

# Not all generics are created equal: Differentiating between ‘do’ and ‘can’ generic statements

Sophie H. Arnold (sophie.arnold@nyu.edu)  
Department of Psychology, New York University

Mohit Mukherji (mohit.mukherji@nyu.edu)  
Department of Psychology, New York University

Marjorie Rhodes (marjorie.rhodes@nyu.edu)  
Department of Psychology, New York University

## Abstract

Generic statements (e.g., “girls wear makeup”) tie properties to groups and are a common way of transmitting stereotypes. One natural but untested way that people might try to undermine these statements is by making a similar statement about salient but not mentioned contrastive groups (e.g., “boys can wear makeup too”). Do *can* generics license the same judgments as *do* generics? Four studies examined how adults judge two novel groups when one group does a property (e.g., Zarpies make pizzas) while the other group can do the property (e.g., Gorps can make pizzas too). Compared to *do* generics, adults consistently judged groups described with *can* generics to be less likely to have, less interested in, less competent at, and for it to be less permissible for them to do the property. Overall, these results suggest that *can* generics are unlikely to be an effective means at equating beliefs about two groups.

**Keywords:** generics, social cognition, pragmatics

## Introduction

Generic statements are pervasive in language, media, and across cultures (DeJesus et al., 2019; Gelman et al., 2008; Gelman et al., 2014; Gelman & Tardiff, 1998). These statements convey that properties are more prevalent among the mentioned category than salient alternative groups (Tessler & Goodman, 2019) and characteristic of category membership in some deep way (e.g., Leslie, 2008; Salomon & Cimpian, 2014). Generics are therefore a common way to transmit stereotypes (Rhodes et al., 2012). As people continue to push for a more equitable society, a natural response to these kinds of statements is to affirm that other groups can have these properties too (e.g., “Boys can wear makeup too.”). In four experiments, we examine the extent to which these kinds of statements accomplish one potential goal of the speaker: Do these statements equate judgments of groups? Or, instead, do they still subtly reinforce the idea that one group is lesser?

When people hear generic statements about a group, they make inferences about both members of the mentioned group and members of relevant alternative contrast categories (e.g., “boys” comes to mind as an obvious salient alternative when statements are made about “girls”). For example, when children and adults are introduced to “Zarpies” and “Gorps” and hear that, “Zarpies make pizzas,” they assume that a novel Zarpie will also make pizza and that a Gorp,

conversely, will *not* (even though whether Gorps make pizza or not is not explicitly communicated by language; Baharloo et al., 2023; Kramer et al., 2021; Moty & Rhodes, 2021). This is one subtle way in which inferences licensed by generic statements extend beyond the semantic content of the statement. How might we interfere with these processes to prevent stereotype transmission? One intuitively plausible and likely common response is to make a generic statement about the unmentioned group, for example, “Gorps can make pizzas too.” It is possible that these generic responses license the same kinds of inferences that the original statements do, equating the pervasiveness of properties across the mentioned groups.

Well-intentioned statements whose goal is to undermine stereotypes can sometimes backfire, however. For example, subject-complement generics like “girls are just as good as boys at math” imply that boys are the group who is typically better at math and therefore the yardstick by which girls’ ability is measured. These statements, though well intentioned, lead both children and adults to uphold the stereotype that girls are naturally better at math than boys (Chestnut & Markman, 2018; Chestnut et al., 2021), particularly among people who are not on the look-out for potential bias in language (Holmes et al., 2022).

These examples show how pragmatic interpretations of counter-stereotypic generics can sometimes lead them to backfire. That is, although the content of such sentences sound egalitarian, they are often not interpreted this way, as listeners account for the broader context and consider the speaker’s choices with respect to alternative things they could have said. In the case of responding to a *do* generic with a *can* generic (e.g., “girls wear makeup,” “boys can wear makeup too”), the listener knows that the speaker(s) could have chosen to say a second *do* generic but instead chose to say a *can* generic (e.g., Degen, 2023; Gibbs & Moise, 1997). This difference in word choice is therefore highlighted and could magnify the difference in semantic meaning between “do” and “can”. While “can” typically licenses weaker inferences than “does,” hearing a *can* generic after a *do* generic could further enhance the difference in judgments between the groups. Such subtle language choices often go undetected and thereby uphold the stereotypical status quo.

