm 1 SE$.

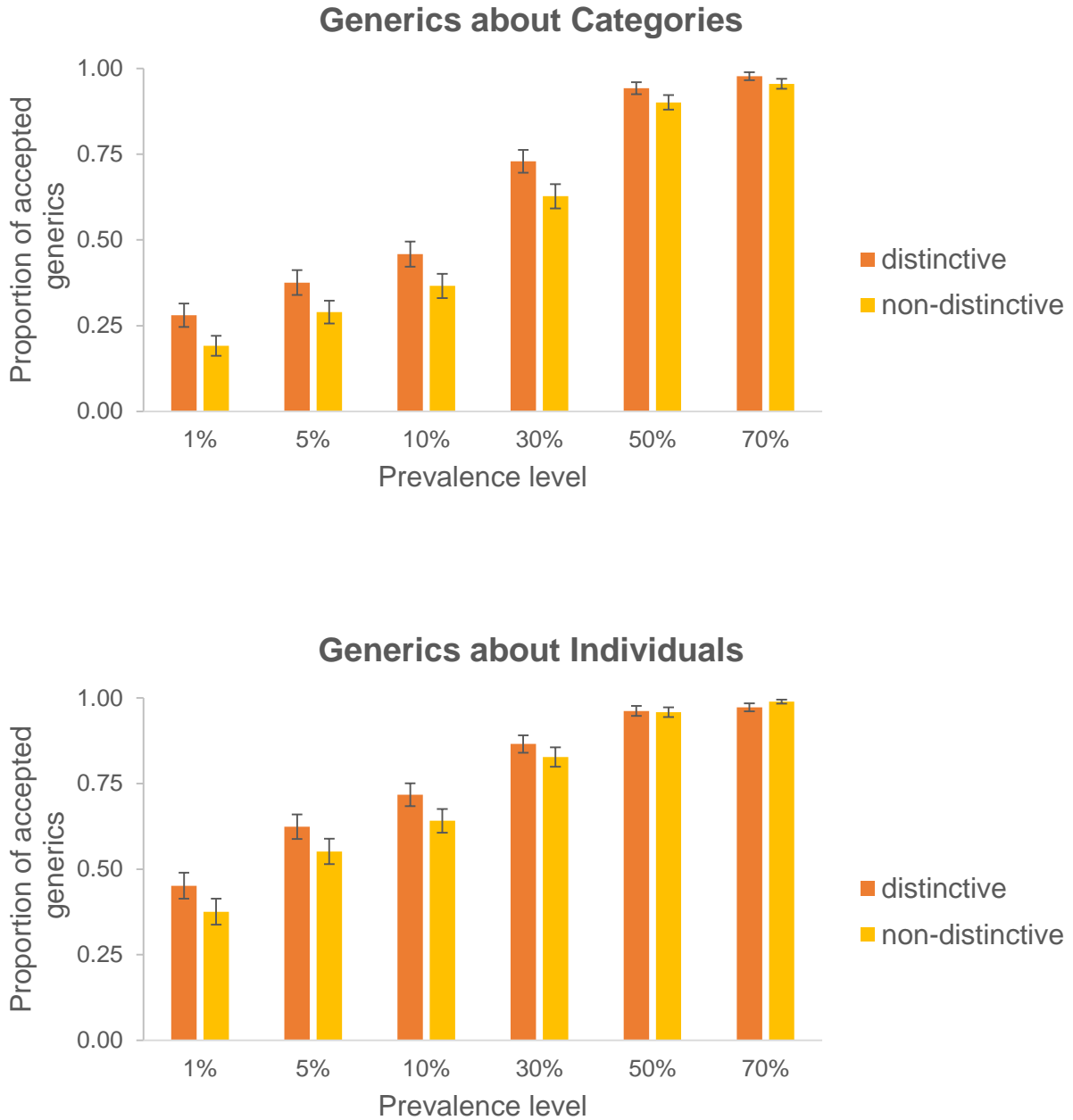


Figure 1R. Average proportion of “true” responses in Experiment 1R, the preregistered replication of Experiment 1, by type of generic (category vs. individual), distinctiveness, and prevalence level. Error bars represent $\pm 1 SE$.

