

Implicit gender biases in the production and comprehension of pronominal references

Veronica Boyce (MIT), Titus von der Malsburg (University of Potsdam), Till Poppels (UCSD), Roger Levy (MIT)

vboyce@mit.edu

Production and comprehension of language draw on a broad range of knowledge and beliefs, including general world knowledge, contextually variable information, and social pragmatics. All of these factors are exemplified in gender processing for role nouns; where the strength of gender stereotypes vary, violations of stereotypical gender elicit surprise (e.g., referring to a surgeon with *she*), and comprehenders can accommodate to non-stereotypical genders within a discourse (Duffy & Kier 2004, Sturt 2003, Osterhout et al. 1997). Von der Malsburg et al. (2017) found a bias against female pronoun usage in descriptions of the next president during the 2016 election cycle: in both production and comprehension, participants were biased more against female descriptions than against male descriptions of the future president, relative to expectations of the individual's likely gender. However, that study leaves open questions of whether their observations are limited to the unique case of the U.S. President, to cases of real-world events of great significance, and/or to cases where the potential referent's identity is restricted to a handful of familiar individuals. Here we address these open questions by studying the relationship between world knowledge and pronoun-use expectations for a wide variety of role nouns. Additionally, we varied vignette tense between past and future, on the hypothesis that future tense might induce conceptualizations of referents as less determinate and reveal the resulting effects on pronoun preferences.

We used a set of 80 role nouns with diverse gender stereotypes (ex. *diplomat, butler, nanny, reporter*). In experiment 1, 300 Mechanical Turk participants each completed 20 cloze tasks designed to elicit pronouns referring to a role noun (ex. *The day before the championships, the gymnast worked out. Before and after working out, the gymnast stretched. . .*). In experiment 2, 366 Mechanical Turk participants did self-paced reading on mini-stories (5 items each), which contained a set-up sentence introducing the role noun, and then two sentences which each used the role noun and a co-referring pronoun (ex. *The day before the championships, the gymnast worked out. Before and after working out, the gymnast stretched her muscles to avoid getting sore. The gymnast monitored his heart rate between each set of exercises.*). If the results of von der Malsburg et al. (2017) generalize, we expected to see a reading time penalty for female pronouns and discourse mismatches. All nouns were normed both alone and in the context used for experiment 1; the gender ratings obtained for each method were well-correlated, and none of our results depend on which norm is used. Materials were constructed such that the only animate referent was the role noun; this meant *they(them,their)* could only have a singular (not plural) interpretation.

In experiment 1, we see that production of gendered pronouns closely tracks the stereotype of the noun, as expected (Fig. 1a). However, *she(her)* is produced less often than *he(him, his)*, for all levels of matching between pronoun and noun stereotype (gap between blue and pink lines, Fig. 1a). Likewise, we see more gender-neutral pronouns produced for female-biased nouns than for nouns with an equal and opposite male bias (Fig. 1b); both results indicate a systematic bias against producing female pronouns. In addition, we see more gender-neutral pronouns for more equibiased nouns and in the future tense (Fig. 1b), suggesting a preference to hedge gender reference when referent characteristics and identity might be conceptualized as less determinate. In experiment 2, we find a slowdown of 200 ms following discourse mismatches (Fig. 2a); however, we do not find an effect on pronouns that mismatch the stereotype of the noun, perhaps due to limited statistical power (Fig. 2b, c.f. Duffy & Keir 2004 and Doherty & Conklin 2017, both of which found a stereotype-mismatch effect on reading times upon encountering first pronominal references).

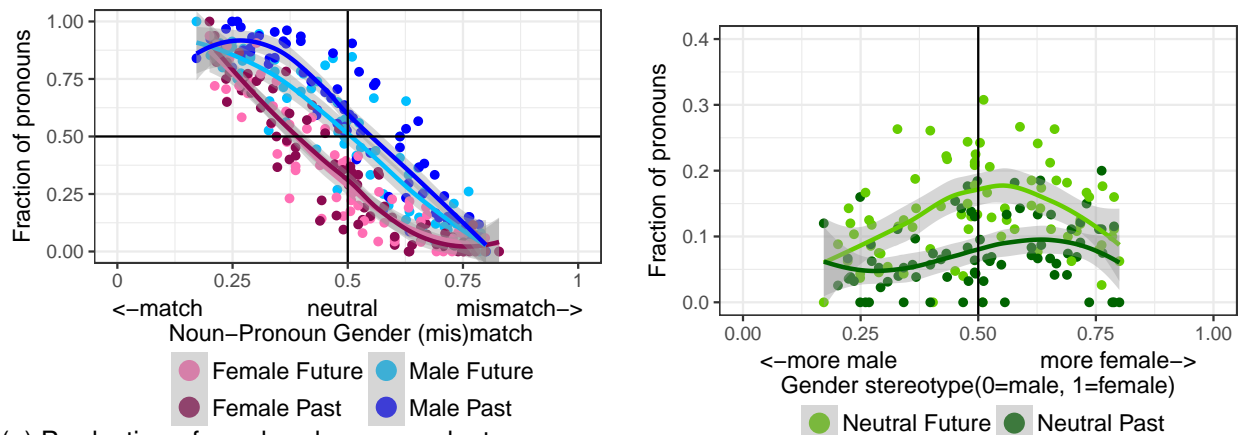
Overall, we show that the bias against female pronouns found in von der Malsburg et al. (2017) generalizes, at least in production, to a wide variety of role nouns.