## The Present Studies

We presented adults with 12 properties (e.g., making pizzas) and asked them to make judgments about two novel groups: how prevalent the property was in each group (Studies 1-4), how interested each group is in the property (Studies 1-4), how competent each group is in the property (Studies 1-4), and how permissible it is for each group to do the property (Studies 3, 4). Across studies, we tested whether this effect was unique to hearing a *can* generic after a *do* generic or if it was simply about hearing a group mentioned second (Study 1), about a *can* generic without hearing a *do* generic (Study 2), about other language that specifies that a group is second (e.g., “too”, Studies 3, 4), and whether it extends to reading these statements as a part of a more naturalistic conversation (Study 4). Across all studies, we predicted that hearing a *can* generic statement after a *do* generic statement would lead to uniquely lower ratings of property prevalence, interest, competence, and permissibility.

## Open and Transparent Science

Research questions, hypotheses, methods, and analyses for all experiments were preregistered on AsPredicted. Study 1: [https://aspredicted.org/Z6P\\_LDF](https://aspredicted.org/Z6P_LDF); Study 2: [https://aspredicted.org/C12\\_93G](https://aspredicted.org/C12_93G); Study 3: [https://aspredicted.org/HV4\\_MCQ](https://aspredicted.org/HV4_MCQ) and [https://aspredicted.org/KYY\\_2GM](https://aspredicted.org/KYY_2GM); Study 4: [https://aspredicted.org/2D8\\_8HN](https://aspredicted.org/2D8_8HN). Any deviations from the preregistrations are marked transparently below.

### Experiment 1

In this first study, we examine what people infer about two groups when they hear that the first mentioned group does a property (e.g., “Zarpies make pizzas”) and the second mentioned group can do the property too (e.g., “Gorps can make pizzas too”). We hypothesize that participants will think the second mentioned group who can do the property will be thought of as less likely to do the property, less interested in the property and less competent at the property. We compare these judgements to a separate condition where participants hear that both the first and second mentioned group do a property (e.g., “Zarpies make pizzas. Gorps make pizzas.”). This allows us to tease apart whether the effects of “can” are unique to that generic construction or whether hearing about any group second would result in lower evaluations of that second group.

## Method

**Participants** 153 university students participated in November 2023. An additional 51 participants were excluded for the following preregistered reasons: failing an attention check ( $n = 15$ ) or noting that they did not pay attention during the study and therefore we should not use their data ( $n = 39$ ).

**Procedure** Participants were introduced to two novel groups, Zarpies and Gorps. Then, participants read 12 statements about Zarpies and Gorps and made judgments about the

groups based on the statements. These statements were about activities (make pizzas, do science, play basketball, create fancy hats), appearances (have long hair, have short fingernails, shave their arms, wear shiny boots) and traits (funny, quiet, smart, hardworking). Participants made up to three property-relevant judgments for each statement for each group: the property prevalence, the property interest, and the property competence. The order of statements and which group (first mentioned, second mentioned) they made judgments about first was randomized across participants. The order of judgments was presented in one of two orders, randomized across participants.

Prior to reading the statements, participants were randomly assigned to one of two conditions: the “do & can” condition where the first mentioned group *does* the property (e.g., “Zarpies make pizzas.”) and the second mentioned group *can* do the property too (e.g., “Gorps can make pizzas too.”) or the “do & do” condition where both the first and second mentioned groups *do* the property (e.g., “Zarpies make pizzas. Gorps make pizzas.”).

**Prevalence** For all statements, participants judged how prevalent the property was in each novel group, “How many [group] do you think [property]?”, on a sliding scale from 0 = *no [group]* to 100 = *all [group]*.

**Interest** For each statement about activities and appearance (but not traits), participants judged their estimation of how interested in the property each group was, “How much do you think [group] like [property]?” on a sliding scale from 0 = *not like it at all* to 100 = *extremely like it*.

**Competence** For each statement about activities and traits (but not appearance), participants judged their estimation of how good at the property each group was, “How good do you think [group] are at [property]?” on a sliding scale from 0 = *not good at all* to 100 = *extremely good*.

**Analytic Strategy** All analyses were conducted in R version 4.3.1 (R Core Team, 2023). We constructed separate linear mixed models for each of our dependent variables: property prevalence, interest, and competence. The fixed and random effect structure for each model was identical. We regressed participants’ judgments on the condition (do & can = 1, do & do = 0), order of group mentioned (second mentioned = 1, first mentioned = 0), and the interaction between the two with crossed random intercepts for participant ID and statement. Our hypothesis would be supported by a significant two-way interaction between the two fixed effects. We ran an ANOVA on the resulting models to test for the significant 2-way interactions. We followed up on all significant interactions with pairwise comparisons from the *emmeans* package (Lenth et al., 2018) adjusting for multiple comparisons.

