Title
Asymmetry Effects in Generic and Quantified Generalizations

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Reuter, Kevin
Neufeld, Eleonore
Del Pinal, Guillermo

Publication Date
2023

Peer reviewed

Asymmetry Effects in Generic and Quantified Generalizations

Kevin Reuter (kevin.reuter@uzh.ch)
Department of Philosophy, Zürichbergstrasse 43
8044 Zürich, Switzerland

Eleonore Neufeld (eneufeld@umass.edu)
Department of Philosophy, University of Massachusetts Amherst
Amherst, MA 01003-9274, United States

Guillermo Del Pinal (gdelpinal@umass.edu)
Department of Philosophy, University of Massachusetts Amherst
Amherst, MA 01003-9274, United States

Abstract

Generic statements ('Tigers have stripes') are pervasive and early-emerging modes of generalization with a distinctive linguistic profile. Previous experimental work found that generics display a unique asymmetry between their acceptance conditions and the implications that are typically drawn from them. This paper presents evidence against the hypothesis that only generics display an asymmetry. Correcting for limitations of previous designs, we found a generalized asymmetry effect across generics, various kinds of explicitly quantified statements ('most', 'some', 'typically', 'usually'), and variations in types of predicated properties (striking vs. neutral). We discuss implications of these results for our understanding of the source of asymmetry effects and whether and in which ways these effects might introduce biased beliefs into social networks.

Keywords: generics; quantifiers; stereotyping; language understanding; asymmetry effect

Introduction

Human natural languages have many vehicles for expressing generalizations, including a rich array of different kinds of quantifiers, adverbs, and modal auxiliaries, each of which can vary in force and degree of precision:

(1) a. Every shark is dangerous.
b. Some/most sharks are dangerous.
c. Usually sharks are dangerous.
d. Exactly twenty sharks are dangerous.

Natural languages can also express generalizations via generic sentences in which the quantificational force is contributed by an unpronounced operator, standardly labelled *Gen*:

(2) Sharks are dangerous.

There is evidence that generics are acquired earlier in development than overtly quantified sentences, and are used as defaults to express and store generalizations, even by adults (Gelman, 2004; Leslie & Gelman, 2012). Despite its psychological primacy, the truth-conditional effect of *Gen* seems to be very complicated, and according to most linguists, is not obviously reducible to any overt quantifier, modal, or adverbial operator (Krifka et al., 1995; Nickel, 2016; van Rooij, 2017).

Previous research has also linked generics to social generalizations, biases, and stereotyping. An important and influential hypothesis, versions of which have frequently been defended in the literature on generics (Cimpian et al., 2010; Goldfarb et al., 2017; Hammond & Cimpian, 2017; Khemlani et al., 2012; Leslie, 2007, 2017; Wodak et al., 2015), is that generic sentences and thought possess a potent capacity to give rise to social biases. This potential is supposed to be unique to generics like (2). Similar sentences and thoughts with quantifiers and adverbs, such as (1-b) and (1-c), do not have those problematic effects, at least to the same extent.

How do generic sentences and thoughts contribute to the creation of biased (social) generalizations? Various influential hypotheses have been advanced. One is that generics, compared to overtly quantified generalizations, lead to a greater degree of essentialization of the subject kinds or groups (Cimpian & Cadena, 2010; Gelman et al., 2003, 2010; Neufeld, 2022; Rhodes et al., 2018, 2012). Another hypothesis, which will be our focus here, is this:

- Asymmetry Generic generalizations display a unique divergence between their introduction and their implications (Brandone et al., 2015; Cella et al., 2022; Cimpian et al., 2010). Suppose that on average, subjects require that a minimum of $n\%$ of sharks have to be dangerous to accept the generic (2). When asked to assume (2) is true and to predict what follows from that concerning the proportion of sharks that are dangerous, on average subjects choose an m such that $m\% >> n\%$. Thus, the prevalence we infer from a generic is much higher than the prevalence needed to accept a generic. Importantly, this divergence between introduction and implication is supposed to be *unique* to generics, compared to analogous statements with overt quantifiers and adverbs like (1-b)-(1-c).

