



Social Psychology

Ladies First or Ladies Last: Do Masculine Generics Evoke a Reduced and Later Retrieval of Female Exemplars?

Nina Keith¹ ^a, Kristine Hartwig¹, Tobias Richter² ^b¹ Department of Psychology, Technical University of Darmstadt, Darmstadt, Germany, ² Educational Psychology, University of Würzburg, Würzburg, Germany

Keywords: androcentrism, sexism, language, gender, Gendersternchen

<https://doi.org/10.1525/collabra.32964>

Collabra: Psychology

Vol. 8, Issue 1, 2022

The use of masculine generics (i.e., grammatically masculine forms that refer to both men and women) is prevalent in many languages but has been criticized for potentially triggering male bias. Empirical evidence for this claim exists but is often based on small and selective samples. This study is a high-powered and pre-registered replication and extension of a 20-year-old study on this biasing effect in German speakers. Under 1 of 4 conditions (masculine generics vs. three gender-inclusive alternatives), 344 participants listed 3 persons of 6 popular occupational categories (e.g., athletes, politicians). Despite 20 years of societal changes, results were remarkably similar, underscoring the high degree of automaticity involved in language comprehension (large effects of 0.71 to 1.12 of a standard deviation). Male bias tended to be particularly pronounced later rather than early in retrieval, suggesting that salient female exemplars may be recalled first but that male exemplars still dominate the overall categorical representations.

Contemplating on the boundary of drama and play, Charles Dickens once opined: “Every writer of fiction, though he may not adopt the dramatic form, writes in effect for the stage” (Crystal, 1997, p. 76). In this statement, was Dickens referring to male writers only or to female and male writers alike? Or is this an idle question because Dickens had no knowledge of any significant female writers? (If so, he would have been unaware of Jane Austen or the Brontë sisters, which may seem rather unlikely.)

As this quote illustrates, generically masculine linguistic expressions, that is, grammatically masculine forms that can refer to both men and women (masculine pronoun *he* in the above example) may at times be ambiguous. Yet, such forms are prevalent in many languages. In English, for example, besides generically used masculine pronoun *he*, some personal nouns explicitly include information on a person’s gender (e.g., queen vs. king, actress vs. actor, witch vs. wizard) while others do not (e.g., consumer, manager, teacher, neighbor). Many other languages, including several Germanic, Slavic, and Italic languages, convey gender information in many personal nouns and in corresponding grammatical structures, with a generic use of masculine forms to denote people of both genders. In German, for example, a male politician is denoted *Politiker* while a female politician is denoted *Politikerin* in singular form, marked with

the morphological affix *-in*. But *Politiker* can also be used generically to refer to a politician of any gender, such as in the phrase *der Beruf des Politikers* (the profession of a politician). Likewise, the plural of the male form, *die Politiker*, may refer to either a group of only male politicians or generically to a mixed group of politicians of any gender (but not to a group of only female politicians, which would be *die Politikerinnen*).

A common assumption is that masculine generics may evoke an increased cognitive accessibility of men to the detriment of women. Consequently, the visibility of women is decreased and, according to feminist reasoning, inequalities that already exist between genders in society will not be counteracted but may even be reinforced by this kind of language use (Tavits & Pérez, 2019). In line with this criticism, official guidelines in many languages and countries recommend the use of gender-inclusive (e.g., *he or she*) or gender-neutral (e.g., singular *they*; *spokesperson* in lieu of *spokesman*) linguistic expressions (for examples, see the bias-free language guidelines of the American Psychological Association, 2020, Ch. 5; or the guidelines on gender-neutral language of the European Parliament, 2018). Such endeavors, however, often remain limited to communication in official documents and regulated communication, while everyday use of explicitly gender-inclusive language

^a <https://www.orgpsy.psychologie.tu-darmstadt.de/>

Correspondence concerning this article should be addressed to Nina Keith, Department of Psychology, Organizational and Business Psychology, Technical University of Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany. Email: keith@psychologie.tu-darmstadt.de

^b <https://go.uniwiue.de/richter>

is less common (Gabriel et al., 2018; MacArthur et al., 2020; Vervecken & Hannover, 2012). Rather, in public opinion, the use of gender-inclusive language is often criticized for being cumbersome, confusing, or even misleading (Sczesny et al., 2016; for overviews of common criticisms, see Gabriel et al., 2018; Parks & Robertson, 1998; Vergoossen et al., 2020).

Empirically, a number of studies have shown influences of masculine generics on cognitive information processing, using a range of methods and in several languages (for reviews, see Henley & Abueg, 2003; Sczesny et al., 2016). Below, we will describe in more theoretical terms why we propose masculine generics to affect cognitive accessibility of female and male exemplars of a category. Before doing so, we will first describe the original study that the present research seeks to replicate and extend. We will also explain why we deemed replication of this particular study by Stahlberg et al. (2001, Study 2) worthwhile—a study which we think used a simple but elegant design, with a realistic and unobtrusive task.

The Original Study by Stahlberg et al. (2001) and why Replication is Worthwhile

The present high-powered and pre-registered study seeks to contribute to research on cognitive effects of masculine generics by replicating and extending one particular study, namely, a roughly 20-years old study on the effects of masculine generics in German speakers (Stahlberg et al., 2001, Study 2; Stahlberg & Sczesny, 2001, Study 3, is the same study). In this study, participants (45 female and 45 male native speakers of German) were asked to name three famous personalities in each of four categories (e.g., singers, politicians, hosts of TV shows). Stahlberg et al. found that significantly more (almost twice as many) female names were listed by participants when women and men were explicitly referenced with the so-called capital-I (*Binnen-I* in German), which replaces lower-case i of feminized forms by a capitalized generic I (e.g., *PolitikerInnen* to refer to both female and male politicians), compared to the masculine generic form (e.g., *Politiker*). In contrast, masculine-feminine word pairs (e.g., *Politiker und Politikerinnen*) did not elicit significantly more female exemplars than the masculine generic form. Participants' gender also had an effect on the number of female names listed (i.e., less women were listed by male than by female participants).

For several theoretical, practical, and methodological reasons, we deemed replication (and extension) of this study worthwhile. First, psychological research, in general, has been criticized for sometimes using small and potentially underpowered samples (for a meta-analysis highlighting this problem, see Stanley et al., 2018). Notwithstanding the strengths of the original study, its sample size was so small (30 participants per condition) that the sensitivity for comparisons of each of the two linguistic forms with masculine generics was rather low, allowing for a reliable detection of large effects only (Cohen's $d > 0.75$, under the assumptions of a Type I error probability of .05 and a power of .90). In fact, in Stahlberg et al. (2001), there was no difference between masculine-feminine word pairs and masculine generics with regard to the female exemplars listed.

This result is inconsistent with studies that do find such a difference (Brauer, 2008; Gabriel, 2008; Gabriel & Mellenberger, 2004). This inconsistency of results may well be due to insufficient power. The present study uses a sample that is more than three times larger than the original one, which allows for the detection of medium to small effects, too.

Second, research on language effects such as masculine generics has been criticized for using highly selective samples, often involving laboratory studies with student participants (Tavits & Pérez, 2019). The study by Stahlberg et al. (2001), too, used a selective sample, with most participants being university students. The present replication used a less selective sample (only 14% of participants were university students).