## Results

Consistent with our hypothesis, participants in the “do & can” condition judged the second mentioned group who *can* do the

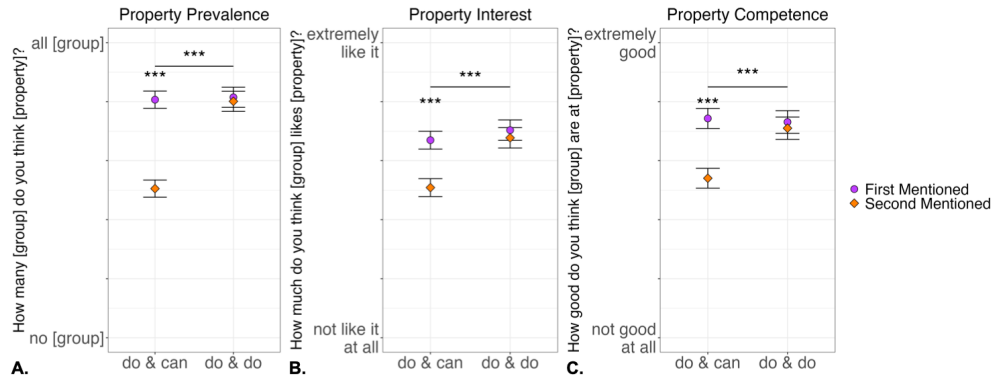


Figure 1: Experiment 1 marginal means with 95% confidence intervals. \*\*\*  $p < .001$ .

property to be less likely to do the property (lower prevalence;  $b = -29.59$ ;  $SE = 2.25$ ,  $p < .001$ ), less interested in the property ( $b = -16.88$ ;  $SE = 2.10$ ,  $p < .001$ ), and less competent at the property ( $b = -16.97$ ;  $SE = 2.19$ ,  $p < .001$ ) than the group who *does* the property (prevalence condition x group interaction: Wald  $X^2(1) = 613.76$ ,  $p < .001$ ; interest condition x group interaction: Wald  $X^2(1) = 90.01$ ,  $p < .001$ ; competence condition x group interaction: Wald  $X^2(1) = 168.47$ ,  $p < .001$ ; see Figure 1). In contrast, participants in the “do & do” condition judged the first and second mentioned groups to have similar property prevalence, interest, and competence,  $ps > .10$ .

## Experiment 2

Experiment 1 found support for our hypothesis that saying that the second mentioned group *can* do a property too leads adults to think that this group has lower property prevalence, interest, and competence, and, importantly, that these effects are not due to being the second mentioned group. In Experiment 2, we now examine whether this effect is simply due to the use of a *can* generic or if it is unique to hearing *can* generic after a *do* generic. In line with the rational speech act framework and work on pragmatics (e.g., Degen, 2023), we hypothesize that hearing a *can* generic after a *do* generic will lead to lower judgments than only hearing a *can* generic without the context of a *do* generic as hearing both statements back-to-back makes it clear that the speaker chose to use a *can* generic rather than a *do* generic.

## Methods

**Participants** 157 university students participated in November 2023. An additional 50 participants were excluded for the following preregistered reasons: failing an attention check ( $n = 12$ ), failing to correctly recall the study purpose ( $n = 1$ ), or noting that they did not pay attention during the study and therefore we should not use their data ( $n = 37$ ).

**Procedure** The procedure was identical to Experiment 1, the only change was the conditions to which participants were assigned. In this study, participants were assigned to one of three conditions: “do & can”, “do”, or “can”. The “do & can”

condition was identical to Experiment 1 except there was no “too” at the end of the “can” statement to make this condition more similar to the comparison “can” condition (e.g., “Zarpies make pizzas. Gorps can make pizzas.”). For the “do” and “can” conditions, participants only read statements about one group. That group either does the property (e.g., “Zarpies make pizzas.”) or can do the property (e.g., “Zarpies can make pizzas.”).

