‘Sally the Congressperson’: The Role of Individual Ideology on the Processing and Production of English Gender-Neutral Role Nouns

Brandon Papineau, Robert J. Podesva, Judith Degen
{branpap, podesva, jdegen}@stanford.edu
Department of Linguistics, 460 Serra Mall
Stanford, CA, 94305

Abstract

Language and gender are inextricably linked; we regularly make reference to the genders of individuals around us, and the language used to do so recursively feeds the biases we hold about gender in the social world. What has been left under-investigated is the role that individual, rather than societally-held, ideologies about gender play in the linguistic system. In two web-based studies, we investigate the processing and production of gender-neutral role nouns such as *congressperson* as a function of individual gender ideology and political alignment. Our results indicate an asymmetry between the processing and production of such nouns: while individuals’ gender ideologies do not modulate processing, they do interact with political party in production tasks such that Democratic participants with more progressive gender ideologies produce more gender-neutral role nouns. We argue that these forms have become linguistic resources for indexing social progressiveness, leading to their use by Democrats and avoidance by Republicans.

Keywords: language and gender; language processing; language production; language and politics; morphology

Introduction

English contains a subset of lexical entries which identify the real-world gender identity of the individuals they pick out, consisting primarily of pronouns, kinship terms, and a limited set of other role nouns. While generally common in discourse, these terms are often socially and politically charged or contested. Consider the famous and contemporary case of English pronouns. While psycholinguistic investigations have indicated that there is a processing advantage found in singular *they* when it is paired with gender-underspecified referents (Ackerman, 2018; Doherty & Conklin, 2017; Foertsch & Gernsbacher, 1997), its usage continues to be debated on the battlefields of style guides, op-eds, and popular discourse, especially as it relates to its use as a pronoun used by non-binary or gender non-conforming individuals.

More conventionally-gendered pronouns have also been the subject of psycholinguistic analysis as they relate to real-world referent gender. In the context of the United States 2016 presidential election, von der Malsburg et al. (2020) found that participant beliefs about whether or not Hillary Clinton would win the presidency had no effect on the production of *she* as a co-referring pronoun with *the future president*, and that *she* induced a processing penalty when read in a context in which it was co-referring with *the future president*. In fact, it was co-referring *they* which increased in frequency as belief in Clinton’s victory increased. However, in the context of the 2017 British General Election, *she* was produced more frequently than *he* when co-referring with *the future Prime Minister*, with a female incumbent (Theresa May). On the other hand, there was no concomitant processing bonus for *she* over *he* until after the results of the election, indicating ‘general gender biases’ (von der Malsburg et al., 2020) in the realm of language processing. These findings are reminiscent of previous work examining the relationship between societal expectations and reading times on gender-anomalous co-referents. For instance, co-referring pronouns are harder to process when they do not align with the stereotypical gender of the role noun in question, such as *he* for *nurse* or *she* for *electrician* (Duffy & Keir, 2004; Foertsch & Gernsbacher, 1997). These findings, taken together, suggest that biases about who performs a particular social role inform the ways we produce and process the pronouns which refer to them.

Beyond the realm of pronominal reference, Pozniak and Burnett (2021) found that respondents who believed that female candidates would win in the 2020 Parisian and Marseille municipal elections were more likely to produce feminine-marked titles (as well as pronouns) to refer to the future politicians, but that masculine-marked forms were still dominant in both locales. Corpus data similarly indicates that referent gender indication is more prevalent when the gender of the referent runs counter to stereotypical assumptions. For example, the Corpus of Contemporary American English (M. Davies, 2008-) contains 165 tokens of *male nurse*, compared to 53 of *female nurse*. These biases are in turn learned by large language models trained on natural language corpora, raising concerns about the perpetuation of societal biases in the realm of automation and language (Bender et al., 2021; Caliskan et al., 2017; Sutton et al., 2018). These findings highlight the role that societally-held beliefs play in the way we choose to linguistically gender individuals in the real world.

While the aforementioned studies have investigated the role of group and societal-level biases and ideologies (which we may define as interactional systems of biases and expectations about the world) in the processing and production of gendered language, this high-level focus raises the question as to the role of *individual* ideologies on this facet of the linguistic system.