Critical reactions to the important body of work that emphasizes the problematic aspects of generics have mostly focused on essentializing. Specifically, recent em-

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

pirical (Noyes & Keil, 2019; Vasilyeva et al., 2018; Vasilyeva & Lombrozo, 2020) and theoretical (Ritchie, 2019; Saul, 2017) work has shown that, under certain conditions, generic sentences can express principled relations that are conceived as depending on structural, rather than inherent or essentialized dispositions. In addition, Hoicka et al. (2021) present evidence that generics do not increase essentialism compared to quantified generalizations. In contrast, there has been less empirical and theoretical work directly on Asymmetry. Indeed, Asymmetry is readily accepted by many philosophers, psychologists, and linguists (Cappelen & Dever, 2019; Leslie, 2017; van Rooij & Schulz, 2020).

The aim of this paper is to assess Asymmetry, a hypothesis which has potentially far-reaching consequences. Suppose Asymmetry is true, and consider its effects on the dynamics of information exchange in a simple social network, with 'experts' and 'recipients'. 'Experts' introduce a generic like (2) when, according to their observations and beliefs, the proportion of sharks that are dangerous meets the minimum threshold for its truth. When that happens, experts can communicate that information to 'recipients' using (2). Yet when the recipients accept (2), they will on average (given Asymmetry) be disposed to infer that the proportion of sharks which are dangerous is significantly greater than the proportion relied on to introduce (2) by the experts. Over time and iterations of this asymmetry effect, we might observe a substantial gap between the actual and perceived prevalences for the properties of many kinds, when that information is communicated using generics. Given these effects of Asymmetry, some theorists have recommended to avoid the use of generics in order to mitigate the formation of stereotypes (Leslie, 2017; Rosola & Cella, 2020).

Due to its influence and potential explanatory power, Asymmetry should be carefully examined. The original finding is due to Cimpian et al. (2010), but their key study has two important limitations, also to be found in subsequent studies which extended asymmetry developmentally (Brandone et al., 2015) and to social categories (Cella et al., 2022). First, the part of the study which tested the asymmetry between introduction conditions (**IC**) and implied prevalences (**IP**) compared generics only with a very limited set of overtly quantified sentences—indeed, only to *most*-sentences—and to no adverbs or modals, which are arguably closer in meaning to *Gen*. This is a substantial limitation, especially since the original finding is often taken to support the view that generic sentences and thoughts are *unique* in reliably triggering an IC-IP asymmetry.

Secondly, Cimpian et al. (2010)'s experimental design might have artificially suppressed the asymmetry effect for *most*. For the IP condition, Cimpian et al. (2010) used a continuous scale: participants could select what is the implied prevalence of Ks that are F, given a generic or *most*-statement. In the *most* IP condition, the average was 78%. In the IC condition, participants were randomly given a statement with a preset prevalence value of either 10%, 30%, 50%, 70%, or 90%, and asked to decide whether, given the specified prevalence, a generic or a *most*-statement is true or acceptable. In the *most* IC condition, the average was 76.9%. But averaging over all of the prevalences a participant accepts as true is ill-suited to assess ICs for quantified sentences. The standard view in linguistics is that *most* means 'more than half', so *most*-sentences are strictly true just above 50% (Heim & Kratzer, 1998). But if participants choose 'true' every time one of the presented prevalence values lies anywhere in the truth-range for *most*, the average will be roughly in the *middle* of that range, which means the ICs will be inflated and not representative of the *minimal* prevalence sufficient for truth. To see this, compare the ICs we would get with this method for the quantifier *at least one*. Participants would choose 'true' at *every* preset value (i.e., 10%, 30%, 50%, 70%, or 90%), and averaging over these choices would yield an average of roughly 50%. But clearly, the ICs of *at least one* are much lower than that, showing that averaging over all the preset values for which a given participant answered 'true' isn't a suitable method to reveal that participant's ICs.