**Analytic Strategy** All analyses were conducted in a similar manner to Experiment 1 with the following change to the fixed effects structure of the model. We regressed participants’ judgments on the statement type (do = 1, can = 0), number of groups mentioned (one group = 1, two groups = 0), and the interaction between the two. Our hypothesis would be supported by a significant two-way interaction between the two fixed effects. We ran an ANOVA on the resulting models to test for the significant 2-way interactions. We followed up on all significant interactions with pairwise comparisons from the *emmeans* package (Lenth et al., 2018) adjusting for multiple comparisons.

## Results

Consistent with our hypothesis, participants in the “do & can” condition judged the group who *can* do the property to be less likely to do the property (lower prevalence;  $b = -33.18$ ;  $SE = 0.94$ ,  $p < .001$ ), less interested in the property ( $b = -14.78$ ;  $SE = 1.14$ ,  $p < .001$ ), and less competent at the property ( $b = -19.35$ ;  $SE = 1.08$ ,  $p < .001$ ) than the group who *does* the property (prevalence condition x group interaction: Wald  $X^2(1) = 42.15$ ,  $p < .001$ ; interest condition x group interaction: Wald  $X^2(1) = 6.39$ ,  $p = .011$ ; competence condition x group interaction: Wald  $X^2(1) = 9.73$ ,  $p = .002$ ; see Figure 2). When looking at the single group conditions, participants in the “can” condition judged the group to be less likely to do the property (lower prevalence;  $b = -14.98$ ;  $SE = 2.64$ ,  $p < .001$ ), less interested in the property ( $b = -8.21$ ;  $SE = 2.34$ ,  $p < .001$ ), and less competent at the property ( $b = -11.47$ ;  $SE = 2.29$ ,  $p < .001$ ) than participants in the “do” condition, though these

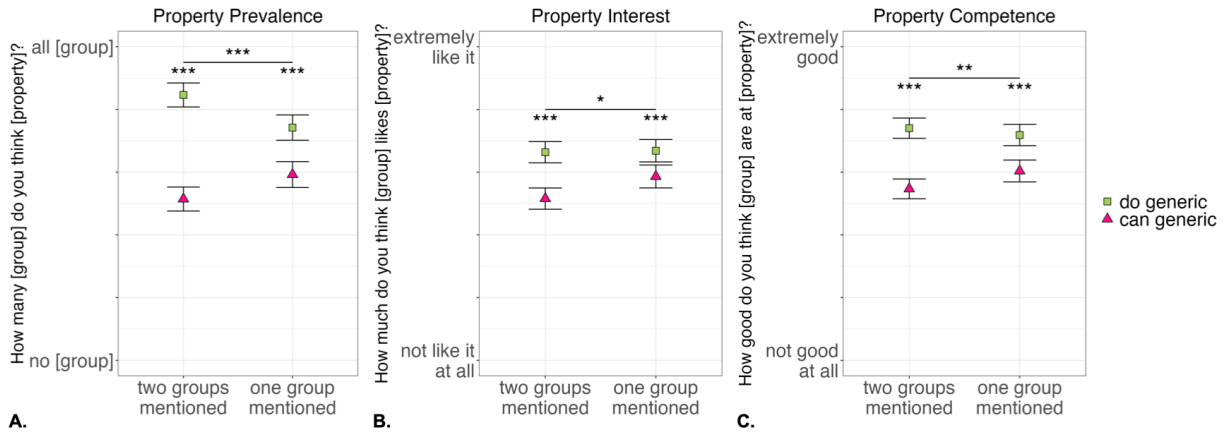


Figure 2: Experiment 2 marginal means with 95% confidence intervals. \*\*\*  $p < .001$ ; \*\*  $p < .01$ ; \*  $p < .05$ .

differences were almost half the magnitude of those in the “do & can” conditions.

### Experiment 3

Experiment 2 provided additional support for our hypothesis that saying that a second mentioned group *can* do a property leads adults to think that this group has lower property prevalence, interest, and competence, and, importantly, that these effects are magnified when hearing a *can* generic after a *do* generic. In Experiment 3, we now examine whether these *can* generics lead to judgments of lower permissibility (in addition to the other variables measured) and test whether this effect is unique to “can” rather the effect of any language that implies that a mentioned group is secondary (e.g., “too”). We hypothesized that “can” would uniquely contribute to lower estimations above and beyond any effect of “too”.

