THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

The influence of stereotypical information on gender inference in Chinese

Citation for published version:
Xu, Z & Sturt, P 2023, 'The influence of stereotypical information on gender inference in Chinese', *SN Social Sciences*, vol. 3, 30. https://doi.org/10.1007/s43545-023-00618-6

Digital Object Identifier (DOI):
10.1007/s43545-023-00618-6

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
SN Social Sciences

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 09. Nov. 2024

Manuscript Click here to access/download;Manuscript;Manuscript.docx

Click here to view linked References

The influence of stereotypical information on gender inference in Chinese

Zhe Xu[1], Patrick Sturt1

The corresponding author: Zhe Xu

E-mail address: elsa_zhexu@outlook.com

Mailing address: 38, Dongyi Xiaoqu, Jixi County, Xuancheng City, Anhui Province, China

Abstract

This study investigates the influence of stereotypical information on the representation of gender in Chinese, applying a sentence evaluation paradigm. Participants were required to decide whether the second sentence about the gender of the characters (e.g., women) was a sensible continuation of the first sentence containing plural phrase denoting a group of people with a role name (e.g., “the surgeons”). Participants' decisions were biased by the stereotypes of the role names, and were male biased when reading neutral role names. Additionally, decision time was relatively longer when there was a conflict between the stereotype and the gender, showing the greatest difference between a feminine-biased role name and a male referential expression. The results support previous research showing that speakers use gender stereotype information in task, in languages that lack grammatical gender.

Keywords: stereotypical information, gender inference, language comprehension, Chinese language

[1] Psychology department, the University of Edinburgh, Edinburgh, Scotland, United Kingdom

Introduction

When people infer the genders of a character mentioned in sentences, they integrate various sources of information, including linguistic knowledge and world experience, to build a comprehensive interpretation. However, people cannot simultaneously process all potentially relevant information, due to cognitive limitations, raising the question of what type of information people mainly rely on when they read. Both grammatical information and stereotypical information are relevant for gender inference. Grammatical information is particularly important in languages that carry grammatical gender marking (e.g., French, German, Spanish), and in many cases, this affects people's inferences about gender. For English, which lacks grammatical gender, stereotypical information plays an important role (see discussion and references below for differences between languages with and without grammatical gender). However, other languages need to be considered to allow generalization of the findings. Thus, the current study aims to fill the gap by testing the effect of stereotypical information on gender inference in referential processing in Chinese, a language which, like English, lacks grammatical gender (Farris, 1988).

There is evidence that Chinese speakers activate gender stereotype information during lexical processing. In a lexical matching task, Wang et al. (2017) showed that Chinese speakers gave more wrong answers, and had slower responses, when a prime and target word were stereotypically unrelated (e.g., female and tie), compared with matching words. Meanwhile, ERP analysis showed that the effect of gender stereotypes started earlier and elicited a larger N400 component relative to the

effect of lexical semantics (Wang et al., 2017). This suggests a high level of automaticity in lexical stereotype activation.

In a related study, Lu et al. (2019) showed an influence of gender stereotype information on the processing of Chinese names. Chinese names use a mix of characters for both girls' and boys' names. However, Chinese people tend to choose written characters with certain meanings for boys or girls. For example, the characters used in boys' names are often related to bravery and intelligence, while those of girls' names are often related to beauty and attraction. This suggests a role of stereotypes in the choice of characters for names. The study also showed that participants considered semantic information when they recognized a random name (Lu et al., 2019),

Evidence from an ERP study also points towards gender stereotypes being activated in Chinese. Su et al (2016) showed that the Mandarin feminine reflexive pronoun 她自己 ("herself") led to elevated P600 amplitude when referring to a stereotypically male antecedent, relative to a stereotypically female one.

As another non-gender-marking language, English shows strong stereotypical effects on gender inference. Many English studies have tested the effect of stereotypes, not only on lexical processing, but also in referential processing and inference. In terms of lexical processing, priming tasks show slower response times for targets words following stereotypically related primes than unrelated primes (Banaji & Hardin 1996). Moreover, reading times are slower in sentences where a role name mismatches in gender stereotype information (e.g. soldier… bikini), compared with

matching information (e.g. housekeeper…bikini) (Garnham et al, 2002). In terms of referential processing, the time taken to judge if two words could refer to the same person is faster when the two words are stereotypically related (e.g. surgeon-brother), compared with when they are not (e.g. surgeon-sister) (Oakhill et al., 2005). Moreover, pronouns and other referring expressions take longer to process if the stereotypical gender of the antecedent mismatches the referring expression (e.g. footballer…she) compared with when there is a match (e.g. footballer…he) (Carreiras et al., 1996; Sturt, 2003; Duffy & Keir, 2004; Gygax et al., 2008; von der Malsburg et al., 2020).