The goal of Studies 1-2 is to examine asymmetry using a design that addresses those problems. Accordingly, our studies include sentences with various other quantifiers and adverbs. This allows us to more directly test whether, and to what extent, the IC-IP asymmetry is unique to generics. In addition, participants used a continuous percentage scale to set both the minimal prevalence level at which they accept a target sentence (ICs) and the implied prevalence that they would infer given the truth of a target sentence (IPs). Other than that, we stayed close to the basic set up and materials used in Cimpian et al. (2010). As we will see, our results call for substantial revision of the asymmetry hypothesis: we observe an IC-IP asymmetry for *all* non-strictly maximal/minimal/exact quantifiers and adverbs. The result holds for sentences that predicate neutral properties, and also for those that predicate striking or dangerous properties. We discuss the implications of those results in the General Discussion.

Study 1

Cimpian et al. (2010) investigate generic and 'most'-statements for plain properties such as 'silver fur', as well as dangerous or distinctive properties such as 'dangerous silver fur', and non-distinctive properties such as 'curly silver fur'. Study 1 focuses solely on plain properties, while Study 2 focuses on generic and quantified statements with dangerous properties. Following previous studies, we aimed to test people's estimated frequencies for generics using invented animal species such as 'lorches' and 'morseths' in order to eliminate the influence of prior knowledge. In contrast to prior studies, we made an effort to align the scale and question in the introductory task as closely as possible with those used in implied prevalence task.

In the IP condition, participants were first asked to take as true a statement of the form "Gen/Quantifier/Adverb Ks are F", e.g. "Lorches have red feathers." or "Usually, sapers have blue spots." Participants were then asked the question "What percentage of Ks [e.g., sapers] do you think have F [e.g. blue spots]?" In the IC condition, participants were asked to consider a statement of the form 'Gen/Quantifier/Adverb Ks are F", and were subsequently asked "What do you think is the minimum percentage of Ks [e.g., morseths] that need to have F [e.g., silver fur] in order for the statement to be true?" In both conditions, participants were asked to rate the estimated frequencies on a continuous scale form 0% to 100% in steps of 1%. This approach avoids the problems of Cimpian et al. (2010)'s design that we highlighted in the Introduction.

The following hypotheses were formulated to establish and test the relation between IC and IP condition:

Hypothesis 1: There are significant IC-IP asymmetries not only for generic statements but also for statements with explicit quantifiers and adverbs.

Hypothesis 2: The differences in IC-IP asymmetries between generic and statements with explicit quantifiers/adverbs are not significant.

The design of our study, incl., hypotheses and methods, were pre-registered with the Open Science Framework. In addition, our raw data are publicly available.

Methods and Stimuli

Study 1 is a 2 x 6 mixed design, with two independent variables: CONDITION and QUANTIFIER. The between-subject factor CONDITION has two levels: introduction condition and implied prevalence. The within-subject factor QUANTIFIER has six levels: Generic, All, Most, Usually, Typically, and Some.

150 Participants were recruited from Prolific Academia and reimbursed for their participation. We set the inclusion criteria to English as first language, 90% success rate on Prolific, exclusion of participants who have participated in our previous studies on generics, and a balanced gender sample. Additionally, the 'All'-condition served as a control condition: Participants who failed to answer 100% for the 'All' statement in either the IC or IP condition were excluded from further analysis. The remaining sample consisted of 114 participants (M_{age} = 38.95 years, 60 females, 54 males, 0 non-binary). Participants were randomly assigned to either the IC or IP condition. Each participant received a statement from each of the six levels of the QUANTIFIER variable.

We used an adapted (from Cimpian et al., 2010) introductory statement telling participants "In this study, we will tell you about six animal species that live on a remote island. This island is very large and has many different animals on it. You will be given some information and asked some questions. Please try to answer our questions intuitively." The six animal species were taken directly from the stimuli used by Cimpian et al.: *cheebas, dorbs, lorches, morseths, sapers, trufts.* The properties were also either taken directly from Cimpian et al. or slightly adapted: *red feathers, yellow scales, green tail, silver fur, blue spots, pink ears.* To avoid possible confounding effects of property and kind names, we used two different versions in which we alternated which kinds and properties were paired with a quantifier.

