Language and Cognition (2022), 14: 1, 47–60
doi:10.1017/langcog.2021.21

ARTICLE

Gender bias in morphological inferences

Sara Finley*, Saara Charania, Tiarra Lewis, Barbara Millward and Stella Wang

Pacific Lutheran University
*Corresponding author. Email: finleysr@plu.edu

(Received 29 January 2021; Revised 17 August 2021; Accepted 19 August 2021)

Abstract

The present study explores how language learners apply gender stereotypes in learning a novel language with grammatical gender. Adult, English-speaking participants were exposed to picture–sound pairs from a miniature language. Each picture was of a matched gendered professional (e.g., male tennis player, female tennis player) with a nonsense form [CVCV-go/gu]. Participants were exposed to 32 picture–sound pairs (16 male, and 16 female, all matched) five times in a randomized order. Following training, participants were given a two-alternative forced-choice test with novel picture–word pairs. Participants were presented with a novel picture paired with two words (e.g., [befegu vs. befebo]) and were asked to choose which word most likely portrayed the meaning conveyed by the picture. These novel items contained gender-matched professions (e.g., male and female chemist), neutral items (office supplies), stereotypically female items (makeup), and stereotypically male items (tools). Participants assigned the appropriate gender to the novel professions, and assigned gender in line with the stereotyped objects at a rate significantly greater than chance (but not for neutral items). These results support the hypothesis that learning a language with a binary grammatical gender might be influenced by gender stereotypes.

Keywords: artificial language learning; morphology; noun classes; gender stereotypes

1. Introduction

Many languages divide nouns up into gender-based semantic classes. While some cases of grammatical gender appear to be arbitrary, many cases are less so. For example, some words show a grammatical gender that is the opposite of its conceptual gender[1] (e.g., German 'the girl' *das Mädchen*), and other languages mark the same objects with the opposite grammatical gender (e.g., *fork* is masculine in Spanish and feminine in French), suggesting that grammatical gender categorization is arbitrary. However, many word classes are semantically driven (e.g., *chica* vs. *chico*

[1]We follow Ackerman's (2019) definitions of *grammatical gender* to indicate the formal properties of the word defined morpho-syntactically, *conceptual gender* to indicate how a language user classifies a given referent, and *biosocial gender* to indicate an individual's gender, based on a mix of biological and sociocultural factors.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

in Spanish to distinguish male from female), suggesting that languages with masculine and feminine noun classes may have roots in non-linguistic gender-based distinctions. Because languages use both semantic and phonological cues to mark grammatical gender, sometimes these cues are in conflict (e.g., *das Mädchen*), which can muddle effects of arbitrary and systematic cues to noun class identification (Culbertson et al., 2017, 2019; Lidz & Gagliardi, 2015; Vuksanović et al., 2015).

While the majority of previous research exploring the relationship between grammatical gender and cognition has focused on the question of linguistic relativity (e.g., whether there is a causal link between grammatical gender and categorization; Bassetti & Nicoladis, 2016; Samuel et al., 2019), the present study explores how language learners apply gender stereotypes when learning a language with a binary grammatical gender classification system that distinguishes between male and female. Previewing our results, adult, English-speaking learners were trained on a miniature artificial language with a masculine/feminine grammatical gender, and then tested on novel objects that were either stereotypically male (tools) or stereotypically female (makeup), or neutral (office supplies). Participants were significantly more likely to assign masculine gender to items that were stereotypically male, and likewise for stereotypically female objects, but not for neutral objects.

Previous research exploring the role of grammatical gender and cognition has suggested that both children and adults show a bias to assign conceptual genders to inanimate objects in line with the grammatical gender of the native language (Forbes et al., 2008; Sera et al., 1994, 2002). In a voice-assignment task, participants were asked to choose whether an animated object should be played by a female voice actor or a male voice actor. Participants were more likely to choose the female actor for objects with feminine gender assignment, and male actors for objects with a masculine gender assignment. These preferences also emerged after exposure to Spanish as an L2 for both native English (Kurinski & Sera, 2011) and native Hungarian speakers (Kurinski et al., 2016). Borodistky and Schmidt (2000) showed that Spanish and German speakers were more likely to remember arbitrary pairings between a name and an inanimate object when the conceptual gender of the name matched the grammatical gender of their language. English speakers' intuitions about conceptual gender for animals aligned with Spanish grammatical gender for these nouns (but not artifacts) (Boroditsky & Schmidt, 2000).