Third, no direct replication of the focal study exists, although the study is well-known and cited in the field. A handful of studies exist that used a somewhat similar task (i.e., to list exemplars of an occupational category). These studies, however, do not constitute replications because they used specific instructions and different manipulations (e.g., participants were asked to list their *favorite vs. least-liked* personalities, personal *heros*, or *good* prime ministers of the political *left vs. right*; Brauer, 2008; Gabriel, 2008; Gabriel & Mellenberger, 2004; see also Sniezek & Jazwinski, 1986). Also, as mentioned above, the studies' results did not fully converge, which may be due to the differences in study designs and manipulations. Another set of studies used still other methods such as sentence evaluations, ratings of sentence correctness, reading times, or eye-tracking (e.g., Gygas et al., 2008; Irmen & Kurovskaja, 2010; Irmen & Roßberg, 2004). These studies are interesting in themselves, as they demonstrate an influence of masculine generics on specific aspects of cognition with different paradigms and measures, but they do not constitute replications of the study by Stahlberg et al. (2001). Successful replications, however, particularly replications of effects that were originally found in underpowered studies, are needed to elevate an effect from single observation to scientific evidence (Zwaan et al., 2018; see also Cesario, 2014).

Fourth, there is a renewed public interest in feminist topics around the world (e.g., Chiu, 2020; MacNicol, 2020) and, at least in some countries, the public debate includes the potentially detrimental role of masculine generics in language use (e.g., Burgen, 2020). Empirical research that systematically analyzes feminists' claims using more recent data may contribute scientifically to this debate.

Fifth, two decades have passed since the publication of the original study and there have been considerable societal changes since then. For example, work force participation of women has increased and so has the number of female leaders in the past decades in the United States and in Europe, although in many countries, women are still largely underrepresented in leading political and economic positions (Begeny et al., 2020; European Commission, 2019; Inter-Parliamentary Union, 2015; U.S. Bureau of Labor Statistics, 2020; Warner et al., 2018). Still, the increased visibility of women in influential positions (e.g., former chancellor *Angela Merkel* in Germany) may have increased the cognitive accessibility of female exemplars and decreased potentially biasing effects of masculine generics. Considering the societal changes, it seems not only important to replicate

the original effects obtained by Stahlberg et al. (2001) but also to examine whether potential biases in the cognitive accessibility exemplars are similar in magnitude to the biases found two decades ago.

Finally, rather than focusing solely on a *quantity* effect of masculine generics (i.e., accessibility of *more* men than women) as has been the focus in previous work, we sought to explore *order* effects as well. If order effects were present, that is, if there was a dynamical change in decreased accessibility of women in the order of exemplars recalled, this may provide some insight as to the processes underlying the effect of masculine generics that has been found in earlier studies.

Language and Cognitive Accessibility of Female and Male Exemplars

Masculine generics may increase the accessibility of male exemplars via mental models that are constructed during language comprehension (Kollmayer et al., 2018). Mental models (also called situation models) are cognitive representations of the situations and events described in a text (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). These representations contain elements of the text, relations among them, as well as their perceptual and other qualities (Johnson-Laird, 1983). For instance, readers of the sentence “Three turtles rested on a floating log, and a fish swam beneath them” (Bransford et al., 1972, p. 195, original text includes italics) would construct a mental model that contains not only three turtles, a log, and a fish, but also, for example, the relative positions of these elements, the speed at which they move, possibly additional qualities of these elements (e.g., colors and shapes of the turtles, log, and fish), and, although not explicitly mentioned in the sentence, a pond or some other stretch of water where the described scenario may plausibly take place.

As this example illustrates, readers, or more generally speaking, comprehenders, use their preexisting knowledge (e.g., that fish swim in water and that the action of floating implies water) when building their mental model. Importantly, they need not consciously or strategically decide to do so, but parts of their preexisting knowledge are activated automatically by verbal cues, words, phrases or sentences in the text, as the reader moves through the text (Cook et al., 1998; Richter & Singer, 2018). As another important point, the preexisting knowledge that is activated and used in mental model construction can be more or less accurate. When people are described in a text, the knowledge used to construct mental models during comprehension may also involve social stereotypes, including stereotypic gender information (Carreiras et al., 1996; Kollmayer et al., 2018). For example, when reading a sentence about a doctor, people may construct a mental model that includes a male doctor, whereas when reading a sentence about a nurse, they more likely construct a mental model that includes a female nurse (Carreiras et al., 1996). As these examples illustrate, cues in the text and internal factors within the comprehenders (e.g., general world knowledge, stereotypes) act in concert to shape their mental models. Masculine generics, then, may function as a linguistic cue that activates associations with male exemplars more than with female exem-

plars. This increased activation of associations with male exemplars occurs irrespective of any intentions the user of masculine generics (e.g., speaker, writer) may or may not hold to refer to women as much as to men. This is because grammatically masculine forms make masculinity particularly salient for comprehenders who, as a consequence, become prone to male bias (i.e., over-association of men), when forming their mental model.

Empirically, a number of studies have demonstrated effects of masculine generics that align with this reasoning. These studies used a range of experimental methods and outcome measures (e.g., reading times, reaction times, participants’ use of pronouns and nouns in stories about fictitious persons, estimations of proportions of women and men in a given population; for an overview, see Sczesny et al., 2016). In line with this research, and with the original study that we seek to replicate, we expect:

Hypothesis 1. Masculine generics lead to a lower cognitive accessibility of female exemplars (compared to male exemplars) than alternative, gender-inclusive forms of linguistic expression.

Gender Differences in Cognitive Accessibility of Female and Male Exemplars

As described above, both external factors (e.g., informational cues such as linguistic expressions) and internal factors (e.g., characteristics of comprehenders) may shape mental models. With regard to cognitive accessibility of female and male exemplars, one such factor may be comprehenders’ gender. Previous research has shown a gender difference in androcentric bias, that is, the tendency to “construe men as more typical than women” (Bailey et al., 2018, p. 313) for human categories that technically include both men and women. Specifically, although androcentrism appears to be universal, men tend to show more androcentric bias than women, at least on some tasks (for an overview, see Bailey et al., 2018, 2020). This gender difference may be a result of ingroup favoritism (Tajfel & Turner, 1979), which leads women more (and men less) to reduce androcentric bias. This gender difference may also be the result of differences in frequency of instantiation (i.e., the phenomenon that frequent exposure to an exemplar of a category leads to this exemplar being judged as more typical; Barsalou, 1985). Specifically, although men and women may principally be exposed to equal numbers of women and men in the public and in the media, in more personal matters (e.g., friendship and close work relationships, preferences in media consumption), men may still be more exposed to men and women more to women (Bailey et al., 2018). In line with this reasoning, and with the findings of the original study that we seek to replicate, we expect:

Hypothesis 2. There is a gender difference in cognitive accessibility of female and male exemplars. Accessibility of female exemplars will be lower in men than in women.

When Does Decreased Cognitive Accessibility of Female Exemplars Occur?

Up to this point, we have focused on theory and findings

Table 1. Modifications of and Extensions to Original Study by Stahlberg et al. (2001, Study 2)

	Original study	Present study
Study design	3 (Linguistic expression of gender: masculine generics, feminine-masculine word pair, capital-I) x 2 design (Participant gender: female, male)	4 x 2 design; additional level of factor Linguistic expression of gender: gender asterisk (*)
Participants	90 participants (45 women, 45 men), most of them university students	344 participants, 55% women, 14% university students
Material	Participants were asked to list 12 persons (3 persons in 4 categories): athletes, singers, politicians, TV hosts	Participants were asked to list 18 persons (3 persons in 6 categories), the two additional categories were: actors, authors
Dependent variable	Number of women listed by participants	(1) Proportion of women listed by participants (relative to overall number of persons listed by participants) (2) Probabilities to list a woman
Statistical analyses	Between-subjects ANOVA, post-hoc comparisons	(1) Between-subjects ANOVA, post-hoc comparisons (2) General Linear Mixed Models (with logit-link) predicting the probability to list a woman

indicating *that* male bias (i.e., a decreased accessibility of female exemplars) may occur; we have not yet discussed at what point in time during retrieval this bias occurs. Given the high degree of automaticity involved in reading and comprehension in general and in the activation of preexisting knowledge during comprehension (Cook et al., 1998; Richter & Singer, 2018), one possibility is that the bias occurs early during retrieval. For example, in the task to come up with names of persons of a certain category (e.g., politicians, musicians), masculine generics may immediately trigger associations with male exemplars, while only at second thought, one may come up with names of women as well. As a consequence, overall, there will be more men than women listed. In other words, the overall over-association of men that was found in earlier studies may in fact be the result of early processes during retrieval that do not necessarily persist throughout later stages of retrieval.