One path through which individual ideologies could affect

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

the processing and production of gendered language is by directly affecting the relative predictability of gendered terms. We can couch this idea in surprisal theory (Hale, 2001; Levy, 2008), under which a word's processing difficulty should be proportional to its surprisal given previous input $w_1, \ldots, w_{i-1}$ and any extralinguistic or extrasentential context C.

$$\text{processing difficulty} \propto -\log P(w_i | w_1, \ldots, w_{i-1}, C) \quad (1)$$

This account has received ample empirical support: for instance, more contextually surprising words incur greater reading times (Aurnhammer & Frank, 2019; Goodkind & Bicknell, 2018; Monsalve et al., 2012; Smith & Levy, 2013) and more negative N400 amplitudes (Delogu et al., 2017; Frank et al., 2013). One might thus explain processing difficulties incurred by co-referring pronouns that do not concord with stereotypical associations of particular occupations by positing that these pronouns are relatively more surprising than would be a stereotype-concordant pronoun, as a result of prior beliefs about gender roles. However, individual expectations for such co-referential terms may vary: for example, an individual with an open-minded attitude towards gender roles might consume different media than an individual with more conservative views on gender; as a result, they may be exposed to more or less gender-neutral language, respectively, and in turn be more or less surprised by its use. This would be consistent with previous findings which indicate individual exposure to lexical items modulate processing (R. A. Davies et al., 2017; Yap et al., 2012), as well as findings that expectations can be experimentally varied in individuals and in turn influence processing (Delogu et al., 2018).

Alternatively, the ideologies themselves might affect language processing and production via an independent pathway. That is, ideologically distinct groups may show effects of ideology above and beyond effects of gendered language surprisal.

We investigated whether individually-held ideologies affect the processing and production of gender-neutral language in two web-based experiments in the domain of 'role nouns.' Role nouns describe individuals' social and professional positions in the world (Misersky et al., 2014). They include both compound forms (n=14) which make a ternary distinction between male, female, and gender-neutral forms, as well as affixed forms which make only a binary distinction (n=6), see (1a) and (1b) for examples.

(1) **Critical Items**

a. **Compound:** *congress{man/woman/person}*

b. **Affix:** *villain, villainess*

This focus reflects ideological associations between such forms and gender-progressivism that have been espoused in public discourse. For example, former Acting Director of National Intelligence Richard Grenell tweeted an image of a cookie with an accompanying display-case card that read "Gingerbread Person". Alongside this was Grenell's caption: 'Stop voting for Democrats.' (Grenell, 2021). Grenell explicitly draws on language ideology to implicitly assert that elected Democrats are responsible for the proliferation of politically-correct language regarding gender. As such, these compound forms offer a fertile ground for investigating gender ideologies in linguistic processes.

Experiment 1 examines whether the *processing* of gender-neutral nouns is modulated by individuals' gender ideology. Experiment 2 examines whether gender ideology affects the *production* of these terms. We conclude with a discussion of how these findings contribute to our understanding of the relationships between ideology, language, and gender.[1]

Experiment 1: Self-Paced Reading

In an experiment similar to that of the processing experiment in von der Malsburg et al. (2020), our first investigation concerned the role that individuals' ideologies about gender play in their processing of gender-neutral role nouns. If participants do exhibit effects of gender ideology, we expect that gender-progressive participants will show faster reading times on gender-neutral terms, either as a result of exposure to the terms or weaker prior beliefs about societal gender roles. If ideology does not modulate processing, we expect there to be no difference between individuals' reading times as a function of ideology.

