COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13209

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13209

Generic Language for Social and Animal Kinds: An Examination of the Asymmetry Between Acceptance and Inferences

Federico Cella,[a] Kristan A. Marchak,[b] Claudia Bianchi,[a] Susan A. Gelman[c]

[a]*Faculty of Philosophy, Vita-Salute San Raffaele University*
[b]*Faculté Saint-Jean, University of Alberta*
[c]*Department of Psychology, University of Michigan*

Received 22 February 2021; received in revised form 16 September 2022; accepted 29 September 2022

Abstract

Generics (e.g., "Ravens are black") express generalizations about categories or their members. Previous research found that generics about animals are interpreted as broadly true of members of a kind, yet also accepted based on minimal evidence. This asymmetry is important for suggesting a mechanism by which unfounded generalizations may flourish; yet, little is known whether this finding extends to generics about groups of people (heretofore, "social generics"). Accordingly, in four preregistered studies ($n = 665$), we tested for an inferential asymmetry for generics regarding novel groups of animals versus people. Participants were randomly assigned to either an Implied Prevalence task (given a generic, asked to estimate the prevalence of a property) or a Truth-Conditions task (given prevalence information, asked whether a generic was true or false). A generic asymmetry was found in both domains, at equivalent levels. The asymmetry also extended to properties varying in valence (dangerous and neutral). Finally, there were differences as a function of property valence in the Implied Prevalence task and a small but consistent interaction between domain and prevalence in the

Correspondence should be sent to Federico Cella, Faculty of Philosophy, Vita-Salute San Raffaele University, Via Olgettina 58, 20132 Milan (MI), Italy. E-mail: f.cella@outlook.com

Truth-Conditions task. We discuss the implications of these results for the semantics of generics, theoretical accounts of the asymmetry, and the relation between generics and stereotyping.

Keywords: Generic language; Social categories; Animal categories; Conceptual domains; Stereotyping

1. Introduction

Generics are sentences that express general claims regarding categories or their members and lack explicit quantifiers, such as "Tigers are striped." They contrast with non-generic sentences such as "This tiger is striped" (which is specific) or "Most tigers are striped" (which is quantified). Generic noun phrases can have different syntactic forms in English, including *bare plural* (as in "Leopards have spots"), *definite singular* (as in "The raven is black"), and *indefinite singular* (as in "A duck lays eggs"). Generics have both puzzling semantic features and distinctive cognitive implications; for these reasons, they have attracted the interest of linguists (e.g., Carlson, 1977; Krifka et al., 1995), philosophers (e.g., Langton, Haslanger, & Anderson, 2012; Leslie, 2008), and psychologists (e.g., Cimpian, Brandone, & Gelman, 2010; Gelman, 2003).

One intriguing finding is that generics about novel animal categories, such as "Morseths have silver fur," are characterized by an *inferential asymmetry*: they are interpreted as referring to nearly all members of the kind, despite being accepted even if the ascribed property is present in relatively few members (e.g., 10%; Cimpian et al., 2010). Among the most far-reaching implications of the asymmetry is its potential for helping to explain the transmission and acceptance of unwarranted generalizations, such as stereotypes about social categories (e.g., Bian & Cimpian, 2017). Yet, to date, little is known about whether the asymmetry would be found in this domain. Accordingly, the current paper investigates the robustness and generalizability of the inferential asymmetry for *social generics*—that is, generics about groups of people. We also examined whether the inferential asymmetry varies as a function of property valence (dangerous and neutral) in two domains (animals and social categories). Below, we briefly review several key background issues, including the semantics of generics, the generic inferential asymmetry, potential variations of the asymmetry as a function of domain and property valence, and theoretical accounts of the asymmetry.

1.1. *Semantics of generics*

Generics are frequent in natural language and play a powerful role in children's development of kind concepts (e.g., Brandone, Cimpian, Leslie, & Gelman, 2012; Gelman, 2003; Gelman & Bloom, 2007; Rhodes, Leslie, Bianchi, & Chalik, 2018a; Segall, Birnbaum, Deeb, & Diesendruck, 2015; Sutherland & Cimpian, 2015). Nonetheless, the semantics of generics is not straightforward. In contrast to sentences including a quantifier such as "some," "most," or "all," which have precise acceptance conditions based on quantity (e.g., the statement "All tigers are striped" is true if and only if every tiger is striped), generics are not simply about quantity (e.g., Abelson & Kanouse, 1966; Brandone, Gelman, & Hedglen, 2015; Carlson, 1977; Cimpian et al., 2010; Gelman, Star, & Flukes, 2002; Krifka et al., 1995; Leslie, 2008). For example, the statement "Ducks lay eggs" is intuitively true, whereas the statement "Ducks are female" is not, even though the number of female ducks is greater than that of egg-laying ducks. Further, generics often gloss over exceptions: "Ravens are black" is intuitively true despite the existence of albino ravens. Generic statements may even be judged to be true when exceptions apply to the vast majority of category members, as in the case of "Mosquitoes carry the West Nile virus," which is intuitively true though less than 1% of mosquitoes have the ascribed property (Cox, 2004).

1.2. The inferential asymmetry

Based on the semantic features of generics, Cimpian et al. (2010) hypothesized that the interpretation of these sentences would elicit an inferential asymmetry between their prevalence implications and acceptance conditions. In the first study carried out by Cimpian et al. (2010), participants were randomly assigned to complete either an *Implied Prevalence* task or a *Truth-Conditions* task. In the Implied Prevalence task, participants were provided with generic statements about a novel animal category (e.g., "Lorches have purple feathers") and were asked to estimate what percentage of category members, from 0 to 100%, possess the ascribed property. In the Truth-Conditions task, participants were told that a certain percentage of the category members had a property (e.g., "30% of lorches have purple feathers") and then were asked whether the corresponding generic statement (e.g., "Lorches have purple feathers") was true or false. The items in this task were presented at the following prevalence levels: 10%, 30%, 50%, 70%, and 90%.

Cimpian et al.'s (2010) findings supported the asymmetry hypothesis: the average score in the *Implied Prevalence* task was significantly higher than the average score in the *Truth-Conditions* task (approximately 95% and 70%, respectively). In other words, the properties ascribed by generics were interpreted as applying to almost all members of the category, whereas the same generic statements were judged to be true even when only a small percentage of the members of the category were said to display the property. Notably, generics were distinctive in this regard: the asymmetry was not found for quantified sentences (e.g., "Most morseths have silver fur"), for which the average scores for acceptance and prevalence implications were equivalent. This asymmetry has been found with children as well as with adults (Brandone et al., 2015), suggesting that it is early emerging in human cognition.

1.3. The role of domain in the interpretation of generics

A key question is whether social generics should behave any differently than generics about animal categories. This is an important question, given that social generics are a common means to express stereotypes (Gelman, Taylor, & Nguyen, 2004). If social generics elicit an inferential asymmetry, they could be especially pernicious and misleading, as a property that is true of only a small number of a group of people may be assumed to be broadly representative. However, as noted earlier, the generic asymmetry has not yet been examined in social categories.

Prior theorizing leaves all possibilities open. We might expect to see the same inferential asymmetry for social categories as was observed for animal kinds, given that people often reason about social categories as if they were natural kinds (Prentice & Miller, 2007; Rhodes & Gelman, 2009; Rothbart & Taylor, 1992). Consistent with this possibility, generic language promotes essentialist reasoning in both domains (Foster-Hanson, Leslie, & Rhodes, 2016, 2019; Gelman, Ware, & Kleinberg, 2010; Leshin et al., 2020; Rhodes, Leslie, & Tworek, 2012, 2018a, 2018b).

A second possibility is that we may find a greater asymmetry in the social domain. An extensive body of research in social psychology demonstrates that stereotypes operate by generalizing well beyond the evidence (e.g., Allport, 1954; Fiske, 1998; Hammond & Cimpian, 2017). For example, people endorse as true the stereotype that "boys don't cry," even though most boys do in fact cry (Wodak, Leslie, & Rhodes, 2015). Accordingly, we may expect people to be especially prone to endorse generic statements about social categories on the basis of minimal evidence, and thus to judge social generics to be true at lower prevalence levels than they judge generics about animals to be true. This would result (all other things being equal) in a greater asymmetry for social generics.

However, a third possibility is that we may instead expect to see less asymmetry for social generics, or even no asymmetry at all. One reason to expect less of an asymmetry may be due to differences in the structure of social and animal categories. In contrast to animal kinds, whose members are often highly similar to one another (e.g., different skunks are highly similar in appearance and behavior), members of a social kind may be more variable (e.g., Brandone, 2017; Nisbett, Krantz, Jepson, & Kunda, 1983). For example, girls may differ from one another in age, race, ethnicity, preferences, dwellings, clothing, dietary preferences, abilities, languages spoken, and so on. If people expect more variability among members of a social category, then generics regarding social categories may elicit lower implied prevalence ratings, and consequently, lead to less of an asymmetry. Another reason to expect less of an asymmetry is that people seem to be more likely to accept negative generic statements regarding animal kinds than social groups (Tasimi, Gelman, Cimpian, & Knobe, 2017). However, Tasimi et al. did not test the asymmetry directly, as they employed the Truth-Conditions task only, and not the Implied Prevalence task.

In short, it is an open question whether the inferential asymmetry holds for social generics, due to the competing theoretical accounts summarized above.

1.4. The role of property valence in the interpretation of generics

Leslie (2008) hypothesized that generics ascribing *distinctive* and *dangerous* properties are more easily accepted than generics ascribing *neutral* properties. Properties that are perceived as distinctive or dangerous have relatively high informational value, and, in turn, might be more prominent in our conceptual knowledge, and so more readily accepted in generic form. To test this hypothesis, Cimpian et al. (2010) examined people's endorsement of properties of novel animal kinds that were neutral (e.g., "have purple feathers"), dangerous (e.g., "have a silver fur that sheds particles that make it impossible to breathe"), or distinctive (e.g., "have distinctive blue scales that are soft, flexible, and very shiny"). Importantly, they found that generics expressing dangerous or distinctive properties of animals were more likely to be judged true than generics expressing neutral properties (see also Bian & Cimpian, 2021). It is unclear, however, whether this tendency would generalize to social categories.

We are aware of only one set of studies that speaks to this issue. Tasimi et al. (2017) found that generics ascribing threatening properties (e.g., "are dangerous") versus generics ascribing non-threatening properties (e.g., "are helpful") were accepted alike when social categories, but not artifact or animal categories, were concerned. Although important, this work was limited in what it can reveal regarding property valence, as the findings were all relative, involving how negatively valenced properties compared to positively valenced properties, rather than property valence differences per se. Moreover, the non-threatening properties tested by Tasimi et al. (2017) did not have a neutral valence but rather were prosocial properties, which may be interpreted differently from neutral properties, when compared with dangerous ones. Thus, it is still an open question whether there is something special about dangerous properties in how they are interpreted in the domain of social kinds.