### Methods

**Participants** 167 adults participated in December 2023 (85 university students, 82 prolific adults). An additional 84 participants were excluded for the following preregistered reasons: failing an attention check ( $n = 14$ ) or failing to correctly recall the study purpose ( $n = 75$ ).

**Procedure** The procedure was identical to Experiment 1, the only changes were an additional condition to which participants could be assigned and an additional dependent variable for all statements (permissibility). Participants were assigned to one of three conditions: “do & can”, “do & too”, or “do & do”. The “do & can” and “do & do” conditions were identical to Experiment 1. The “do & too” condition was the same as the “do & can” condition, just without the word “can” (e.g., “Zarpies make pizzas. Gorps make pizzas too.”).

**Permissibility** For all statements, participants judged how permissible it was for each group to do the property, “How okay is it for [group] to [property]?”, on a sliding scale from 0 = *very not okay* to 100 = *very okay*.

**Analytic Strategy** All analyses were conducted in a similar manner to prior experiments (e.g., one linear mixed model per dependent variable) with the following changes to the fixed effects structure of the model. We regressed participants’ judgments on the condition (do & can, do & too, do & do), order of group mentioned (first mentioned, second mentioned), and their interaction. We additionally included a covariate for sample type (Prolific vs. University Study Pool). Our hypothesis would be supported by a significant two-way interaction between the two fixed effects. We ran an ANOVA on the resulting models to test for the significant 2-way interactions. We followed up on all significant interactions with pairwise comparisons from the *emmeans* package (Lenth et al., 2018) adjusting for multiple comparisons.

### Results

Consistent with our hypothesis, participants in the “do & can” condition judged the second mentioned group who *can* do the property too to be less likely to do the property (lower prevalence;  $b = -20.55$ ;  $SE = 0.85$ ,  $p < .001$ ), less interested in the property ( $b = -12.48$ ;  $SE = 0.99$ ,  $p < .001$ ), less competent at the property ( $b = -12.31$ ;  $SE = 1.06$ ,  $p < .001$ ), and thought it was less permissible for them to do the property ( $b = -6.36$ ;  $SE = 0.61$ ,  $p < .001$ ) than the first mentioned group who *does* the property (prevalence condition x group interaction:  $\text{Wald } \chi^2(2) = 346.07$ ,  $p < .001$ ; interest condition x group interaction:  $\text{Wald } \chi^2(2) = 91.77$ ,  $p < .001$ ; competence condition x group interaction:  $\text{Wald } \chi^2(2) = 79.75$ ,  $p < .001$ ; permissibility condition x group interaction:  $\text{Wald } \chi^2(2) = 46.25$ ,  $p < .001$ ; see Figure 3).

As a further support of our hypothesis, the differences in the “do & too” condition were only sometimes significant and when they were, these differences were less than half the magnitude of the differences in the “do & can” condition. In particular, the “do & too” condition, the second mentioned group was judged to do the property less (lower prevalence,  $b = -1.97$ ;  $SE = 0.88$ ,  $p = .025$ ) and for it to be less permissible for them to do the property (lower permissibility,  $b = -2.24$ ;  $SE = 0.64$ ,  $p < .001$ ) than the first mentioned group. However,

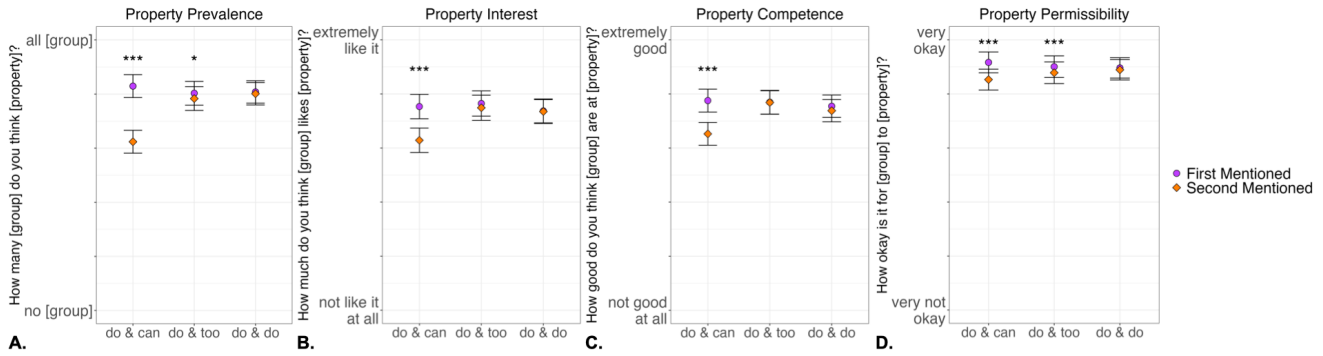


Figure 3: Experiment 3 marginal means with 95% confidence intervals. \*\*\*  $p < .001$ ; \*  $p < .05$ .