Methods

Participants 298 participants (mean age: 33.6) were recruited through the online recruitment platform "Prolific" (2014), excluding any participants who failed to correctly respond to at least 85% of attention check questions (n=19).[2] All participants additionally self-identified as L1 English speakers and as having been born in and currently residing in the United States. See Table 1 for participant demographics.[3]

Stimuli & Procedure In a web-based self-paced reading task, participants saw a series of 20 sentence sets of the form "[NAME] is a(n) [TITLE] from [STATE]. S/he likes [ACTIVITY]", where "[TITLE]" stands in for the critical item of gendered role noun. The states and activities were randomized at the stimuli creation stage so that they remained constant for all participants. Names varied such that each participant saw 10 vignettes with male-coded names and 10 with female-coded names. Role nouns were then distributed so

[1] It is important to note that many of the assumptions in our designs, such as the decision to use 'male' and 'female' names, implicitly endorse or perpetuate the notion of gender as a binary. We would like to highlight that these decisions in no way reflect the beliefs or values of the authors.

[2] 200 participants were initially recruited, and an additional 98 Republican participants were subsequently recruited after the original sample revealed a heavy skew towards Democrat-identifying participants.

[3] In both studies, 'Non-Partisan' participants were recruited as either Democrats or Republicans, but reported a centrist identity in the post-experimental questionnaire.

Table 1: Experiment 1 and 2 Participant Demographics (Democrat/Republican/Non-Partisan)

	Experiment 1	Experiment 2
Female	64/41/34	82/62/25
Male	46/59/25	42/46/10
Other	3/0/0	4/0/0
Decline to state	0/3/1	1/0/1

that 5 of the female names co-occurred with female-marked forms and the other 5 with neutral forms; the same was true for the male names, but with male-marked forms. We intentionally avoided gender-incongruent forms such as 'David is a congresswoman', for fear that doing so would bring too much attention to the research question regarding gender. The resulting conditions are presented in (2); participants saw each of the four combinations five times, followed by activity preferences, for a total of twenty trials. Each name and title occurred only once, such that, for example, no participant saw both *congressman* and *congressperson*.

(2) ***Stimuli Sentences***

a. Sally is a congress{woman/person} from Kansas.

a. David is a congress{man/person} from Kansas.

In order to attain sufficiently-gendered names, the twenty most popular male and female names were selected from the lists of most popular names for boys and girls in 1998 according to the United States Social Security Administration (2021). Names which appeared within the top 100 entries on both lists (e.g. Taylor, Ryan) were excluded.

Participants proceeded through these sentences one word at a time by pressing the spacebar to the reveal the next word and hide the previous; measurements of reading time were taken for each word in the sentence as a proxy for processing difficulty or effort, as has been standardized in the field (Forster et al., 2009). At the end of each trial, participants were asked about properties of the character described, providing a 'yes' or 'no' answer to questions about their home state (*Is Sally from Kansas?*) or about their preferred activities (*Does David enjoy skiing?*); these questions served both to distract from the principal question under investigation, and as attention checks. Participants were provided with an example that did not mark gender before proceeding to the main set of 20 vignettes.

Post-Experimental Survey Upon completing the reading task, participants proceeded to the post-experimental survey.

In order to assess the participants' ideologies towards gender, we employ the Social Roles Questionnaire developed by Baber and Tucker (2006). This survey consists of 13 questions which are designed to elicit both implicit and explicit ideologies about gender, including the notions of gender as an immutable fact vs gender as a social construct (what Baber and Tucker term 'gender transcendence'), as well as about the societal roles performed by the (binary) genders ('gender linking').

Each of the 13 questionnaire items was presented alongside a sliding scale from 'strongly disagree' to 'strongly agree', which corresponded to numerical values of 0 and 100, respectively. The questions related to 'gender linking' were inversely coded and then converted to the same scaling as the 'gender transcendence' subscale. Participants were then assigned a gender ideology score from 0 to 100 by taking the mean of their individual responses; a score of 0 indicated a maximally open-minded approach to gender.

Finally, participants filled out an optional post-experimental demographic survey, including questions about their own gender, political affiliations, and age. Participants who declined to indicate their age or political orientation were excluded from analysis.[4][5]

Unigram Surprisal In order to account for effects of word surprisal, the unigram surprisal of each of the twenty critical items' neutral forms was computed from the 'Spoken' (news media) section of COCA (M. Davies, 2008-). The decision to use unigram, contextless surprisal values was due to the difficulty in obtaining surprisal values for very infrequent terms, such as *foreperson*. The decision to use the same surprisal values for all participants stems from the high correlation between unigram surprisal values in the right- and left-wing sources in COCA (cor = .83).