A related issue concerns people's willingness to generalize a property based on whether the generic property is dangerous or neutral. Prasada, Khemlani, Leslie, and Glucksberg (2013) found that dangerous generic properties about familiar natural and artifact categories were interpreted as referring to common *dispositions* rather than to prevalence per se. For example, the interpretation of a generic like "Ticks carry Lyme disease" is related to the shared biological structure of ticks, which cause them to be disposed to carry the relevant disease. Such an interpretation does not take into account how prevalent the ascribed property is; indeed, the actual proportion of category members displaying the dangerous property depends on whether determined enabling conditions are present (e.g., whether ticks feed on infected animals), which may be rare or even absent. Such statements led to lower prevalence estimates than other generic properties. Similarly, Lazaridou-Chatzigoga, Katsos, and Stockall (2019) found that both adults and children were more willing to extend the properties ascribed by neutral generic properties about novel creatures (e.g., "Ackles love to play with toys") to a new member of the category than the properties ascribed by dangerous generic properties (e.g., "Ackles love to play with fire"). However, as Lazaridou-Chatzigoga et al. (2019) noted, the properties they tested were more child-friendly than those of Cimpian et al. (2010) and thus potentially less salient. Moreover, many of their dangerous properties were likely to be perceived as dangerous for the category members only. It also may be that baseline assumptions played a role (e.g., in general, playing with toys is considered more enjoyable than playing with fire). In conclusion, another open question is whether the prevalence estimations elicited by dangerous generic properties differ from those based on neutral generic properties.

1.5. Theoretical accounts of the asymmetry

Two main proposals have been put forward to explain the processes that underlie the generic asymmetry. The first proposal by van Rooij and Schulz (2020) states that generics are judged to have nearly universal prevalence implications because people often confuse the *representativeness*, or *stereotypicality*, of a genericized feature for a particular kind with its probability. In contrast, the flexible truth-conditions of generics should be analyzed in terms of three different factors (*property typicality* measured via *relative difference*, *co-alternative* *features*, and representativeness of the feature for the category) with representativeness being most relevant to the asymmetry. Similar to judgments of generics' implied prevalence, one reason generics are accepted based on even weak evidence is that their acceptance conditions should also be analyzed in terms of the representativeness, or stereotypicality, of the genericized feature for the relevant kind: if representativeness is high enough, the relevant generic is accepted even if most members of the kind lack the ascribed feature.

The second proposal for the asymmetry comes from Tessler and Goodman (2019a). They contend that the nearly universal implications of generics should be explained by what they call the "*interpretation* model." According to this model, when a speaker utters a generic, the listener interprets the utterance as concerning a prevalence level higher than a threshold θ. Whereas quantified generalizations such as "Most Ks have F" have a fixed θ (in this example, the sentence is true only if more than 50% of Ks have F), generics have a vague or underspecified θ, which is contextually determined by the probabilistic world knowledge of the listeners. When interpreting generics, listeners usually assume that θ is high, unless their world knowledge about the ascribed feature suggests a lower θ (e.g., the use of accidental/temporary sounding properties; see also Cimpian et al., 2010; Tessler & Goodman, 2019b). For this reason, prevalence estimates based on generics tend to be high.

Tessler and Goodman (2019a) further argue that the flexible truth-conditions of generics are explained by what they call the "*endorsement* model." They characterize endorsement as the decision of a speaker to produce or not to produce a generic for a naïve listener. Such a decision is made based on the assumption that the listener would use the interpretation model to evaluate the utterance of a generic. Following the utterance of a generic, listeners update their prior beliefs about the prevalence of the ascribed property among the relevant category members. Consequently, before using a generic, a speaker needs to reason about whether the actual prevalence of the property is more consistent with (1) the listener's prior prevalence estimate of the ascribed property or with (2) the listener's posterior prevalence estimate of the same property after the utterance. If the speaker thinks that the actual prevalence of the property is more consistent with (2), the generic is endorsed and produced. Consider one of the examples provided by Tessler and Goodman (2019a): the dangerous generic "Mosquitoes carry malaria." Many animal kinds lack the property "carrying malaria," and even among animals that do carry malaria, very few individuals display the ascribed property. Although the prior prevalence for "carrying malaria" tends to be very low, the speaker might think that the listener's posterior prevalence estimate will be more consistent with the actual proportion of malaria-carrying mosquitoes than the listener's prior prevalence estimate. In that case, the generic "Mosquitoes carry malaria" is endorsed and produced.

Investigating potential variations of the interpretation of generics as a function of domain and property valence would inform theoretical accounts of the generic asymmetry. More specifically, examining different domains and properties would allow us to test how "priors" affect the asymmetry. First, consider the role of domain in the interpretation of generics. As previously discussed, prior research leaves open whether the finding of the asymmetry extends to the interpretation of social generics, since animal kinds are perceived as more homogenous than social categories. For example, if prevalence expectations differ across domains, but the generic asymmetry does not differ across domains, then the analysis put forward by Tessler and Goodman (2019a) would require updating. On the contrary, if the prior prevalence expectations about domain affect the inferential asymmetry, the results of the current investigation would support Tessler and Goodman's (2019a) account.

Investigating the role of property valence in the interpretation of generics would also allow us to test theoretical accounts. van Rooij and Schulz's (2020) account of the asymmetry predicts that dangerous generics should elicit higher prevalence estimates than neutral generics (as the former have higher representativeness). Given that features that are perceived as dangerous, striking, or fear-inducing have high representativeness, they suggest that the emotional impact of such information will often lead people to think that the relevant features are widespread among the category members. In contrast, Tessler and Goodman's (2019a) account predicts that dangerous generics should elicit lower estimates than neutral generics (as the prior prevalence for dangerous properties tends to be very low). For this reason, investigating whether and how prevalence estimates elicited by dangerous generics differ from those based on neutral generics could provide data that partly speak to these two accounts.

1.6. *The present studies*

In the present work, we conducted four preregistered studies to test whether the inferential asymmetry found by Cimpian et al. (2010) with generics about animals extends to social generics. Furthermore, we examined whether property valence (neutral vs. dangerous) affects people's judgments. In Study 1, we investigated whether the finding of the inferential asymmetry obtained with generics about animal categories replicates using an improved methodology and a larger online sample. In Study 2, we tested whether the generic asymmetry differs as a function of domain (animals vs. people). In Studies 3a and 3b, we tested whether the generic asymmetry differs as a function of domain (animals vs. people) and property valence (neutral vs. dangerous).

2. Study 1

In Study 1, we examined whether the finding of the asymmetry replicates with a larger online sample. Furthermore, we examined whether this finding holds when a 100% prevalence level was included in the *Truth-Conditions* task, as the prevalence levels used by Cimpian et al. (2010) in their *Truth-Conditions* task did not include 100%. The inclusion of a 100% level in our study simplified the interpretation of participants' responses, because we did not need to impute a score of 100% to participants who judged all items to be false, as was done in Cimpian et al. (2010). The preregistration of this study is available at https://aspredicted.org/blind.php?x=mw5pn8.

2.1. *Method*

2.1.1. *Participants*

One-hundred and twelve adults from the United States (60 men, 52 women; Mean age = 42.99 years; range = 24–75 years) completed the study online via Amazon Mechanical

Table 1
Sample item from Study 1 in the *Implied Prevalence* and *Truth-Conditions* tasks.

Implied Prevalence task	*Truth-Conditions task*
Information: MORSETHS have silver fur.	Information: 30% of MORSETHS have silver fur.
Question: What percentage of MORSETHS have silver fur?	Question: Is the following sentence true or false? **MORSETHS have silver fur.**

Turk (MTurk) for \$0.40.[1] Participants were randomly assigned to either the *Implied Prevalence* task ($n = 55$) or the *Truth-Conditions* task ($n = 57$). In this and each subsequent study, participants had been granted Master Worker status by MTurk, had a US IP address, 1000+ approved HITs, and a 99%+ HIT approval rate. Participants were 83% White, 8% Asian or Asian American, 5% Black or African American, 2% Latino or Hispanic, and 2% Multiracial/Multiethnic. Five additional participants were tested and excluded from the final sample for having non-US IP addresses. One other participant was tested and excluded for having a duplicate IP address.

2.1.2. Materials and procedure

In all studies, novel categories and novel labels were used, to ensure that participants were not simply retrieving learned facts. We created a list of 12 items, each consisting of a novel label for an animal category and a property that described the color of a body part of the animal (e.g., “MORSETHS have silver fur”). The list consisted of the 10 plain properties from Cimpian et al. (2010) and two additional items that we created for this study. (This was done to include two items at the 100% prevalence level; for a complete list of items, see Appendix A.) Also, in contrast to Cimpian et al. (2010), the assignment of labels to properties was randomized for each participant. Moreover, in contrast to Cimpian et al., the labels for the novel categories in our studies were presented entirely in uppercase letters. This permitted consistent presentation for animals and people in subsequent studies, as otherwise category labels for animals typically begin with a lowercase letter (e.g., dogs), whereas category labels for people typically begin with a capital letter (e.g., Canadians). For an example of an item as it would appear to participants, see Table 1.

At the beginning of the survey, each participant read the same introductory text as presented in Cimpian et al. (2010):

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

In the *Implied Prevalence* task, participants were presented with a generic statement and then asked to estimate the implied prevalence of the property described in the statement, from 0 to 100%. Participants in this task received 10 items randomly selected from the list of 12 items. (We provided 10 items instead of 12 in order to equate the items with that of the original Truth-Conditions task mentioned in footnote 1 and presented in the Supplementary Online Materials (SOM); see "Study S1.") In the *Truth-Conditions* task, participants judged whether generic statements were "true" or "false" based on the prevalence level of an ascribed property, with each of the following prevalence levels presented twice: 10%, 30%, 50%, 70%, 90%, and 100%. In this task, participants received all 12 items. The order of the items in both tasks was randomized for each participant.

2.1.3. Open data

The raw data for this and subsequent studies are available on the Open Science Framework (OSF): https://osf.io/ru9t7/.

2.2. Results and discussion

2.2.1. Data coding

To enable a comparison of participants' responses in the *Implied Prevalence* task to those in the *Truth-Conditions* task, in this and subsequent studies, we calculated mean prevalence scores for participants in each task (based on the coding scheme used in Cimpian et al., 2010). In the *Implied Prevalence* task, we averaged participants' responses out of 100% across the 10 items. In the *Truth-Conditions* task, we converted participants' "true"/"false" judgments to a mean prevalence level that led to "true" responses. To calculate this score, we added the percentage level of items that a participant judged to be "true" and divided this score by the total number of items that the participant judged to be "true." For example, a participant's score would be 80% if they selected "true" on two items, one where the prevalence was 70% and the other where the prevalence level was 90%.

2.2.2. Did the generic asymmetry replicate with a larger online sample and when we included a 100% level in the Truth-Conditions task?

We conducted an independent sample *t*-test of the mean prevalence ratings in the two tasks. As in Cimpian et al. (2010), we observed an asymmetry between ratings, with participants providing higher mean ratings in the *Implied Prevalence* task than in the *Truth-Conditions* task, $t(110) = 10.49$, $p < .001$, $d = 1.98$ (see Fig. 1).

2.2.3. Conclusions

The results of Study 1 provide clear evidence of a generic asymmetry about physical features of animal categories when the methods of Cimpian et al. (2010) were replicated with a larger sample, including a replication of the large effect size found in prior work ($\eta_p{}^2 = .39$ or the equivalent of $d = 1.60$). In the *Implied Prevalence* task, participants interpreted generics as referring to nearly all members of the kind. In the *Truth-Conditions* task, instead, they judged these same generics to be true at substantially lower prevalence levels. These results suggest that the generic asymmetry is robust, even with an improved methodology that includes a 100% level in the Truth-Conditions task.