Sample Stimuli (highlighting indicates regions of reading-time analysis):

Experiment 1: *After the shop on High Street [closed/closes] for the night, a baker [stayed/will stay] to tidy up. Before the baker [took/takes] out the trash, . . .*

Experiment 2: *After the shop on High Street [closed/closes] for the night, a baker [stayed/will stay] to tidy up. Before the baker [took/takes] out the trash, [she/he/they] [swept/will sweep] the floor and [wiped/wipe] down the counter. After counting the cash in the register, the baker [pulled/will pull] a key out of [her/his/their] pocket and [locked/lock] the money in the store safe.*

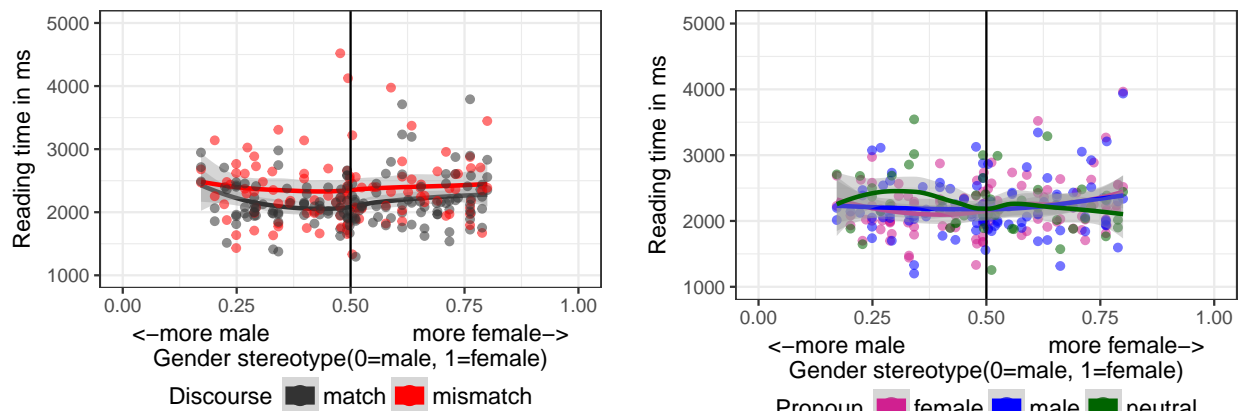
Figure 1: Cloze task results



(a) Production of gendered pronouns by tense, gender, and degree of match with noun stereotype norms. (In all figures, noun stereotype norms are compressed because norming participants avoided extreme ratings.) Across the range of stereotype norms, we find a consistent bias against producing female pronouns.

(b) Production of gender-neutral pronouns by tense and noun stereotype. We find that neutral pronouns are produced more in future tense and for neutral nouns, and also that there are more neutral pronouns produced for female-biased nouns than for male-biased nouns.

Figure 2: Self paced reading results



(a) Summed reading times at and immediately after the second pronoun, by discourse and noun stereotype. We find that second pronouns whose gender mismatches that of the first pronoun of a different gender cause increased reading times.

(b) Summed reading times at and immediately after the first pronoun, by pronoun gender and noun stereotype. We do not see a clear relationship between noun gender stereotype and reading behavior upon encountering the first pronoun.

References: Doherty, A., & Conklin, K. (2017). *Q J Exp Psychol*, 70(4), 718-735. • Duffy, S. A., & Keir, J. A. (2004). *Mem Cognit*, 32(4), 551-559. • Osterhout, L., Bersick, M., & McLaughlin, J. (1997). *Mem Cognit*, 25(3), 273-285. • Sturt, P. (2003). *J Mem Lang*, 48,542-562. • von der Malsburg, T., Poppels, T., & Levy, R. (2017). *Talk at CUNY 33*.