Gender Bias in Picture Noun Phrase Reflexive Resolution

Yuhang Xu, Carly Eisen, Lauryn Fluellen, Yuyi Zhou, Nicholas Ringhoff, Rachel Coons and Jeffrey Runner (University of Rochester) yuhang.xu@rochester.edu

Background. It is well documented (*1*, *2*) that people use various types of constraints during anaphor resolution. (*1*) examined so-called "picture noun phrase" (PNP) reflexives showing that, besides a syntactic constraint (e.g., binding principle A), people are also sensitive to other constraints such as "source of information". Recently, interest has grown in understanding what kinds of constraints are used and how those constraints interact during anaphor resolution. One challenge is that these constraints often co-occur, e.g., both gender (feminine) and syntactic (reflexive) information are encoded on "herself", making it difficult to tease apart their effects. Here we try to "eliminate" one critical constraint on the reflexive (i.e., the gender cue) and test whether people are still sensitive to it, and how it interacts with other constraints in PNP reflexive resolution.

Exp1. We employed a single-trial "broken text" paradigm using PNP stimuli taken from (*1*), with gender information obscured by random characters on the reflexive (e.g., "Steve told Amy about≠the picture of h??self on the wall"). MTurk participants (N=160) read a sentence, completed an antecedent choice task, and rated their confidence on a 7-point scale. We manipulated the gender of the target/subject (male/female name), the gender pair (same/differ pair) and the source of information (verb: "tell" vs. "hear"). **Results** (Fig.1): We replicated previous findings (*1*) that the source of information had strong effects on PNP reflexive resolution ($\hat{\beta}$=1.43, z=3.00): people tended to choose the source of the information as the target (i.e., "tell" conditions); confidence ratings were also higher when the target was the source of the information ($\hat{\beta}$=1.04, z=4.73). There was also a main effect of the gender pair manipulation ($\hat{\beta}$=.99, z=2.1): people had fewer target choices when a different gender competitor was presented, which was mainly driven by the conditions where the target was female with a male competitor as the source of information.

Explanations. The gender bias favoring a male as the source of information provides evidence for context sensitive coordination of multiple constraints during reference resolution. These constraints do not exert influence separately but are intertwined: the effect of verb semantics ("hear" vs. "tell") relies on both the gender of the source of the information according to people's language experiences (that there may be a bias to interpret the source of information as male, hypothetically related to the "mansplaining" phenomenon (*3*)) and on the current context such that different gender pairs influence the weights of these effects. *Alternatively*, the male bias may simply be due to the token frequency of "himself" being higher than "herself" in usage (Google Books Ngram 1990~2008). A separate **control experiment** (Exp2) tests whether the pure frequency of the pro-form has this effect on people's interpretation.

Exp2 used the same paradigm and tasks testing how people (N=82) interpreted argument reflexives in sentences like "Mike said≠that Steve hurt h??self" again manipulating the gender of two antecedents. If people's interpretations of h??self are driven by the token frequency (himself>herself), then we should still observe a gender bias such that people have more target/matrix subject choices when it's a male name. Alternatively, we predict that because argument reflexives are lacking the sensitivity to non-syntactic constraints (*4*), there should be no gender effects in people's choices. **Results** (Fig.2) were consistent with our predication that no effects of gender would be observed (ps>.9); in fact, the male target conditions had even numerically fewer choices than female target conditions.

Together, our results are consistent with the "multiple constraints" approach (*1*) on anaphor resolution such that people are sensitive to all kinds of information even when it is unavailable (i.e., the gender information on the reflexive). More importantly, we further show that different constraints are not used independently but are intertwined and interact with each other during resolution in a context sensitive fashion.

Method:

- *"Broken-text" paradigm*
  - *"We accidentally opened some text in the wrong editor, and as a result, have some broken text we need help identifying and some related comprehension questions that we need help answering. You will first read a sentence, then answer some questions about what event occurred in the sentence and how sure you are about your answers."*
- *Single trial experiment*

Tasks:

1. Antecedent choice: Who is in the picture? (Exp1) Who was hurt?(Exp2)
2. Confidence rating: How confident are you about your choice? (7-point scale)

Exp1: PNP reflexives

Gender masked reflexives with gender stereotyped names (*5*)

Name1- {told/heard from} - Name2 - about the picture of h[?][?]self on the wall.

Manipulations:

- Target (subject) gender: male/female names
- Competitor (object) gender: same/differ gender with the target
- Source of information: tell vs. hear

Exp2: Argument reflexives

Gender masked reflexives with gender stereotyped names (*5*)

Name1 said[≠]that Name2 hurt h[?][?]self.

Manipulations:

- Target (subject) gender: male/female names
- Competitor (object) gender: same/differ gender with the target

Fig.1 Results of Exp1

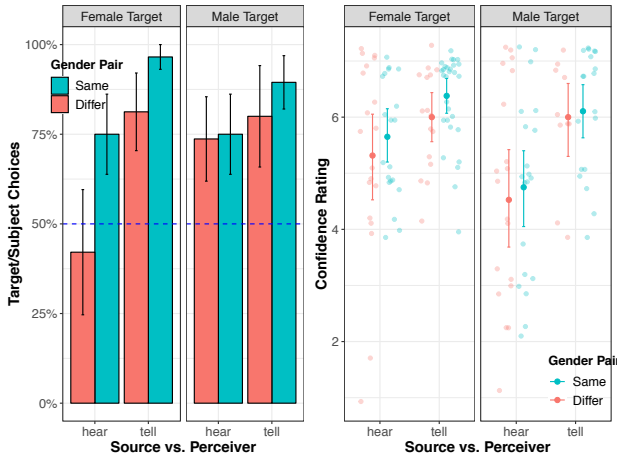
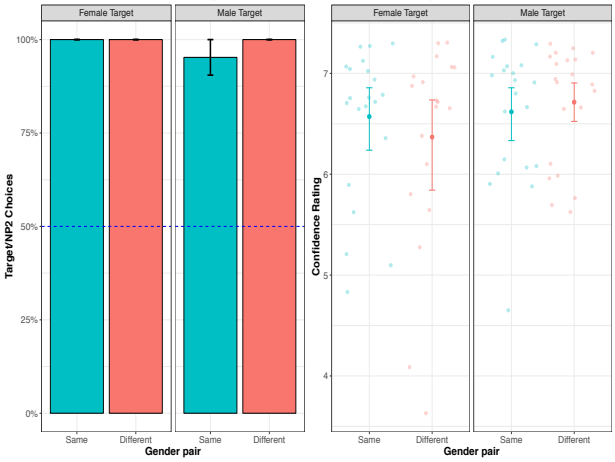


Fig.2 Results of Exp2



References:

1. E. Kaiser, J. T. Runner, R. S. Sussman, M. K. Tanenhaus, Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*. **112**, 55–80 (2009).
2. J. E. Arnold, J. G. Eisenband, S. Brown-Schmidt, J. C. Trueswell, The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*. **76**, B13–B26 (2000).
3. L. Rothman, A cultural history of mansplaining. *Atl.* **1** (2012).
4. P. Sturt, The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* (2003), , doi:10.1016/S0749-596X(02)00536-3.
5. A. Caliskan, J. J. Bryson, A. Narayanan, Semantics derived automatically from language corpora contain human-like biases. *Science (80-.).* (2016), doi:10.1126/science.aal4230.