

Gender stereotype subversion across English *wh* islands

Presented by:

Dennis Ryan Storoshenko – University of Calgary

Jesse Weir – Simon Fraser University

CLA 2021 Virtual Conference



UNIVERSITY OF
CALGARY



Issue

- Kotek et al. (2020) points out some important concerns about the content of example sentences in syntax
 - Male gendered arguments:
 - Appear more often
 - More likely to be subjects
 - Have more varied occupations
 - Female gendered arguments:
 - More likely to be referred to with kinship terms (X's wife/mother)
 - More likely to have their appearances described
- How does this translate to the experimental setting?

Background

- Stimulus sentences in experimental research can often perpetuate stereotypes:

The teacher bought math workbooks for 10th graders for her advanced ninth grade students (Arnold et al, 2004)

After the billionaire and his wife paid the mistress asked for more money (Patson & Ferreira 2009)

- Subverting gender stereotypes in processing studies will lead to slower reading times. (Carreriras et al., 1996; Garnham et al., 2002)

Does stereotyped content influence studies where that content is unrelated to the research question?

Study Design

- English *wh*-questions where a gender stereotyped object was extracted from a *wh*-adjunct island.
 - Subject – Gender stereotyped profession (from a pre-test).
 - Extracted object would either uphold (**Match**) or subvert this stereotype (**Mismatch**).
 - Gendered pronoun within the island would follow the same alternation.
- 4 conditions:

Match-Match	Which <u>tie</u> did the general know where <u>he</u> had left after the parade?
Match-Mismatch	Which <u>tie</u> did the general know where <u>she</u> had left after the parade?
Mismatch-Match	Which <u>lipstick</u> did the general know where <u>he</u> had left after the parade?
Mismatch-Mismatch	Which <u>lipstick</u> did the general know where <u>she</u> had left after the parade?

- Also controlled for male/female stereotype of the profession making a 2x2x2 design

Study Design

Match-Match	Which <u>tie</u> did the general know where <u>he</u> had left after the parade?
Match-Mismatch	Which <u>tie</u> did the general know where <u>she</u> had left after the parade?
Mismatch-Match	Which <u>lipstick</u> did the general know where <u>he</u> had left after the parade?
Mismatch-Mismatch	Which <u>lipstick</u> did the general know where <u>she</u> had left after the parade?