Another possibility is that while female exemplars may be less frequent and less typical in the representation of a category, they are not necessarily less salient. Female exemplars, despite being less frequent, might even be more salient because of their atypicality, as women's minority status can contribute to increased saliency (Heilman, 2012; Kanter, 1977). If this reasoning holds, one might expect a male bias in the overall frequency of listed male vs. female exemplars but not necessarily an order effect with male exemplars listed first. Rather, female exemplars may even be listed first due to their increased saliency. We conclude that based on existing theory and research, we cannot draw definite conclusions regarding potential order effects during retrieval. We, therefore, put forth open research questions:

Open Research Question 1. When does decreased cognitive accessibility of female exemplars occur (e.g., early or late in the process of retrieval)?

Open Research Question 2. If decreased cognitive accessibility of female exemplars changes over time, does this change interact with (a) effects of masculine gener-

ics or (b) gender differences in accessibility?

Method

This study is a replication and extension of a study that was conducted about 20 years ago (Stahlberg et al., 2001, Study 2) and we will describe the methods with reference to the original study. Table 1 additionally lists the major similarities and differences between the original study and this replication. This study was preregistered on the platform of the Open Science Framework (<https://osf.io/te9c3>). The preregistration includes the description of the expected effects, details on the planning of the required sample size, criteria for excluding participants, analysis plan, and other methodological details. This method section discloses all measures and manipulations used. Data analyses did not commence until data collection was completed. The data and analysis scripts are available at https://osf.io/rn96f/?view_only=4436a99bf52740ef84a97a4951c414eb.

Design

The study used a 4 x 2 between-subjects design. The first, experimentally manipulated, factor was *Linguistic expression of gender*. The first three levels of this factor were identical to those used in the original study, namely, (1) Masculine generics (e.g., grammatically masculine *Politiker* to refer to both female and male politicians); (2) feminine-masculine word pairs (e.g., *Politikerinnen*_{feminine} or *Politiker*_{masculine}); and (3) Capital-I (*Binnen-I* in German), which replaces lower-case i of feminized forms by a capitalized generic I (e.g., *PolitikerInnen* to refer to both female and male politicians). These forms all constitute formally correct or at least accepted (and sometimes criticized) forms of expression of gender information that are probably known by all (Forms 1 and 2) or by most (Form 3) German speakers. As a fourth condition, we included the so-called gender asterisk (*Gendersternchen*), which is a relatively recent creation that places an asterisk between the grammatically

male stem of a noun and its feminization (e.g., *Politiker*innen* to refer to both female and male politicians). This form is explicitly meant to include not only women and men but also people who do not identify with binary gender. The term Gendersternchen has attracted some attention in German media, as part of the debate on gender-fair language use. In 2018, the term was awarded the publicly well-known German Anglicism-of-the-year award by an independent initiative, chaired by a German linguist of Freie Universität Berlin (Gendersternchen is an Anglicism because 'Gender' is not a German but an English term; 'Sternchen', which refers to the asterisk, is a German term which roughly translates *little star* or *starlet*). The second factor was *Participants' gender*, an organismic factor with the two levels male and female participants. For additional analyses regarding the exploratory research questions, we also included the *Position* at which a name was mentioned by participants as additional within-subject factor with three levels (Position 1, Position 2, Position 3).

Participants, Sample Size, and Power

The original study included 90 participants (45 women and 45 men), mostly university students. In the present study, we sought to employ a larger and more heterogeneous sample, with a higher inclusion of non-student participants. For this purpose, we approached persons at public places in the center of a mid-sized German city and additionally recruited participants via acquaintances and other personal contacts of one of the authors. Participation was voluntary and was not compensated except for the possibility to win one of five vouchers for a movie theater.

When determining the sample size, we sought, first, to arrive at a sample that is substantially larger and less selective than the sample of the original study and, second, to have enough power to detect a medium to small effect. In particular, we sought to achieve reasonable power for analyses involving simple contrasts between experimental conditions (masculine generics vs. other linguistic expressions of gender). For detecting a medium to small effect for these contrasts ($d = 0.40$), with a Type I error probability of $\alpha = .05$ and a power of $1 - \beta = .80$, a sample size of 312 participants (78 per group) is required (according to G*Power; Faul et al., 2007). This sample size is also sufficient to detect a medium to small effect ($f = .19$) for the overall effect of linguistic expression of gender. We opted to recruit a larger sample, however, to account for potential data losses (which we expected, as described in the preregistration).

Of the originally recruited 400 participants, the data of 3 participants could not be used due to incompleteness. Additional 27 participants (6.75%) correctly guessed the study purpose, 5 participants of the generic-masculinum condition deliberately rewrote the instructions to include grammatically female forms, and 1 participant of the generic-masculinum condition erroneously thought that the generic masculinum referred to men only (which may be a consequence of the use of masculine generics; Miller & James, 2009). As preset in our preregistration, these participants' data were also deleted from further analyses, resulting in a sample size of 364. Again as preset in our preregistration, we further excluded data sets of 20 participants of the cap-

ital-I condition who exclusively named women. We did so because these participants may have misread the capital I as a lower-case i, in which case they would have erroneously thought that only women are referred to in the experimental material. (Note that this exclusion works against our hypotheses; we also reran analyses without this exclusion and found the pattern of results to be the same.)

The resulting sample of 344 participants was aged between 14 and 79 years, with 55% women and 44% men. Two participants did not disclose their gender and two participants did not identify with binary gender. These participants' data were excluded from all analyses that involved participant gender, resulting in a sample size of 340 for all analyses involving participant gender. Only 14% of participants were university students; 44.7% indicated to hold a university degree; 81.1% reported to be employed. According to a sensitivity power analysis with G*Power (Faul et al., 2007), this sample was sufficiently large to detect a small effect both for the experimentally manipulated factor (i.e., linguistic expression of gender, as stated in Hypothesis 1, $f = .18$) and for the organismic factor (i.e., participant gender, as stated in Hypothesis 2, $f = .15$) (with a Type I error probability of $\alpha = .05$ and a power of $1 - \beta = .80$).

Material and Procedure

The study was a paper-and-pencil survey that consisted of three parts. In Part 1, participants were asked to respond to eleven questions concerning their media behavior (e.g., how often they watch TV, how often they watch videos online, how often they read newspapers, etc.). This was done to conceal the actual study goal from participants, that is, to give participants the impression that this study was about media consumption (the same procedure was used in the original study). Part 2 was the focal part of the study in which the manipulation of the experimental variable and assessment of the dependent variable took place. Under 1 of the 4 experimental conditions, participants were asked to list those three athletes, singers, actors, TV hosts, politicians, and authors who would come to their minds first. The original study did not include actors and authors but only four categories (Table 1). We included the two additional categories to achieve a broader and more reliable measurement of the dependent variable. Participants were randomly assigned to experimental conditions by placing the surveys in random order before distributing them. As a result, experimenters were blind to conditions as well. All experimenters distributed surveys of all experimental conditions. In Part 3, additional variables were collected for exploratory purposes and demographic data was ascertained.