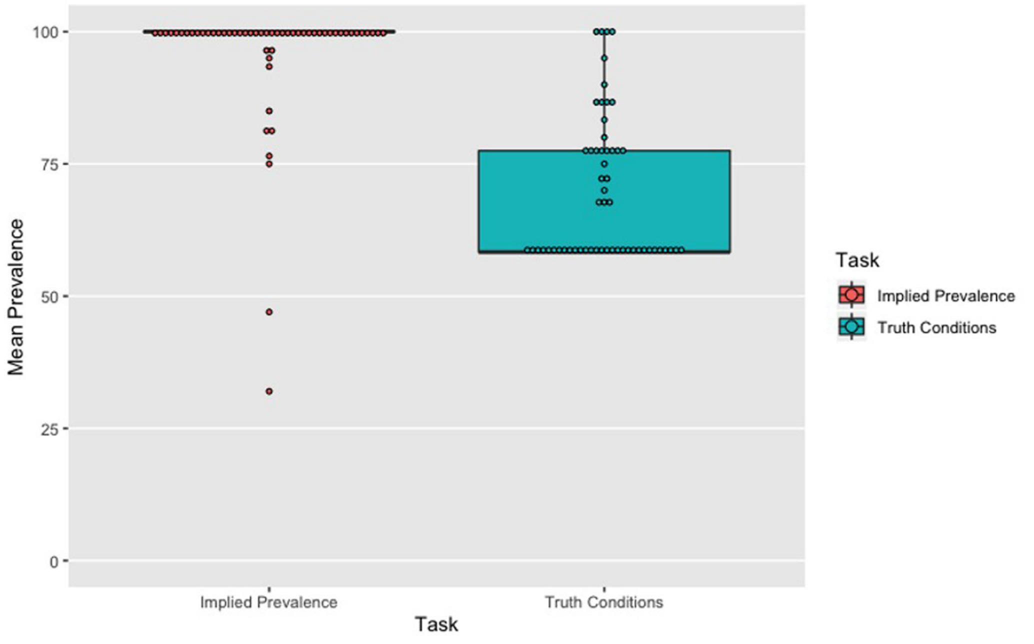

Fig. 1. Study 1, dot plots (with box plot overlays) of mean prevalence ratings, plotted by condition (Implied Prevalence vs. Truth Conditions). Each dot represents the mean prevalence rating for a participant, which was computed by averaging responses across the items they rated. The solid line represents the median.

3. Study 2

In Study 2, we investigated whether a generic asymmetry would be found for social generics, as well as for generics about animal categories. In the *Truth-Conditions* task, we specifically examined whether domain affected judgments of generics' acceptability at different prevalence levels. The preregistration of this study is available at https://aspredicted.org/blind.php?x=ru2pa8.

3.1. Method

3.1.1. Participants

Two-hundred and thirteen adults from the United States (96 men, 116 women, 1 undisclosed; Mean age = 40.72 years; range = 21–68 years) completed the study on MTurk and were paid $1. Participants were randomly assigned to one of four conditions: *Animals–Implied Prevalence* ($n = 52$), *People–Implied Prevalence* ($n = 53$), *Animals–Truth-Conditions* ($n = 56$), or *People–Truth-Conditions* ($n = 52$). Participants were 75% White, 8% Asian or Asian American, 7% Black or African American, 6% Multiracial/Multiethnic, 3% Latino or Hispanic, 0.5% Ashkenazi Jew, 0.5% not listed/other, and 0.5% undisclosed. Twenty-one additional participants were tested and excluded because they failed the manipulation check ($n = 10$ in the *People–Implied Prevalence* task and $n = 11$ in the

People–Truth-Conditions task; see "Materials and Procedure"). Three participants were excluded from the final sample for having non-US IP addresses. Two other participants were tested and excluded from the final sample since they had already participated.[2] One participant was also tested and excluded from the final sample for not being a native speaker of English.

3.1.2. Materials and procedure

We created a list of 24 items. For the sake of generality, we included 12 items that described a physical property of the category (e.g., "Xs have large tonsils"), and 12 items that described a non-physical property (e.g., "Xs sleep under trees"). For each participant, the properties were randomly assigned to one of 24 labels (see Appendix B for a complete list of the labels and items used in this study).

We adapted the introductory text from Cimpian et al. (2010) and Study 1 in the following way: "In this study, we will tell you about some **[animals]/[people]** that live on a remote island. This island is very large, and has many different **[animals]/[people]** on it. It is roughly the size of Alaska, and has a lot of geographical, climatic, and environmental variety. For each item, you will be given some information and asked a question. Please try to answer our questions to the best of your ability. Remember: the following questions are about **[animals]/[people]** on an island."

We described the island as being "roughly the size of Alaska" and having "a lot of geographical, climatic, and environmental variety" to avoid the possibility that participants would imagine a stereotypical, small tropical island. We were specifically concerned about the influence of beliefs about the size of the island on responses in the *People* conditions because (1) participants may already have pre-existing beliefs about groups of people living on tropical islands, and (2) they may also reject the premise that numerous different groups of people could live on a small island.

As in Study 1, participants completed either the *Implied Prevalence* task or the *Truth-Conditions* task. In both tasks, participants were presented with all 24 items, but the order of the items was randomized for each participant. In the *Truth-Conditions* task, participants judged two physical properties and two non-physical properties at each of the following prevalence levels: 10%, 30%, 50%, 70%, 90%, and 100%.

After each task, we also included a manipulation check to determine whether participants remembered whether the questions were about animals or people (depending on the condition they were in). We reminded them that they would see a sentence that they had originally seen at the beginning of the study, and asked them to fill in the blank in the following sentence: "Remember: the following questions are about ______ on an island." Participants who failed this manipulation check were excluded from the final sample (see "Participants").

At the end of the study, we asked participants to complete an abbreviated 13-item social desirability measure (Reynolds, 1982) to examine whether responses on this measure were predictive of people's responses in the *People* conditions in the main task. We included this measure because we wanted to make sure that participants in these conditions did not provide socially desirable answers to not appear biased toward new groups of people.

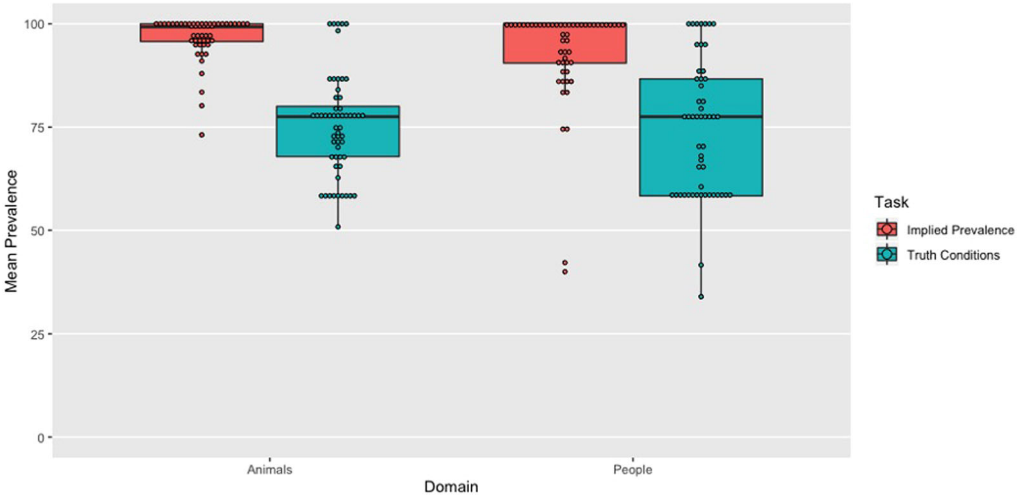

Fig. 2. Study 2, dot plots (with box plot overlays) of mean prevalence ratings, plotted by condition (Implied Prevalence vs. Truth Conditions). Each dot represents the mean prevalence rating for a participant, which was computed by averaging responses across the 24 items they rated. The solid line represents the median.

3.2. Results and discussion

3.2.1. Did the generic asymmetry vary as a function of domain?

To explore this question, we conducted a two-way analysis of variance (ANOVA) on the mean prevalence ratings, with task (Implied Prevalence vs. Truth-Conditions) and domain (animals vs. people) as between-subjects factors (see Fig. 2). We observed a main effect of task, with ratings overall higher in the *Implied Prevalence* task than the *Truth-Conditions* task, $F(1, 209) = 141.96$, $p < .001$, ${\eta_p}^2 = .40$. No other main effects or interactions were significant.

We did not observe a significant correlation between a person's score on the social desirability scale and their mean prevalence score in either the *People–Implied Prevalence* or *People–Truth-Conditions* tasks, ps $> .05$. The absence of correlation indicates that participants' judgments in our task were not based on their attempts to provide socially desirable responses.

3.2.2. Did domain affect people's judgments of generics' acceptability at different prevalence levels in the Truth-Conditions tasks?

We submitted participants' "true/false" responses in this task to a logistic mixed-effects model using the glmer command in the lme4 package in R (Bates, 2007). This analysis was not preregistered, but was included in this and subsequent studies in order to compare our results with those of Cimpian et al. (2010). In this model, we included domain (animals = 0; people = 1; between-subject), prevalence (.1, .3, .5, .7, .9, 1; within-subject), and their interactions as predictors (see Table 2). All predictors were mean-centered. We also included *participant* as a random intercept.[3] We observed a main effect of prevalence, indicating that

Table 2
Logistic regression predicting "true"/"false" judgments, based on domain, prevalence, and their interactions in Study 2.

Fixed effects	**Estimate**	***SE***	***p*-value**
(Intercept)	2.30	0.39	<.001
Domain	0.09	0.75	.90
Prevalence	9.21	0.45	< .001
Domain x Prevalence	–2.29	0.82	.005
Random effect		***SD***	
Participant	Intercept	3.68	

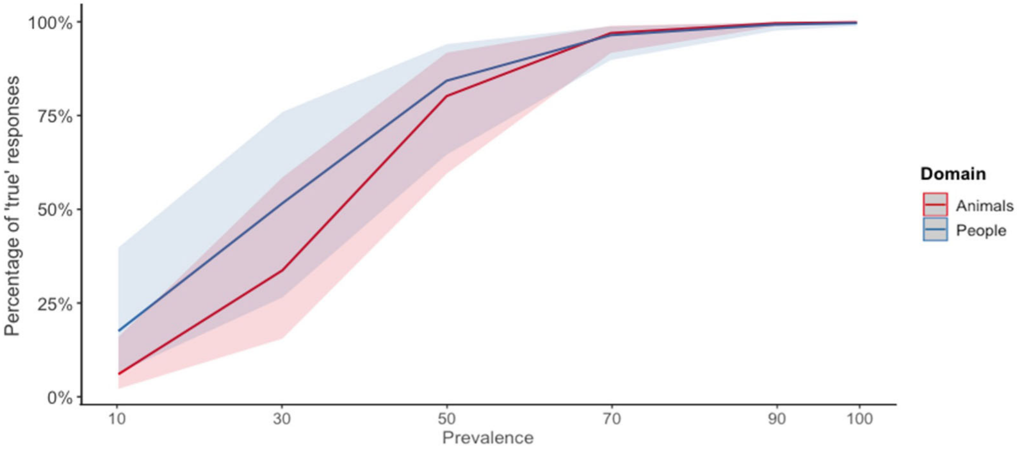

Fig. 3. Mean percentage of "true" responses in Study 2 (Truth-Conditions task) by domain and prevalence level.

generic sentences were more likely to be judged to be true for higher than lower prevalence levels. We additionally observed an interaction between domain and prevalence.