Results

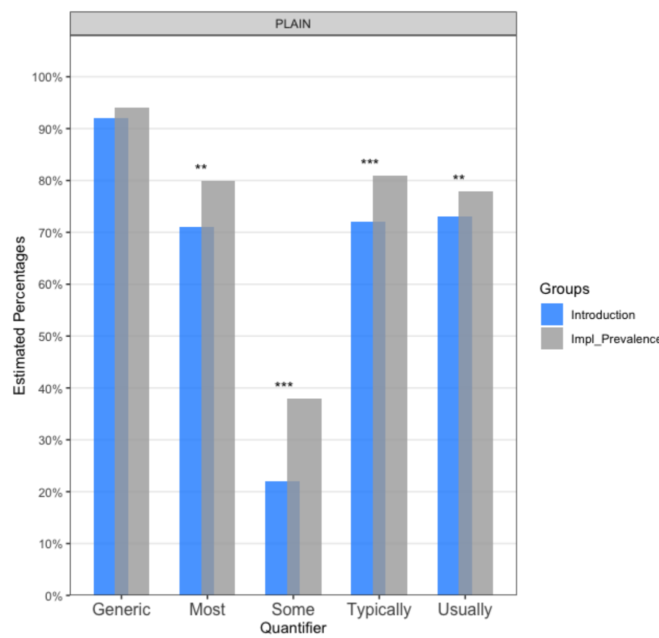

Figure 1: IC and IP frequencies for five different generalized statement types featuring plain properties.

A repeated measures ANOVA was performed with participants' frequency ratings as the dependent measure, the within-subject factor QUANTIFIER and the between-subject factor CONDITION. The results indicated that the within-subject factor QUANTIFIER, $F(4,448) = 376.41, p < 0.001, \eta^2 = 0.77$, was significant. There was also a significant interaction between QUANTIFIER and CONDITION, $F(4,448) = 4.76, p < 0.001, \eta^2 = 0.04$. Thus, Hypothesis 2 was not supported by our data. The between-subject factor CONDITION was significant, $F(1,112) = 28.03, p < 0.001, \eta^2 = 0.20$, which means that we cannot reject Hypothesis 1. The average frequency ratings are displayed in Figure 1. We also conducted independent samples t-tests examining differences for each quantifier type. Only in the generics condition was the difference not significant, $t(112) = -.743, p = 0.230$.

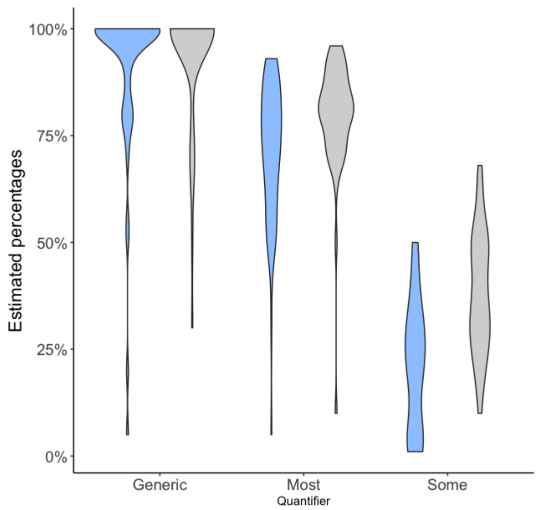

Figure 2: Violin plots for the Generic, Most and Some statements. The blue plots show the distribution in the IC conditions, the grey plots in the IP conditions.

Discussion

The results of Study 1 do not match the results of previous studies in two ways. First, while there was a difference, in the expected direction, between the IC and IP condition for the generic statements, this difference was not significant (but see Study 2). Second, and most importantly, we found a significant IC-IP asymmetry for all tested sentences with explicit quantifiers (*most, some*) and adverbs (*usually, typically*). Differences between IC frequencies and IP frequencies were around 10%.