Measures

Dependent Variable

In the original study, the dependent variable was the overall number of women listed by participants. This variable, in the present study, has a theoretical range from 0 to 18 (the empirical range was 0 to 16; $M = 5.54$, $SD = 2.93$), because participants could list a maximum of 18 persons (i.e., 3 persons x 6 occupational categories). Yet, we decided not

Table 2. Means, Standard Deviations, and Intercorrelations of Study Variables

	M	SD	1	2	3	4	5	6
<i>Demographics</i>								
1 Age	39.12	16.04						
2 Gender ^a	0.56	0.50	-.12*					
3 Student ^a	0.14	0.35	-.40**	.07				
4 First language ^a	0.93	0.25	-.02	.09	.04			
<i>Experimental factor</i>								
5 Gender-inclusive ^a	0.75	0.44	-.03	.04	.02	-.02		
<i>Dependent variable</i>								
6 Proportion of women (in %)	33.37	17.19	-.03	.38**	-.01	.05	.35**	
7 Number of women	5.53	2.93	.02	.37**	.02	.08	.34**	.93**

Note. 340 < N < 344 (of 344 participants, 2 did not disclose their gender and another 2 did not identify with either female or male gender, resulting in 340 participants for all analyses involving gender).

^a Variables were coded as follows: Gender was coded 1 for women and 0 for men; Student was coded 1 for student and 0 for non-student participant; First language was coded 1 for German and 0 for other language; Experimental factor was coded 1 for gender-inclusive conditions (i.e., feminine-masculine word pairs, capital-I, or gender asterisk) and 0 for masculine generics.

* $p < .05$. ** $p < .01$.

to use the overall number of women listed by participants as main dependent variable but the ratio of women listed by participants, relative to all persons listed by participants. We did so for two reasons. First, we observed that not all participants were able to come up with 18 persons (for example, some were able to list only 2 actors or only 1 author; although the average was close to 18, at 16.50, $SD = 2.59$) and the ratio (rather than the number) of women appears to be the more accurate measure in these cases. Second, the ratio or percentage is readily interpretable. For example, a value of 33.3% indicates that one-third of the persons listed were women. (We also reran all analyses with the number of women as dependent variable and found highly similar results, which is not surprising because number and ratio are partially redundant and empirically correlated at .93, $p < .01$, see [Table 2](#).)

Additional Variables

For exploratory purposes, we assessed participants' self-reported gender role (German version of the Bem Sex Role Inventory by Troche & Rammsayer, 2011) and attitude towards the use of gender-fair language (self-developed single item measure). We will not report on these variables in the remainder of this manuscript because they did not reveal any noteworthy insights.

Results

We analyzed the data using two analytical strategies. The first analytical strategy closely resembled the one used in the original study and uses classical between-subjects ANOVA as well as post-hoc comparisons between groups; the unit of analysis was the participant and independent variables were linguistic expression of gender and participants' gender. The second analytical strategy treated participants' binary responses to the occupational roles (list-

ings of female vs. male exemplars) as nested within participants and participants nested within occupational roles and, hence, used a Generalized Linear Mixed Model (GLMM) with a logit link function and crossed random effects (Dixon, 2008; Jaeger, 2008). Being based on item-by-item analyses rather than aggregated data, the GLMM allows for estimating the effects of the position (first, second, or third position) at which a male or female exemplar was listed, in addition to the effects of linguistic expression of gender and participants' gender. By including the random effects of participants and occupational roles, the GLMM also accounts for the fact that both participants and the occupational roles are samples drawn from larger populations, avoiding a problem known as "language-as-fixed-effect fallacy" (Clark, 1973) in psycholinguistics and enhancing the validity of statistical inferences (Baayen et al., 2008). For the GLMM analysis, we used the packages lme4 (Bates et al., 2015) and lsmeans in the R-environment (R Core Team, 2020). In the following, we will report the results of these two analytical strategies consecutively. For descriptive statistics and intercorrelations of study variables, see [Table 2](#).

Effects of Linguistic Expression of Gender and of Participant Gender on Proportion of Listed Women

Hypothesis 1 predicted linguistic expression of gender to affect the proportion of women listed by participants; in particular, we expected participants of the masculine-generics condition to list less women than participants of the three gender-inclusive conditions (i.e., feminine-masculine word pair, capital-I, or gender asterisk). We further sought to explore whether this effect is stronger for the capital-I than the word-pair conditions, as has been found in the original study by Stahlberg et al. (2001). Hypothesis 2 further predicted participant gender to affect the propor-

Table 3. Means and Standard Deviations of Dependent Variable (Percentage of Women Listed) in Experimental Groups as well as Post-hoc Comparisons with Masculine Generics (Experimental Condition) and With Female Participants (Participant Gender)

	<i>M</i>	<i>SD</i>	<i>N</i>	<i>t(df)</i>	<i>d</i>
<i>Overall</i>	33.37	17.19	344		
<i>By experimental condition</i>					
Masculine generics	23.04	12.66	87	—	—
Feminine-masculine word pairs	32.46	13.67	95	4.00(181)***	0.71
Capital-I	40.58	18.71	72	6.93(159)***	1.12
Gender asterisk	38.55	18.24	90	6.49(161)***	0.99
<i>By participant gender</i>					
Female Participants	39.21	16.59	189	—	—
Male Participants	26.01	14.95	151	8.42(339)***	0.84

Note. *N* = 340 for analyses involving participant gender.

*** $p < .001$.

tion of women listed, in that female participants would list more women than would male participants.

We first tested Hypothesis 1 in a one-factorial between-subjects ANOVA¹, with four levels of Factor *Linguistic expression of gender* (masculine generics, feminine-masculine word pair, capital-I, and gender asterisk). We found a significant and large main effect, $F(3,340) = 20.50$, $p < .001$, $\eta_{\text{partial}}^2 = .15$. In line with what we expected based on the original study by Stahlberg et al. (2001), participants of the conditions with gender-inclusive expressions listed more women than those of the masculine-generics condition, $t(343) = 7.17$, $p < .001$, Cohen's $d = 0.86$. The effect is illustrated in Figure 1. Post-hoc analyses further revealed all comparisons with the masculine-generics condition to be substantial, with Cohen's d effect sizes ranging from 0.71 to 1.12 (Table 3). Among the gender-inclusive expressions, the capital-I and the gender-asterisk condition led to an equally high proportion of women listed by participants, $t(161) = 0.81$, $p = .419$, with both conditions differing from the feminine-masculine word-pair condition, $t(256) = 3.45$, $p < .001$. In sum, masculine generics affected the proportion of women listed by participants in the expected way, with the effect being more pronounced for the capital-I and gender-asterisk conditions than for the word-pair condition.

On an additional note, in all experimental conditions, the proportion of women listed was substantially below a hypothetical parity of 50%. This is illustrated by the parity line in Figure 1 (all values and standard errors fall substantially below the parity line). This observation was also confirmed in one-sample t -tests that tested the values in experimental conditions against the hypothetical parity value of

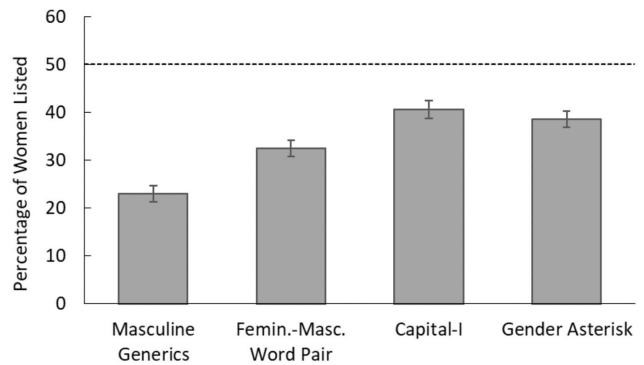


Figure 1. Percentage of Women Listed as a Function of Linguistic Expression of Gender (Masculine Generics vs. Three Gender-Inclusive Forms)

Note. Femin.-Masc. = Feminine-Masculine. Error bars represent standard errors. Dashed line denotes parity (50%).

50% (for all comparisons, $p < .0001$).