While many studies point to a relationship between grammatical gender and conceptual gender, the results are often mixed and idiosyncratic, and appear to be largely task-dependent (Samuel et al., 2019). First, the effects for German speakers, and languages with a neuter gender may show a weaker effect than languages with a binary masculine–feminine distinction (Sera et al., 2002). Second, speakers of multiple languages may show weaker effects on gender assignment tasks (Bassetti & Nicoladis, 2016), suggesting that exposure to multiple noun categorization schemas highlights the arbitrariness of grammatical gender marking in their languages. Third, some studies showed stronger effects for feminine forms (Bassetti, 2007; Bassetti & Nicoladis, 2016; Segel & Boroditsky, 2011). For example, Segel and Boroditsky (2011) found that 84% of grammatically feminine items were personified as female, but 60% of grammatically masculine items were personified as male (though both were statistically different from 50% chance). This has largely been attributed to the fact that masculine is often the 'default' grammatical gender in many languages with a masculine–feminine distinction, meaning that more gender-neutral terms are marked as masculine, diluting the correspondence

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

between masculine as a grammatical gender and masculine attributes. Fourth, some effects may be task-dependent (Samuel et al., 2019), such as whether the stimuli makes use of written words or pictures (Ramos & Roberson, 2011; Vigliocco et al., 2005), as effects are weaker when the stimuli do not contain written words. This suggests that links between language and cognition require direct activation of linguistic structures. Finally, some studies have failed to replicate (Mickan et al., 2014), or may be subject to alternative explanations such as cultural stereotypes (Beller et al., 2015).

While task effects may play a role in some of the differences in effects across studies (Samuel et al., 2019), another factor related to these mixed results is that the link between grammatical gender and conceptual gender may be mediated by a variety of factors, such as the grammar of the language, phonology, and gender stereotypes. It is possible that, rather than language having a direct effect on thought, cultural attitudes and gender roles may have played a role in the evolution of languages with grammatical gender systems. The links between grammatical gender and conceptual gender may therefore reflect speakers' non-linguistic factors related to gender classification (Bassetti & Nicoladis, 2016). For example, some studies have shown that participants are more likely to assign feminine gender to natural kinds, and masculine gender to artifacts, even when grammatical gender was controlled for (Forbes et al., 2008; Sera et al., 1994). Other studies have also shown that natural items are more likely to be described as female, while artifacts are more likely to be described as male, even in languages without grammatical gender (Martinez & Shatz, 1996; Ramos & Roberson, 2011; Sera et al., 1994).

Some studies have shown correlations between grammatical gender and gender stereotypes (Lewis & Lupyan, 2020). While Lewis and Lupyan argue that the direction of the correlation suggests a causal link between the words and structures in a language and implicit stereotypes, they do not rule out the possibility that gender stereotypes influence language use. For example, someone with strong gender associations might be more likely to use gendered language (such as *waitress* and *waiter*), while someone with weaker stereotypical associations may be more likely to use non-gendered language (such as *server*). Meagher (2017) found that participants who held stronger beliefs in gender roles showed stronger preferences to classify inanimate objects as either masculine or feminine.

Several studies have shown support for the effect of gender stereotypes in language use (Flaherty, 2001; Sato & Athanasopoulos, 2018; Segel & Boroditsky, 2011; Yorkston & De Mello, 2005). In languages without grammatical gender, such as English, participants frequently assign gender roles based on cultural stereotypes (Flaherty, 2001; Sato & Athanasopoulos, 2018). When asked to assign attributes to objects, English-speaking adults (and young Spanish-speaking children) assigned names on the basis of gender-based stereotypes, suggesting a role for stereotyped information in cognition.

Yorkston and De Mello (2005) asked Spanish- and English-speaking participants to rate and recall various product names. Spanish-speaking participants were more likely to recall product names when there was agreement between the grammatical gender of the product and the name of the product, but semantic agreement (e.g., stereotyped as more female/male) was associated with greater likability of the product. English-speaking participants' recall and likeability judgments were based on semantic agreement only, as the English gender system is limited to pronouns. This suggests that language users may store grammatical gender as part of the

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

representation of individual words, and use it for recall when the language contains grammatical gender; otherwise speakers may default to gender stereotypes in forming such associations.

One way of addressing some of the issues related to how linguistic representations are affected by gender stereotypes is to study the role of gender stereotypes in language learning. In addition to studies focusing on learning a second language in the classroom (Kurinski et al., 2016; Kurinski & Sera, 2011), the artificial grammar learning paradigm can be helpful to understand the role of learning in the relationship between gender and cognition. In an artificial grammar learning paradigm, it is possible to control for many aspects of the language in question, and test for effects after a very brief exposure phase. Several artificial language learning studies showed an effect of brief exposure to a language with grammatical gender can result in changes to performance on gender-based categorization tasks, such as the voice assignment task (Sera et al., 2002) or a similarity task (Eberhard et al., 2005; Phillips & Boroditsky, 2003). Sera et al. (2002) showed that participants who were exposed to a binary (Spanish-like) gender pattern were more likely to assign voices in line with the grammatical gender of the artificial language, like native Spanish speakers, but participants who were exposed to a language with three grammatical genders (German-like) did not show the same effect, similar to the behavior of native German speakers.

Eberhard et al. (2005) trained native English participants on a miniature language with a binary gender-based noun class mapping (e.g., some nouns were marked with 'oos' and others were marked with 'sou'). Once participants successfully learned this mapping, they were given a similarity rating task, where they rated the similarity of each object to a picture of a face. Overall, participants showed consistency with the grammatical gender of the noun in the artificial language and the sex of the face, but the effect was most likely carried by participants who learned the language the fastest. These results are consistent with Phillips and Boroditsky's (2003) experiment that used a similar method.