Looking at the violin plots in Figure 2, we can appreciate the importance of using a continuous scale for the ICs for *most* and *some*. For *most*, note (blue plot) that a substantial number of participants choose minimum prevalence levels greater than 50% and lower than 70%. Yet those participants would have been modeled as answering 'true' only at 70% or above if they had to provide their ratings (or set the minimum value for truth) using a non-continuous scale that jumped from 50% ('most' literally false) to 70% and then 90%. In addition, if we had averaged over participants' true answers on a scale like Cimpian et al.'s, then that average (approximately 80%) wouldn't have reflected the minimum prevalence level for truth for that large group of participants.

Finally, although our study did not target differences in frequency estimates for different explicit quantifiers and adverbs, our results indicate that frequency estimates for 'most', 'usually, and 'typically' statements are very similar in both conditions. Those average frequency estimates were substantially below the IC and IP frequencies for generics. This result challenges accounts which assume that *Gen* can be reduced to operators like 'most', 'typically', or 'usually'.

Study 2

In Study 1, we explored statements such as "Lorches have red feathers" which predicate a plain property to a fictitious animal species. In Study 2, we examine statements that predicate striking/dangerous properties, such as "Lorches have deadly red feathers." Given the similarity in design, we pre-registered analogous hypotheses:

Hypothesis 3: There are significant IC-IP asymmetries not only for generic statements but also for statements with explicit quantifiers and adverbs.

Hypothesis 4: The differences in IC-IP asymmetries between generic and explicit quantifier/adverb statements are not significant.

Our study design, hypotheses, and methods were preregistered with the Open Science Framework

Methods and Stimuli

Our study employed a 2 x 6 mixed design, with two levels of the independent variable CONDITION (introduction and implied prevalence) and six levels of the independent variable QUANTIFIER (Gen, All, Most, Usually, Typically, Some). A total of 150 participants were recruited from Prolific Academic and randomly assigned to either IC or IP condition.[1] We excluded participants who failed to answer '100%' in the control condition ('All-statements'). The remaining participant pool consisted of 104 participants (M_{age} = 38.97 years, 52 females, 52 males, 0 non-binary).

The methodology employed in Study 2 was similar to that of Study 1, with one key difference: the statements presented to participants featured dangerous properties rather than plain properties. Participants were presented with six statements, each featuring one of the following six danger properties: *dangerous*, *deadly*, *hazardous*, *lethal*, *toxic*, and *vicious*. In order to control for potential confounding effects of danger attributes, animal species, and properties, two versions were used in which the pairing of danger properties, animal species, and properties with quantifiers were alternated.

Results

A repeated measures analysis of variance (ANOVA) was conducted to examine the effect of the within-subject factor QUANTIFIER and the between-subject factor CONDITION on participants' frequency ratings. The results revealed that the within-subject factor QUANTIFIER, $F(4,408) = 308.36, p < 0.001, \eta^2 = 0.75$, was found

[1] Same inclusion criteria were employed as in Study 1.

to be statistically significant. There was no significant interaction between QUANTIFIER and CONDITION, $F(4,448) = 1.15, p = 0.334, \eta^2 = 0.01$, thereby supporting the prediction outlined in hypothesis 4. The between-subject factor CONDITION was found to be statistically significant, $F(1,102) = 28.21, p < 0.001, \eta^2 = 0.22$, suggesting that hypothesis 3 is supported by our data. The average frequency ratings are depicted in Figure 3. Furthermore, we conducted independent samples t-tests to examine differences for each quantifier type, which revealed that for each quantifier, the differences were all significant.[2]