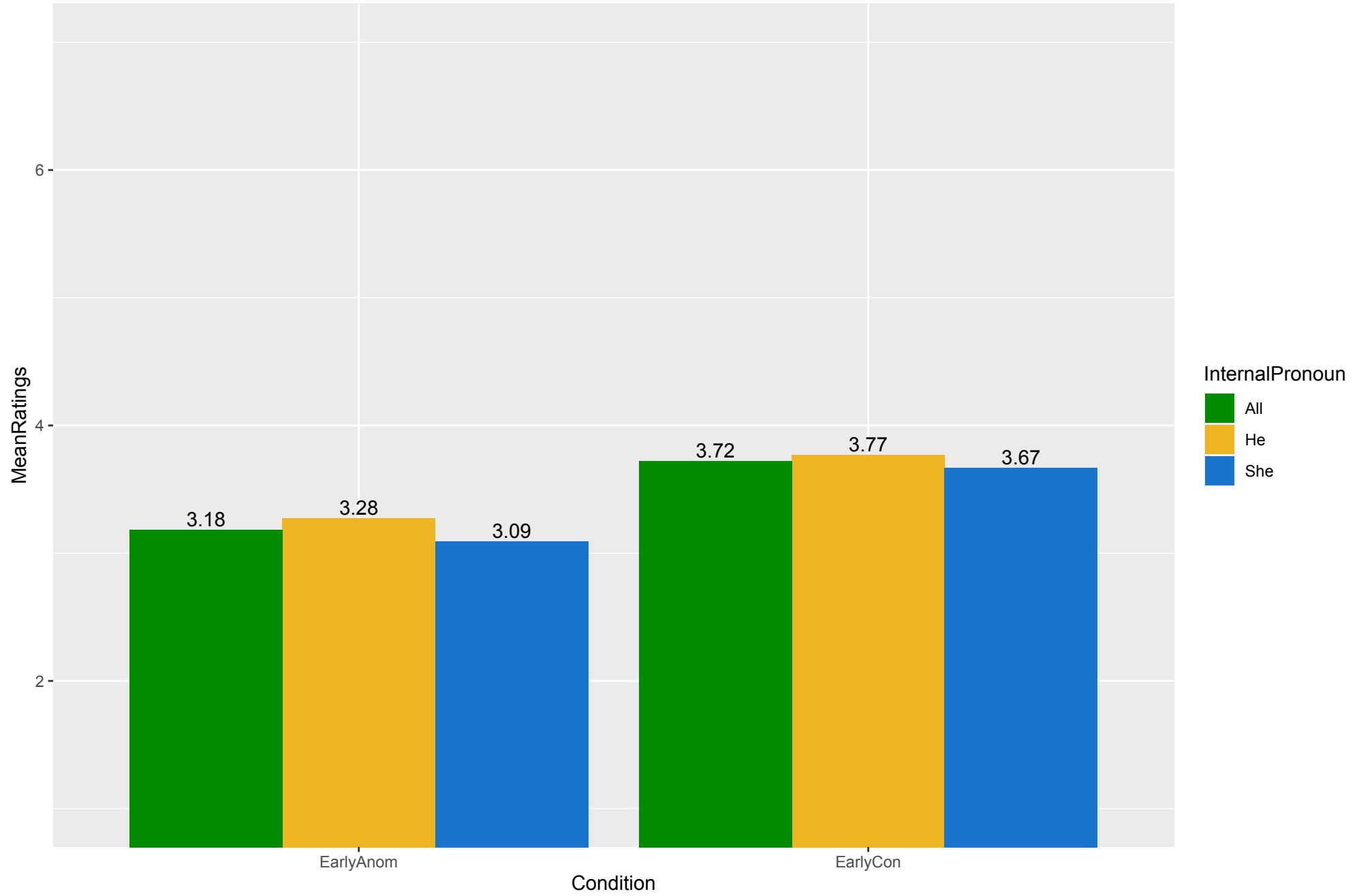
- Participants rate each sentence on a 7-point Likert scale for acceptability. Reaction times also recorded.
- Participants see four items from each condition (16 trials) along with 58 distractor fillers (including some control items).
- 64 monolingual English speakers recruited online using Prolific Academic; data collection in PsychoPy3 (Pierce et al., 2019).