While artificial grammar learning experiments have explored the relationship between language and cognition, few studies have used this paradigm to examine the role of gender stereotypes on language learning. Exploring this effect in natural languages is challenging because many seemingly arbitrary noun class categories may have an historical basis that could be effected by stereotypes that are no longer present (Sera et al., 1994). Using an artificial language allows the researcher to control for many aspects of the language and manipulate stimuli based on current gender stereotypes. The present study uses an artificial grammar learning experiment to explore how gender stereotypes might be applied to learning a novel language. Participants were exposed to a miniature artificial language where gender was explicitly marked via a suffix ([-go] for females and [-gu] for males) for matched professions (via pictures). Following exposure, participants were tested via a two alternative forced-choice test with novel objects, choosing between a word ending in [-gu] and a word ending in [-go] to best describe the picture. Participants were tested on novel professions, as well as items that were either stereotypically associated with males (tools) or stereotypically associated with females (makeup), or gender neutral (office supplies). If participants apply gender stereotypes in language learning, then we expect that participants will be more likely to assign [-gu] (male) endings to tools and [-go] (female) endings to makeup items.

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

2. Method

2.1. Participants

All participants were adult English speakers who were recruited from a small liberal arts college in the Pacific Northwest of the United States, and received course credit for their participation. A total of 35 participants completed the experiment, but data from seven participants were excluded because their data was uninterpretable (they pressed the incorrect response button on all test trials); data from 28 participants are therefore included in the analyses.

2.2. Materials

Participants were trained on words from an artificial language that consisted of 32 images derived from 16 professions matched for gender (e.g., male chef and female chef). All images were drawings found on Google Image search, labeled for non-commercial reuse. Images were vetted by the authors for transparency and ease of interpretation. If an image was not clearly interpretable, it was replaced.

Each image was paired with an auditorily presented trisyllabic word that ended in [-gu] for masculine and [-go] for feminine. All words were of the form CVCVCV (where C is a consonant and V is a vowel), and were drawn from the consonant inventory of [p, t, k, b, d, g, m, n] and the vowel inventory [i, e, o, u]. All words were non-English words that did not bear close resemblance to a known English word (e.g., words like *kelephant* would not be included). Examples of training stimuli can be found in Table 1. Full stimuli lists can be found in the supplementary materials (available at: http://doi.org/10.1017/langcog.2021.21). Additional materials, analysis code, and anonymized data are available on the OSF (Open Science Framework) at <https://osf.io/bm54e/>.

Learning and generalization to novel items was assessed via a two alternative forced choice test that pitted the masculine form (/-gu/) against the feminine form (/-go/). There were five different sets of test items (with eight of each), which included novel professions (Professions_Male, Professions_Female), stereotypically female items (Makeup), and stereotypically male items (Tools).

The professions were matched so that there were eight male versions, and eight female versions of each profession, and the corresponding audio choices were the same (e.g., the chemist was paired with [benego] vs. [benegu] for both the male and female choices). The eight stereotypically female objects were all makeup items, and included items such as a mirror, hair curler, and lipstick. The eight stereotypically male objects were all different types of tools, and included items such as an ax, lawnmower, and ladder. The eight neutral objects were all office supplies, and included items such as a computer, eraser, and paperclip. The materials for the test items were selected in the same manner as the pictures used in the exposure phase;

Table 1. Examples of training items

Profession	Masculine	Feminine
pharmacist	bidigu	bidigo
newscaster	dikugu	dikugo
housecleaner	ketegu	ketego
singer	nopegu	nopego

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

Table 2. Examples of test items (note that order of presentation of audio files was counterbalanced across items)

Test Item	Picture	Audio
Profession-Masculine	chemist (male)	benego vs. benegu
	farmer (male)	pudugo vs. pudugu
	wait-staff (male)	kugigo vs. kugigu
Profession-Feminine	chemist (female)	benego vs. benegu
	farmer (female)	pudugo vs. pudugu
	wait-staff (female)	kugigo vs. kugigu
Office Supplies	computer	podogo vs. podogu
	desk	nugigo vs. nugigu
	scissors	bimugo vs. bimugu
Makeup	lipstick	dukugo vs. dukugu
	mirror	nukugo vs. nukugu
	curler	bemego vs. bemegu
Tools	ax	dutugo vs. dutugu
	hammer	dekego vs. dekegu
	ladder	petego vs. petegu

audio files had the same CVCVCV format and consonant and vowel inventory, and pictures were drawings found from Google Image (with permissions for reuse). Examples of test items can be found in Table 2.