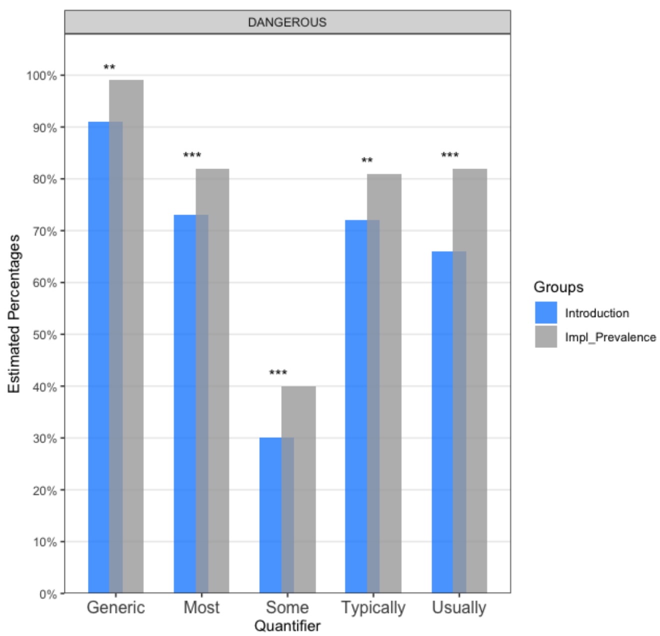

Figure 3: IC and IP frequencies for five different generalized statement types featuring dangerous properties.

Discussion

The results of Study 2 paint a clear picture in favor of general IC-IP asymmetries for both generics and statements with explicit quantifiers or adverbs: IC and IP frequencies were different for *Gen* and for each explicit quantifier and adverb. As in Study 1, the average frequency estimates were roughly the same for 'most', 'typically' and 'usually' statements, and lower compared to frequency estimates for generic statements.

In the literature, scholars often argue that generic statements featuring a dangerous property are considered true at specially low frequencies (Bian & Cimpian, 2021; Cimpian et al., 2010; Leslie, 2017). Our Study 2 did not confirm such a trend. Instead, the estimated frequencies for generics and quantified statements did not differ markedly in that direction between plain statements and statements featuring a dangerous property.

[2] The results of both studies remained largely unaffected when the threshold for All-statements was reduced to 90% to accommodate pragmatically-driven responses.

General Discussion

Studies 1-2 support two main hypotheses. First, generics tend to trigger an IC-IP asymmetry, yet that effect is also triggered by sentences with overt quantifiers and adverbs. Thus, IC-IP asymmetry effects are not unique to generics. Second, the asymmetry effect was observed—and its magnitude was not interestingly different—for predications with neutral and dangerous properties. So at least in IC-IP type reasoning tasks, striking properties (compared to neutral ones) do not seem to increase the likelihood that biased beliefs will be introduced into social networks for any of the tested quantifiers. We now consider some implications and open questions.

Source of the asymmetry effects

Cimpian et al. (2010) consider two possible explanations for ASYMMETRY, both focusing on why IPs are inflated. One is that typical uses of generics are supported by high-prevalence situations. So when they do not have relevant background information, listeners tend to assume that a novel generic is likely supported by a high-prevalence situation, even if that is not strictly required for its truth. The other explanation is Gricean. The basic idea is that if a speaker S of a generic *Ks are F* meant to convey that F applies to just a subset of the Ks, then S would have said that, using alternatives with, e.g., *some* or *many*. As we will see, our result that asymmetry is a general effect of (non-maximal/minimal/exact) generalizations helps adjudicate between these accounts.

Consider first the Gricean explanation. Even applied just to generics, the proposal is incomplete. Generics typically convey information about principled and not merely accidental connections. So when figuring out why a speaker S of a generic didn't choose an overtly quantified counterpart, a reasonable bet would be that S wants to highlight the principled connection. In addition, this explanation doesn't generalize. Take the asymmetry with *most*-statements. The analogous way of explaining why the IPs are inflated would be that if S meant to convey a lower prevalence level, S would have chosen a weaker quantifier. Yet here there are no obvious alternatives: e.g., the fact that *half* was not selected only helps listeners guess above 50%, which is anyways entailed by the strict truth-conditions of *most*. Can we appeal to specific complex quantifiers, e.g., *exactly 90%*? Even if such quantifiers are assumed, implausibly, to be salient alternatives of simple lexicalized ones, there are candidates that entail both higher and lower prevalances than what we observed for the IPs of *most*.