To explore the interaction (see Fig. 3), we conducted post-hoc tests that revealed that participants were numerically, but not significantly, more likely to endorse social generics than generics about animals at lower prevalence levels: 10% level (Average Marginal Effect (AME) = 0.07, SE = 0.05, p = .12, 95% CI = –0.02, 0.17), 30% level (AME = 0.08, SE = 0.08, p = .32, 95% CI = –0.07, 0.22), and 50% level (AME = 0.03, SE = 0.08, p = .71, 95% CI = –0.13, 0.19). In contrast, we observed that participants were numerically, but not significantly, more likely to endorse generics about animals than social generics at higher prevalence levels: 70% level (AME = –0.01, SE = 0.05, p = .81, 95% CI = –0.12, 0.09), 90% level (AME = –0.03, SE = 0.04, p = .45, 95% CI = –0.12, 0.05), and 100% level (AME = –0.04, SE = 0.04, p = .32, 95% CI = –0.11, 0.04). For a complete overview of the mean endorsement percentages at each prevalence level by domain and property type, see "Study 2: Mean Endorsement Truth-Conditions" in the SOM.

3.2.3. Conclusions

Overall, the results of Study 2 support the robustness and generalizability of the generic asymmetry. In particular, these results show that there is an inferential asymmetry for both social generics and generics about animals. Nonetheless, the acceptability of generics in the Truth-Conditions task was affected by a small but measurable interaction between domain and prevalence: social generics were more likely than generics about animals to be accepted at the lowest prevalence levels, whereas generics about animals were more likely than social generics to be accepted at the highest prevalence levels. The reasons why we observed this effect are unclear, as we did not observe any effect related to domain in the Implied Prevalence task. Thus, this result should be replicated before it is ascribed much significance.

Finally, we conducted a supplementary study to obtain baseline data on the homogeneity of the social and animal categories presented in Study 2, in the absence of generic information. Making use of the task developed by Nisbett et al. (1983), we told participants that three instances of a category had a property and then asked them how broadly they would generalize this property to other members of the same category (see "Study S2: Supplementary Study" in the SOM). Participants in this baseline study judged the animal categories in our study to be more homogeneous than the social categories (e.g., they judged members of an animal category to be more alike than members of a social category). This result is consistent with prior work showing that animal categories are assumed to be more homogeneous than social categories (e.g., Brandone, 2017; Nisbett et al., 1983). Moreover, this result highlights the power of generic language for fostering broad inferences. That is, participants in the Implied Prevalence task of Study 2 interpreted generic language in similar ways for social categories and animal kinds, despite expectations of greater heterogeneity for members of social kinds in the absence of generic information.

4. Study 3a

In Study 3a, we again tested for a generic asymmetry in two domains (people and animals), but this time we also varied property valence (dangerous vs. neutral). Prior theorizing and empirical work have suggested that generics expressing dangerous properties may be more likely to be judged true than generics expressing neutral properties (Bian & Cimpian, 2021; Cimpian et al., 2010; Leslie, 2008). Accordingly, we wished to examine how this would affect the asymmetry in the two domains. In the *Truth-Conditions* task, we also examined whether domain and property valence affected generics' acceptability at different prevalence levels. We used a similar methodology to Study 2. The preregistration of this study is available at https://aspredicted.org/blind.php?x=zr3sy3.

Cimpian et al. (2010) originally examined whether the inferential asymmetry and the acceptance conditions of generics about animals vary as a function of property valence. In addition to the inclusion of social generics, our methodology differs from that of Cimpian et al. (2010) in a number of important respects.

First, in Cimpian et al.'s (2010) studies, the interpretation of neutral generics was compared to dangerous properties that were either (1) both dangerous *and* distinctive (e.g., "Reesles have blue scales that secrete a strong venom that kills you on the spot. No other animals have this kind of scales") or (2) intermingled with generics that varied in distinctiveness, potentially priming participants to think about this particular factor throughout the study. For this reason, in the present study, we tested only dangerous and neutral properties. Our properties were also pretested to validate their status as dangerous or neutral.

Second, in Cimpian et al.'s (2010) studies, the neutral generic statements were substantially shorter than the other types of generic statements. For this reason, in their first study, Cimpian et al. (2010) additionally included non-distinctive control generic sentences of approximately the same length as the dangerous/distinctive sentences to examine whether simply providing more information could affect participants' evaluations. Cimpian et al. (2010) observed that dangerous/distinctive generics were more likely to be accepted than both neutral and non-distinctive generics. For this reason, they concluded that participants' responses were due to the informational value of items rather than the items' length. However, this represented only an indirect control of the impact of item length; indeed, Cimpian et al. (2010) needed to infer from the observation that dangerous/distinctive generics were more likely to be accepted than both neutral and non-distinctive generics that the length difference between dangerous/distinctive generics and neutral generics did not explain their findings. In the present study, we tested dangerous and neutral generic sentences that were roughly equivalent in length, with predicates ranging from 2 to 6 words to provide more direct control of item length.

Third, we took care to ensure that neither dangerous nor neutral items in the present study included generic-you, given that this expression conveys norms and broad generalizations (Orvell, Kross, & Gelman, 2017, 2019, 2020). This is in contrast to Cimpian et al. (2010), which included generic-you for dangerous but not neutral generic statements (e.g., "have red scales that secrete a strong venom that kills *you* on the spot"; italics added).

4.1. Method

4.1.1. Participants

Two-hundred and twenty-nine adults from the United States (93 men, 133 women, 2 undisclosed, and 1 gender-fluid; Mean age = 41.00 years; range = 24–72 years) completed the study online and were paid \$1. Participants were randomly assigned to one of four conditions: *Animals–Implied Prevalence* ($n = 54$), *People–Implied Prevalence* ($n = 66$), *Animals–Truth-Conditions* ($n = 57$), or *People–Truth-Conditions* ($n = 52$). Participants were 81% White, 6% Black or African American, 5% Multiethnic, 5% Asian or Asian American, 2% Latino or Hispanic, 0.4% Native Hawaiian or Other Pacific Islander, and 0.4% not listed. Nineteen other participants were tested and excluded from the final sample because they failed the manipulation check ($n = 2$ in the *Animals–Implied Prevalence*, $n = 2$ in the *People–Implied Prevalence*, $n = 3$ in the *Animals–Truth-Conditions*, and $n = 12$ in the *People–Truth-Conditions*, tasks). Four participants were tested and excluded from the final sample for having duplicate IP addresses. Three participants were tested and excluded from the final sample because they were not native speakers of English. One additional participant was tested and excluded for having a non-US IP address.

4.1.2. Materials and procedure

An additional 97 participants completed a norming study to allow us to select dangerous and neutral items. We presented participants with the same introductory text as in Study 2. Half of the participants ($n = 50$) were asked about animal categories and half ($n = 47$) were asked about people categories.[4] Participants were presented with 60 items: 20 intended to be dangerous (e.g., "Xs hunt strangers"), 20 intended to be neutral (e.g., "Xs sleep under trees"), and 20 intended to be safe (e.g., "Xs help others"). They were then asked to rate how dangerous the category is (e.g., "How dangerous are Xs?") on a scale of 1 (safe) to 7 (dangerous), with the midpoint (4) as neutral. Each item consisted of a novel label paired with one of the 60 properties; furthermore, the order of these items was randomized for each participant. Based on the norming study, we selected 24 items for use in the main experiment: 12 neutral items (e.g., "Xs hide underground"; M rating = 3.58, SD = .75) and 12 dangerous items (e.g., "Xs carry a deadly virus"; M rating = 6.71, SD = .47), $t(96) = 33.18$, $p < .001$, $d = 3.37$. See Appendix C for a complete list of the items. For both the dangerous and neutral properties, half of the items included descriptions of traits (e.g., "are messy," "are dangerous") and the other half included descriptions of behaviors (e.g., "sleep under trees," "hunt strangers").

Unlike the dangerous properties used by Cimpian et al. (2010), our dangerous properties had approximately the same number of words as the neutral properties. For each participant, the properties were randomly associated with one of 24 labels from Study 2 (see Appendix B for a list of the labels). The order of the items was randomized for each participant.

Participants were presented with the same introductory text as in Study 2 and then completed either the *Implied Prevalence* task or the *Truth-Conditions* task. After the task, participants were also asked to complete the same manipulation check as in Study 2 and Reynolds' (1982) social desirability measure.

4.2. Results and discussion

4.2.1. Did the generic asymmetry vary as a function of domain and/or property valence?

We conducted a repeated measures ANOVA focused on the mean prevalence ratings, with task (Implied Prevalence vs. Truth-Conditions) and domain (animals vs. people) as between-subjects factors, and property valence (dangerous vs. neutral) as a within-subjects factor (see Fig. 4). We found a main effect of task, $F(1, 225) = 51.07$, $p < .001$, $\eta_p{}^2 = .19$, a main effect of property valence, $F(1, 225) = 27.30$, $p < .001$, $\eta_p{}^2 = .11$, and an interaction between task and property valence, $F(1, 225) = 23.41$, $p < .001$, $\eta_p{}^2 = .09$.

Given this interaction, we examined the simple main effects of property valence within the *Implied Prevalence* and *Truth-Conditions* tasks separately. In the *Implied Prevalence* task, the mean prevalence implied by generic statements was higher for neutral than dangerous properties, $F(1, 225) = 52.97$, $p < .001$, $\eta_p{}^2 = .19$. In contrast, in the *Truth-Conditions* task, the mean prevalence that led participants to accept the generic statements did not differ based on property valence, $F(1, 225) = 0.07$, $p = .79$, $\eta_p{}^2 = .00$.[5]

As in Study 2, we examined whether there was a significant correlation between a participant's score on Reynolds' (1982) social desirability scale and their mean prevalence score.