The items were chosen based on gender schema theory (Bem, 1981), and fit into the classifications of gender roles described in Bem (1974). The authors used their own intuitions about gender stereotypes and consulted friends and roommates for verification of the stereotypical nature of the stimuli. The gender associations of the items used in the experiment were also verified by cross-referencing with studies that either made explicit use of gender-based or gender-neutral stimuli (Baer et al., 2006; Conry-Murray, 2015; Sato & Athanasopoulos, 2018; Sereno & O'Donnell, 2009), or explicitly asked participants to provide a gendered association for items (Blakemore & Centers, 2005; Crawford et al., 2004; Meagher, 2017; Scott et al., 2019). In all the studies surveyed, makeup items were classified as feminine, tool items were classified as masculine, and office supplies were classified as gender neutral. We also checked each item with the use or ratings in these studies, and found all the items that had data were coded in line with the assigned gender stereotype. There were some items that could not be identified exactly, but similar items could be found, and the ratings for these were used for verification. This included 'makeup' for 'compact', 'hair straightener' for 'curler', 'file' and 'cabinet' for 'file cabinet' (Scott et al., 2019), 'mascara' for 'eye shadow', and 'saw' for 'chainsaw'. Discussions of curlers ad chainsaws as feminine and masculine items, respectively, can be found in Dyer (2018). All items fit within the gender norms of these studies.

The experiment made use of auditory materials that had been previously used for a different artificial language learning study. These auditory stimuli were recorded in a sound-attenuated booth by an adult male native English speaker. While the speaker was aware that the stimuli were to be used for an artificial grammar learning study, the speaker was unaware of the hypothesis of the study. Tokens were individually recorded. Each token was spoken four times in list format. A single token was chosen from the second or third element of the set in order to keep the prosody as uniform as possible. The first and last elements were avoided to avoid the intonation of first and last elements in a list. The speaker was told to speak each word as clearly and

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

accurately as possible (without reducing vowels). Primary stress was placed on the initial syllable, but all vowels were fully pronounced. All stimuli were normalized to 70dB using Praat software (Boersma & Weenink, 2017).

It was hypothesized that participants would select the feminine gender for the stereotypically female items at a rate greater than 50% chance, while participants would select the masculine gender for the stereotypically male items at a rate greater than 50% chance. If participants treat masculine gender as the 'default' gender, then learners may show a slight preference for the masculine gender for the gender-neutral items; otherwise it is expected that participants should select the masculine form at a rate no greater than 50% chance.

2.3. Procedure

The experiment was conducted using SuperLab 5 software (Heller et al., 2014). Participants were told that they would be listening to words from a language that they had never heard before, and that their task was to remember the word with the picture, as best they could. Each image was displayed on a standard Dell monitor with 200×200 pixels, and centered on the screen. The images were paired with an audio file that played as the picture was displayed on the screen. Participants were instructed to press the space bar to move on to the next trial. The 32 (16 professions for each gender) were displayed five times each, for a total of 160 trials in the exposure phase.

The training was followed by a two-alternative forced-choice test that contained the 40 items discussed above (in a random order). Participants were told that they would hear two words, one corresponding to the meaning of the picture presented, and the other not corresponding to the meaning of the picture. If they believed the first word best captured the meaning of the picture, they were to press the 'a' key; if they believed the second word was a better portrayal of the picture's meaning, they were to press the 'l' key. Order of presentation of items was counterbalanced such that half of the trials presented the masculine option first, and the other half presented the feminine option first. Participants were told to respond as quickly and accurately as possible, but no time limit was given for responses. Participants were given a verbal debriefing upon completion of the experiment, which took approximately 15 minutes to complete.

3. Statistical analysis and results

The data were entered into generalized mixed effects models using the glmer function in the lme4 package (Bates et al., 2015) (version 1.127.1) in R (R Development Core Team, 2018) (version 4.1.0) in RStudio (RStudio Team, 2020) (version 1.4.1717), with the maximal model that would converge, as justified by the experimental design (Barr et al., 2013). This included random intercepts for items and subjects (but models with random slopes failed to converge). To test whether each of the five types of test items were significantly different from chance, a separate model was run with the relevant test item set as the baseline, with a significant intercept interpreted as a significant difference from chance. In addition, pairwise comparisons for masculine responses were made for relevant pairs of items using office supplies as the baseline and the lsmeans function in the emeans package (version 1.6.1) (Lenth, 2016, 2018), with Tukey corrections for multiple comparisons.

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

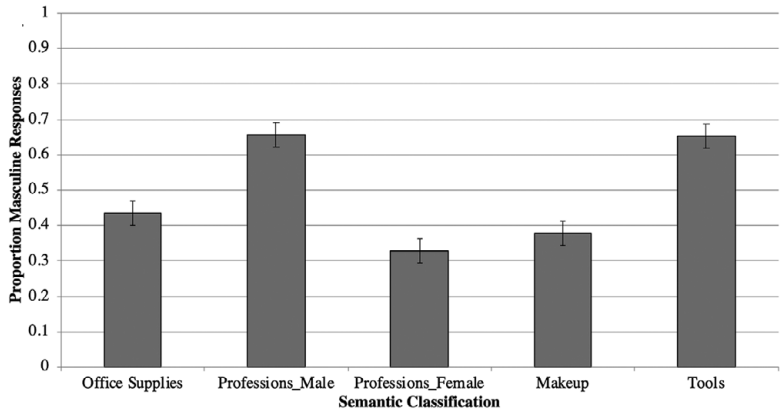


Fig. 1. Mean proportion of masculine responses (error bars indicate standard error of the mean).