both groups were judged to be similarly interested in ( $b = -1.59$ ;  $SE = 1.03$ ,  $p = .12$ ) and competent at ( $b = -0.10$ ;  $SE = 1.10$ ,  $p = .93$ ) the property. There were no significant differences in judgments between the first and second mentioned groups in the “do & do” condition,  $ps > .10$ .

### Experiment 4

In Experiment 3, we found that lower judgments of the second mentioned group were not due to emphasizing the group was mentioned second (e.g., “too”) but more so due to being paired with a *can* generic. In Experiment 4, we provide a more naturalistic test of our hypothesis to examine whether a conversation structure with two speakers elucidates the same pattern of results as two sentences have in the previous studies. We hypothesize that in the context of a conversation, participants will continue to judge a second mentioned group that can do the property to be less likely to do the property, less interested in the property, less competent at the property, and for it to be less permissible for them to do the property than a group that does the property.

### Methods

**Participants** 188 Prolific adults participated in December 2023. An additional 12 participants were excluded for the preregistered reason of failing 3 or more of the 5 attention checks (e.g., “Do not answer this question.”).

**Procedure** We adapted the procedure of prior studies to a conversation format. Participants were first introduced to two middle school students who were walking around a town on a field trip. We chose to clarify the age of the speakers to make the simple language more plausible. Participants were all told that while these students didn’t live in the town, they had previous experience with the people who lived there. Then, participants read 12 conversations about the same 12 properties in prior studies. For each conversation, there was an introductory sentence (e.g., “Lowen and Remy stop by a pizzeria for lunch.”) followed by two statements made by the protagonists about the groups (e.g., “Lowen says: “Zarpies make pizzas.” Remy says: “Gorps can make pizzas too.”). There were two conditions in this study: the “do & can” and “do & too” conditions from Experiment 3. To ensure that

participants did not think that one child was the “Zarpie expert” and the other the “Gorp Expert,” Lowen said the first statement for half of each property types (e.g., activities) and Remy said the first statement for the other half. After each statement, participants reported their judgments of the groups on the same dependent variables as in Experiment 3.

**Analytic Strategy** All analyses were conducted in a similar manner to prior experiments (e.g., one linear mixed model per dependent variable) with the following change to the fixed effects structure of the model. We regressed participants’ judgments on the condition (do & can = 1, do & too = 0), order of group mentioned (second mentioned = 1, first mentioned = 0), and the interaction between the two. Our hypothesis would be supported by a significant two-way interaction between the two fixed effects. We ran an ANOVA on the resulting models to test for the significant 2-way interactions. We followed up on all significant interactions with pairwise comparisons from the *emmeans* package (Lenth et al., 2018) adjusting for multiple comparisons.

### Results

Consistent with our hypothesis, participants in the “do & can” condition judged the group who *can* do the property too to be less likely to do the property (lower prevalence;  $b = -18.31$ ;  $SE = 0.62$ ,  $p < .001$ ), less interested in the property ( $b = -13.23$ ;  $SE = 0.65$ ,  $p < .001$ ), less competent at the property ( $b = -13.51$ ;  $SE = 0.67$ ,  $p < .001$ ), and thought it was less permissible for them to do the property ( $b = -6.17$ ;  $SE = 0.39$ ,  $p < .001$ ) than the group who *does* the property (prevalence condition x group interaction: Wald  $\chi^2(1) = 360.34$ ,  $p < .001$ ; interest condition x group interaction: Wald  $\chi^2(1) = 149.57$ ,  $p < .001$ ; competence condition x group interaction: Wald  $\chi^2(1) = 156.25$ ,  $p < .001$ ; permissibility condition x group interaction: Wald  $\chi^2(1) = 83.84$ ,  $p < .001$ ; see Figure 4). The difference in judgments of the first and second mentioned groups for the “do & can” condition were between 5.8 and 10.8 times larger than the difference in judgments for the “do & too” conditions, suggesting that “can” is providing most of the inferential power (too–do lower prevalence,  $b = -1.69$ ;  $SE = 0.62$ ,  $p = .006$ ; lower interest,  $b = -1.91$ ;  $SE = 0.66$ ,  $p =$