We tested Hypothesis 2 in a two-factorial between-subjects ANOVA, with participant gender as second factor, in addition to the experimental factor Linguistic expression of gender. As in the one-factorial ANOVA reported above¹, we found a significant and large effect of the experimental factor, $F(3,332) = 23.99$, $p < .001$, $\eta_{\text{partial}}^2 = .18$. We also found a significant and large effect of participant gender, $F(1,332) = 70.92$, $p < .001$, $\eta_{\text{partial}}^2 = .18$, but no interaction of the two factors, $F(3,332) = 1.18$, $p = .316$, $\eta_{\text{partial}}^2 = .01$. In line with what we expected based on the original study by Stahlberg et al. (2001), female participants listed more

¹ A two-factorial ANOVA with the second factor participant gender (see test of Hypothesis 2) yielded the same results. We still report the results for the one-factorial ANOVA because they are based on the full sample available (because two participants did not disclose their gender and another two did not identify with binary gender, all analyses involving participant gender are based on a reduced sample of 340 instead of 344 participants).

women than male participants did, with a Cohen's d effect size of 0.84 (Table 3).

Taken together, our results replicate those found by Stahlberg et al. (2001). In particular, (1) the linguistic expression of gender affected the extent to which women were listed by participants, with masculine generics leading to the lowest percentage of women listed; (2) this effect was more pronounced for the capital-I and for the gender asterisks conditions (the latter was not included in the original study) than for the feminine-masculine word-pair condition; (3) there was a large effect of participant gender (female participants listed more women than did male participants); and (4) the two factors did not interact.

As an additional comparison between the original and the present study, we compared the absolute numbers of women listed as a function of experimental condition, based on the four occupational categories included in the original study. We did so by calculating 95% confidence intervals around the mean number of women listed in conditions and by determining whether the values reported in the original study² fell within or outside these confidence intervals, with the following results. In the masculine-generics (95% CI [2.11, 2.91]) and the capital-I (95% CI [4.24, 5.12]) conditions, the values of the original study fell within the confidence intervals (2.37 and 4.72, respectively). In the condition of feminine-masculine word pairs (95% CI [3.20, 3.96]), however, the value of the original study fell substantially below the lower bound of the confidence interval (2.67), indicating a significant difference (i.e., in the original study, there were fewer women listed in the word-pairs condition than in the present study).

Additional Analyses of Order Effects

In addition to testing the hypotheses on the effects of linguistic expression of gender and of participant gender, we explored at what point in time the decreased cognitive accessibility of female exemplars occurs (e.g., early or late in the process of retrieval; Open Research Question 1) and whether the time course of retrieving female exemplars changes as a function of linguistic expression of gender or of participant gender (i.e., interaction effect; Open Research Question 2). To examine these research questions, we first estimated a model that contained only the main effects of linguistic expression of gender, of participants' gender, and of position as fixed effects, plus the random effects (random intercepts) of participants and occupations (Model 1) (9 parameters, AIC = 6678, BIC = 6738, deviance = 6660). We then estimated a model that additionally included the two-way interactions of these predictors (Model 2) (26 parameters, AIC = 6699, BIC = 6871, deviance = 6646). Despite being far more complex than Model 1, Model 2 did not exhibit a better model fit than Model 1, χ^2 ($df = 17$) = 13.66, $p = .691$ (Likelihood-Ratio Test). Moreover, none of the interaction terms in Model 2, including the interactions of position with linguistic expression of gender and with participant gender, was significantly different from zero (for all

tests of interaction terms, $|z| < 1.33$, $p > .183$). We also estimated two simpler models (Models 1a and 2a) in which linguistic expression of gender was included as a two-level factor, masculine generics vs. all other conditions combined. Again, Model 2a that included the interaction terms did not exhibit a better model fit than Model 1a that included only the main effects, χ^2 ($df = 7$) = 7.10, $p = .409$ (Likelihood-Ratio Test) and none of the interaction terms was significant (for all tests of interaction terms: $|z| < 1.84$, $p > .067$). We therefore conclude with regard to Open Research Question 2 that there is no evidence that the time course of retrieving female exemplars is affected by linguistic expression of gender or by participants' gender (i.e., no interaction effect).

To answer Open Research Question 1, we proceeded with the more parsimonious Model 1, which was structured as follows:

$$\ln\left(\frac{p_{female}}{p_{male}}\right) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_4 \cdot X_4 + \beta_5 \cdot X_5 + \beta_6 \cdot X_6 + u_{0Part} + u_{0Occ}$$

In this model, β_0 represents the fixed effect estimate of the intercept. The predictors X_1 , X_2 , and X_3 are dummy-coded variables representing the levels of linguistic expression of gender that compare the reference category masculine generics with masculine-feminine word pairs (X_1), capital-I (X_2), and gender asterisk (X_3), respectively. The coefficients β_1 , β_2 , and β_3 represent the fixed effects of these predictors. The predictor X_4 is participants' gender (dummy-coded, male participants as reference category), and the coefficients β_4 is the fixed effect of this predictor. The predictors X_5 and X_6 are dummy-coded variables representing the position on list that compare Position 2 to Position 1 (X_5) and Position 3 to Position 1 (X_6). The error terms u_{0Part} and u_{0Occ} represent the random effects (random intercepts) of participants and occupations, respectively.

Table 4 (left-hand side) displays the parameter estimates of this model. The main effect of linguistic expression of gender predicted by Hypothesis 1 was again significant, χ^2 ($df = 3$) = 76.39, $p < .001$. Moreover, the main effect of participants' gender predicted by Hypothesis 2 was again significant, χ^2 ($df = 1$) = 60.74, $p < .001$. Additionally, a main effect of position emerged, χ^2 ($df = 2$) = 68.95, $p < .001$. Female exemplars were more likely to be listed at the first position (*Probability* = .390, *SE* = .042), compared to the second position (*Probability* = .289, *SE* = .037) or the third position (*Probability* = .272, *SE* = .036). The effects of the three factors (linguistic expression of gender, participants' gender, and position on list) on probabilities are shown in Figure 2. As this figure illustrates, the probabilities varied widely, ranging from .17 (Position 3 for male participants in masculine-generics condition) to .58 (Position 1 for female participants in capital-I condition). With few exceptions, probabilities were far below parity (i.e., far below .5).

To explore whether the effects of the independent variables generalize across occupations, we ran a variant of

2 We used the values reported in the original study because the study's raw data, collected some 20 years ago, were no longer available.

Table 4. Parameter Estimates for the Generalized Linear Mixed Model (With Logit-Link) Examining the Effects of Linguistic Expression of Gender, Participant Gender and Position on the Probability to List a Woman

Effect	Parameter	Model 1		Model 1 (random slopes)	
		Est. (SE)	z	Est. (SE)	z
<i>Fixed effects</i>					
Intercept	β_0	-1.340 (0.204)	-6.59***	-1.414 (0.381)	-3.72***
Linguistic Expression of Gender					
Feminine-masculine word pairs vs. masculine generics (X_1)	β_1	0.494 (0.106)	4.65***	0.522 (0.133)	3.92***
Capital-I vs. masculine generics (X_2)	β_2	0.932 (0.113)	8.22***	0.961 (0.195)	4.92***
Gender asterisk vs. masculine generics (X_3)	β_3	0.804 (0.107)	7.51***	0.806 (0.161)	5.01***
Participants' gender (female vs. male) (X_4)	β_4	0.616 (0.077)	8.01***	0.654 (0.142)	4.60***
Position					
Position 2 vs. Position 1 (X_5)	β_5	-0.470 (0.072)	-6.54***	-0.438 (0.239)	-1.84 ⁺
Position 3 vs. Position 1 (X_6)	β_6	-0.558 (0.074)	-7.56***	-0.528 (0.306)	-1.73 ⁺
<i>Random effects (variances)</i>					
Participants:					
Random intercept		0.168		0.196	
Occupations:					
Random intercept		0.191		0.803	
Random Slope X_1				0.032	
Random Slope X_2				0.143	
Random Slope X_3				0.078	
Random Slope X_4				0.082	
Random Slope X_5				0.309	
Random Slope X_6				0.527	
<i>Model fit</i>					
AIC		6678		6594	
BIC		6738		6833	
Deviance		6660		6522	

Note. $N_{\text{observations}} = 5606$. $N_{\text{participants}} = 340$. Linguistic expression of gender: Three dummy-coded predictors, with masculine generics as reference category (coded as 0); Participants' gender: dummy-coded, male participants as reference category (coded as 0); Position: dummy-coded, Position 1 as reference category (coded as 0).