Means (and standard errors) of the analyzed data can be found in Figure 1. Due to experimenter error, some participants did not receive all trials. Of the 28 participants whose data were analyzed, nine participants received only 39 test trials, 10 participants received only 38 test trials, one participant received only 37 trials, and one participant received only 36 trials. In addition, some trials were removed because the participant failed to give an appropriate response, usually pressing the space bar instead of either the 'a' or 'l' key as instructed. This occurred for 101 trials across seven participants. Only interpretable data are included in the analysis.

Feminine responses to Professions_Female (mean = 0.67, $CI \pm 0.069$), $\beta = 0.88$, $SE = 0.27$, $z = 3.22$, $p = 0.0013$, and masculine responses to Professions_Male (mean = 0.66, $CI \pm 0.068$), $\beta = 0.81$, $SE = 0.27$, $z = 2.99$, $p = 0.0028$, were both selected at rates significantly different from 50% chance, suggesting that participants applied the gendered suffix appropriately to novel professions. In addition, masculine responses to Tools (mean = 0.65, $CI \pm 0.066$), $\beta = 0.78$, $SE = 0.26$, $z = 2.94$, $p = 0.0033$, and feminine responses to Makeup (mean = 0.62 $CI \pm 0.067$), $\beta = 0.61$, $SE = 0.26$, $z = 2.34$, $p = 0.019$, were also selected at a rate significantly greater than 50% chance, suggesting that participants made some use of gender stereotypes in selecting their responses. However, masculine responses to Office Supplies (mean = 0.44, $CI \pm 0.069$), were not selected at a rate significantly different from chance, $\beta = -0.30$, $SE = 0.26$, $z = -1.145$, $p = 0.25$, suggesting no overall preference for masculine or feminine forms for these items.

Participants were more likely to select the masculine response for Tools compared to Makeup, $\beta = 1.16$, $SE = 0.24$, $z = 4.825$, $p < 0.001$, and were more likely to select the masculine response for Professions_Male compared to Professions_Female, $\beta = 1.39$, $SE = 0.25$, $z = 5.50$, $p < 0.001$. Participants were also more likely to select the masculine items for Tools compared to Office Supplies, $\beta = 0.91$, $SE = 0.24$, $z = 3.80$, $p = 0.0013$, and Professions_Male compared to Office Supplies, $\beta = 0.93$, $SE = 0.24$, $z = 3.82$, $p = 0.0013$. However, there were no significant differences between Makeup and Office Supplies, $\beta = 0.25$, $SE = 0.24$, $z = 1.08$, $p = 0.82$, or Professions_Female and Office Supplies, $\beta = 0.47$, $SE = 0.24$, $z = 1.90$, $p = 0.32$.

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

4. Discussion

The results support the hypothesis that, after exposure to a language with a binary grammatical distinction for gender, learners may apply that grammatical distinction to novel words that meet gender stereotypes at a rate greater than chance. Participants applied the feminine suffix to novel professions with a female actor, as well as to makeup – items that are stereotypically female – at rates greater than chance. Participants also applied the masculine suffix to novel professions with a male actor, as well as to tools – items that are stereotypically masculine – at rates greater than chance.

Participants also appear to be more likely to apply the masculine affix to masculine-stereotyped categories than feminine-stereotyped categories. However, masculine responses to gender-neutral items were only significantly different from stereotypically masculine categories. This may have been driven by a small (but not significant) tendency for participants to select the feminine response for Office Supplies. In addition, there were several items in the Office Supplies condition in which masculine gender was selected less than 40% of the time. These included the eraser (mean = 0.35, $CI \pm 0.20$), the paperclip (mean = 0.27, $CI \pm 0.18$), and the envelope (mean = 0.32, $CI \pm 0.20$). This suggests a possibility that participants might show a feminine bias for some office supplies, possibly due to a bias to think of office workers (e.g., administrative assistants or secretaries) as female, or because many erasers (including the image used in this experiment) are pink, which is commonly associated with femininity in the United States. With these three items removed, the mean masculine responses was 0.51 ($CI \pm 0.088$), and the differences between the Office Supplies and Makeup, $\beta = 0.56$, $SE = 0.25$, $z = 2.26$, $p = 0.024$, and Office Supplies and Professions_Female, $\beta = 0.77$, $SE = 0.26$, $z = 3.02$, $p = 0.0025$, were both significant.

In addition to the outliers in the Office Supplies condition, there were some items that did not pattern with the others in their category. There was a bias towards the masculine gender for the image of the female chemist (mean = 0.19, $CI \pm 0.21$). There are several possibilities for this. One is that participants were biased to view the scientist as a masculine profession, and noted the picture as such. However, the male chemist did not show a large bias (mean = 0.58, $CI \pm 0.33$), and other male-dominated professions did not show the same bias (e.g., participants generally applied the feminine ending to the female doctor, mean = 0.80, $CI \pm 0.17$). Another possibility is that the drawing of the female chemist featured a woman with short hair, which could have been interpreted by some to be a picture of a man. The combination of a bias to view scientists as male along with a less stereotypical female (e.g., short hair) could have driven the interpretation.