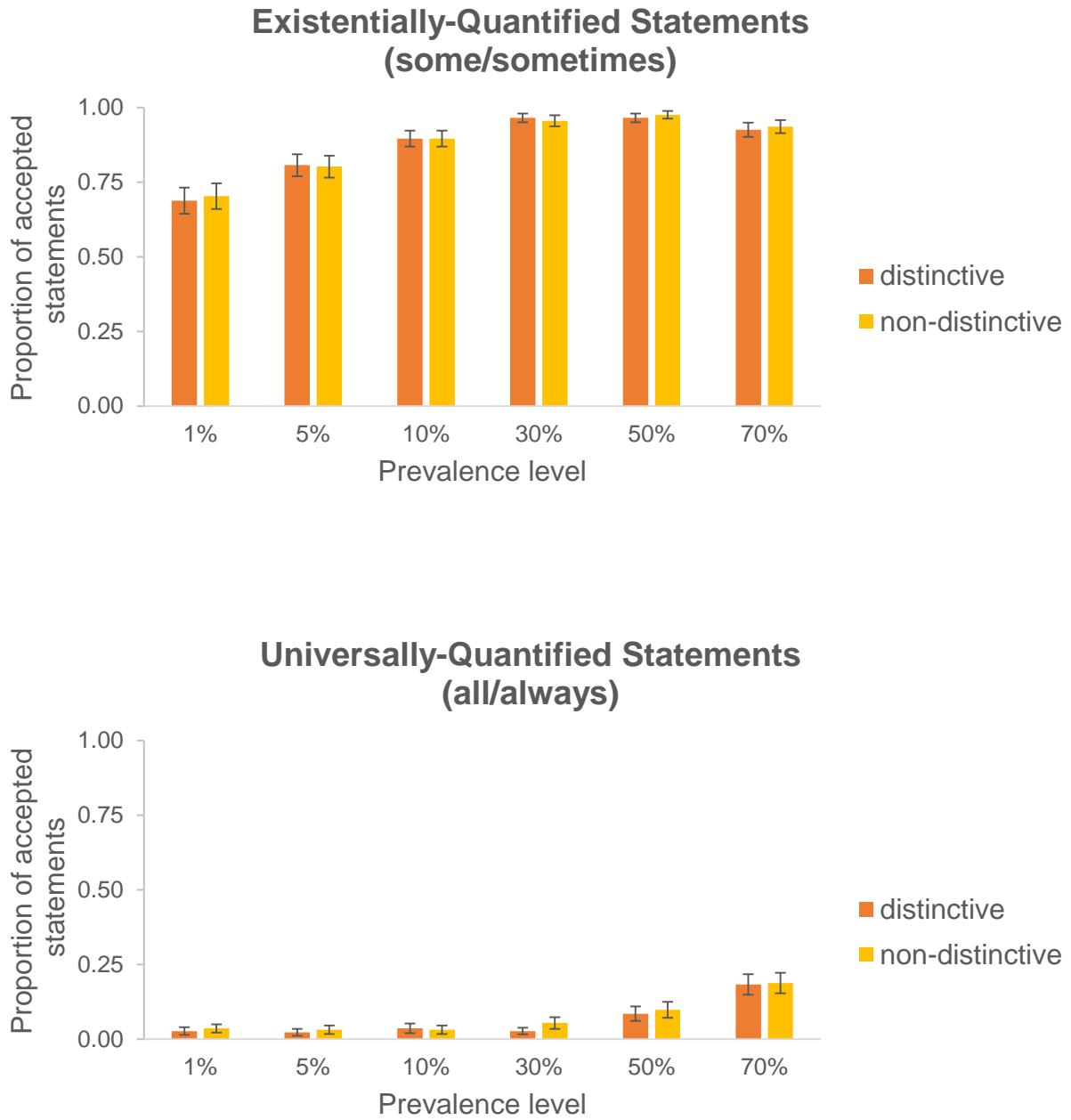


Figure 2. Average proportion of “true” responses in Experiment 2, by type of quantified statement, distinctiveness, and prevalence level. Error bars represent $\pm 1 SE$.