Gender stereotypes appear to be activated automatically, at least in some circumstances. In Oakhill et al.'s (2005) study, the inflated reaction times for deciding whether a "surgeon" and "sister" could refer to the same person were found even in a version of the experiment where the instructions explicitly pointed out that "yes" was the correct answer. In addition, Banaji & Hardin (1996)'s priming effect was found even in a version of the task that did not involve making a judgement about gender. Despite this, in referential processing, the gender stereotype effect can be overridden if specific gender information is introduced in the context (Duffy & Keir, 2004).

There have been various proposals to explain these findings. Carreiras et al. (1996) claimed that people's mental models incorporate stereotypical information and that these models are constantly updated in response to incoming information during reading. When conflicts occur, it takes additional time to update the mental model, so longer reaction times are found at mismatching words (Carrieiras et al., 1996). On the other hand, on the basis of very early evidence of a slow-down following the

reading of a mismatching reflexive (e.g. “surgeon….herself”) Sturt (2003) suggested that the effect might be lexically based.

One important question is that of how gender stereotype information influences processing in languages that have grammatical gender. Some research highlights the influence of stereotypes in grammatical gender languages (Cacciari & Padovani, 2007; Gabriel & Gygax, 2008; Molinaro et al., 2016; Casado et al., 2021; Pozniak & Burnett, 2021). For example, using a task similar to Banaji and Hardin (1996), Cacciari and Padovani (2007) showed a gender stereotype in reaction times, for Italian, a grammatical gender language. Other studies have shown the importance of stereotypical gender in Spanish (Molinaro et al., 2016), including evidence that gender stereotypes can interfere with the processing of Spanish grammatical gender (Casado et al., 2021). Additionally, Gabriel and Gygax (2008) examined the influence of gender stereotype information in Norwegian. This is an interesting case because Norway’s gender neutralization language policy has resulted in the loss of a feminine form of plural role nouns, with the previously masculine form now being used for all genders. In their study, they found effects of both stereotypical information and grammatical information, supporting the idea that both of these factors are important in gender inference.

The studies discussed above highlight an important role for gender stereotypes, both in languages that include grammatical gender, and in those that do not. However, a study reported by Gygax et al (2008) shows that these two types of language differ sharply in their use of gender stereotypes in some circumstances. In their study,

Gygax et al (2008) ran an experiment in French and German (which both include grammatical gender), as well as in English (which lacks grammatical gender). In the English version of the task, they tested sentences like 1a,b below:

(1a) The social workers were walking through the station.

(1b) Since sunny weather was forecast several of the women weren't wearing a coat.

The experiment manipulated whether the second sentence mentioned "women" (as in 1b) or "men", and also manipulated whether the plural role name of the group mentioned in the first sentence used a female stereotype (like "social workers" in 1a), a neutral stereotype (like "neighbours"), or a male stereotype (like "technicians"). In the English version of the study, participants' judgments were strongly affected by stereotypes, with more positive responses when the form used in the second sentence matched the stereotype of the group mentioned in the first sentence. For example, as "social workers" was stereotypically female, there were more positive judgments when it was followed by "some of the women" than when it was followed by "some of the men", with the opposite tendency for the male stereotypes, and an equal preference for stereotypically neutral nouns.

Gygax et al (2008) also ran equivalent studies in both French and German, and here the results were very different from the English study. Crucially, the French and German equivalents of "the social workers" (i.e. "les assistants sociaux" (French) / "die Sozialarbeiter" (German)), are grammatically masculine, yet similarly to the English equivalents, they may refer either to groups composed exclusively of males,

or to mixed groups of males and females. The results showed that in both French and German, participants showed a strong bias towards interpreting these phrases as referring to groups of men, with only a very weak influence of stereotypes. This pattern shows a sharp contrast with the English results, and indicates that grammatical gender plays an important role in gender inference in languages such as French and German, that include grammatical gender, largely overriding the influence of stereotypes. In contrast, stereotypes play an important role in English, which lacks grammatical gender.