Finally, some Makeup and Tool items did not show any clear bias. These include the curler (mean = 0.44, $CI \pm 0.20$), the lawnmower (mean = 0.48, $CI \pm 0.21$), and the wrench (mean = 0.54, $CI \pm 0.21$). In addition, there was no bias for the male volleyball player (mean = 0.46, $CI \pm 0.21$). A full list of the means and confidence intervals for all test items can be found in the supplementary materials.

Participants in the present experiment applied their knowledge of the grammatical gender of a novel language to novel items, based on gender stereotypes; participants appeared to be more likely to select the feminine suffix for female-stereotyped items and the masculine suffix for masculine-stereotyped items, but did not show a statistically significant bias for gender-neutral items. Participants also appeared to

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

apply the gendered suffix appropriately to novel items that were similar to the training set (professions) demonstrating that they had encoded the gendered terms beyond rote memorization. These results support the hypothesis that gender stereotypes may play a role in language learning and development.

4.1. Limitations and directions for future research

While the present study shows some support of the notion of gender biases in language learning, there are important limitations to discuss. Many of these limitations could be addressed in future research. First, it is important to note that while the results reached statistical significance in most tests, there was variability across items and participants. This suggests that if the evolution of grammatical gender assignment in language makes some use of gender stereotypes, that this may be just one of many factors that played a role, and would be unlikely to explain the entire distribution of noun classes in gendered language. Gender assignment for inanimate objects can be arbitrary, as well as based on phonological and semantic factors (Culbertson et al., 2017; Gagliardi & Lidz, 2014). Future research could address this by exploring some of the factors that may lead to variation across participants and items.

It is possible that some of the individual differences in the present results could be explained in terms of individual differences in implicit gender biases. Lewis and Lupyan (2020) showed a correlation across languages between implicit sexism in the IAT (Implicit Associations Test; Greenwald et al., 1998). Meagher (2017) also showed that sexist beliefs correlated with participants' willingness to assign a gender to inanimate objects based on their gendered associations. These results suggest that participants who show stronger gendered associations might be more likely to hold more traditional, stereotyped beliefs about gender roles. Future research could explore whether differences between participants on how much they applied gender stereotypes in this language learning task are correlated with scores on the IAT or scales of gender stereotyping. It is possible that participants who hold stronger gender biases or more internalization of gender stereotypes might show a stronger effect in this artificial grammar learning task.

Eberhard et al. (2005) found that the size of the effect of gender on a similarity-rating task was dependent on the speed of learning the grammatical gender of the language; fast learners showed a larger effect than slower learners. Because the present study did not measure speed of learning the grammatical gender, and all participants received the same number of exposure trials, we do not know if variation between participants was mediated by ease of learning. Future research could test whether participants who are faster learners show a stronger effect on this task.

It is important to note that while the artificial language made use of binary grammatical categories, biosocial gender is not binary. It has been suggested that the use of binary grammatical gender categories may force speakers to view conceptual gender in similarly binary terms, which may explain why languages with binary grammatical gender categories show larger effects in studies of gendered language on cognition (Beller et al., 2015; Bender et al., 2011; Sera et al., 2002). In addition, it has been suggested that speakers of languages with binary grammatical gender categories show higher rates of sexist language, and possible negative effects for gender equity (Bassetti & Nicoladis, 2016). Future research could address this concern by

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

manipulating the grammatical gender system (e.g., binary vs. ternary systems). Participants trained on a binary system (as in the present experiment) may show stronger effects than participants trained on a ternary system.

The effects of a binary grammatical gender system may have been exaggerated with the use of a two-alternative forced-choice test, where the only difference between alternatives was the suffix, either masculine or feminine. Previous research has shown that when the task does not have a forced choice between two genders, the effects are less consistent (Bassetti & Nicoladis, 2016). Thus, the task itself may have helped to steer participants towards gendered effects. Since participants had no other knowledge of the miniature language besides its use to indicate conceptual gender of professionals, participants may have used the gender as a tool to figure out the task.

Relatedly, it is possible that demand characteristics may have affected the results. If participants noticed that some of the items were stereotypically male or female, they may have surmised that this is what the study was testing, and adjusted their responses accordingly. While this generally provides support of the hypothesis (linking stereotypes to grammatical gender), it might have consequences for the external validity of the study. There are several ways in which future research could test these effects. First, one could directly ask participants after completion of the study if they had noticed that some of the items were stereotypically masculine or feminine, and whether this affected their responses. Second, future research could ask participants to produce novel forms, to test whether learners use a novel suffix, or apply the known suffixes in line with gender stereotypes.

In the present study, all of the novel items were from specific categories (tools, makeup, and office supplies). Future studies could test items from different categories, or more heterogeneous items. Participants' responses to gender-stereotyped items suggest that learning the grammatical gender of objects should be easier if the stereotyped gender association matches the grammatical gender, since initial guesses were in line with this association. This assumption could be tested, where participants are specifically trained on items that either match or mismatch the stereotyped associations.

In order to engage in the future research described above, a much larger sample size would be required. While the sample size used in this study was similar to those used in previous artificial grammar learning studies, and was large enough for statistically significant effects for most comparisons, the sample size is generally too small to make more fine-grained comparisons related to the individual variation across participants. The use of a convenience sample from college students may not be generalizable, and may not be large enough to make broad conclusions about language learning in the general population. Thus, future research should make use of a larger and broader sample size to better address the questions that this study raises about the role of gender stereotypes in language learning.