Results

Exclusions In addition to the aforementioned participant exclusions, 238 trials (4.2%) with response times more than 2.5 standard deviations from that lexical item's mean reading time were excluded.

Model Structure A linear mixed effects model predicted length-residualized log-transformed reading time on neutral terms from dummy-coded fixed effects of political party (reference level: "Democrat") and referent gender (reference level: "female"), and centered fixed effects of participant age, gender ideology, and unigram surprisal; as well as the interactions between ideology and age, surprisal and party, age and surprisal, age and party, ideology and party, and the three-way interaction between age, surprisal, and party. These interactions were included as a result of initial investigations which revealed a significant modulation of surprisal effects by age (Fig. 2). The random effects structure included random by-participant and by-lexeme intercepts.

Gender Ideology There was no effect of gender ideology for Democrats (β = -0.00, SE = 0.00, t = -0.11, $p > 0.5$), or in the higher-level interactions for Republicans (β = -0.00, SE = 0.00, t = -1.51, $p > 0.1$) or Non-Partisans (β = -0.00, SE

[4]Experimental stimuli, data, and analysis are available at https://github.com/BranPap/gender_ideology.

[5]Pre-registrations are available at https://osf.io/yaqwx/?view_only=41020ff80b0440d09bd26f9e47cf768c.

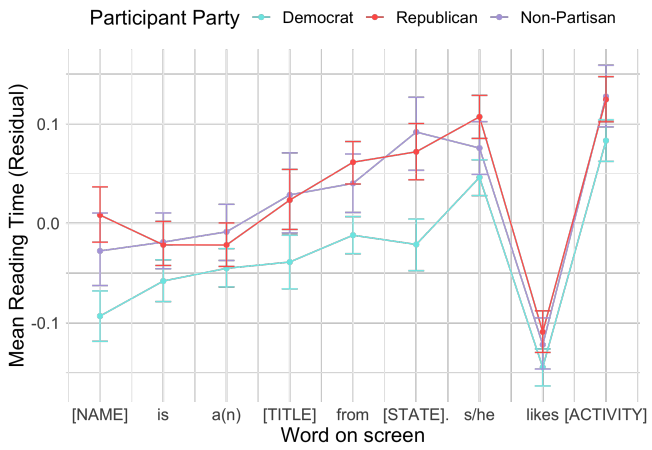

Figure 1: Residualized log reading time by word in sentence. "[TITLE]" indicates the location of the critical items.

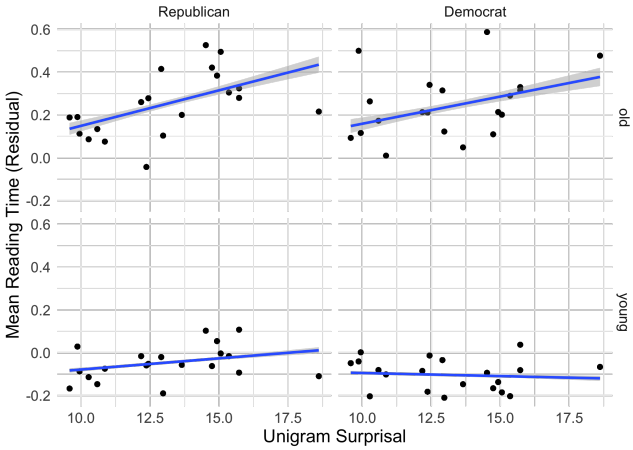

Figure 2: Residualised log reading time on critical words by word surprisal, for Republicans (left) vs. Democrats (right) and for older (top, >40 years) vs. younger (bottom, ≤40 years) participants. Each point indicates a lexeme.

= 0.00, t = -1.34, $p > 0.1$). There was thus no evidence that ideological beliefs about gender and its binary social roles modulate the processing of gender-neutral language. This is similar to von der Malsburg's results, which found no processing advantage for the pronoun which co-referred with the real-world gender of the expected election winner (von der Malsburg et al., 2020).