Below, we report an experiment examining the role of stereotypical gender in Mandarin Chinese. The experiment is a conceptual replication of Gygax et al (2008)'s English, German and French study described above. Mandarin Chinese is typologically distinct from the Indo-European languages examined by Gygax et al, but it shares with English the lack of grammatical gender. There are at least two reasons why we believe our Mandarin replication makes an important contribution. The first is simply that replication plays an important role in ensuring the robustness and generalizability of previous research findings. The second reason is that by testing Mandarin Chinese, we can gain more insight into the competing roles of grammatical and stereotypical gender across languages. In Gygax et al's (2008) study, there was a strong male-bias in the interpretation of French and German equivalents of phrases like "the social workers", while in English, this male bias was absent, and responses were instead dominated by stereotypes. We agree with Gygax et al (2008) that this contrast is likely to be due to the fact that French and German have grammatical

gender while English does not. However, it remains theoretically possible that the contrast is due to some other unforeseen difference between French and German on one hand and English on the other. By testing Mandarin Chinese, a language that is unrelated to all three languages studied by Gygax et al, we can shed further light on this issue. If Gygax et al's contrast is due to the presence vs. absence of grammatical gender, we expect Mandarin Chinese to behave like English in being dominated by stereotype information, because Mandarin Chinese, like English, lacks grammatical gender. If, on the other hand, Gygax et al are wrong to conclude that their contrast was related to grammatical gender, then the Mandarin Chinese results may instead pattern like those of French or German. This would be consistent with a general preference by speakers of these languages to treat groups as consisting of men, in the absence of conclusive disambiguating information. Finally, we believe that there is value in generalizing psycholinguistic findings to as wide a typological range of languages as possible.

Methods

Participants

Thirty-six participants (eighteen males, eighteen females) from the University of Edinburgh took part in this experiment. All of them were native Mandarin Chinese speakers. Each participant was paid £5. The participants were aged between 21 and 31 years (mean = 23).

Materials

The materials and design of the current experiment were based on the study of Gygax et al. (2008). The materials consisted of 36 experimental passages and 36 filler passages that included two sentences each. The first sentence introduced a group of people using a role name. The second sentence mentioned the gender of this group, and the situation or action for some of them. An example of a passage is (2a) followed by (2b):

(2a) The nurses were walking through the station.

(2b) Since sunny weather was forecast several of the women weren't wearing a coat.

For this example, the corresponding pair in Chinese is shown in (3a and 3b):

(3a) 这些护士正在穿过车站。

(3b) 因为天气预报说有晴朗的天气，所以女人们中的一些人没有穿外套。

Out of the 36 passages, twelve used stereotypically female role names, twelve used stereotypically male role names and twelve used neutral role names. Thus, the stereotypical gender of the role name (female vs. male vs. neutral) was a between items factor. These role names were selected from the database of a Chinese norming study (Lei et al., 2019), which adopted the method of Misersky et al. (2014), using an 11-point Likert scale to record participants' ratings of the perceived relative percentages of males vs. females for each role name. The Likert scores for the role

names were transformed into proportions, such that 0 represented 100% men and 0% women, .5 represented 50% men 50% women, and 1 represented 100% women and 0% men. Accordingly, we selected 12 male biased role names with a mean perceived proportion of .17 (range: 14-18); 12 neutral role names with a mean proportion of .42 (range .41-.43); and 12 female-biased role names with a mean proportion of .78 (range .72-86). The overall mean rating of was .42.

The second sentence was manipulated for gender within items, mentioning either *men* or *women.* Thus, for the items that included a stereotypically female role name, the *women* continuation matched the stereotypical gender, and the *men* continuation mismatched the stereotypical gender. The second sentences qualified the '*men*' or the '*women*' with one of *the majority of*, *most of*, *several of*, *some of*, *few of*, *one of the men/ women*. In addition, the detailed content of the first sentences also differed. The content was translated from the English materials used in Gygax et al. (2008) into Chinese. Gygax et al. (2008) chose six different content categories, such as (1) *coming out of a place*, (2) *going into a place*, (3) *waiting somewhere*, (4) *being somewhere*, (5) *walking across a place*, and (6) *going across a place*. Six role names were randomly assigned to each of six content types. After each sentence pair, a question was presented (see procedure section), asking participants to judge whether the second sentence was a sensible continuation of the first. The correct answer to this question was "yes" for the experimental sentences.

To ensure that the participants carefully read the passages, 36 filler passages were made, requiring a "no" response to the question of whether the second sentence was a

sensible continuation of the first (note that for experimental sentences, “yes” is the correct response). The filler pairs were very similar to the experimental pairs, except that the second sentence did not make a sensible continuation of the first sentence. An English example of filler passages (4a and 4b) and the corresponding Chinese pair are (5a and 5b):

(4a) The entertainers boarded the train.