While participants were primarily monolingual English speakers, many speakers did have exposure to Spanish, which makes use of a noun class system similar to the one used in the present experiment. In Spanish, words that end in [o] are more likely to be masculine. In the present experiment [-go] was used to mark the feminine. Thus, if participants made direct use of Spanish knowledge in the task, then we would expect a bias towards [-go] to be assigned to masculine nouns, but no such bias occurred.

It is possible that knowledge of a language like Spanish, French, or German, which all have gendered noun-class systems, may have make participants more willing to apply a gendered system to a novel language, which could have eased participants'

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

learning. However, there are several reasons to believe that effects of exposure to a language with a noun-class system are minimal in the present study. First, the items in the test were balanced for grammatical gender in Spanish (at least half of the makeup items were masculine, and at least half of the tool items were feminine in Spanish). Second, the results did not reflect Spanish gender. Third, very few participants had more than a beginner's level of exposure to the Spanish language, meaning that participants would be unlikely to know the grammatical genders of many of the nouns used in this study. Fourth, it is possible that speakers might have used their knowledge of the relative arbitrariness of grammatical gender assignment to assume that grammatical gender applied in an arbitrary manner, which could have influenced the results in the opposite direction. Fifth, if participants used general knowledge that grammatical gender can be arbitrarily assigned, this would have pushed the results in the opposite direction than was found. Finally, the outliers in the Office Supply condition include only one item that is feminine in Spanish (the eraser). The other items (paperclip, and envelope) are masculine in Spanish, meaning that the gender of Spanish is unlikely to have substantially influenced the present results.

While the effects of prior knowledge of Spanish grammatical gender is likely to be minimal, one potential solution to this would be to remove any participant with even a rudimentary knowledge of Spanish. However, this would remove the majority of the participants, since Spanish is one of the most frequently spoken languages other than English in the United States, and is one of the most popular choices to fulfill language learning requirements at the high school and college levels. In addition, the large number of Spanish speakers in the Pacific Northwest of the United States would make it unlikely that any participant had not been exposed to some Spanish in their daily lives.

5. Conclusion

The present study used an artificial grammar learning experiment to explore gender biases in language learning. Participants were exposed to a language with a binary grammatical gender pattern (different suffixes were assigned to male and female professionals). When presented with novel items, participants appeared to choose the suffix that matched the stereotype of the word at a rate greater than chance. These results support the hypothesis that gender stereotypes may play a role in language learning.

Acknowledgements. We are grateful to the participants at Pacific Lutheran University for their cooperation and assistance, as well as to Wendelyn Shore, Michelle Ceynar, Marianne Taylor, and two anonymous reviewers for helpful comments related to this research. Many thanks to Ariel Goldberg for help with stimuli creation. We resume responsibility for any and all errors.

Conflicts of Interest. We have no conflicts of interest to declare.

Supplementary Materials. To view supplementary material for this article, please visit http://dx.doi.org/10.1017/langcog.2021.21.

References

Ackerman, L. (2019). Syntactic and cognitive issues in investigating gendered coreference. *Glossa: A Journal of General Linguistics* 4(1), 1–27.

Baer, A., Trumpeter, N. & Weathington, B. (2006). Gender differences in memory recall. *Modern Psychological Studies* 12(1), 11–16.

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* 68(3), 255–278.

Bassetti, B. A. L. (2007). Bilingualism and thought: grammatical gender and concepts of objects in Italian–German bilingual children. *International Journal of Bilingualism* 11(3), 251–273.

Bassetti, B. A. L. & Nicoladis, E. (2016). Research on grammatical gender and thought in early and emergent bilinguals. *International Journal of Bilingualism* 20(1), 3–16.

Bates, D., Mächler,M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software* 67, 1–48.

Beller, S., Brattebø, K. F., Lavik, K. O., Reigstad, R. D. & Bender, A. (2015). Culture or language: What drives effects of grammatical gender? *Cognitive Linguistics* 26(2), 331–359.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology* 42(2), 155–162.

Bem, S. L. (1981). Gender Schema Theory: a cognitive account of sex typing. *Psychological Review* 88(4), 354–364.

Bender, A., Beller, S. & Klauer, K. C. (2011). Grammatical gender in German: A case for linguistic relativity? *Quarterly Journal of Experimental Psychology* 64(9), 1821–1835.

Blakemore, J. E. O. & Centers, R. E. (2005). Characteristics of boys' and girls' toys. *Sex Roles* 53(9/10), 619–633.

Boersma, P. & Weenink, D. (2017). *Praat: doing phonetics by computer.* Online <http://praat.org>.

Boroditsky, L. & Schmidt, L. A. (2000). Sex, syntax, and semantics. *Proceedings of the Annual Meeting of the Cognitive Science Society* 22(22), 1–7.

Conry-Murray, C. (2015). Children's judgments of inequitable distributions that conform to gender norms. *Merrill-Palmer Quarterly* 61(3), 319–344.