*** $p < .001$, + $p < .10$ (two-tailed).

Model 1 that additionally included random slopes of linguistic expression of gender, of participants' gender, and of position (i.e., the effects of these factors were assumed to vary across occupations). In comparison with the baseline Model 1 described above, this model additionally includes the random slopes u_{1Occ} , u_{2Occ} , u_{3Occ} , u_{4Occ} , u_{5Occ} , and u_{6Occ} in the linear combination.

The parameter estimates of this model are displayed in Table 4 (right-hand side). In this model, the fixed effects of linguistic expression of gender and of participants' gender remained significant ($p < .001$), despite the increased standard errors. However, including the random slopes of occupations increased the standard errors for the fixed effects of position to an extent that these effects were significant

only at a liberal level of $\alpha < .10$. This result indicates that the random variation of the position effect between different occupations was large, leading to a decreased reliability of the fixed effect that reflect the mean effect of position across occupations. In other words, the position effects differ between occupations, suggesting that the position effects might not generalize to other occupations.

Discussion

The present high-powered and pre-registered study sought to replicate the detrimental effect of masculine generics on the accessibility of female exemplars of a given occupational category—an effect that was found about two

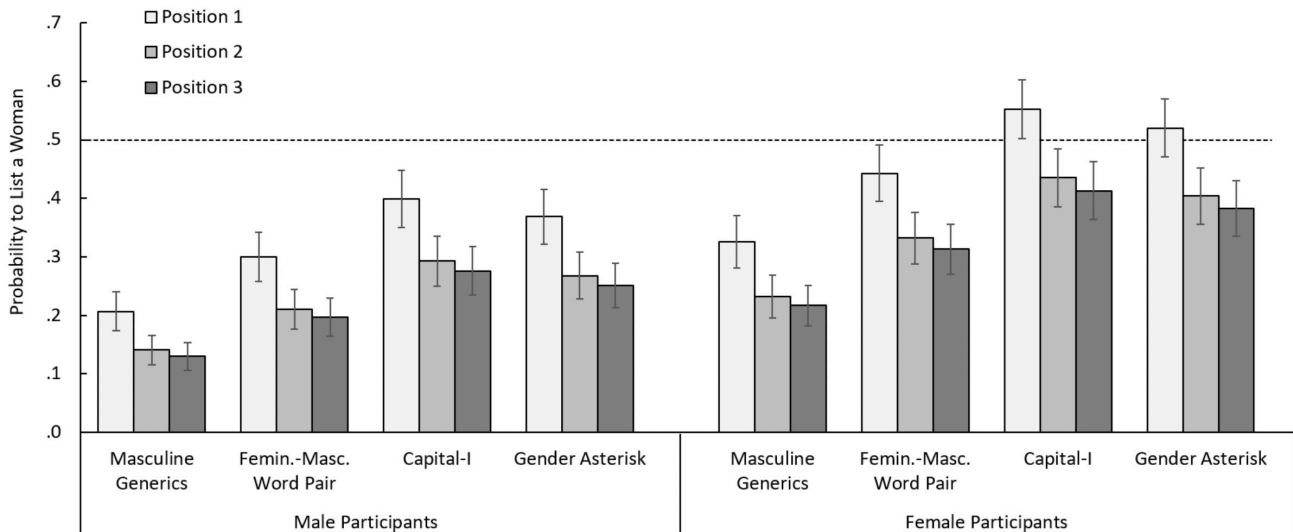


Figure 2. Probabilities to List a Woman as a Function of Linguistic Expression of Gender (i.e., Masculine Generics vs. Three Gender-Inclusive Forms), Participants' Gender (Male vs. Female Participants), and Position in List (Positions 1, 2, or 3) (Back-Transformed from the Logits Estimated in the GLMM, Model 1)

Note. Error bars represent standard errors. Dashed line denotes parity (probability of .5).

decades ago in a sample of university students (Stahlberg et al., 2001, Study 2; Stahlberg & Sczesny, 2001, Study 3). We tested the effect using an almost four times larger and less selective participant sample (only 14% university students) and found virtually the same results. This convergence of findings is remarkable, given the societal changes that have occurred in the past decades in terms of, for example, increases in workforce participation of women and visibility of female political leaders. Apparently, a seemingly minor variation of linguistic expression, namely, whether masculine generics or explicitly gender-inclusive forms are used, can have strong effects on recipients' cognitive inclusion of female exemplars. This convergence of findings did not only occur for the experimental effect (i.e., same pattern of main effects in the present and in the original study) but also in absolute terms at least in 2 of the 3 experimental conditions. That is, for both the experimental condition that produced the lowest inclusion of women (i.e., masculine generics) and the highest inclusion of women (capital-I), the number of women listed by participants in the present study and by participants two decades ago did not differ statistically. Although not a true experimental test, this convergence in absolute numbers indicates that the changes regarding the visibility and increased power of women in society did not lead to an overall increased cognitive accessibility of women. Only for masculine-feminine word pairs, our results differed from the original study. In our replication, significantly more women were listed in this condition than in the masculine generics condition, whereas this contrast was not significant in the original study. Accordingly, the number of women listed in this condition was significantly larger in the present study than in the original study. A likely explanation for this deviation is the relatively small sample size of the original study, which led to relatively low statistical power and potentially instable estimates of the population means.

Theoretical Contributions

The present research informs theory in several ways. First, the convergence of findings between the present study and the 20-years-old one suggests that the effects of masculine generics are based on highly automatized processes during text comprehension, as recipients use linguistic cues when constructing their mental model. These processes appear to be independent of societal changes regarding the public visibility of influential men and women—or the societal changes may simply be still too incremental to evoke measurable changes in highly automatized cognitive information processing.

Second, not only linguistic cues (i.e., masculine generics) can produce considerable effects on cognitive inclusion of women; the effect was just as strong for a participant characteristic, namely, their gender, and this occurred independently of the linguistic cue (i.e., no interaction between the two factors). Notwithstanding the large effects produced by the linguistic cue (masculine generics) and participant characteristic (participant gender), under none of the conditions, a numerical 50%-parity was achieved (i.e., even under the condition with the highest proportion of listed women, the average proportion was only slightly above 40%, see Table 3 and one-sample *t*-tests). Thus, if we define male bias as any deviation from 50%, male bias was present under all conditions and only the degree of male bias varied by condition. For example, being a female participant reduced, but did not eliminate male bias. Similarly, gender-inclusive linguistic cues reduced, but did not eliminate male bias. These findings once again underscore the prevalence and persistence of male bias, in terms of a male dominance in cognitive representations and accessibility, as suggested by various lines of research (Bailey et al., 2018; Tavits & Pérez, 2019).

Third, the order effect we found in additional analyses indicated that male bias dynamically changed during re-

trieval, in that it was least pronounced early during retrieval. That is, the probability to list a woman was highest for the first position and then the probability dropped for Positions 2 and 3. This pattern was consistent across linguistic cues (i.e., masculine generics vs. gender-inclusive forms) and participant gender (i.e., no interaction of position with these factors). This position effect suggests that female exemplars do exist in the minds of many respondents and that these exemplars are, in principle, cognitively accessible. However, if at all, these salient exemplars come to one's mind first, whereas later retrieval of female exemplars becomes far less likely. For example, when asked to list three politicians, many Germans may come up with the name of longtime chancellor *Angela Merkel* but a second or even third woman would not easily come to their mind (to rule out the possibility that the position effect was actually driven by the category of politicians and chancellor Merkel's saliency, we reran analyses to the exclusion of this category and found the pattern of results to remain unchanged). This saliency interpretation of the position effect is consistent with tokenism theory, according to which "social group members who are numerically underrepresented" (Watkins et al., 2019, p. 334), such as female leaders, become all the more salient (Heilman, 2012; Kanter, 1977).