- Participant recruitment controlled at 32 male + 32 female, but no significant effects for this variable emerge

Baseline – No Stereotype

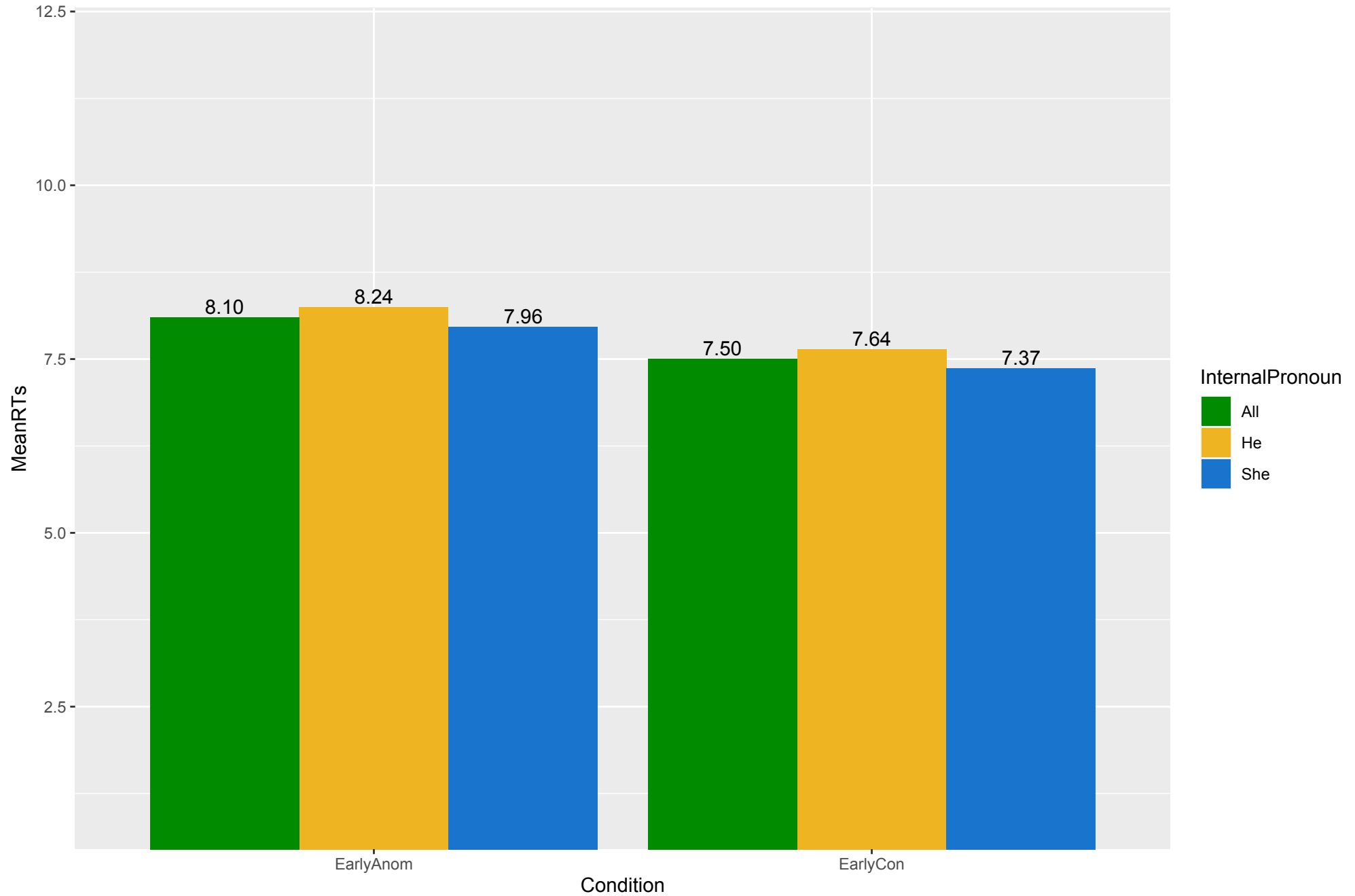
- 8 profession nouns pre-tested to have no gender stereotype
- Each paired with an object that is either **congruent** or **anomalous**
- Each participant sees 8 of these, counterbalanced between conditions and gender of the pronoun inside the island