(4b) Because of the heat most of the men were wearing a coat.

(5a) 这些艺人登上了火车。

(5b) 因为高温，男人们中的大部分人都穿了外套。

The filler passages were the same in the materials shown to all participants, and they were randomly interspersed with the experimental pairs.

Procedure

The stimuli were divided into two lists using a Latin Square procedure. Each list contained one continuation condition of each item (either *men* or *women*), with an equal number of *men* and *women* continuations overall. This resulted in 36 experimental passages per list. Thus, each participant saw 72 pairs of sentences (36 experimental passages and 36 filler passages) on the screen. The first sentence was shown on the screen initially, and participants were required to press the space bar

after they had read it. Then, the second sentence was shown on the screen, and participants needed to decide whether the continuation of these two sentences was *sensible* (In Chinese, the term was translated into '合理的', which is the most suitable translation here) or not. Participants were required to put the index finger of the left hand on the 'z' key, and the index finger of the right hand on the 'm' key. If they thought the continuation was sensible, they needed to press the 'm' key; otherwise they needed to press the 'z' key. They were told to respond promptly and honestly. There was a trial session with 4 passages before the experiment.

Results

Continuation judgments

The means and standard errors of the percentages of positive judgments for the experimental passages are shown in Table 1. All the judgment data were analyzed using Logistic Linear mixed effect regression, considering participants and items as random intercepts. The participant random effect structure also included random slopes for continuation and stereotype, and their interaction, while the item random effect structure also included random slopes for continuation (n.b. stereotype was manipulated between item, so the slope parameter is not included in the item random effect structure). Random correlation parameters were excluded from the model, to aid convergence. For the fixed effects, the 2-level 'continuation' factor was mean centered, and the 3-level 'stereotype' factor was treatment coded, with the neutral stereotype as the reference level.

A model comparison using a log-likelihood chi-square showed an interaction between stereotype and continuation, χ^2 (2) = 23.862 ***, p < .001. This shows that the effect of continuation depended on the stereotype. Table 1 shows that participants' responses differed by continuation under each stereotype. When given feminine stereotypical information, participants made more positive responses if the continuation mentioned *women* (.79), compared with when it mentioned *men* (.65). Conversely, in the masculine stereotype condition, there were more positive responses when the continuation mentioned *men* (.71) than when it mentioned *women* (.44). Furthermore, participants' responses also varied in the neutral stereotype condition. More positive response occurred when the second sentence mentioned *men* (.76) than when it mentioned *women* (.67). In addition, two separate 2x2 interactions were analysed. These used the neutral stereotype as a baseline. The first interaction tests whether the effect of continuation was different for the feminine stereotype compared with the neutral, and the second interaction tests whether the effect of continuation was different for the masculine stereotype compared with the neutral. The interaction between continuation and neutral vs. feminine stereotype was significant, b = 1.537, Z = 3.43***, p < .001. Meanwhile, the interaction between continuation and neutral vs. masculine stereotype was also significant, b = -1.000, Z = -2.45*, p < .05.

Pairwise comparisons were also calculated separately for the neutral, female and male stereotypes. The pairwise comparison between male and female continuations was significant for the neutral stereotype, Z = -2.239*, p < .05, for the masculine stereotype, Z = -5.071***, p < .001, and also for the feminine stereotype, Z = 2.601**,

$p < .01$. The results therefore showed a similar pattern to those of the English study reported by Gygax et al. (2008). Moreover, one noteable result was that participants tended to make positive judgements when a neutral role name was followed by a continuation of *men,* relative to when the continuation mentioned *women*, which is unusual to emerge in a non-gender-marked language.

Table 1

Means and standard errors of the proportion of positive judgments

Stereotypes	Continuations	Means	Standard errors
Feminine	Men	.65	.05
	Women	.79	.04
Masculine	Men	.71	.04
	Women	.44	.05
Neutral	Men	.76	.04
	Women	.67	.04

Reaction times

Table 2 shows the means and standard errors of the reaction times for the second sentence judgment, excluding cases where the participant made a negative judgment. Log-transformed reaction times for positive responses were analysed in a Linear Mixed effect model using a 2 (continuation: men versus women) × 3 (stereotype:

feminine versus masculine versus neutral) design, with coding and random effects analogous to the logistic model reported above.