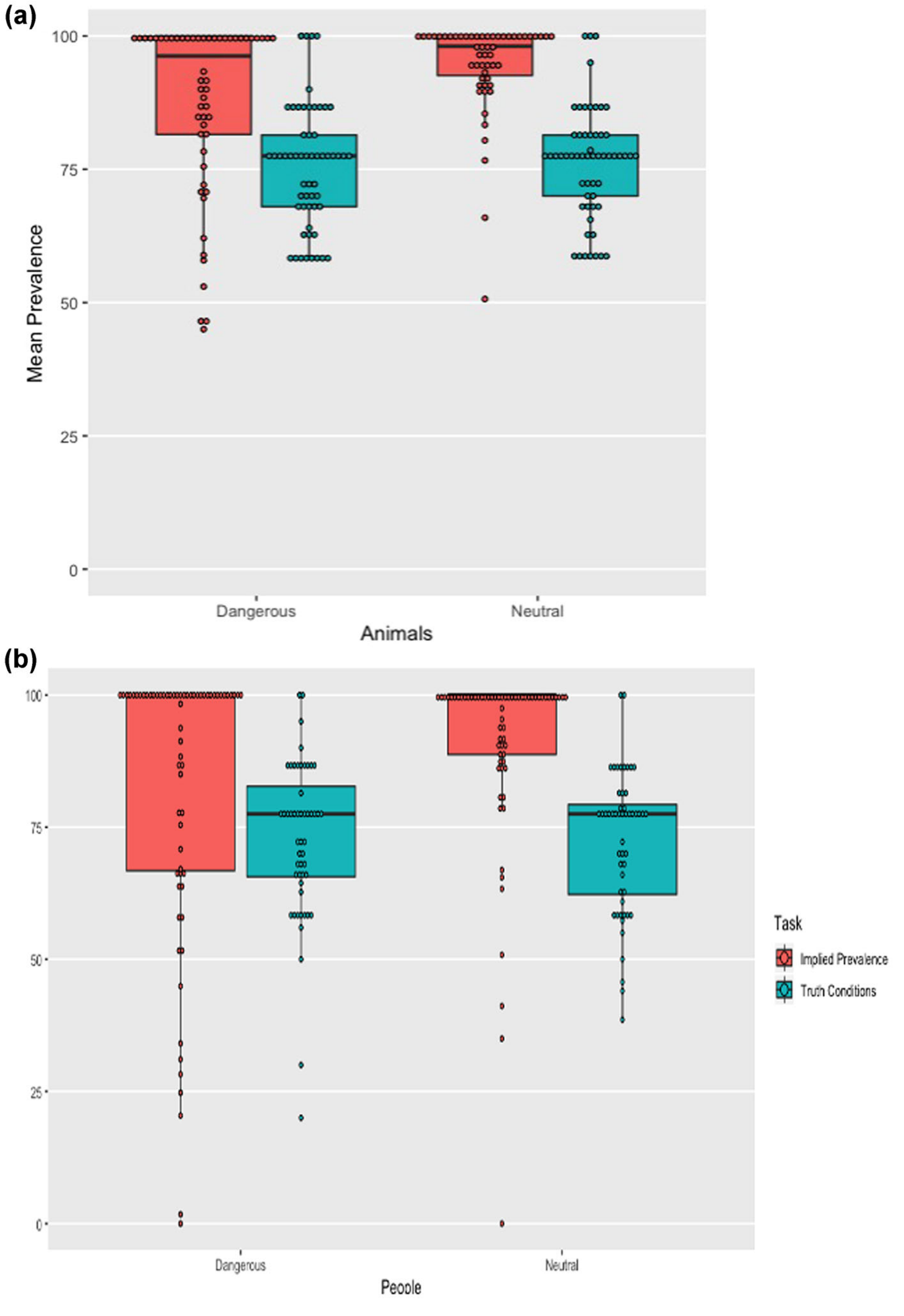

Fig. 4. Study 3a, dot plots (with box plot overlays) of mean prevalence ratings, plotted by condition (Implied Prevalence vs. Truth Conditions). Each dot represents the mean prevalence rating for a participant, which was computed by averaging responses across the 12 dangerous and 12 neutral items they rated. The solid line represents the median.

Table 3
Logistic regression predicting "true"/"false" judgments, based on domain, property valence, prevalence, and their interactions in Study 3a.

Fixed effects	**Estimate**	***SE***	***p*-value**
(Intercept)	2.25	0.31	<.001
Domain	–1.12	0.60	.06
Property valence	0.10	0.17	.55
Prevalence	10.44	0.55	<.001
Domain x Property Valence	–0.58	0.33	.08
Domain x Prevalence	–6.92	1.01	<.001
Property Valence x Prevalence	–0.21	0.59	.72
Domain x Property Valence x Prevalence	0.52	1.16	.65
Random effect		**SD**	
Participant	Intercept	2.89	

As in Study 2, we found no significant correlations in either the *People–Implied Prevalence* or *People–Truth-Conditions* tasks for both neutral and dangerous properties, *p*s > .05.

4.2.2. Did domain and property valence affect generics' acceptability at different prevalence levels?

The second set of analyses examined whether responses in the *Truth-Conditions* task differed based on prevalence level. We submitted participants' responses to a logistic mixed-effects model using the glmer command in the lme4 package in R (Bates, 2007). In this model, we included domain (animals = 0; people = 1; between-subject), property valence (neutral = 0; dangerous = 1; within-subject), prevalence (.1, .3, .5, .7, .9, 1; within-subject), and their interactions as predictors (see Table 3). All predictors were mean-centered. We also included *participant* as a random intercept. We observed a main effect of prevalence, indicating that generic sentences were more likely to be judged true for higher than lower prevalence levels. However, this main effect needs to be interpreted within the context of a significant domain by prevalence interaction (see Fig. 5).

Post-hoc tests revealed that participants were more likely to endorse social generics than generics about animals at the 10% level (Average Marginal Effect (AME) = 0.17, *SE* = 0.06, *p* = .006, 95% CI = 0.05, 0.29). At the 30% and 50% levels, there was no difference between endorsement of social generics and generics about animals (AME = 0.11, *SE* = 0.08, *p* = .16, 95% CI = –0.04, 0.26; AME = –0.07, *SE* = 0.08, *p* = .36, 95% CI = –0.22, 0.08, respectively). We observed that participants were more likely to endorse generics about animals than social generics at higher prevalence levels: 70% level (AME = –0.13, *SE* = 0.05, *p* = .01, 95% CI = –0.24, –0.03), 90% level (AME = –0.12, *SE* = 0.04, *p* = .006, 95% CI = –0.20, –0.03), and 100% level (AME = –0.09, *SE* = 0.03, *p* = .003, 95% CI = –0.15, –0.03). For a complete overview of the mean endorsement percentages at each prevalence level by domain and property valence, see "Study 3a: Mean Endorsment Truth-Conditions" in the SOM.

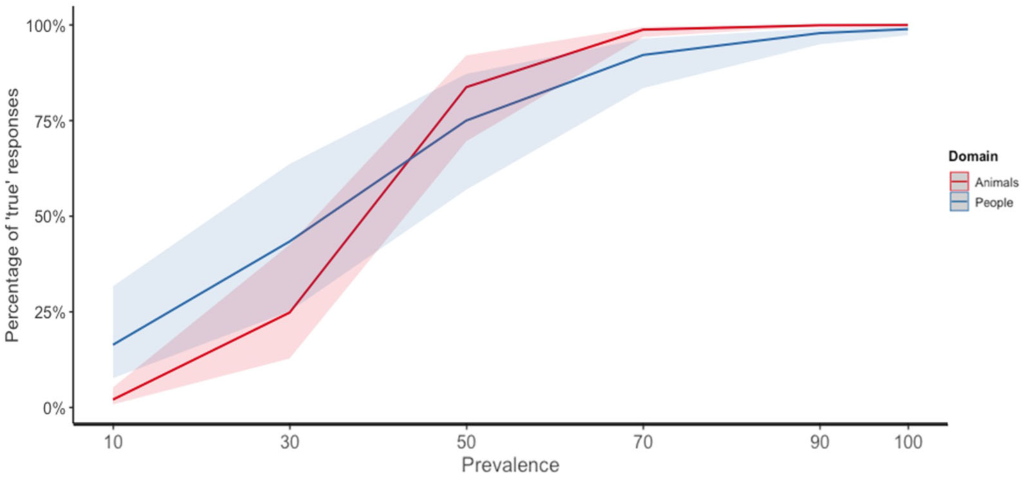

Fig. 5. Mean percentage of “true” responses in Study 3a (Truth-Conditions task) by domain and prevalence level.

4.2.3. Conclusions

The results of Study 3a provide additional support for the robustness and generalizability of the generic asymmetry. In particular, there was an inferential asymmetry for both social generics and generics about animals, and this held for both neutral and dangerous properties. This suggests that the interpretation of social generics and that of generics about animals is very similar and is consistent with previous studies that documented how generic language elicits essentialist reasoning in both domains (Foster-Hanson et al., 2016, 2019; Gelman et al., 2010; Leshin et al., 2020; Rhodes et al., 2012, 2018b).

Despite the lack of overall domain effects, we did obtain effects of valence: neutral generics yielded higher levels of implied prevalence than dangerous generics. This result differs from Cimpian et al. (2010), who reported no variation in the asymmetry as a function of property valence. A first possibility is that participants were reluctant to report that nearly all the members of a social category present cruel and harmful properties like the ones used in this study, for reasons of social desirability. However, we did not find a significant correlation between participants’ score on the social desirability scale and their responses.

A second potential contributor to this effect may be that people have developed a set of expectations about the use of generics; in particular, based on their experience with *familiar* dangerous versus novel properties, people might form the overhypothesis that generics concerning dangerous properties signal low prevalence. It has been observed that rare but dangerous properties may readily be expressed with generics (e.g., “Sharks attack swimmers”; “Mosquitoes carry the West Nile Virus”), and this may paradoxically lead to the inverse expectation that a *novel* generic about a dangerous property indicates that it is likely to be relatively rare. Importantly, this possibility is predicted by the interpretation model put forward by Tessler and Goodman (2019a; see Section 1.5). According to this model, naïve listeners interpret generics as concerning a prevalence level higher than a threshold θ, which is contextually determined by their probabilistic world knowledge. In Tessler and Goodman’s (2019a) view, the prior distribution expectations concerning dangerous properties are likely to be very low, as these properties are rarely found in the environment. For this reason, a naïve listener might assume that θ is low when interpreting dangerous generics, leading to a lower implied prevalence than that elicited by neutral generics.

This possibility is also consistent with previous evidence showing that dangerous generic properties, compared to other generic properties, were interpreted as referring to common dispositions (i.e., properties that Ks are commonly disposed to have but are displayed only under certain circumstances) rather than prevalent properties (i.e., properties actually displayed by a high proportion of Ks; Prasada et al., 2013), potentially leading to the expectation that the relevant property is less generalizable (Lazaridou-Chatzigoga et al., 2019). Thus, this seems to be the most plausible explanation for this result.

Although the inferential asymmetry elicited by social generics did not differ from that elicited by generics about animals, we replicated the interaction between domain and prevalence observed in the *Truth-Conditions* task of Study 2. Participants were more likely to accept social generics at the 10% level, whereas generics about animals were more likely to be accepted at the 70%, 90%, and 100% levels. In both Study 2 and Study 3a, the interaction between prevalence level and domain was a small effect. Nevertheless, these results indicate that people were sensitive to domain at different prevalence levels in the Truth-Conditions task. As previously noted, stereotypes operate by generalizing well beyond the evidence. For this reason, social generics might be easily accepted even based on minimal evidence, as compared to generics about animals. Animal categories, instead, may be perceived as less variable than social categories (e.g., Brandone, 2017; Nisbett et al., 1983). As a consequence, if a small percentage of members of an animal category display a certain property, people might be less willing to accept the corresponding generalization. This possibility would also explain why, conversely, participants were less conservative in accepting generics about animals at the highest prevalence levels: because animal categories are perceived as more homogeneous than social categories, people might be more willing to accept the corresponding generalizations about animals in cases where a vast proportion of category members share the ascribed properties.