Congruent	Which <u>helmet</u> did the cyclist forget when <u>he</u> bought for the race?
Anomalous	Which <u>flippers</u> did the cyclist forget when <u>he</u> bought for the race?

Mean Condition Ratings in Semantic Anomaly Items



Mean Condition RTs in Semantic Anomaly Items



Baseline Item Statistical Analysis

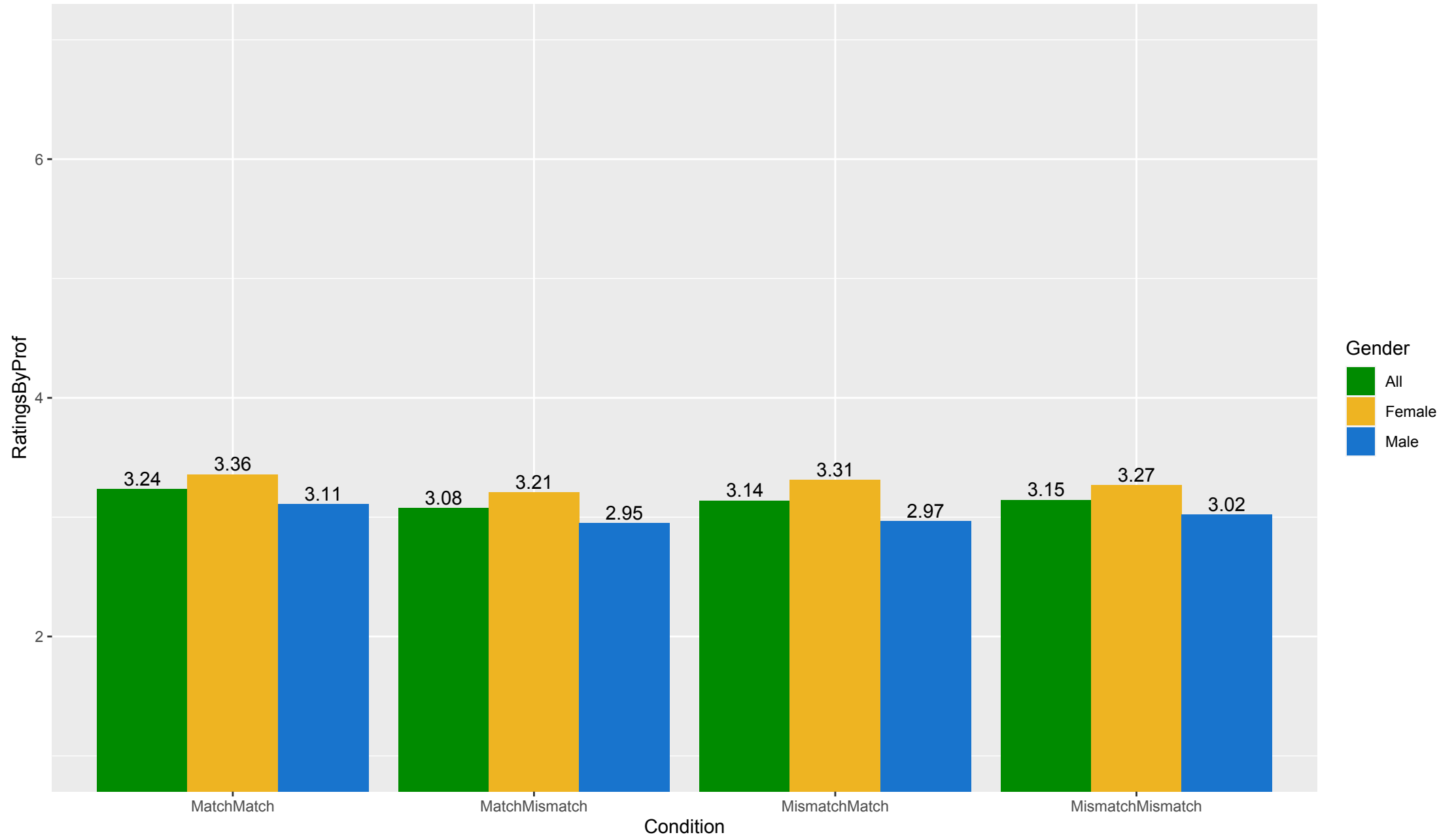
Formula: Rating ~ BeforeIsland + (1 | ItemID) + (1 | PartID)

Fixed effects:

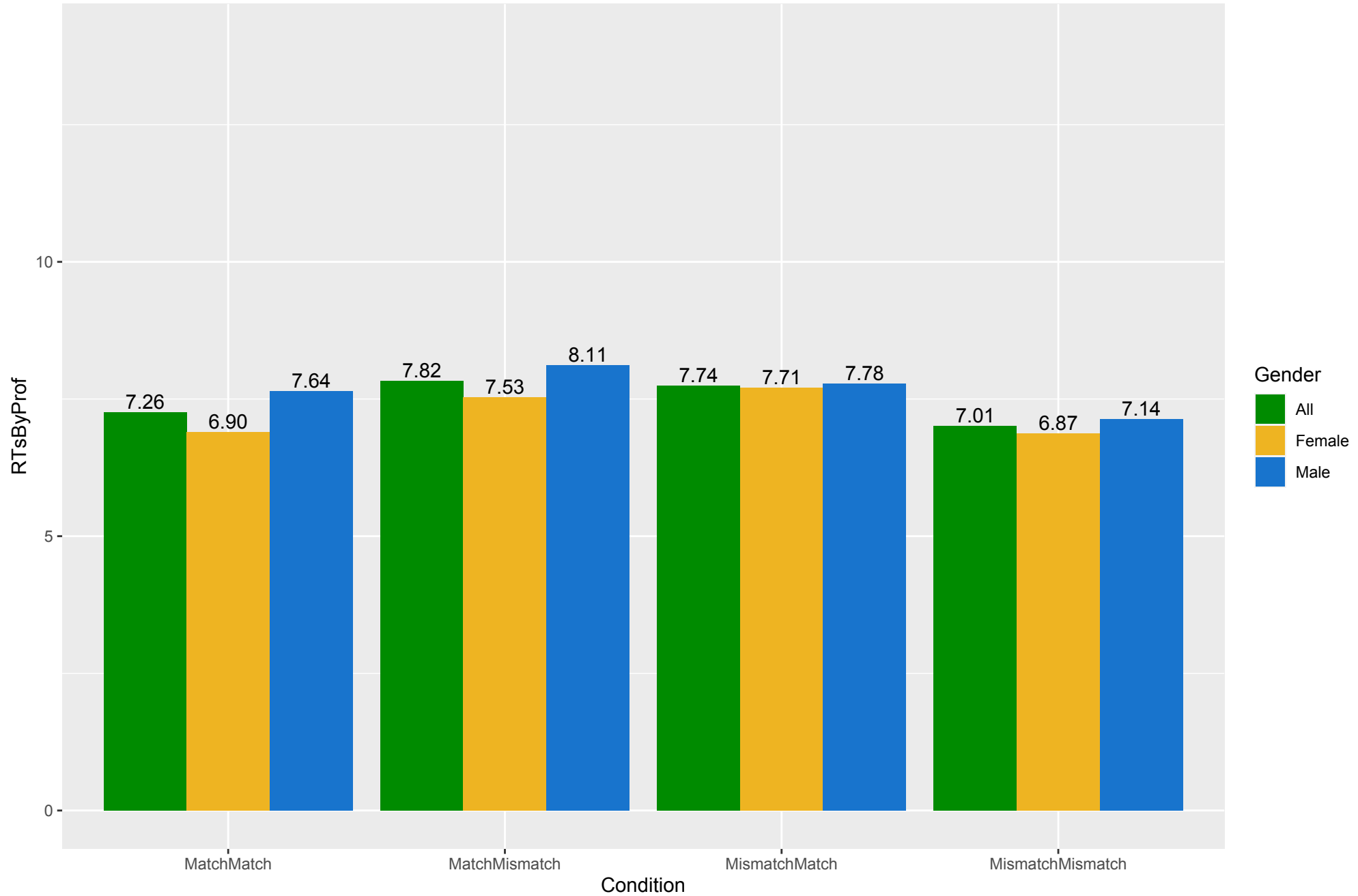
	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	3.1840	0.1615	60.5000	19.71	< 2e-16 ***
BeforeIslandEarlyCongruent	0.5331	0.1650	28.7500	3.23	0.00309 **