There was an interaction between stereotype and continuation in the reaction times, χ^2 (2) = 6.1278 *, p < .05. That shows that people's reaction times to the male and female continuations were affected differently depending on the stereotype. Table 2 shows that participants' reaction time patterns differed by stereotype. Following female stereotype nouns, reaction times were longer when the continuation mentioned *men* relative to when it mentioned *women* (5237 vs 4667msec). However, following a male stereotype noun, positive judgements were shorter for continuations mentioning *men* (3943msec) compared to those mentioning *women* (4382msec). In the neutral stereotypes, there were shorter reaction times when the second sentence mentioned *men* (4505msec), relative to when it mentioned *women* (4810msec). Furthermore, the 2x2 interactions comparing the effect of continuation for neutral versus feminine, and neutral versus masculine stereotype were analysed, and an effect was interpreted as significant when |t| > 2. The effect of continuation for neutral versus feminine stereotype was significant (t = -2.223), but the continuation for neutral versus masculine stereotype was not significant (t = -0.036).

Table 2

Means and standard errors of the reaction time (in milliseconds) of positive judgments (second sentence judgment)

Stereotypes	Continuations	Means	Standard errors
Feminine	Men	5237	382
	Women	4667	440
Masculine	Men	3943	374
	Women	4382	295
Neutral	Men	4505	328
	Women	4810	256

Discussion

Participants made more positive judgements when the stereotypes of the role nouns matched the continuations, relative to the mismatched conditions, and this pattern of results resembles the findings of the English study (Gygax et al., 2008). This similarity was expected, given the two languages do not employ grammatical gender (Farris, 1988). In Chinese, gender inference of role names should be solely based on stereotypes (i.e., if firefighters are stereotyped as male, reading it would lead to a male-biased gender representation). This finding improves the validity of the supposition that the bias of gender inference in non-gender-marked languages is caused by stereotype.

The second finding is that reaction times of positive judgements were shorter at matches of stereotypes and continuations than mismatches. This means that participants consumed more cognitive effort when they decided the mismatches to be sensible, especially when they tried to understand a man worked on a stereotypically female job. Similar results were reported by Oakhill et al. (2005) who claimed that

people spent longer time when they attempted to counteract the effect of stereotypes after they realized the effect, although completely overcoming the effect was hard. In the current study, participants were not informed of the effect before the experiments, but reaction times suggested that overcoming the effect cost. Thus, stereotypes affect people's interpretations of role names automatically, and it takes time to rectify the biased representations. Additionally, the longest reaction time, which was found at the mismatches of feminine-biased occupations and male referents, was possibly due to the unbalanced improvement of both genders in labor market. Contrary to girls' less stereotyped belief of occupational competence, boys only believed their competence in masculine-stereotyped jobs, perhaps because girls felt less pressure when they showed gender non-conformity (Canessa-Pollard et al., 2022).

One unexpected finding was that neutral role names also show male-biased representations. Participants decided male referents to be more sensible relative to female referents of neutral role names. Thus, the male inference of a neutral role name (e.g., client) would occur more frequently in Chinese. Reviewing the stereotype norms, the mean of neutral stereotypes for our items was .42, which was slightly below .5 (i.e., fully balanced neutral stereotypes), possibly causing the slightly male biased results.

This study suggests several directions for future research. First, the influence of stereotypes should be tested in more languages. Different languages, as well as different cultures, carry various specialties, so the findings based on a limited sample of languages cannot be applied to other societies where other languages are used. For

example, the role name ‘golfer’ is a masculine role name in English, French (*Golfeurs*), German (*Golfspieler*) and Norwegian (*Golfspillerne*), but this is not a stereotyped role name in Chinese, thereby not appropriate in Chinese research. Second, similar research should be tested in a more authentic and diverse environment. Researchers could recruit more local participants from different communities to obtain more representative results. Third, more Chinese language features could be tested, such as plural personal pronouns. Personal pronouns in Chinese show the similar grammatical features to that in grammatical languages. In Chinese personal pronouns, ‘她们’ with the feminine mark refers to a group of women only, while ‘他们’ with the masculine mark refers to a group of men or a group of both sexes (i.e., generic). That is different from the feature of the pronoun *they* in English which acts as a generic pronoun and shows no grammatical mark. Gygax et al. (2008) claimed that grammatical information overrode stereotypical information in grammatical language, and the masculine plurals were not interpreted as generic when the sex of the group is ambiguous. Therefore, personal pronouns in Chinese may show the similar results as grammatical languages, even though Chinese is a stereotypical language. The current study only investigated the situation in nouns referring genders, so it omitted how personal pronouns would interact with gender stereotypes, which could be tested in the future. That can help researchers to know more specifically about the relationship between language and stereotype, which may help researchers to find effective ways to reduce the negative influence of stereotypes.