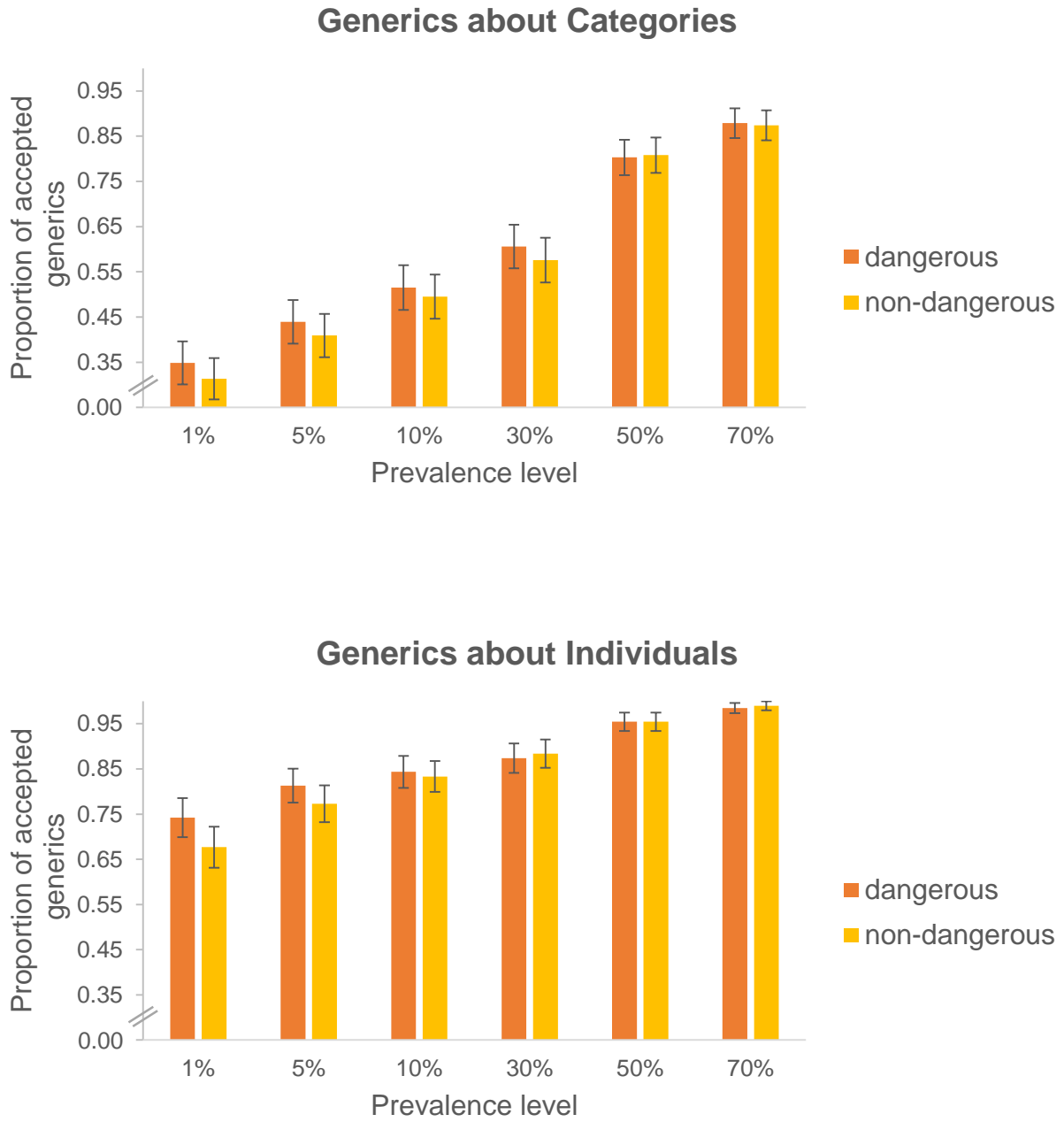


Figure 3. Average proportion of “true” responses in Experiment 3, by type of generic (category vs. individual), dangerousness, and prevalence level. Error bars represent $\pm 1 SE$.