Political Affiliation At the party-level, we observe no significant difference in reading times on neutral items (location 4 in Fig. 1) between Democrats and Non-Partisans (β = -0.09, SE = 0.4, t = -.26, $p > 0.5$), or between Democrats and Republicans (β = 0.11, SE = 0.33, t = 0.33, $p > 0.5$). These results suggest that party affiliation does not significantly modulate processing of gender-neutral role nouns, either as a result of exposure or attitude.

Unigram Surprisal Mean residualized reading times on neutral terms are shown as a function of political affiliation, age, and surprisal in Fig. 2. More surprising words were read only marginally more slowly overall (β = -0.02, SE = 0.01, t = -1.825, p = 0.07). However, there was a significant two-way interaction between surprisal and participant age, such that older participants showed sensitivity to word surprisal in the expected direction, while young participants did not (β = 0.00, SE = 0.00, t = 2.38, p = 0.018). This may indicate that the frequency values obtained from COCA are not representative of the linguistic input experienced by younger Americans, or that younger participants are not as sensitive to surprisal effects on ideologically charged terms. A 3-way interaction between age, surprisal, and the Non-Partisan party contrast suggests that Non-partisan participants were not sensitive to surprisal (β = -0.00, SE = 0.00, t = -2.54, p = 0.01).

Interim Summary

In our investigation of gender-neutral role noun processing, we found that individual gender ideology did not significantly impact the processing of gender-neutral role nouns. We did find that there were interactions between participant party and age that modulated sensitivity to word surprisal, such that younger participants showed significantly less sensitivity to unigram surprisal than older participants.

Experiment 2: Forced-Choice Production

We next investigate the role of gender ideology on the *production* of gender-neutral role nouns. In a forced-choice task, participants selected the form of the lexeme they felt best completed the vignettes from Experiment 1. An effect of gender ideology on choice of role noun gender would indicate that gender-neutral forms are being drawn on as linguistic resources with which individuals can create outward-facing personae with gender-progressive stances. If ideology affects lexical choice, gender-progressive participants should produce a higher rate of gender-neutral role nouns than their more gender-conservative counterparts.

Methods

Participants 301 participants (mean age: 24.6) were recruited using Prolific, with the same criteria as Experiment 1[6]. Participants who failed to correctly respond to 80% of attention checks were excluded (n=25). See Table 1 for participant demographics.

Stimuli & Procedure All items in the experiment consisted of a complete sentence missing a single word, using the same sentence frames and critical items as in Experiment 1. Partic-

[6]100 Democrats and 100 Republicans were recruited initially, in order to maintain a political balance. An additional 100 male-identifying participants were subsequently recruited due to a significant gender imbalance in the initial participant population (13.4% male-identifying participants in the original population), as a result of an influx of female participants after Prolific went viral on social media app TikTok (Charalambides, 2021).

ipants were asked to select the word which best completed the sentence, by choosing from pool of possible sentence-completing words. The decision to use a forced-choice task was brought about by the difficulty in eliciting the forms under investigation, many of which are extremely infrequent. On critical trials, the choice was between the words investigated in Experiment 1.

Filler items took one of two forms; semantic fillers and grammatical fillers. Semantic fillers had no prescriptively correct answer, as in (3).

(3) Revati is a (writer/journalist/author) from India.

Grammatical fillers, on the other hand, had prescriptively correct answers, and employed grammatical processes such as demonstrative selection (4), verb agreement, or preposition selection, among others. These items served a secondary purpose as attention check questions.

(4) She is typing on (**the**/these/those) computer.

The presented order of response possibilities was shuffled between participants. There were a total of 80 trials, with 20 critical items and 60 filler items.

Trial order was randomized. After completing the experiment, participants completed the same post-experimental questionnaire as in Experiment 1.