Crawford, J. T., Leynes, P. A., Mayhorn, C. B. & Bink, M. L. (2004). Champagne, beer, or coffee? A corpus of gender-related and neutral words. *Behavior Research Methods, Instruments, and Computers* 36(3), 444–458.

Culbertson, J., Gagliardi, A. & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language* 92, 343–358.

Culbertson, J., Jarvinen, H., Haggarty, F. & Smith, K. (2019). Children's sensitivity to phonological and semantic cues during noun class learning: evidence for a phonological bias. *Language* 95(2), 268–293.

Dyer, J. (2018). From curlers to chainsaws: women and their machines. *Women's Studies* 47(6), 679–680.

Eberhard, K. M., Scheutz, M. & Heilman, M. (2005). An empirical and computational test of linguistic relativity. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* 27, 618–623.

Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology* 32(1), 18–31.

Forbes, J. N., Poulin-Dubois, D., Rivero, M. R. & Sera, M. D. (2008). Grammatical gender affects bilinguals' conceptual gender: implications for linguistic relativity and decision making. *Open Applied Linguistics Journal* 1(1), 68–76.

Gagliardi, A & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90(1), 58–89.

Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology* 74(6), 1464–1480.

Heller, K., Matsak, E., Abboud, H., Schultz, H. & Zeitlin, V. (2014). *SuperLab 5.* Cedrus Corporation.

Kurinski, E., Jambor, E. & Sera, M. D. (2016). Spanish grammatical gender: its effects on categorization in native Hungarian speakers. *International Journal of Bilingualism* 20(1), 76–93.

Kurinski, E. & Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism* 14(2), 203–220.

Lenth, R. V. (2016). Least-Squares Means: the R package lsmeans. *Journal of Statistical Software* 69(1), 1–33.

Lenth, R. V. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means.*

Lewis, M. & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour* 4(10), 1021–1028.

Lidz, J. & Gagliardi, A. (2015). How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics* 1(1), 333–353.

Martinez, I. M. & Shatz, M. (1996). Linguistic influences on categorization in preschool children. *Journal of Child Language* 23(3), 529–545.

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press

Meagher, B. R. (2017). Judging the gender of the inanimate: benevolent sexism and gender stereotypes guide impressions of physical objects. *British Journal of Social Psychology* 56(3), 537–560.

Mickan, A., Schiefke, M. & Stefanowitsch, A. (2014). Key is a llave is a Schlüssel: a failure to replicate an experiment from Boroditsky et al. 2003. *Yearbook of the German Cognitive Linguistics Association* 2(1), 39–50.

Phillips, W. & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, 928–933.

R Development Core Team, R. (2018). R: a language and environment for statistical computing. In R. D. C. Team (ed.), *R Foundation for Statistical Computing* (Vol. 1, Issue 2.11.1, p. 409). R Foundation for Statistical Computing.

Ramos, S. & Roberson, D. (2011). What constrains grammatical gender effects on semantic judgements? Evidence from Portuguese. *Journal of Cognitive Psychology* 23(1), 102–111.

RStudio Team (2020). *RStudio: Integrated Development for R.* 2020. Online <http://www.rstudio.com/>.

Samuel, S., Cole, G. & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: a systematic review. *Psychonomic Bulletin and Review* 26(6), 1767–1786.

Sato, S. & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: evidence for the structural-feedback hypothesis. *Cognition* 176, 220–231.

Scott, G. G., Keitel, A., Becirspahic, M., Donnell, P. J. O. & Sereno, S. C. (2019). The Glasgow Norms: ratings of 5,500 words on nine scales. *Behavior Research Methods* 51(3), 1258–1270.

Segel, E. & Boroditsky, L. (2011). Grammar in art. *Frontiers in Psychology* 1, e2010.00244.

Sera, M. D., Berge, C. A. H. & Pintado, J. del C. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development* 9(3), 261–292.

Sera, M. D., Elieff, C., Burch, M. C., Forbes, J., Rodríguez, W. & Dubois, D. P. (2002). When language affects cognition and when it does not: an analysis of grammatical gender and classification. *Journal of Experimental Psychology: General* 131(3), 377–397.

Sereno, S. C. & O'Donnell, P. J. (2009). Participant and word gender in age of acquisition effects: the role of gender socialization. *Sex Roles* 61(7/8), 510–518.

Vigliocco, G., Vinson, D. P., Paganelli, F. & Dworzynski, K. (2005). Grammatical gender effects on cognition: implications for language learning and language use. *Journal of Experimental Psychology: General* 134(4), 501–520.

Vuksanović, J., Bjekić, J. & Radivojević, N. (2015). Grammatical gender and mental representation of object: the case of musical instruments. *Journal of Psycholinguistic Research* 44(4), 383–397.

Yorkston, E. & De Mello, G. E. (2005). Linguistic gender marking and categorization. *Journal of Consumer Research* 32(2), 224–234.

Cite this article: Finley, S., Charania, S., Lewis, T., Millward, B. & Wang, S. (2022). Gender bias in morphological inferences *Language and Cognition* 14: 47–60, https://doi.org/10.1017/langcog.2021.21

https://doi.org/10.1017/langcog.2021.21 Published online by Cambridge University Press