As in Study 2, we also wanted to examine whether there were baseline differences in the ratings of the generalizability of properties as a function of domain. We thus conducted a supplementary study to assess this question about the property/category pairings in Study 3a in the absence of generic language (see "Study S3a: Supplementary Study" in the SOM). Participants judged the properties in Study 3a to be more homogeneous for animal than social categories (consistent with prior research, e.g., Brandone, 2017; Nisbett et al., 1983). Furthermore, in the baseline study, neutral properties were judged to be more generalizable than dangerous properties, consistent with the argument that dangerous properties would be assumed to be relatively rare. Given these domain and property valence baseline differences, one might argue that our results are not due to the asymmetry per se; rather, these baseline differences would indicate that our results are due to the different category/property pairings we used. However, we are not claiming that our data indicate that the difference between neutral and dangerous properties shows up only when testing generics. Rather, we are claiming the opposite—namely, that our data indicate that the baseline expectations about the category/property pairings being considered change the interpretation of generics too. Given that neutral properties are considerably more generalizable, it is notable that the asymmetry shows up testing both neutral and dangerous properties. The interaction between domain and property valence found in the supplementary study indicates that such baseline differences cannot explain the interaction between prevalence and domain found in the Truth-Conditions tasks of our studies. Furthermore, given that participants in the Implied Prevalence task of both Study 2 and Study 3a interpreted generic language as granting broad generalizations without domain differences despite lower baselines for social categories, these results further highlight the power of generic language to license broad inferences.

5. Study 3b

In Study 3b, we aimed to better understand an unexpected finding of Study 3a. Specifically, in contrast to Cimpian et al. (2010), we found that people rated dangerous and neutral generics alike in the *Truth-Conditions* task. This may have been due to the properties we used, which differed from those of Cimpian et al. (2010). In Study 3b, we examined this by presenting participants with the same animal items as Cimpian et al. (2010), focusing exclusively on the *Truth-Conditions* task. We also examined whether the same pattern of responses was observed in the *People* condition, using an analogous set of items for groups of people. The preregistration of this study is available at: https://aspredicted.org/blind.php?x=6kh579.

5.1. Method

5.1.1. Participants

One-hundred and eleven adults from the United States (48 men, 62 women, 1 non-binary/third gender; Mean Age = 41.01 years; range = 24–74 years) completed the study online and were paid $0.75. Participants were randomly assigned to either the *Animals–Truth-Conditions* ($n = 58$) or the *People–Truth-Conditions* ($n = 53$) tasks. Participants were 80% White, 8% Black or African American, 6% Multiethnic, 3% Asian or Asian American, and 3% Latino or Hispanic. Seven participants were tested and excluded from the final sample because they failed the same manipulation check used in the previous studies ($n = 4$ in the *People–Truth-Conditions task* and $n = 3$ in the *Animals–Truth-Conditions*). One participant was tested and excluded from the final sample for having a non-US IP address. Finally, one participant was tested and excluded from the final sample for not being a native speaker of English.

5.1.2. Materials and procedure

As in Study 3a, we created a list of 24 items with 12 neutral properties (e.g., “Xs have silver fur”) and 12 dangerous properties (e.g., “Xs have dangerous silver fur. This fur sheds particles that get lodged in your lungs and make it impossible to breathe”). See Appendix D for a complete list of the items. For the animal items, we used the 10 neutral and dangerous properties from Cimpian et al. (2010), and added two additional properties for each type of

Table 4
Logistic regression predicting "true"/"false" judgments, based on domain, property valence, prevalence, and their interactions in Study 3b.

Fixed effects	**Estimate**	***SE***	***p*-value**
(Intercept)	5.26	0.64	<.001
Domain	2.83	1.12	.01
Property valence	0.32	0.20	.12
Prevalence	10.67	0.66	<.001
Domain x Property Valence	0.12	0.41	.77
Domain x Prevalence	0.92	1.20	.45
Property Valence x Prevalence	–0.95	0.61	.12
Domain x Property Valence x Prevalence	1.31	1.23	.29
Random effect		**SD**	
Participant	Intercept	4.96	

property. (This was done to add two items at the 100% prevalence level, which Cimpian et al. (2010) did not include; see Study 1.) For the people items, we were concerned that ascribing the same properties to groups of people would have resulted in the perception of these groups as non-human. For this reason, we decided to describe objects that these groups of people use, rather than describing body parts (e.g., "Xs use dangerous silver sprayers. These sprayers shed particles that get lodged in your lungs and make it impossible to breathe"). In both conditions, for each participant, the neutral and dangerous properties were randomly assigned to the 24 labels used in Experiment 2 (see Appendix B for a list of the labels). The order of the items was randomized for each participant.

Participants read the same introductory text as in Studies 2 and 3a. All participants in this study completed the *Truth-Conditions* task. After the task, participants were also asked to complete the same manipulation check as in Studies 2 and 3a. However, we did not include Reynolds' (1982) measure of social desirability in this study, due to the lack of significant correlations observed in previous studies.

5.2. Results and discussion

As in the *Truth-Conditions* of Study 3a, we submitted participants' responses to a logistic mixed-effects model.[6] In this model, we included domain (animals = 0; people = 1; between-subject); property valence (neutral = 0; dangerous = 1; within-subject), prevalence (.1, .3, .5, .7, .9, 1; within-subject), and their interactions as predictors (see Table 4). All predictors were mean-centered. We also included *participant* as a random intercept. We observed a main effect of domain, indicating that participants were more likely to judge social generics than generics about animals to be true. We also observed a main effect of prevalence. For a complete overview of the mean endorsement percentages at each prevalence level by domain and property valence, see "Study 3b: Mean Endorsment Truth-Conditions" in the SOM.

The results of Study 3b show that participants endorsed dangerous and neutral generics at equivalent rates. Moreover, although overall the proportion of "true" responses was higher for social generics than for generics about animals, this time we did not observe an interaction between domain and prevalence level. On the one hand, these findings are consistent with those of Study 3a, where we observed that property valence did not affect participants' responses in the Truth-Conditions task. On the other hand, these findings differ from those of Study 2 and Study 3a, where we observed no domain effects and a small interaction between prevalence level and domain. This discrepancy is likely due to differences in the items between these studies.

Together with the results of Study 3a, these findings suggest that danger alone might not be sufficient to make dangerous generic predications easier to accept; something else might be required to observe this effect. Differently from our studies, Cimpian et al. (2010) compared the interpretation of neutral generics to dangerous generic predications that were either (1) both dangerous and distinctive or (2) intermingled with generics that varied in distinctiveness. This context might have primed participants to think about distinctiveness throughout the experiment. If so, then dangerous generic predications may need to be perceived as distinctive to be accepted more easily than neutral generic predications. Due to similar concerns, Bian and Cimpian (2021) compared items that were either dangerous + distinctive, or non-dangerous + distinctive (thus controlling for distinctiveness), and found that generics expressing dangerous properties were more likely to be accepted than those expressing non-dangerous properties. However, this result is still consistent with the possibility that it is not danger per se that is operative here, but rather danger combined with distinctiveness. For this reason, further research is needed to test this hypothesis.

As in Studies 2 and 3a, we conducted a supplementary study to assess the baseline generalizability of the property/category pairings in Study 3b in the absence of generic language (see "Study S3b: Supplementary Study" in the SOM). As in the prior studies, participants judged these properties to be more generalizable in animal than social categories, and judged neutral properties to be more generalizable than dangerous properties in the People condition. This supplementary study replicates the previously observed baseline domain differences in the absence of generic language.

6. General discussion

The main goal of the present studies was to investigate the robustness of the inferential asymmetry of generics—that is, the discrepancy between people's judgments about the prevalence of a property expressed in a generic statement and the prevalence of a property among members of a category that led them to accept the sentence as being true. Specifically, we wanted to test whether the asymmetry observed for generics about animal categories extended to the interpretation of social generics. We were also interested in investigating whether the asymmetry varied as a function of property valence (dangerous vs. neutral).

Overall, our studies demonstrated that the asymmetry between generics' prevalence implications and acceptance conditions is robust: the inferential asymmetry elicited by generics about animal categories was also observed for social generics and held for properties varying in valence. However, property valence affected generics' prevalence implications.

Specifically, neutral generic predications yielded higher prevalence estimates than dangerous generic predications. In the Truth-Conditions task of both Study 2 and Study 3a, we also observed an interaction between domain and prevalence level. Specifically, social generics were more likely to be accepted than generics about animals at the lowest prevalence levels. Generics about animals, instead, were more likely to be accepted than social generics at the highest prevalence levels. Below, we discuss these results in more detail.

6.1. The robustness and generalizability of the generic asymmetry

Overall, our findings provide further confirmation that the inferential asymmetry found by Cimpian et al. (2010) is robust and central to the interpretation of generics. Specifically, Study 1 showed that the generic asymmetry replicates with a larger online sample and the inclusion of the 100% level while testing generics' acceptability. Importantly, the results of Studies 2 and 3a show that the inferential asymmetry elicited by generics about animals does not differ from that elicited by social generics.

This result was not a foregone conclusion; as previously discussed, we could have found a different generic asymmetry across domains. On the one hand, given that social categories may be perceived as more variable than animal categories (e.g., Brandone, 2017; Nisbett et al., 1983; as well as our own baseline data), we could have observed less asymmetry for social generics, or even no asymmetry at all. On the other hand, given that stereotypes are often accepted based on minimal evidence (e.g., Allport, 1954), we could have observed (all other things being equal) a greater asymmetry for social generics. On the contrary, our results suggest that the interpretation of social generics and the interpretation of generics about animals are very similar. This finding is consistent with previous evidence showing that people often reason about social categories as if they were natural kinds (Prentice & Miller, 2007; Rhodes & Gelman, 2009; Rothbart & Taylor, 1992) and that generic language promotes essentialist reasoning in both domains (Foster-Hanson et al., 2016, 2019; Gelman et al., 2010; Leshin et al., 2020; Rhodes et al., 2012, 2018a, 2018b).

Our finding that this asymmetry extends to the interpretation of social generics suggests that this way of interpreting generics is domain-general. To this end, investigating whether the generic asymmetry also extends to the interpretation of generics about artifacts and in languages other than English is an important direction for future research. By testing whether this asymmetry arises in multiple domains and languages, we would be able to better understand the extent to which this phenomenon depends on fundamental cognitive processes rather than domain-specific representations and linguistic variation.

In our studies, we observed the generic asymmetry across all of our manipulations. Our data have thus not established boundary conditions on this effect. However, one known factor that limits the generic asymmetry is the use of accidental/temporary properties, such as "have muddy feathers" and "have broken legs" (Cimpian et al., 2010), consistent with other work showing that generics are not used to express this type of property (e.g., Cimpian & Markman, 2008; Gelman, 1988; Sutherland, Cimpian, Leslie, & Gelman, 2015; Tessler & Goodman, 2019b). Identifying other factors that limit the generic asymmetry remains a question for future research.