- Adding a semantic anomaly significantly reduces the rating
- Nothing significant in the RT data: semantic anomalies do not seem to impact processing of the island violation

Mean Condition Ratings By Stereotype of Profession



Mean Condition RTs By Stereotype of Profession



Rating Model – Test Items

Formula: Rating ~ Stereotype + (1 | ItemID) + (1 | PartID)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	3.2934	0.1680	103.0300	19.603	<2e-16 ***
StereotypeMale	-0.2767	0.1887	60.4300	-1.466	0.148

- There's really nothing here, but this has a lower AIC than any of the other fixed effects
- In a post-hoc *t*-test, the Match-Match items are rated significantly lower than the Congruent items ($p=0.002$), while there is no difference between the Mismatch-Match and the Anomalous items ($p=0.78$)

RT Model – Test Items

RT ~ Insielnd * Beforeland + (1 | ItemID) + (1 | PartID)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	7.2754	0.4242	105.2200	17.152	<2e-16 ***
InsielndLateMismatch	0.5671	0.4449	57.5000	1.275	0.2075
BeforelandEarlyMismatch	0.4677	0.4439	57.0400	1.053	0.2966
LateMismatch:EarlyMismatch	-1.3001	0.6283	57.2100	-2.069	0.0431 *

- Pairwise comparisons show the MismatchMismatch is significantly faster than both single Mismatch conditions

Three takeaways (and Promissory Note)

1. Having stereotypes at all seems to lower the ratings to the same degree as a semantic anomaly
 2. Subverting the stereotypes does nothing to significantly shift those ratings
 3. Using and subverting stereotypes impacts RT in strange and unexpected ways (So, it is probably best to avoid these in anything where you are gathering fine timing data)
- Adapting this to self-paced reading is next on our list of things to do

Thanks for visiting!

Comments?