Expectation of Neutrality To control for the possibility that participants simply produce predictable words when faced with a choice, we calculated a neutrality expectation score for each item. Because participants were presented with both gendered and gender-neutral options, we calculated this expectation as the log-transformed relative probability of a neutral over a gendered noun occurring, relative to the gender of the sentential subject referent:

$$\text{neutrality expectation} = \log \frac{P(w_{\text{neutral}})}{P(w_{\text{gendered}})} \qquad (2)$$

For example, in the sentence 'Sally is a congress[person/woman/man]', the expectation for 'congressperson' is calculated based on the relative probability of 'congressperson' over 'congresswoman'. In contrast, for 'David is a congress[person/woman/man]', the computation is based on the relative probability of 'congressperson' over 'congressman'.

Results

The proportion of neutral and gendered (male, female) noun roles selected are shown in Fig. 3.

Exclusions 241 responses were excluded from analysis for being incongruent with the names that appeared in the vignettes, such as 'David is a congresswoman' or 'Sally is a congressman'. For completeness, these responses are included in Fig. 3.[7]

[7]It is worth noting that participants were more likely to produce incongruent forms of the type 'Sally is a congressman' than of the type 'David is a congressman'. This incongruency may reflect the diachronic pathway by which masculine forms are re-interpreted as gender-neutral forms; this is the same pathway by which forms such as 'actor' and 'villain' have come to represent ostensible gender-neutrality. We leave a fuller exploration of this finding for future work.

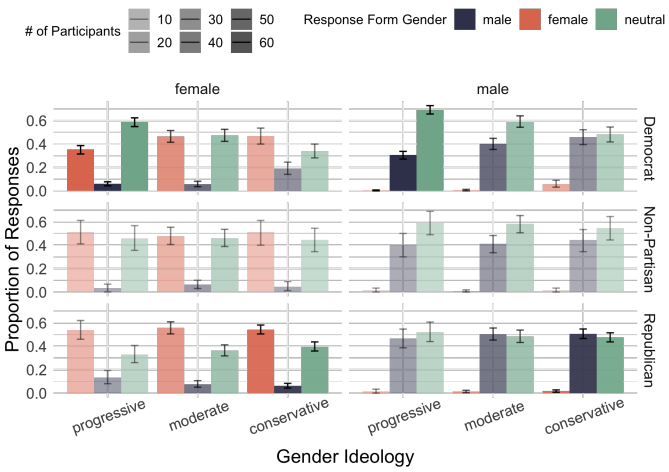

Figure 3: Proportion of neutral and gendered (male, female) responses selected in Experiment 2 as a function of participant gender ideology, separately by referent gender (left: female referents; right: male referents) and participant political affiliation (rows). Gender ideology bins are for illustrative purposes only; ideology was coded as a continuous variable.

Model Structure We fit separate logistic mixed effects models for each of the political parties, for the sake of interpretability of interaction terms. These models predicted neutral over gendered responses from fixed effects of neutrality expectation (centered), gender ideology (centered), and referent gender (dummy-coded, centered and scaled, reference level: "male") and the interaction between gender ideology and referent gender. We also included random by-participant and by-lexeme intercepts. The interaction between ideology and referent gender did not reach significance at $p < .05$ for any of the parties. The remaining effects are shown in Table 2.

Gender Ideology More gender progressive Democrats were more likely to produce gender-neutral role nouns than their less progressive counterparts (Table 2, Row 2). Republicans and Non-Partisans showed no such modulation by gender ideology. This suggests that Democrats have recruited gender-neutral role nouns as a semiotic resource with which to construct progressive personae.

Moreover, a mixed effects model on the whole dataset predicting neutral selections from only a fixed effect of political affiliation, with random intercepts for participant and lexical item, showed that Democrats had a higher base production rate of gender-neutral role nouns than their Non-Partisan (β = -0.38, SE = 0.19, z = -2.021, p = 0.04) and Republican (β = -0.83, SE = 0.13, z = -6.4, $p < .001$) counterparts. While

Table 2: Model outputs for each fixed effect (rows) for each of the political macrocategories.