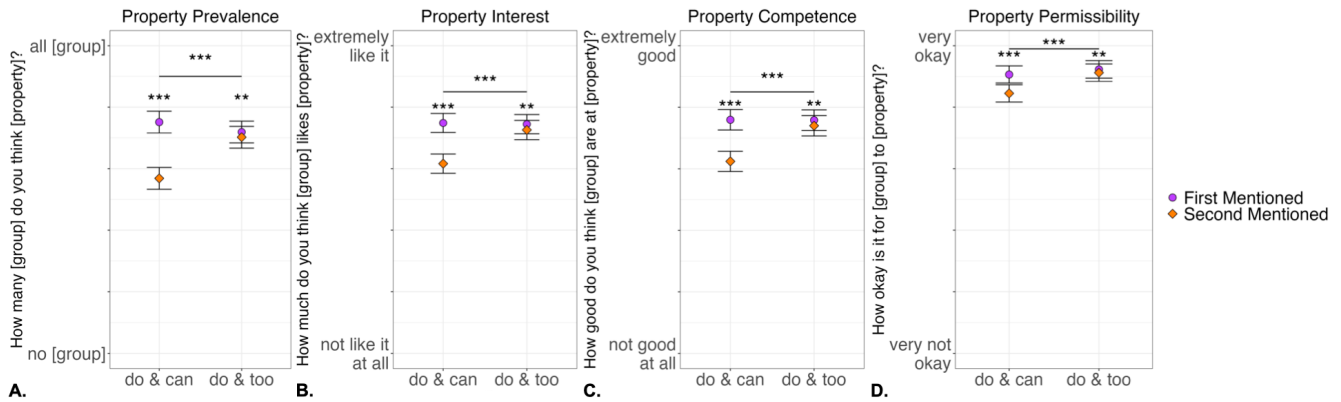


Figure 4: Experiment 4 marginal means with 95% confidence intervals. \*\*\*  $p < .001$ ; \*\*  $p < .01$ .

.004; lower competence,  $b = -1.81$ ;  $SE = 0.68$ ,  $p = .007$ ; lower permissibility,  $b = -1.06$ ;  $SE = 0.40$ ,  $p = .008$ ).

## Discussion

Generic statements convey that properties are characteristically true of the groups they reference and are a common way of communicating stereotypes. Here, we examined whether one natural response to these statements—that the unmentioned group can also do the property (e.g., “boys can wear makeup too”)—equates judgments about how often the groups do the property, how interested they are in the property, how competent they are at the property, and how permissible it is for them to do the property. Across all four studies, we found that they do not. Indeed, hearing a *can* generic after a *do* generic resulted in lower judgments of property prevalence, interest, competence, and permissibility. This difference was not due to the order in which groups are mentioned—when the property-group links are tied with *do* generics for both groups, there are little-to-no differences in judgments between the first and second mentioned groups.

We found that *can* generics are generally interpreted as describing weaker property-group links than *do* generics. Key to our argument, this difference was magnified when adults heard a *can* generic after a *do* generic. This is likely because the option to have used a *do* generic for both groups is salient, making the choice not to do so appear informative (e.g., Degen, 2023). These data suggest that in the effort to undermine stereotypes, *can* generic statements in response to stereotypical *do* generic statements instead highlight the difference between groups. An important caveat is that all these data were collected with novel groups, so adults did not hold prior beliefs about property-group links for either group. In cases where adults already hold stereotypes (e.g., “girls wear makeup”), the stereotype may function in a similar way as the *do* generic—a baseline yardstick by which the *can* generic is measured. Or, alternatively, *do* generic statements about known property-group ties (e.g., long held stereotypes), may not be able to be changed by a single equivalent generic statement for the second mentioned group

as responding to one generic statement with another generic statement can be fraught (e.g., Foster-Hanson & Rhodes, 2022). Further work is required to understand how these statements are interpreted in real-world contexts where people already hold beliefs about property-group links.

Interestingly, the lower judgments for groups described with *can* generics rather than *do* generics held across a variety of types of properties (activities, appearances, traits) and when presented as simple sentences or as conversations between two knowledgeable speakers, highlighting the potential generalizability of the current phenomenon.