6.2. The effect of domain on generics' acceptability

Although the inferential asymmetry did not differ by domain, we found that people's endorsement of generics in the Truth-Conditions task differed for animal versus social categories at different prevalence levels. Specifically, in Studies 2 and 3a, we observed that social generics led to a higher proportion of "true" responses than generics about animals at the lowest prevalence levels; generics about animals, instead, led to a higher proportion of "true" responses than social generics at the highest prevalence levels. We hypothesize that because stereotypes generalize beyond the available evidence (e.g., Allport, 1954; Wodak et al., 2015), social generics might be easily accepted even at low prevalence levels. Conversely, because animal categories are often perceived as more homogeneous than social categories (e.g., Brandone, 2017; Nisbett et al., 1983; our own baseline data), people might be less willing to accept the corresponding generalizations when only a few members of an animal category display the same properties. Furthermore, because animal categories are often perceived as more homogenous (again, see our baseline data), people might be more willing to accept generics about animals when high proportions of category members share the ascribed properties.

This interaction was not observed in Study 3b; in that study, instead, we found that overall social generics were more likely to be accepted than generics about animals. We hypothesize that this discrepancy is due to differences in the items across these studies. Whereas in Study 2 and Study 3a we tested the same properties across domains, the design of Study 3b required that we test different properties for people and animals.

6.3. The effect of property valence on the interpretation of generics

The results of Study 3a show that the generic asymmetry held for different property valences (dangerous vs. neutral). These results provide an additional demonstration of the robustness and centrality of the generic asymmetry. However, we also observed that neutral generic properties yielded higher prevalence estimates than dangerous generic properties in the Implied Prevalence task. We hypothesize that this difference was due to specific expectations that people develop about dangerous generic properties. Dangerous generics tend to be readily accepted despite the rarity of the relevant property (e.g., "Sharks attack swimmers"); in turn, this might have led to the inverse expectation that novel dangerous generic properties have a lower implied prevalence than other novel generic properties. This possibility is compatible with the "*interpretation* model" proposed by Tessler and Goodman (2019a), which predicts that prior knowledge about dangerous properties might lead naïve listeners to assume a low implied prevalence for generics ascribing this type of property. Consistent with this hypothesis, previous studies indicate that dangerous generic properties are expected to refer to common dispositions rather than widespread properties, leading children and adults to interpret dangerous properties as less generalizable than other types of properties (Lazaridou-Chatzigoga et al., 2019; Prasada et al., 2013).

Contrary to Leslie's hypothesis (2008) and the findings of Cimpian et al. (2010) and Bian and Cimpian (2021), we did not find any consistent tendency to endorse dangerous generics at a higher rate than neutral generics. In Study 3a, we tested items that were pre-validated to test their status as dangerous or neutral, had approximately the same number of words, and did not include generic-you expressions. In Study 3b, we tested the same neutral and dangerous animal items as Cimpian et al. (2010) had used, together with an analogous set of people items. Of course, the present findings do not refute Leslie's (2008) hypothesis; more systematic and direct evidence would be needed to elucidate the relation between property valence and generic endorsement. Nonetheless, based on our results, we hypothesize that danger might not be sufficient per se to observe the effect theorized by Leslie (2008) and found by Cimpian et al. (2010) and Bian and Cimpian (2021), and instead distinctiveness might be required to increase the informational value of dangerous generic predications. Recall, for example, that Cimpian et al. (2010) compared the interpretation of neutral generics to dangerous generic predications that were either (1) both dangerous and distinctive or (2) intermingled with generics that varied in distinctiveness, potentially priming participants to think about *distinctiveness* throughout. Exploring this possibility is a potential direction for future research.

Finally, the lack of interaction between property valence and domain in the Truth-Conditions task partly supports and partly contrasts with Tasimi et al.'s (2017) results. On the one hand, we replicated Tasimi et al.'s (2017) finding that, in the social domain, threatening generics (e.g., "Xs are dangerous") and non-threatening generics (e.g., "Xs are helpful") were accepted equally often. On the other hand, Tasimi et al. (2017) also observed that for artifacts and animals, threatening generics were accepted more often than non-threatening generics. These differences may be due to variations in the stimuli (e.g., Tasimi et al. employed positive rather than neutral properties), but more research is needed to determine whether this is the case. A wider comparison between the interpretation of dangerous, distinctive, neutral, and prosocial generic predications in different domains is another important goal for future research.

6.4. Implications for the semantics of generics

The observation that generics about both animals and social categories are interpreted as broadly true, yet also accepted based on weak evidence, provides additional support for the view that the semantics of these generalizations is not simply based on statistical prevalence (e.g., Abelson & Kanouse, 1966; Brandone et al., 2015; Carlson, 1977; Cimpian et al., 2010; Gelman et al., 2002; Krifka et al., 1995; Leslie, 2008). On the contrary, these results further reinforce the view that the interpretation of generics is based on the interaction between our conceptual knowledge and the valence of the property being ascribed (e.g., Cimpian et al., 2010; Leslie, 2008; Prasada et al., 2013). Indeed, the results of Study 3a indicate that the prevalence estimates based on neutral generic predications were higher than those based on dangerous generic predications. Furthermore, in both Study 2 and Study 3a, we observed that domain affected generics' acceptability at different prevalence levels.

These results additionally inform the theories that have been put forward to explain the possible causes of the generic asymmetry. van Rooij and Schulz (2020) predict that dangerous generics should be accepted more easily and elicit higher prevalence estimates compared to neutral generics, as the representativeness of dangerous properties is higher than that of neutral properties. However, our findings do not support this prediction, as we found no difference in the generics' acceptance conditions linked to property valence. Furthermore, whereas Cimpian et al. (2010) reported no variation in prevalence estimations as a function of property valence, we found that neutral generics yielded higher levels of implied prevalence than dangerous generics, consistent with previous evidence (Lazaridou-Chatzigoga et al., 2019; Prasada et al., 2013).

In contrast, according to Tessler and Goodman (2019a), generics might elicit an inferential asymmetry due to the interaction between *interpretation* and *endorsement*. Given that listeners often interpret generics as referring to nearly all the relevant category members, due to the typically high θ, speakers may assume that listeners' posterior prevalence estimates are more consistent with the actual prevalence than the prior prevalence estimates, which are typically very low. Consequently, speakers might endorse generics even based on low actual prevalence levels. Nonetheless, a complication arises in adopting Tessler and Goodman's (2019a) account. According to the interpretation model, θ is contextually determined by the listeners' prior distribution expectations. When θ is assumed to be low, the corresponding generics lead to low prevalence estimates. On the contrary, when θ is high, the corresponding generics lead to high prevalence estimates. Our supplementary studies concerning participants' baseline expectations consistently show that their prior beliefs about prevalence were significantly higher for animal categories. Consequently, following the interpretation model, the different baseline expectations about domain should have affected generics' implied prevalence across domains. However, the present studies show that the prevalence estimates based on generics did not differ for animal versus social categories. As a result, our studies are novel in providing evidence against an account of the generic asymmetry that is too closely tied to assumptions regarding priors related to the distribution of properties within a category.

6.5. *Implications for stereotyping*

Our finding that the inferential asymmetry extends to the interpretation of social generics also has important implications for stereotyping. Our results demonstrate that these statements are assumed to be broadly representative of the category despite being accepted even at low prevalence levels, thus potentially perpetuating stereotypes. Consequently, they could be particularly pernicious and misleading in communication. Indeed, social generics are a common means to express stereotypes (Gelman et al., 2004), and the current results help explain a way in which they might contribute to the transmission and flourishing of unfounded generalizations. Indeed, several scholars have assumed that the generic asymmetry concerning animal categories would extend to the interpretation of social generics. Based on this assumption, they argued that social generics might profoundly affect our social cognition. For example, both Cimpian et al. (2010) and Brandone et al. (2015) suggest that generics like "Girls are bad at math" would be easily accepted based on little evidence and despite the existence of substantial counterevidence; however, once accepted, such generics may be interpreted as referring to nearly all the category members. Consequently, some scholars have argued that generics may strongly impact perceptions, beliefs, and behavior (see also Bian & Cimpian, 2017; Hammond & Cimpian, 2017; Rosola & Cella, 2020).

Importantly, our studies are the first to investigate whether the generic asymmetry actually extends to the interpretation of social generics. Our studies showed that the inferential asymmetry elicited by generics about animal categories does not differ from the one elicited by social generics. Overall, our findings increase the validity of previous work built on Cimpian et al.'s (2010) results, together with our understanding of the relation between generics and social cognition. However, it is an open question whether the potential negative consequences for stereotyping would hold equally for familiar as for novel social categories.

Finally, the present results complicate the debate concerning whether social generics are more readily accepted when ascribing dangerous properties than neutral properties. Based on Cimpian et al.'s (2010) original findings, several authors (e.g., Langton et al., 2012; Leslie, 2017) assumed that social generics ascribing dangerous properties (e.g., "Italians are mobsters") are especially pernicious because they are accepted more easily than neutral social generics. For example, Leslie (2017) argues that "just as it takes but a few instances of sharks attacking bathers [...] for us to make the corresponding category-wide generalization, so also a strikingly negative action of a few members of a racial, ethnic, or religious minority may lead others to form a general belief concerning their entire group" (p. 399). However, although our studies demonstrated the robustness and generalizability of the inferential asymmetry for social generics, it remains unclear under what circumstances dangerous generic properties about social categories are more easily accepted than others. These possibilities are worthy of future investigation.

7. Conclusions

Findings from four preregistered studies showed that people's interpretations of generics robustly demonstrate an inferential asymmetry, in both the animal and social domains. Furthermore, our results show that the asymmetry held for different property valences (dangerous vs. neutral). Finally, these results show that generics' implied prevalence is affected by property valence, whereas their acceptance conditions are affected by domain at different prevalence levels. Overall, our findings provide further support for the view that generics are a linguistic outlet of our conceptual knowledge and that they are a powerful means to convey information about animal and social categories in the world.

Open Research Badges

This article has earned Open Data and Preregistered badges. Data are available at https://osf.io/ru9t7/ and Preregistered Study 1: preregistration available at https://aspredicted.org/blind.php?x=mw5pn8, Study 2: preregistration available at https://aspredicted.org/blind.php?x=ru2pa8, Study 3a: preregistration available at https://aspredicted.org/blind.php?x=zr3sy3, and Study 3b: preregistration available at https://aspredicted.org/blind.php?x=6kh579.

Acknowledgments

This research was supported by funds from Vita-Salute San Raffaele University to Cella and from the University of Michigan to Gelman. Between 2018 and 2019, this research was presented in its early stages at the Department of Psychology and the Department of Philosophy at the University of Michigan, and at the Faculty of Philosophy of Vita-Salute San Raffaele University. Earlier versions of this manuscript were presented at the XXVIII Italian National Conference of the Ph.D. Programs in Philosophy, at the 16th Annual Conference of the Italian Association of Cognitive Science, at the 2nd European Experimental Philosophy Conference, and at the 2022 Mid-Term Conference of the Italian Society for Analytic Philosophy. We would like to thank Andrei Cimpian and Marjorie Rhodes, who were external reviewers for Cella's Ph.D. dissertation, which includes the main studies presented here. We would also like to thank the committee members of Cella's Ph.D. viva for the helpful discussion: Amanda Brandone, Bianca Cepollaro, and Jennifer Saul. Finally, we would like to thank four anonymous reviewers for their comments on previous drafts of the manuscript; Filippo Domaneschi, Sarah Probst, and Martina Rosola for their insightful suggestions; and Nicole Cuneo, Zach Schudson, and the U-M Conceptual Development Lab team for their assistance in collecting the data.