The position effect may also be a manifestation of an androcentric bias to conceive of men as being more typical exemplars of a human category than women (Bailey et al., 2018, 2020). Specifically, after listing a woman, people may consciously or unconsciously be motivated not to list a second or even a third woman, because a two-women-one-man or even a three-women-no-man list would somewhat just not feel right (i.e., not representative for the category as a whole), whereas a one-women-two-men list feels appropriate (i.e., representative for a human category such as politicians, authors, or athletes). Obviously, based on the present data we can only speculate about the causes of the position effect we found and we encourage future research to replicate the effect and to identify its underlying mechanisms. In that regard, it must be noted that we also found a considerable variation between occupational categories, which, when included as a random slope, greatly increased the standard errors of the position effect and lowered its reliability. Thus, position effects affecting the probability of naming male or female exemplars seem to depend considerably on the particular occupational category which is provided as additional retrieval cue. In other words, the position effects that we found might not generalize to occupations other than those used in the present study.

Practical Implications

Practical implications of these findings are quite straightforward: If we agree upon the goal to increase cognitive inclusion of women and to decrease male bias, a direct recommendation would be to avoid masculine generics and to replace them by gender-inclusive linguistic expressions. This seems all the more appropriate when considering that cognitive biases may contribute to biased behavior (Bailey et al., 2018). However, our results suggest that parity will still not be achieved even when abandoning masculine

generics from language use. Also, it should be kept in mind that avoiding masculine generics in favor of more gender-inclusive forms may also have undesirable side effects. For example, for some forms of feminizations and gender-inclusion (e.g., explicitly referencing both male and female engineers), research indicates that explicit female referencing may negatively affect the perceived social status and evaluation of the referenced person and decrease the perceived difficulty of an occupation (Formanowicz et al., 2013; Gabriel et al., 2018; Verweken & Hannover, 2015). Also, in some languages (e.g., Italian), feminized forms may be pejorative. As a recent English-language example, Hollywood star Cate Blanchett, who headed the Venice film festival in 2020, was cited for saying that she would rather be called an actor than an actress because of the pejorative sense often implied in the word actress (Agence France-Presse, 2020).

Limitations and Directions for Future Research

We would like to point out two limitations. First, this research was conducted in one particular context and tested one particular form of masculine generics in German language. Obviously, we do not know to what extent these findings generalize to other contexts, including, among others, other occupational roles, other forms of text and discourse, as well as other forms of masculine generics. Given the substantial size of effects in the present study and other available evidence (from different paradigms in different languages), the ubiquity of linguistic gender asymmetries across many languages (Sczesny et al., 2016), and the universality of androcentrism (Bailey et al., 2018, 2020), we would expect similar effects of masculine generics in a range of situations. Moreover, the effects of linguistic expression of gender and of participants' gender were significant and strong in the GLMM, too, that included random effects of participants and occupations, accounting for the fact that not only participants but also the occupations were drawn from larger populations. Nevertheless, future research may investigate systematically what context characteristics may modulate effects of masculine generics. For example, future research may systematically vary the included occupations with regard to how many female exemplars plausibly exist or are visibly represented in public media. If only few or less representative female exemplars exist in a category, accessibility of female exemplars may become even more unlikely (Stahlberg & Sczesny, 2001, Study 2).

Second, it remains unclear to what extent our findings are manifestations of conscious/controlled vs. unconscious/automatic processes. On the one hand, based on research in reading, it seems safe to assume that processes involved in text comprehension are highly automatized (e.g., Cook et al., 1998; Richter & Singer, 2018). Unlike many other studies in the field, due to its simplicity, the paradigm introduced by Stahlberg et al. (2001) is likely to induce little strategic processing. On the other hand, the task we used, namely, to list three exemplars of a given category, is a task that is potentially controllable and strategic. It is possible that the gender-inclusive forms triggered participants' motivation to deliberately and consciously come up with at least a few female exemplars, at least in the beginning of retrieval, as the position effect suggests. Gender-exclu-

sive masculine generics, in contrast, might not motivate explicit retrieval of women, with the consequence of male bias remaining uncorrected and particularly strong under this condition.

Conclusion

Our findings clearly indicate that masculine generics decrease cognitive accessibility of female exemplars. If we agree that this form of male bias is undesirable, then the conclusion is straightforward, namely, that masculine generics should be avoided and replaced by gender-inclusive forms of linguistic expression—even though the former may be considered by some to be grammatically correct and the latter cumbersome. Referring back to the opening quote from a speech by Charles Dickens: It is not the point that we cannot be sure what writers of what gender Dickens was referring to. The point is that, irrespective of Dickens' intentions, when listening to or reading his speech, chances are that many people will mentally include only few female writers.

Author Note

Portions of these findings were presented as a poster at the 11th conference of the Division of Work, Organizational, and Business Psychology of the German Psychological Society (DGPs) in Braunschweig, Germany. The poster was awarded the conference's Best Student Poster award (awardee: KH).

Acknowledgements

We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG – German Research Foundation) and the Open Access Publishing Fund of Technical University of Darmstadt.

Contributions

Contributed to conception and design: NK, KH, TR
 Contributed to acquisition of data: KH
 Contributed to analysis and interpretation of data: NK, KH, TR
 Drafted and/or revised the article: NK, TR
 Approved the submitted version for publication: NK, TR

Competing Interests

The authors declare no conflict of interest.

Data Accessibility Statement

The preregistration of this study can be found at: <https://osf.io/te9c3>
 The data and analysis scripts are available at https://osf.io/rn96f/?view_only=4436a99bf52740ef84a97a4951c414eb

Submitted: December 04, 2021 PST, Accepted: February 10, 2022 PST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