The present work contributes to the growing body of literature examining how well-meaning ways of increasing equality often backfire (Caleo & Heilman, 2019; Chestnut & Markman, 2018; Chestnut et al., 2021; Rhodes et al., 2020).

While we proposed that speakers use *can* generics in an attempt to equate two groups, it is also possible that people selectively use *can* generics to purposefully convey weaker property-group links. For example, a speaker who says that “boys can wear makeup too” may intend to convey that it is less common for boys to wear makeup, that boys are less interested in makeup, that boys are less competent at makeup and that it is less permissible for them to wear makeup than girls. Future work is required to understand when and why speakers use *can* generics.

The current findings lay the foundation for future developmental work. Generic statements are commonly said to children (e.g., Gelman et al., 2008; Gelman et al., 2014), and children reason pragmatically about similar kinds of generic statements (Chestnut et al., 2021; Moty & Rhodes, 2021). It is therefore plausible that hearing *can* generics in response to *do* generics is another way in which children learn stereotypes. Therefore it is important to understand the frequency of *can* generics in response to *do* generics in child-directed speech and media.

## Acknowledgments

This research was supported in part by a National Science Foundation Graduate Research Fellowship under Grant DGE-2234660 awarded to SHA.

## References

- Baharloo, R., Vasil, N., Ellwood-Lowe, M. E., & Srinivasan, M. (2023). Children's use of pragmatic inference to learn about the social world. *Developmental Science*, 26(3), e13333.
- Caleo, S., & Heilman, M. E. (2019). What could go wrong? Some unintended consequences of gender bias interventions. *Archives of Scientific Psychology*, 7(1), 71.
- Chestnut, E. K., & Markman, E. M. (2018). "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive Science*, 42(7), 2229-2249.
- Chestnut, E. K., Zhang, M. Y., & Markman, E. M. (2021). "Just as good": Learning gender stereotypes from attempts to counteract them. *Developmental Psychology*, 57(1), 114.
- Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, 9, 519-540.
- DeJesus, J. M., Callanan, M. A., Solis, G., & Gelman, S. A. (2019). Generic language in scientific communication. *Proceedings of the National Academy of Sciences*, 116(37), 18370-18377.
- Foster-Hanson, E., Leslie, S. J., & Rhodes, M. (2022). Speaking of kinds: How correcting generic statements can shape children's concepts. *Cognitive science*, 46(12), e13223.
- Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development*, 4(1), 1-31.
- Gelman, S. A., & Tardif, T. (1998). A cross-linguistic comparison of generic noun phrases in English and Mandarin. *Cognition*, 66(3), 215-248.
- Gelman, S. A., Ware, E. A., Kleinberg, F., Manczak, E. M., & Stilwell, S. M. (2014). Individual differences in children's and parents' generic language. *Child Development*, 85(3), 924-940.
- Gibbs Jr, R. W., & Moise, J. F. (1997). Pragmatics in understanding what is said. *Cognition*, 62(1), 51-74.
- Holmes, K. J., Doherty, E. M., & Flusberg, S. J. (2022). How and when does syntax perpetuate stereotypes? Probing the framing effects of subject-complement statements of equality. *Thinking & Reasoning*, 28(2), 226-260.
- Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive psychology*, 58(3), 376-415.
- Kramer, H. J., Goldfarb, D., Tashjian, S. M., & Lagattuta, K. H. (2021). Dichotomous thinking about social groups: Learning about one group can activate opposite beliefs about another group. *Cognitive Psychology*, 129, 101408.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package "Emmeans". R Package Version 4.0-3. <http://cran.r-project.org/package=emmeans>
- Leslie, S. J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, 117(1), 1-47.
- Moty, K., & Rhodes, M. (2021). The unintended consequences of the things we say: What generic statements communicate to children about unmentioned categories. *Psychological Science*, 32(2), 189-203.
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rhodes, M., Cardarelli, A., & Leslie, S. J. (2020). Asking young children to "do science" instead of "be scientists" increases science engagement in a randomized field experiment. *Proceedings of the National Academy of Sciences*, 117(18), 9808-9814.
- Rhodes, M., Leslie, S.J., & Tworek, C. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, 109: 13526-13531.
- Salomon, E., & Cimpian, A. (2014). The inherence heuristic as a source of essentialist thought. *Personality and Social Psychology Bulletin*, 40(10), 1297-1315.
- Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological review*, 126(3), 395.