Notes

1 In the preregistration of subsequent studies, we reported that 175 participants took part in Study 1. There are two reasons for the discrepancy between the number of participants reported here and in the preregistrations. First, 59 participants completed a Truth-Conditions task that matched the original design of Cimpian, Brandone, and Gelman (2010) (i.e., without a 100% prevalence level). We report the results of that task in the Supplementary Online Materials (see "Study S1"). Second, we initially believed that only one participant in this study had a non-US IP address. However, we realized after completing the preregistration that four additional participants in Study 1 had a non-US IP address.

2 One participant had completed the norming study that we conducted to select the items for Study 3. (The norming study was the first study that we conducted.) The other participant had already participated in Study 1.

3 We additionally fit a model including item as a random intercept; however, we found that the estimate for the SD of the intercept for item was zero, so we omitted item as a random intercept in the final model in this and subsequent studies.

4 In initial analyses, we explored whether there were domain differences in ratings. However, we observed no significant impact of domain; thus, we collapsed across this factor in subsequent analyses of data from the norming study.

5 Although the length differences in our items were minimal, we also conducted a supplementary analysis by dropping the two shortest dangerous behavior items and the two longest neutral behavior items. We replicate our results in these analyses when we control for predicate length (see "Study 3a: Length Differences" in the SOM).

6 In the preregistration for this study, we indicated that we would use ANOVAs to analyze the data. However, because the data are binary, it was necessary to analyze the data using nonparametric statistics; we thus opted for logistic regression.

References

Abelson, R. P., & Kanouse, D. E. (1966). Subjective acceptance of verbal generalizations. In S., Feldman (Ed.), *Cognitive consistency: Motivational antecedents and behavioral consequents* (pp. 171–197). New York: Academic Press.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Bates, D. (2007). *Linear mixed model implementation in lme4*. University of Wisconsin–Madison.

Bian, L., & Cimpian, A. (2017). Are stereotypes accurate? A perspective from the cognitive science of concepts. *Behavioral and Brain Sciences*, *40*, 22–24.

Bian, L., & Cimpian, A. (2021). Generics about categories and generics about individuals: Same phenomenon or different? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(11), 1836–1855.

Brandone, A. C. (2017). Changes in beliefs about category homogeneity and variability across childhood. *Child Development*, *88*(3), 846–866.

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

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

Carlson, G. N. (1977). *Reference to kinds in English* [Doctoral dissertation]. New York: University of Massachusetts, Amherst. Published in 1980 by Garland.

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., & Markman, E. M. (2008). Preschool children's use of cues to generic meaning. *Cognition*, *107*(1), 19–53.

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

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. T. Fiske., & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. *2*, pp. 357–411). New York: McGraw-Hill.

Foster-Hanson, E., Leslie, S. J., & Rhodes, M. (2016). How does generic language elicit essentialist beliefs? In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1541–1546). Cognitive Science Society.

Foster-Hanson, E., Leslie, S. J., & Rhodes, M. (2019). Speaking of kinds: How generic language shapes the development of category representations. PsyArXiv. doi: 10.31234/osf.io/28qf7

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*. Oxford, England: Oxford University Press.

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

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

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother–child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*(1), 1–142.

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*(3), 273–301.

Hammond, M. D., & Cimpian, A. (2017). Investigating the cognitive structure of stereotypes: Generic beliefs about groups predict social judgments better than statistical beliefs. *Journal of Experimental Psychology: General*, *146*(5), 607–614.

Krifka, M., Pelletier, F. J., Carlson, G. N., Meulen Ter, A., Chierchia, G., & Link, G. (1995). Genericity: An introduction. In G. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 1–124). Chicago, IL: Chicago University Press.

Langton, R., Haslanger, S., & Anderson, L. (2012). Language and race. In G. Russell & D. Graff Fara (Eds.), *The Routledge companion to the philosophy of language* (pp. 753–767). London: Routledge.

Lazaridou-Chatzigoga, D., Katsos, N., & Stockall, L. (2019). Generalising about striking properties: Do glippets love to play with fire? *Frontiers in Psychology*, *10*, 1–14.

Leshin, R., 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. PsyArXiv. doi: 10.31234/osf.io/nb6ys

Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, *117*(1), 1–47.

Leslie, S. J. (2017). The original sin of cognition: Fear, prejudice, and generalization. *Journal of Philosophy*, *114*(8), 393–421.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*(4), 339–363.

Orvell, A., Kross, E., & Gelman, S. A. (2017). How "you" makes meaning. *Science*, *355*, 1299–1302.

Orvell, A., Kross, E., & Gelman, S. A. (2019). "You" and "I" in a foreign land: The persuasive force of generic-you. *Journal of Experimental Social Psychology*, *85*, 1–9.

Orvell, A., Kross, E., & Gelman, S. A. (2020). "You" speaks to me: Effects of generic-you in creating resonance between people and ideas. *Proceedings of the National Academy of Sciences*, *117*(49), 31038–31045.

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

Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, *16*(4), 202–206.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe–Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*, 119–125.

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), 244–274.

Rhodes, M., Leslie, S. J., Bianchi, L., & Chalik, L. (2018a). The role of generic language in the early development of social categorization. *Child Development*, *89*(1), 148–155.

Rhodes, M., Leslie, S. J., Saunders, K., Dunham, Y., & Cimpian, A. (2018b). How does social essentialism affect the development of inter-group relations? *Developmental Science*, *21*, 1–15.

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.

Rosola, M., & Cella, F. (2020). Generics and epistemic injustice. *Ethical Theory and Moral Practice*, *23*(5), 739–754.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In: G. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). London: Sage.

Segall, G., Birnbaum, D., Deeb, I., & Diesendruck, G. (2015). The intergenerational transmission of ethnic essentialism: How parents talk counts the most. *Developmental Science*, *18*(4), 543–555.

Sutherland, S. L., & Cimpian, A. (2015). Children show heightened knew-it-all-along errors when learning new facts about kinds: Evidence for the power of kind representations in children's thinking. *Developmental Psychology*, *51*(8), 1115–1130.

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.

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., & Goodman, N. (2019a). The language of generalization. *Psychological Review*, *126*(3), 395–436.

Tessler, M., & Goodman, N. (2019b). Learning from generic language. *PsyArXiv*. doi: Retrieved from https://psyarxiv.com/hnm8p

van Rooij, R., & Schulz, K. (2020). Generics and typicality: A bounded rationality approach. *Linguistics and Philosophy*, *43*(1), 83–117.

Wodak, D., Leslie, S. J., & Rhodes, M. (2015). What a loaded generalization: Generics and social cognition. *Philosophy Compass*, *10*(9), 625–635.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A

Items used in Study 1

Items from Cimpian et al. (2010)	MORSETHS have silver fur BLINS have red scales ZORBS have orange tails DAITHS have gold stripes MOXES have green shells LUDINOS have yellow legs ELLEPS have pink ears LORCHES have purple feathers GLIPPETS have copper spots THUPS have blue teeth
New items	FRAMS have gray antennae JOTES have brown wings

Appendix B

Labels and items used in Study 2

Labels		Physical properties	Nonphysical properties
MORSETHS	HUXANS	have silver hair/fur	sleep under trees
FRAMS	BEMES	have gray skin	stamp their feet to greet others
LUZAKS	HEABS	have twelve toes	bury their leftover food
JOTES	NOJAS	have red eyes	live underground
KERNS	OGETS	have brown tongues	run in a zigzag pattern
NARES	TEMBAS	have long eyelashes	swim when it is raining
STADES	UPOS	have pointy ears	are nocturnal
TARBS	WHEZAS	have hairy noses	eat insects
ZAVS	WUPTAS	have yellow freckles	dance to communicate
ANEDS	YOXAS	have an extra row of teeth	gather leaves for warmth
EDERS	TOABOS	have large tonsils	are nomadic
GEZOS	VIRDEXES	have short bones	hang upside down

Appendix C

Items used in Study 3a

Neutral properties	Dangerous properties
sleep under trees	carry a deadly virus
stamp their feet to greet others	kill for fun
leave their leftovers on the ground	attack people
hide underground	kidnap babies
run in zigzag patterns	assault for no reason
swim when it is raining	hunt strangers
are nocturnal	are brutal
are short	are dangerous
are messy	are ruthless
are skillful	are bloodthirsty
are nomadic	are violent
are hairy	are vicious

Appendix D

Items used in Study 3b

Neutral properties	Dangerous properties
Animals condition	
have silver fur	have dangerous silver fur. This fur sheds particles that get lodged in your lungs and make it impossible to breathe.
have red scales	have dangerous red scales. These scales secrete a strong venom that kills you on the spot.
have orange tails	have dangerous orange tails. These tails are so long and muscular that they can suffocate you in a matter of minutes.
have gold stripes	have dangerous gold stripes. These stripes deliver a powerful electric shock that's deadly to anyone within a few feet.
have green shells	have dangerous green shells. These shells are so very heavy that they would immediately crush your bones.
have yellow legs	have dangerous yellow legs. These legs are so powerful that a single blow could kill you.
have pink ears	have dangerous pink ears. These ears are home to dangerous parasites that can make you go deaf.
have purple feathers	have dangerous purple feathers. These feathers are as sharp as needles and can easily get lodged in you, causing massive bleeding.
have copper spots	have dangerous copper spots. These spots are home to a contagious fungus that is deadly to anyone who becomes infected with it.
have blue teeth	have dangerous blue teeth. These teeth are razor-sharp and so powerful that a single bite can be lethal.
have gray antennae	have dangerous gray antennae. These antennae are covered with deadly bacteria that attack the body immediately.
have brown wings	have dangerous brown wings. These wings have sharp and long spikes that can easily tear you apart.
People condition	
use silver sprayers	use dangerous silver sprayers. These sprayers shed particles that get lodged in your lungs and make it impossible to breathe.
brew red drinks	brew dangerous red drinks. These drinks are made with a strong venom that kills you on the spot.
braid orange ropes	braid dangerous orange ropes. These ropes are so thick and strong that they can suffocate you in a matter of minutes.
wear gold wristbands	wear dangerous gold wristbands. These wristbands deliver a powerful electric shock that's deadly to anyone within a few feet.
throw green sticks	throw dangerous green sticks. These sticks are so very heavy that they would immediately crush your bones.
wear yellow shoes	wear dangerous yellow shoes. These shoes contain a hidden knife so sharp that a single kick could kill you.
prepare pink ointment	prepare dangerous pink ointment. This ointment contains dangerous parasites that can make you go deaf.

(*Continued*)

Neutral properties	Dangerous properties
toss purple stones	toss dangerous purple stones. These stones are as sharp as needles and can easily get lodged in you, causing massive bleeding.
use copper dyes	use dangerous copper dyes. These dyes contain a contagious fungus that is deadly to anyone who becomes infected with it.
have blue tools	have dangerous blue tools. These tools are razor-sharp and so powerful that a single stab can be lethal.
build gray fences	build dangerous gray fences. These fences are covered with deadly bacteria that attack the body immediately.
have brown gloves	have dangerous brown gloves. These gloves have sharp and long spikes that can easily tear you apart.