In contrast, a version of the first hypothesis of Cimpian et al. (2010)—i.e., appealing to typical conditions of use—is more promising. When participants set the IC, their task is to select the *minimum* prevalence level which guarantees the truth of instances of [Quantifier/Adverb/*Gen*] *Ks are F*. Yet it doesn't follow that it is usually a good interpretative strategy to set the IP level at that same IC threshold. To illustrate, imagine you are at the supermarket and your roommate texts you, 'please get me some apples, lemons and cookies'. You proceed to bring them exactly one of each, on the grounds that it is strictly compatible with the request, and then notice in puzzlement that your roommate is not too happy at the result. That is the kind of misalignment which we would systematically be subject to if as listeners we by default interpret implied prevalences at the strict minimum level required for the truth of the communicated generalization.

So the minimum prevalence levels required for the strict acceptance of generic or (non-maximal/minimal) quantified sentences are, in many domains and contexts, much lower than the average prevalence levels observed in the situations in which those sentences are *typically* used. Listeners of those generalizations who set the implied prevalence level so as to match the minimum levels for their strict truth, would generally underestimate the amount which a cooperative speaker intended to convey. Cases that suggest otherwise are typically legalistic or adversarial, non-collaborative information exchanges. Suppose a company advertised 'some of our cereal boxes have a toy', yet it was discovered, in a law suit by angry parents, that just 2 out of every 100,000 boxes have a toy. One can imagine a successful defense on the grounds that 'some' strictly requires just 'at least one'. Still, we can also understand why parents (and the public) may agree that the company was intentionally misleading. For when 'some'-statements are used, the proportion is typically much greater than 2/100,000. So it is reasonable to interpret the company's novel 'some'-statement as suggesting a prevalence significantly greater than 2/100,000, even if still lower than 'half' or 'most'.

Implications for social and political speech

Studies 1-2 show that there is an asymmetry effect for different kinds of generalizations, and for both neutral and striking properties. While our results do not undermine the hypothesis that generics are psychologically unique in multiple respects, they do call for refinement of some of the implications which have been drawn, assuming the supposed uniqueness of ASYMMETRY for generics, for political discourse and social stereotyping (Brandone et al., 2015; Cella et al., 2022; Cimpian et al., 2010; Leslie, 2017; McKeever & Sterken, 2021; Rosola & Cella, 2020). At the same time, our results shed light on the kinds of studies one can run to determine whether and to what extent generalized asymmetry effects may introduce biases into social networks.

According to Cimpian et al. (2010), since generics are "legitimized even by scant evidence, their truth is rarely questioned. Yet, after they become part of the accepted discourse, they take a life of their own, turning what may have originally been a nuanced, contextualized fact into a definite pronouncement" (p. 1473). From a few incidents of nuclear plant accidents, people may accept, *Nuclear plants are dangerous*, yet after that generic is accepted and communicated across social networks, it may be taken to convey near-universal facts about nuclear plants. Since this asymmetry is thought to be distinctive of generics, it can be tempting, from this perspective, to recommend that we should try to avoid their use to express political and social generalizations, especially those that may affect vulnerable groups (Leslie, 2017; Rhodes et al., 2012; Wodak et al., 2015). In addition, since Cimpian et al. (2010) found that striking properties modulated the asymmetry effect, one might be tempted to emphasize this recommendation for striking property generalizations.

Yet given our results, efforts to avoid using generics would likely not reduce social biases caused by the asymmetry effect, insofar as people continue to use, as seems unavoidable, generalizations with quantifiers or adverbs like *some*, *most*, *usually* or *typically*. Generalizations with those overt quantifiers and adverbs show an asymmetry effect which is at least as substantial as that triggered by their generic counterparts. Recommendations to avoid using any quantifiers or adverbs with relatively open ended or vague implications is impracticable. Still, the improved understanding which we now have on the source of the general asymmetry effect, can inform us about its possible social and political implications, and eventually also about possible interventions.