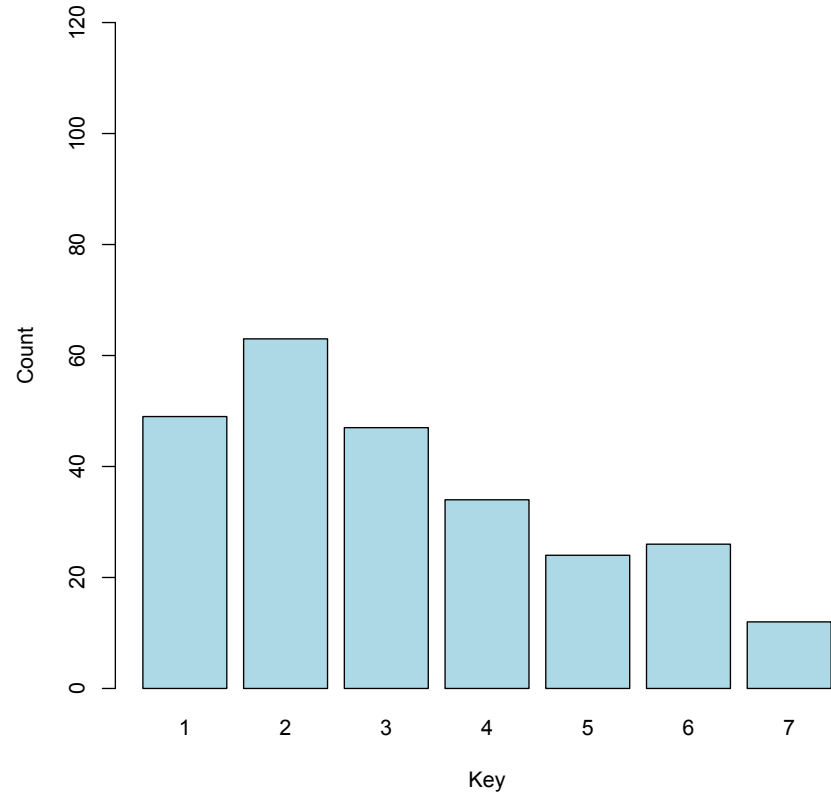
Questions?

References

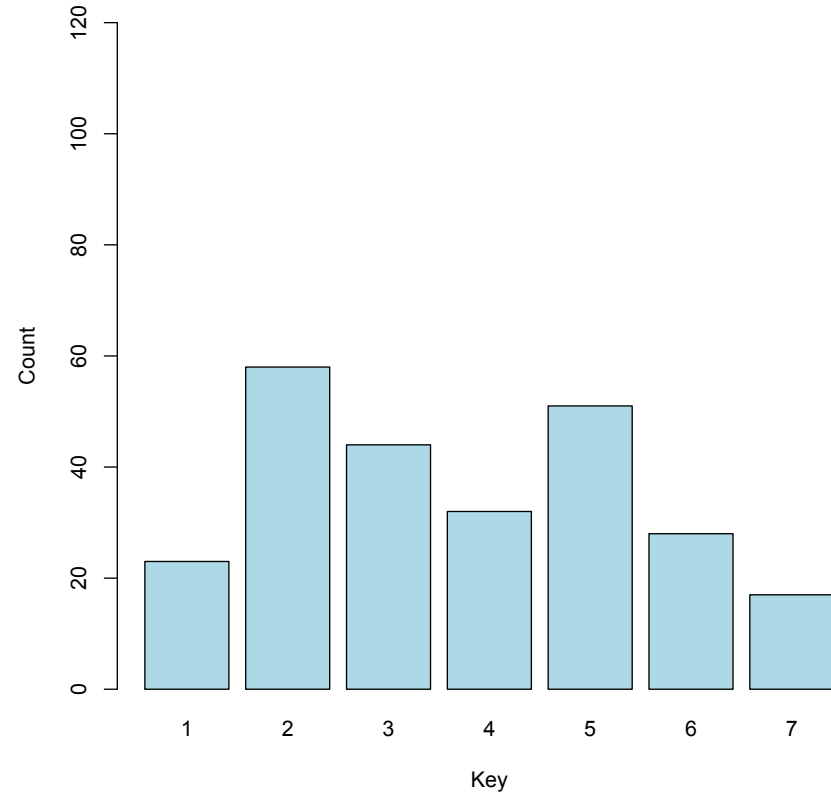
- Arnold, Jennifer E., Thomas Wasow, Ash Asudeh, and Peter Alrenga. 2004. Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language* 51:55–71.
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1-48.
- Carreiras, Manuel, Alan Garnham, Jane Oakhill, and Kate Cain. 1996. The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology* 49 A: 639–663.
- Garnham, Alan, Jane Oakhill, and David Reynolds. 2002. Are inferences from stereotyped rolenames to characters' gender made elaboratively? *Memory and Cognition* 30: 439-446.
- Kotek, Hadas, Rikker Dockum, Sarah Babinski, and Christopher Geissler. 2020. Gender bias and stereotype in linguistic example sentences. URL: lingbuzz/005367.
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H.B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1-28.
- Patson, Nikole D., and Fernanda Ferreira. 2009. Conceptual plural information is used to guide early parsing decisions: Evidence from garden-path sentences with reciprocal verbs. *Journal of Memory and Language* 60:464–486.
- Pierce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hi-royukli Sogo, Erik Kastman, and Jonas Kristoffer Lindelov. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51: 195-203.