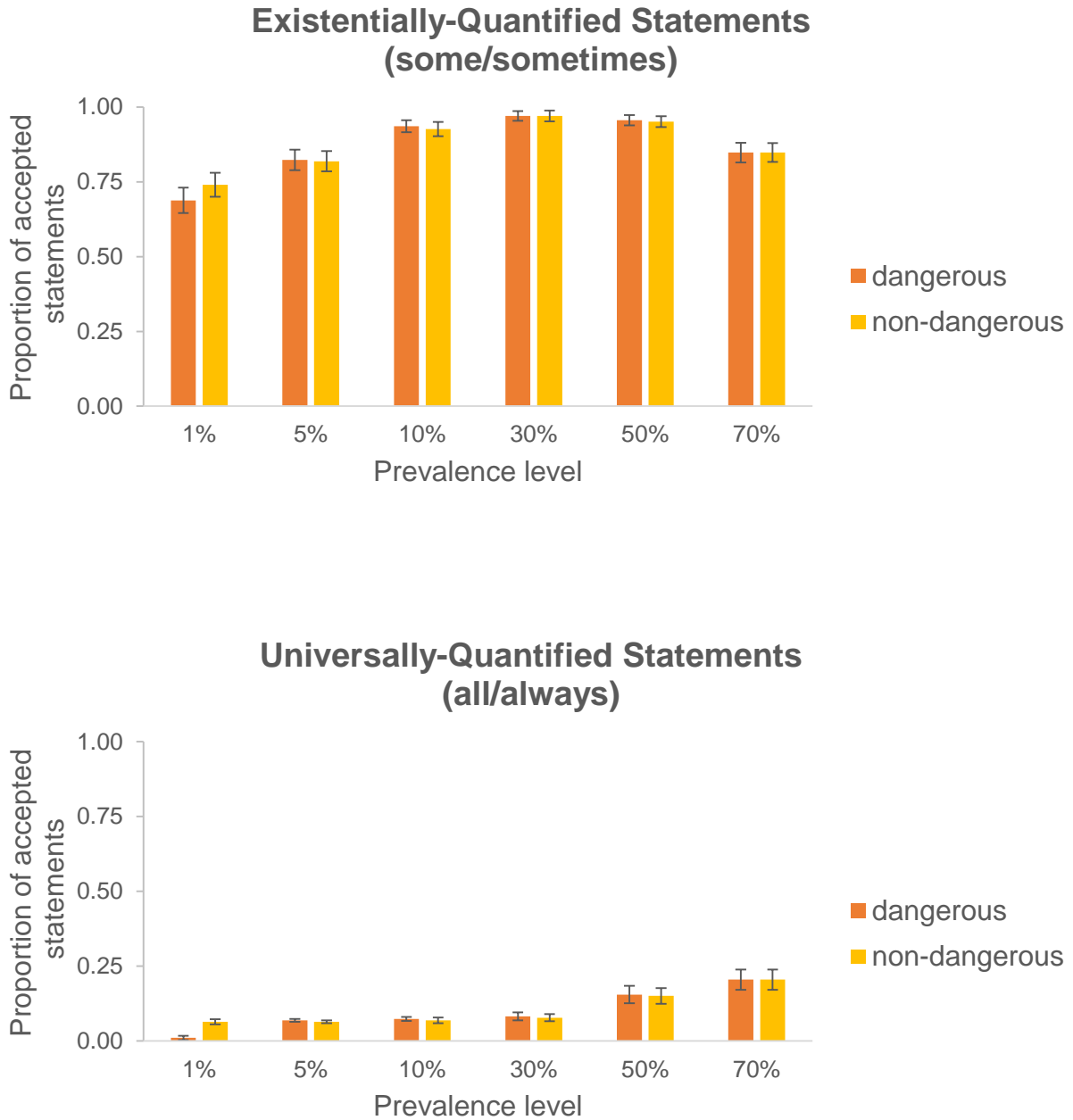


Figure 4. Average proportion of “true” responses in Experiment 4, by type of quantified statement (category vs. individual), dangerousness, and prevalence level. Error bars represent ± 1 SE.