References

Banaji, M.R., & Hardin, C.D. (1996). Automatic Stereotyping. *Psychological Science*, 7(3), 136-141. URL: https://www.jstor.org/stable/40062930

Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, 28(2), 277-293. DOI: 10.1017/S0142716407070142

Canessa-Pollard, V., Reby, D., Banerjee, R., Oakhill, J., & Garnham, A. (2022). The development of explicit occupational gender stereotypes in children: Comparing perceived gender ratios and competence beliefs. *Journal of Vocational Behavior*, 134, 1. DOI: 10.1016/j.jvb.2022.103703

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The Use of Stereotypical Gender Information in Constructing a Mental Model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49(3), 639-663. DOI: 10.1080/713755647

Casado, A., Palma, A., & Paolieri, D. (2021). The scope of grammatical gender in Spanish: Transference to the conceptual level. *Acta psychological*, 218, 103361-103361. DOI: 10.1016/j.actpsy.2021.103361

Duffy, S.A., & Keir, J.A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory and Cognition*, 32, 551-559. DOI: 10.3758/BF03195846

Farris, C.S. (1988). Gender and Grammar in Chinese: With Implications for Language Universals. *Modern China*, 14(3), 277-308. DOI: 10.1177/009770048801400302

Gabriel, U., & Gygax, P. (2008). Can societal language amendments change gender representation? The case of Norway. *Scandinavian Journal of Psychology*, 49, 451-457. DOI: 10.1111/j.1467-9450.2008.00650.x

Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory and Cognition*, 30(3), 439-446. DOI: 10.3758/BF03194944

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and cognitive processes*, 23:3, 464-485. DOI: 10.1080/01690960701702035

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., Oakhill, J. (2012). Gender Representation in Different Languages and Grammatical Marking on Pronouns: When Beauticians, Musicians, and Mechanics Remain Men. *Discourse Processes*, 49:6, 481-500, DOI: 10.1080/0163853X.2012.688184

Lei, C., Su, J.-J., Öttl, A., Gygax, P., & Gabriel, U. (2019). [Norms on the gender perception of role nouns in Chinese]. Unpublished data.

Lu, A., Zhang, J., Zhang, Y., Li, M., Hong, X., Zheng, D., & Deng, R. (2019). The Role of Semantic Gender in Name Comprehension : An Event-Related

Potential Study. *Journal of psycholinguistic research*, 49(1), 175-185. DOI: 10.1007/s10936-019-09677-4

Misersky, J., Gygax, P.M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulikova, A., Öttl, A., Valdrova, J., Stockhausen, L.V., & Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behaviour Research Methods*, 46(3), 841-871. DOI: 10.3758/s13428-013-0409-z

Molinaro, N., Su, J.-J., & Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and language*, 155-156, 36-43. DOI: 10.1016/j.bandl.2016.03.002

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, 33(6), 972-983. DOI: 10.3758/BF03193206

Pozniak, C., & Burnett, H. (2021). Failures of Gricean reasoning and the role of stereotypes in the production of gender marking in French. *Glossa (London)*, 6(1), 1-13. DOI: 10.5334/gjgl.1310

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562. DOI: 10.1016/S0749-596X(02)00536-3

Su, J.-J., Molinaro, N., Gillon-Downes, M., Tsai, P-S., Wu, D.H., & Carreiras, M., (2016). When “he” can also be “she”: An ERP study of reflexive pronoun

resolution in written Mandarin Chinese. *Frontiers in Psychology*, DOI: 10.3389/fpsyg.2016.0015

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. DOI: 10.1177/0956797619890619

Wang, P., Yang, Y.-P., Tan, C.-H., Chen, Q.-W., & Cantfort, T. (2017). Gender Stereotype Activation versus Lexical Semantic Activation: An ERP Study. *The Journal of general psychology*, 144(4), 283-308. DOI: 10.1080/00221309.2017.1310714

Statements and Declarations

Contribution statement

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Zhe Xu and Patrick Sturt. The first draft of the manuscript was written by Zhe Xu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

This study was funded by the Psychology department in THE UNIVERSITY OF EDINBURGH. We thank Dr. Anton Öttl and his colleagues for offering their data of the norming study.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding: This study was funded by the Psychology department in the University of Edinburgh.