	Democrats				Non-Partisans				Republicans			
	β	SE	z	p	β	SE	z	p	β	SE	z	p
referent gender	0.86	0.12	6.95	<0.001	1.04	0.22	4.67	<0.001	1.27	0.14	8.92	<0.001
ideology	-0.03	0.01	-4.64	<0.001	-0.01	0.02	-0.37	0.71	0.00	.01	.14	0.89
neutrality	9.23	2.22	4.16	<0.001	15.24	4.62	3.3	<0.001	14.6	2.32	6.3	<0.001

Democrats selected the gender-neutral forms 59.6% of the time, Republicans selected them only 45.1% of the time. The Non-Partisans selected the neutral forms at an intermediate rate, 53% of the time. This underscores the use of gender-neutral language as a marker of progressive gender ideology.

Referent Gender There was a main effect of sentential referent gender on production rates of gender-neutral titles, such that participants of all three political macrocategories were more likely to produce gender-neutral forms when picking a role title that co-referred with a male name (Table 2, Row 1). Gender-neutral forms were produced 57% of the time with male names, compared to only 48.7% of the time with female names. This may be a case of marked gender-role configurations receiving marked descriptors, as the female forms are generally either less frequent (frequency-marked) or morphologically more complex (morphologically marked).

Neutrality Expectation Finally, there was an effect of neutrality expectation in the expected direction for all three political parties, such that items with an a priori more frequently used neutral form elicited more neutral responses (Table 2, Row 3). For example, 'police officer' was more frequent than 'police woman' in the corpus, which would predict a neutral response in the female referent gender vignettes. In contrast, 'businessman' was more frequent than 'businessperson', predicting a gendered response on the male referent gender trials.

General Discussion

We observed no effect of gender ideology on the processing of gender-neutral role titles when they co-referred with gendered names. This is reminiscent of the findings of von der Malsburg et al. (2020), wherein co-referring *she* with *president* incurred a processing penalty despite societal expectations that Hillary Clinton would win the 2016 election. Our data similarly indicates individually-held beliefs about gender do not modulate the processing of gender-neutral role nouns.

However, we did observe a difference in processing as a function of age and word surprisal, such that young participants showed less sensitivity to surprisal effects than older participants. This runs counter to previous findings, which have found stronger effects of word predictability in younger participants than in older ones (Moers et al., 2017; Rayner et al., 2006; Steen-Baker et al., 2017). This finding may mean our surprisal values are not accurate for the younger participants in our study, possibly reflecting exposure discrepancies.

When selecting a role noun which co-referred with a gendered name, gender-progressive Democrats were more likely to select the gender neutral version than their more conservative counterparts. This was true both group-internally (i.e., progressive Democrats used neutral terms more than conservative Democrats) and group-externally (i.e., Democrats used more neutral terms than Republicans or Non-Partisans).

To explain this discrepancy, we argue that the 'indexical nature of morphosyntactic variables' (Eckert, 2019) enables the use of gender-neutral forms of morphologically-gendered items as semiotic resources upon which users of English can draw to convey relative social progressiveness. Democrats, then, can use these forms to index their progressive stances towards gender. Moreover, if the use of these neutral terms has come to be associated with the Democratic party or political correctness as a higher order of indexicality (Silverstein, 2003), as the aforementioned tweet from Richard Grenell indicates is the case, then even Republicans with gender-progressive stances might avoid these terms for fear of projecting an ideological system associated with their political opponents. This arugment is lent additional support by the full set of ideology scores collected, which show Democrats to be generally more socially progressive with regards to gender than Republicans, although Republicans show greater intra-group variability than do Democrats.[8]

Taken together, our results highlight an incongruity in the processing and production of gender-neutral role nouns. This incongruity, we argue, stems from the fact that individuals' ideologies about social phenomena are able to be integrated in *production*, where language users maintain a degree of agency, while the relatively limited agency maintained in *processing* inhibits such an integration. Finally, the fact that this incongruity is found at the individual level calls for a greater degree of granularity in our investigations of biases in the linguistic system, which are critical in the development of fair and inclusive language. We hope this work will encourage others to pursue such work with the individual and their ideologies in mind.

[8] See Supplementary Materials in the GitHub repository.