*Research approved by the University of Calgary Conjoint REB, and funded by a URGCC SSHRC Seed Grant to Storoshenko

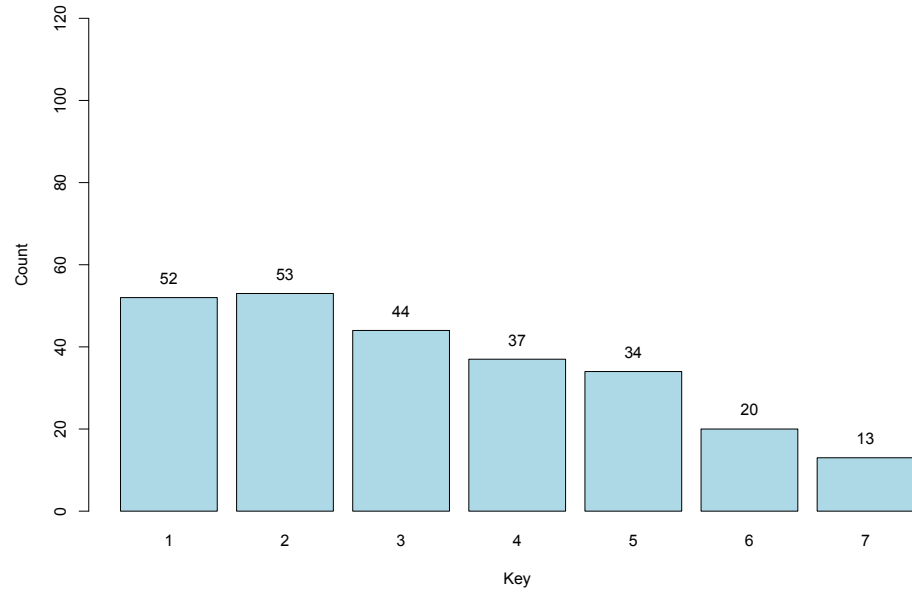
Frequency of Keypresses in Early Anomalous Condition



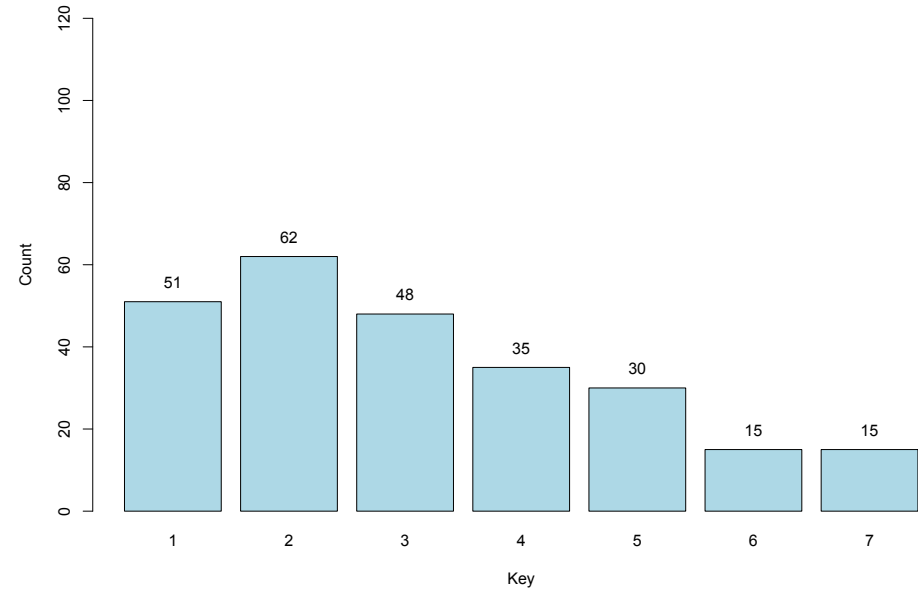
Frequency of Keypresses in Early Congruent Condition