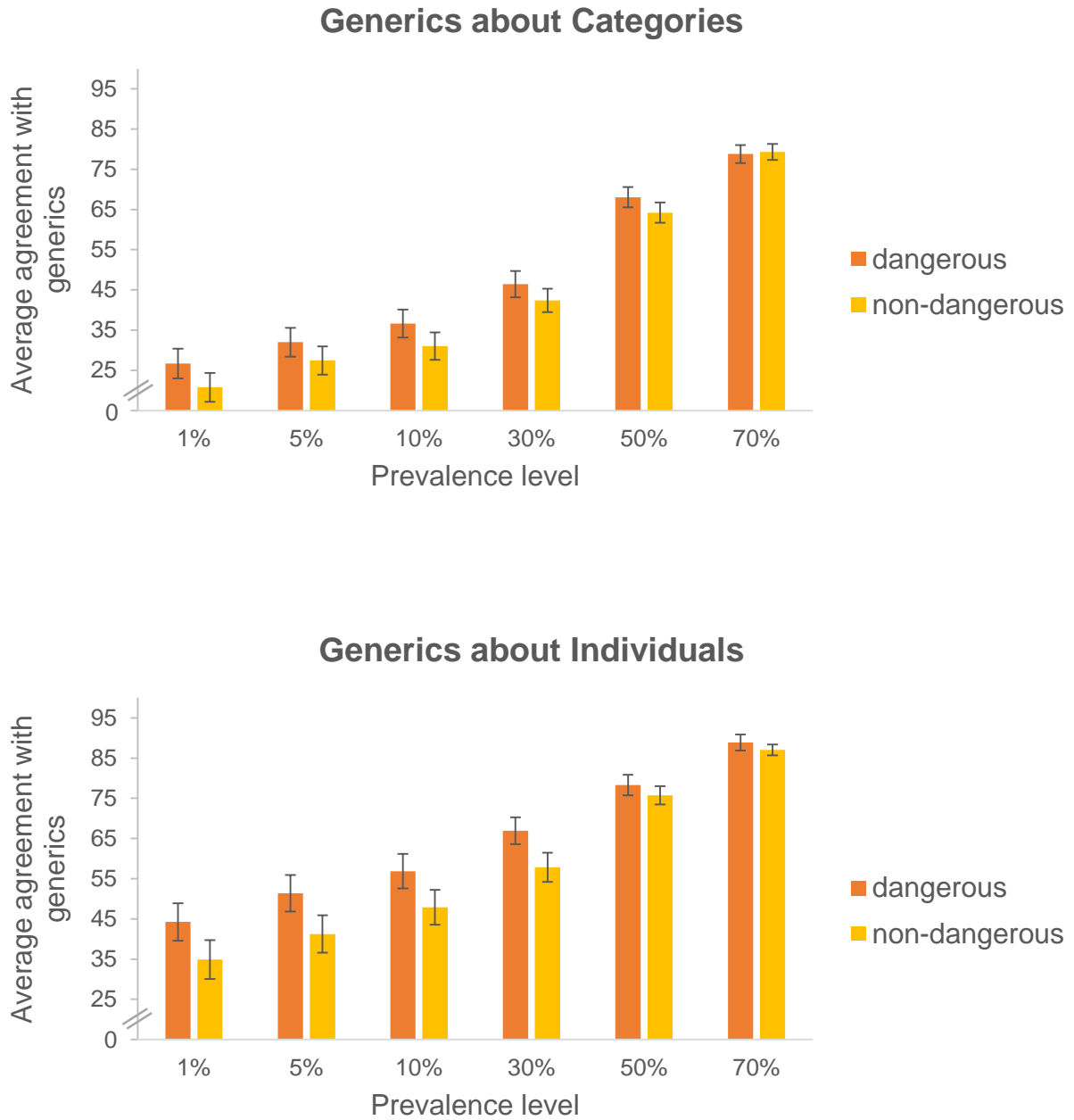


Figure 5. Average agreement level in Experiment 5, by type of generic (category vs. individual), dangerousness, and prevalence level. Error bars represent $\pm 1 SE$.