Suppose, as we suggested, that listeners infer IP levels for generic and quantified generalizations based on the prevalence levels associated with their typical uses. Whether this interpretative procedure introduces biases depends on how well it traces, on average, the prevalence levels that speakers had in mind, for different novel combinations of kinds and properties. It also depends on how speakers select which sentences to use given particular situations and prevalences. For example, speakers can agree that, strictly, *some Labradors attacked children* is true if two Labradors did that, yet still disprefer using that sentence to describe that specific situation based on the expectation that a typical listener will likely infer a much higher proportion (e.g., a 'very few'-sentence might be a better choice). IC tasks ask subjects to set the minimal prevalences at which specific generalizing sentences are strictly true, but not which sentence, given a choice between various options, they would choose as the best way to communicate that/a specific prevalence information. To determine if the IC-IP asymmetry introduces biases, via information exchanges, into social networks, we need to determine how well speakers and hearers coordinate in this kind of scenarios. We plan to experimentally explore this in future work on this project.

Acknowledgments

We would like to thank Joshua Knobe as well as three anonymous reviewers for their very helpful comments on previous versions of the manuscript. This research was funded by the Swiss National Science Foundation (grant number: PCEFP1_181082).

References

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.

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.

Cappelen, H., & Dever, J. (2019). *Bad language*. Oxford, UK: Oxford University Press.

Cella, F., Marchak, K. A., Bianchi, C., & Gelman, S. A. (2022). Generic language for social and animal kinds: An examination of the asymmetry between acceptance and inferences. *Cognitive Science*, *46*(12), e13209.

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., & Cadena, C. (2010). Why are dunkels sticky? preschoolers infer functionality and intentional creation for artifact properties learned from generic language. *Cognition*, *117*(1), 62–68.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in cognitive sciences*, *8*(9), 404–409.

Gelman, S. A., et al. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford Cognitive Development.

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

Goldfarb, D., Lagattuta, K. H., Kramer, H. J., Kennedy, K., & Tashjian, S. M. (2017). When your kind cannot live here: How generic language and criminal sanctions shape social categorization. *Psychological Science*, *28*(11), 1597–1609.

Hammond, D., Matthew, & 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.

Heim, I., & Kratzer, A. (1998). *Semantics in Generative Grammar*. Oxford, UK: Blackwell Publishers Ltd.

Hoicka, E., Saul, J., Prouten, E., Whitehead, L., & Sterken, R. (2021). Language signaling high proportions and generics lead to generalizing, but not essentializing, for novel social kinds. *Cognitive Science*, *45*(11), e13051.

Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Language and Cognitive Processes*, *27*(6), 887–900.

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

Leslie, S.-J. (2007). Generics and the structure of the mind. *Philosophical Perspectives*, *21*(1), 375–403.

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

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

McKeever, M., & Sterken, R. (2021). Social and political aspects of generic language and speech. In *The routledge handbook of social and political philosophy of language* (pp. 259–280). Routledge.

Neufeld, E. (2022). Psychological essentialism and the structure of concepts. *Philosophy Compass*, *17*(5), e12823.

Nickel, B. (2016). *Between logic and the world: An integrated theory of generics*. Oxford University Press.

Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359.

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

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.

Ritchie, K. (2019). Should we use racial and gender generics? *Thought: A Journal of Philosophy*, *8*(1), 33–41.

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

Saul, J. (2017). Are generics especially pernicious? *Inquiry*, 1–18.

van Rooij, R. (2017). Generics and typicality. In *Proceedings of* Sinn und Beudeutung.

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

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology*, *54*(9), 1735.

Vasilyeva, N., & Lombrozo, T. (2020). Structural thinking about social categories: Evidence from formal explanations, generics, and generalization. *Cognition*, *204*, 104383.

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