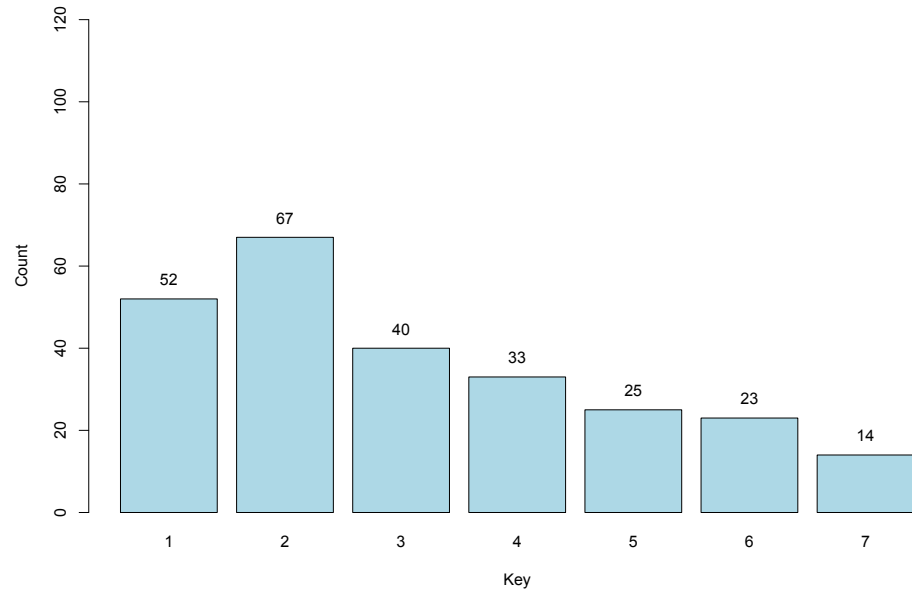
Frequency of Keypresses in Match-Match Condition



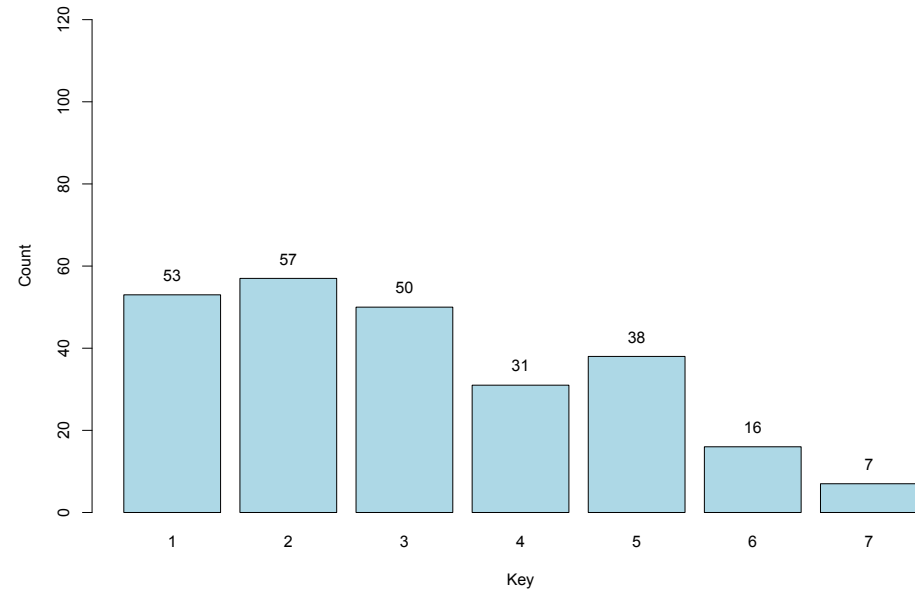
Frequency of Keypresses in Mismatch-Match Condition



Frequency of Keypresses in Mismatch-Mismatch Condition



Frequency of Keypresses in Prolific Match-Mismatch Condition



Professions Used

Female Stereotype

- Escort
- Secretary
- Nurse
- Feminist
- Flight Attendant
- Stripper
- Model
- Manicurist

Male Stereotype

- Goalkeeper
- Pilot
- Plumber
- Electrician
- Barber
- General
- Reporter
- Gamer

No Stereotype

- Artist
- Counselor
- Cook
- Marketer
- Cyclist
- Meteorologist
- Personal Trainer
- Podcaster

- Drawn from a prior study testing 100 profession nouns testing for gender stereotype by comparing preferences for a gendered object reflexive

The pilot saw herself in the mirror



The pilot saw himself in the mirror