Acknowledgments

We would like to extend our thanks to Meghan Sumner, the members of the ALPS and Interactional Sociophonetics Labs at Stanford for their insights, and the four anonymous CogSci reviewers whose helpful feedback has shaped this paper. We would also like to extend our thanks to Adolfo Hermosillo, Alexia Hernandez, Anthony Velasquez, Bonnie Krejci, Brandon Waldon, Chantal Gratton, Christian Brickhouse, Elisa Kreiss, Evelyn Fernández-Lizárraga, Lewis Esposito, Madelaine O'Reilly-Brown, Revati Thatte, Sahba Mobini, Stefan Pophristic, and Sarang Jeong. While all of these folks have contributed to this paper in some way, shape, or form, any remaining mistakes or shortcomings are our own.

References

Ackerman, L. (2018). Processing singular *they* with generic and specific antecedents. *Architectures and Mechanisms of Language Processing*.

Administration, S. S. (2021). Popular names in 1998. https://www.ssa.gov/cgi-bin/popularnames.cgi

Aurnhammer, C., & Frank, S. L. (2019). Evaluating information-theoretic measures of word prediction in naturalistic sentence reading. *Neuropsychologia*, *134*, 107198.

Baber, K. M., & Tucker, C. J. (2006). The social roles questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles*, *54*(7-8), 459–467.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Charalambides, N. (2021). We recently went viral on tiktok - here's what we learned. https://blog.prolific.co/we-recently-went-%20viral-on-tiktok-heres-what-we-learned/

Davies, M. (2008-). The corpus of contemporary american english (coca). https://www.english-corpora.org/coca/

Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1298.

Delogu, F., Crocker, M. W., & Drenhaus, H. (2017). Teasing apart coercion and surprisal: Evidence from eye-movements and erps. *Cognition*, *161*, 46–59.

Delogu, F., Drenhaus, H., & Crocker, M. W. (2018). On the predictability of event boundaries in discourse: An erp investigation. *Memory & cognition*, *46*(2), 315–325.

Doherty, A., & Conklin, K. (2017). How gender-expectancy affects the processing of "them". *Quarterly Journal of Experimental Psychology*, *70*(4), 718–735.

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*(4), 551–559.

Eckert, P. (2019). The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language*, *95*(4), 751–776.

Foertsch, J., & Gernsbacher, M. A. (1997). In search of gender neutrality: Is singular they a cognitively efficient substitute for generic he? *Psychological science*, *8*(2), 106–111.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior research methods*, *41*(1), 163–171.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts n400 amplitude during reading.

Gal, S., & Irvine, J. T. (1995). The boundaries of languages and disciplines: How ideologies construct difference. *Social research*, 967–1001.

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. *Proceedings of the 8th workshop on cognitive modeling and computational linguistics (CMCL 2018)*, 10–18.

Grenell, R. (2021). Stop voting for democrats. https://twitter.com/richardgrenell/status/1471502835682480128?s=21

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second meeting of the north American chapter of the association for computational linguistics*.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulikova, A., Öttl, A., et al. (2014). Norms on the gender perception of role nouns in czech, english, french, german, italian, norwegian, and slovak. *Behavior research methods*, *46*(3), 841–871.

Moers, C., Meyer, A., & Janse, E. (2017). Effects of word frequency and transitional probability on word reading durations of younger and older speakers. *Language and Speech*, *60*(2), 289–317.

Monsalve, I. F., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398–408.

Pozniak, C., & Burnett, H. (2021). Failures of gricean reasoning and the role of stereotypes in the production

of gender marking in french. *Glossa: a journal of general linguistics*, *6*(1).

Prolific. (2014). https://www.prolific.co

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and aging*, *21*(3), 448.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & communication*, *23*(3-4), 193–229.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Steen-Baker, A. A., Ng, S., Payne, B. R., Anderson, C. J., Federmeier, K. D., & Stine-Morrow, E. A. (2017). The effects of context on processing words during sentence reading among adults varying in age and literacy skill. *Psychology and aging*, *32*(5), 460.

Sutton, A., Lansdall-Welfare, T., & Cristianini, N. (2018). Biased embeddings from wild data: Measuring, understanding and removing.

von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 united states and 2017 united kingdom elections. *Psychological science*, *31*(2), 115–128.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the english lexicon project. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 53.