- Agence France-Presse. (2020, September 3). *Cate Blanchett says she would rather be called an actor than an actress*. <https://www.theguardian.com/film/2020/sep/03/cate-blanchett-says-she-would-rather-be-called-an-actor-than-an-actress>
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). <https://doi.org/10.1037/0000165-000>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2018). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review*, 23(4), 307–331. <https://doi.org/10.1177/1088868318782>
- Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2020). Implicit androcentrism: Men are human, women are gendered. *Journal of Experimental Social Psychology*, 89, 103980. <https://doi.org/10.1016/j.jesp.2020.103980>
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629–654. <https://doi.org/10.1037/0278-7393.11.1-4.629>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Begeny, C. T., Moss-Racusin, C. A., Ryan, M. K., & Ravetz, G. (2020). In some professions, women have become well represented, yet gender bias persists—perpetuated by those who think it is not happening. *Science Advances*, 6(26), eaba7814. <https://doi.org/10.1126/sciadv.aba7814>
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3(2), 193–209. [https://doi.org/10.1016/0010-0285\(72\)90003-5](https://doi.org/10.1016/0010-0285(72)90003-5)
- Brauer, M. (2008). Un ministre peut-il tomber enceinte? L'impact du générique masculin sur les représentations mentales. *L'Année Psychologique*, 108(2), 243–272. <https://doi.org/10.4074/s0003503308002030>
- Burgen, S. (2020, January 19). Masculine, feminist or neutral? The language battle that has split Spain. *The Guardian*. <https://www.theguardian.com/world/2020/jan/19/gender-neutral-language-battle-spain>
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, 49(3), 639–663. <https://doi.org/10.1080/13755647>
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, 9(1), 40–48. <https://doi.org/10.1177/1745691613513470>
- Chiu, B. (2020, March 8). 2020s mark a new wave of feminist mobilization. *Forbes*. <https://www.forbes.com/sites/bonniechiu/2020/03/08/2020s-mark-a-new-wave-of-feminist-mobilization/?sh=23490f2e485e>
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. [https://doi.org/10.1016/s0022-5371\(73\)80014-3](https://doi.org/10.1016/s0022-5371(73)80014-3)
- Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory - based view of text processing. *Discourse Processes*, 26(2–3), 109–129. <https://doi.org/10.1080/01638539809545041>
- Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.). Cambridge University Press.
- Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59(4), 447–456. <https://doi.org/10.1016/j.jml.2007.1.004>
- European Commission. (2019). *2019 report on equality between women and men in the EU*. [Online document]. Publications Office of the European Union. <https://doi.org/10.2838/395144>
- European Parliament. (2018). *Gender-neutral language in the European Parliament* [Online document]. https://www.europarl.europa.eu/cmsdata/151780/GNL_Guidelines_EN.pdf
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Formanowicz, M., Bedynska, S., Cisiak, A., Braun, F., & Sczesny, S. (2013). Side effects of gender-fair language: How feminine job titles influence the evaluation of female applicants. *European Journal of Social Psychology*, 43(1), 62–71. <https://doi.org/10.1002/ejsp.1924>
- Gabriel, U. (2008). Language policies and in-group favoritism: The malleability of the interpretation of generically intended masculine forms. *Social Psychology*, 39(2), 103–107. <https://doi.org/10.1027/1864-9335.39.2.103>
- Gabriel, U., Gygas, P. M., & Kuhn, E. A. (2018). Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations*, 21(5), 844–858. <https://doi.org/10.1177/1368430218771742>
- Gabriel, U., & Mellenberger, F. (2004). Exchanging the generic masculine for gender-balanced forms – the impact of context valence. *Swiss Journal of Psychology*, 63(4), 273–278. <https://doi.org/10.1024/1421-0185.63.4.273>

- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, 23(3), 464–485. <https://doi.org/10.1080/01690960701702035>
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32, 113–135. <https://doi.org/10.1016/j.riob.2012.11.003>
- Henley, N. M., & Abueg, J. (2003). A review and synthesis of research on comprehension of the masculine as a generic form in English. *Sociolinguistic Studies*, 4(2), 427–454. <https://doi.org/10.1558/sols.v4.i2.427>
- Inter-Parliamentary Union. (2015). *Women in parliament: 20 years in review* [Online document]. Inter-Parliamentary Union. <http://archive.ipu.org/english/perdcls.htm#wmn-year>
- Irmen, L., & Kurovskaja, J. (2010). On the semantic content of grammatical gender and its impact on the representation of human referents. *Experimental Psychology*, 57(1), 367–375. <https://doi.org/10.1027/1618-3169/a000044>
- Irmen, L., & Roßberg, N. (2004). Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology*, 23(3), 272–307. <https://doi.org/10.1177/0261927x04266810>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.1.007>
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge University Press.
- Kanter, R. M. (1977). *Men and women of the corporation*. Basic Books.
- Kollmayer, M., Pfaffel, A., Schober, B., & Brandt, L. (2018). Breaking away from the male stereotype of a specialist: Gendered language affects performance in a thinking task. *Frontiers in Psychology*, 9, 985.
- MacArthur, H. J., Cundiff, J. L., & Mehl, M. R. (2020). Estimating the prevalence of gender-biased language in undergraduates' everyday speech. *Sex Roles*, 82(3), 81–93. <https://doi.org/10.1007/s11199-019-01033-z>
- MacNicol, G. (2020, December 4). 50 years on, the Feminist Press is radical and relevant. *The New York Times*. <https://www.nytimes.com/2020/12/04/us/50-years-the-feminist-press-gloria-steinem-florence-howe.html>
- Miller, M. M., & James, L. E. (2009). Is the generic pronoun he still comprehended as excluding women? *The American Journal of Psychology*, 122(4), 483–496.
- Parks, J. B., & Robertson, M. A. (1998). Contemporary arguments against nonsexist language: Blaubeergs (1980) revisited. *Sex Roles*, 39(5/6), 445–461. <https://doi.org/10.1023/a:1018827227128>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Richter, T., & Singer, M. (2018). Discourse updating: Acquiring and revising knowledge through discourse. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Routledge handbooks in linguistics. The Routledge handbook of discourse processes* (pp. 167–190). Routledge/Taylor & Francis Group.
- Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, 7, 25. <https://doi.org/10.3389/fpsyg.2016.00025>
- Sniezek, J. A., & Jazwinski, C. H. (1986). Gender bias in English: In search of fair language. *Journal of Applied Social Psychology*, 16(7), 642–662. <https://doi.org/10.1111/j.1559-1816.1986.tb01165.x>
- Stahlberg, D., & Sczesny, S. (2001). Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen [Effects of the generic masculinum and of alternative forms of expression on the mental inclusion of women]. *Psychologische Rundschau*, 52(3), 131–140. <https://doi.org/10.1026/0033-3042.52.3.131>
- Stahlberg, D., Sczesny, S., & Braun, F. (2001). Name your favorite musician: Effects of masculine generics and of their alternatives in German. *Journal of Language and Social Psychology*, 20(4), 464–469. <https://doi.org/10.1177/0261927x01020004004>
- Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, 144(12), 1325–1346. <https://doi.org/10.1037/bul0000169>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.
- Tavits, M., & Pérez, E. O. (2019). Language influences mass opinion toward gender and LGBT equality. *Proceedings of the National Academy of Sciences*, 116(34), 16781–16786. <https://doi.org/10.1073/pnas.1908156116>
- Troche, S., & Rammsayer, T. H. (2011). Eine Revision des deutschsprachigen Bem Sex-Role Inventory [A revision of the German Bem Sex-Role Inventory]. *Klinische Diagnostik und Evaluation*, 4(3), 262–283.
- U.S. Bureau of Labor Statistics. (2020). *Labor force statistics from the current population survey*. <https://www.bls.gov/cps/cpsaat11.htm>
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Vergoossen, H. P., Renström, E. A., Lindqvist, A., & Sendén, M. G. (2020). Four dimensions of criticism against gender-fair language. *Sex Roles*, 83(5–6), 328–337. <https://doi.org/10.1007/s11199-019-01108-x>
- Vervecken, D., & Hannover, B. (2012). Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts perception of the speaker. *European Journal of Social Psychology*, 42(6), 754–762. <https://doi.org/10.1002/ejsp.1893>

- Vervecken, D., & Hannover, B. (2015). Yes I can! Effects of gender fair job descriptions on children's perceptions of job status, job difficulty, and vocational self-efficacy. *Social Psychology, 46*(2), 76–92. <https://doi.org/10.1027/1864-9335/a000229>
- Warner, J., Ellman, N., & Boesch, D. (2018). *The women's leadership gap: Women's leadership by the numbers* [Online document]. Center for American Progress. <https://cdn.americanprogress.org/content/uploads/2018/11/19121654/WomensLeadershipFactSheet.pdf>
- Watkins, M. B., Simmons, A., & Umphress, E. (2019). It's not black and white: Toward a contingency perspective on the consequences of being a token. *Academy of Management Perspectives, 33*(3), 334–365. <https://doi.org/10.5465/amp.2015.0154>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*, e120. <https://doi.org/10.1017/s0140525x17001972>

SUPPLEMENTARY MATERIALS

Peer Review History

Download: https://collabra.scholasticahq.com/article/32964-ladies-first-or-ladies-last-do-masculine-generics-evoke-a-reduced-and-later-retrieval-of-female-exemplars/attachment/83011.docx?auth_token=sWwIThafvsRPI1sOA2U
