CrossMark

Applied Psycholinguistics **36** (2015), 977–1006
doi:10.1017/S0142716414000010

# Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements

CHIARA REALI, YULIA ESAULOVA, and LISA VON STOCKHAUSEN
*University of Duisburg–Essen*

Received: May 3, 2012 Accepted for publication: October 7, 2013

ADDRESS FOR CORRESPONDENCE
Chiara Reali, Department of Psychology, University of Duisburg–Essen, Berliner Platz 6–8, Essen 45127, Germany. E-mail: chiara.reali@uni-duisburg-essen.de

ABSTRACT
The present study investigates the effects of stereotypical gender during anaphor resolution in German. The study aims at isolating the effects of gender-stereotypical cues from the effects of grammatical gender. Experiment 1 employs descriptions of typically male, female, and neutral occupations that contain no grammatical cue to the referent gender, followed by a masculine or feminine role noun, in a reaction time priming paradigm. Experiment 2 uses eye-tracking methodology to examine how the gender typicality of these descriptions affects the resolution of a matching or mismatching anaphoric pronoun. Results show a mismatch effect manifest at very early stages of processing. Both experiments also reveal asymmetries in the processing of the two genders suggesting that the representation of female rather than male referents is more flexible in counterstereotypical contexts. No systematic relation is found between eye movements and individual gender attitude measures, whereas a reliable correlation is found with gender typicality ratings.

The present study investigates the influence of gender stereotypes on sentence comprehension in German. In grammatical gender languages, the effect of stereotypical cues is commonly investigated in interaction with grammatical gender cues (Carreiras, Garnham, Oakhill, & Cain, 1996; Gygax, Gabriel, Sarrasin, Oakhill, & Garnham, 2008; Irmen, 2007). Our approach aims at isolating the effect of gender-stereotypical cues, while excluding the confounding influence of grammatical gender.

In contrast to natural gender languages, such as English, human role nouns in grammatical gender languages usually contain morphological markings that indicate the gender of the referent. For example, while in English a *surgeon* can be either a man or a woman, the corresponding German role noun *Chirurg/Chirurgin* "surgeon$_{\text{masculine}}$/surgeon$_{\text{feminine}}$" specifies whether or not the referent is a woman through the presence or the absence of the suffix *–in*. This characteristic can be challenging for the study of gender stereotypes, because morphological cues of

© Cambridge University Press 2014 0142-7164/14 $15.00

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

the stimuli may reveal referential gender and/or override the gender-typical representation of the role. For example, the typically male representation associated with the professional role "surgeon" may be partially or totally concealed when the role is presented in the feminine grammatical form.

In German, feminine role nouns are almost exclusively derived by the suffix *–in*, which, in most cases, is added to existing masculine terms, for example, *Maler/Malerin*, "(male/female) painter," and *Sportler/Sportlerin*, "(male/female) athlete." The feminine terms are female specific. The masculine terms are gender specific but may, in addition, be used in a generic function to designate both male and female referents. Recent observations describe a tendency toward a closer association of grammatical and lexical/referential gender, as masculine personal nouns are losing some of their "generic" potential and becoming more male specific (Bußmann & Hellinger, 2003). In comparison to role nouns in natural gender languages, therefore, German role nouns contain an additional source of gender information, which must be controlled for when testing stereotypical gender.

Stereotypes are cognitive structures that contain perceivers' knowledge, beliefs, and expectancies about a given group of persons (Hamilton & Trolier, 1986, p. 133). In the case of gender stereotypes, the reference groups are men and women. Gender-stereotypical representations may result from the perception of actual distributions of women and men in different occupations; in Germany, for example, an engineer is more likely to be a man than a woman (cf. International Labour Organization of the United Nations, 2000). This purely descriptive aspect of stereotypes may nevertheless have relevant behavioral consequences when it frames our expectation of how reality should be, for example, when it affects the decision of hiring a man or a woman in correspondence with this representation. In cognitive psychology and psycholinguistics, gender stereotypes and their influence on language processing have been studied mostly through priming paradigms and reference resolution paradigms, respectively. We will focus our review of existing research on those studies that investigate the influence of gender stereotypes with the help of the paradigm employed in the eye-tracking experiment of the present study, namely, reference resolution during sentence reading.[1]

In languages without grammatical gender, for example, in English (for overviews on gender systems, see, e.g., Cacciari & Cubelli, 2003; Corbett, 1991; Stahlberg, Braun, Irmen, & Sczesny, 2007), the effects of gender typicality are commonly investigated through the use of role nouns, which are usually unmarked for gender (morphological gender marking, as in *actr-ess* or *waitr-ess*, is rare). Studies on these languages have shown the activation of gender stereotypes conveyed through social and occupational role nouns. This effect is reflected in a disruption of the anaphor resolution process in the condition of mismatch between antecedent and referent gender; the influence of stereotypical cues has been documented with various methods of investigation.

In a reading time study, Kennison and Trofe (2003) analyzed the influence of gender stereotypes on pronoun resolution. Participants were presented with pairs of sentences. The grammatical subject of the first sentence was a typically male or female role noun; the subject of the second sentence was a pronoun (*he/she*) that referred back to the role noun (e.g., *The executive . . . she . . .*). Results showed

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

longer reading times in the condition of mismatch between gender typicality of the role noun and the gender of the personal pronoun. The mismatch effect occurred in the region following the pronoun. A similar paradigm was used by Duffy and Keir (2004) in an eye-tracking study. Participants read sentences containing a typically male or female role noun, followed by a gender-congruent or incongruent reflexive pronoun (*himself/herself*). In addition, the target sentences were partly preceded by a context where referent gender was specified (e.g., *The electrician was a cautious woman*). Results showed that in the absence of a disambiguating context, gender stereotypes were activated and that they caused longer fixation times on the pronoun and the spillover region in the gender-incongruent condition. In contrast, the specification of the referent gender in a preceding context eliminated the mismatch effect between role noun typicality and gender of the reflexive pronoun. This shows that the activation of stereotypes can be modulated by a manipulation of context information.

Role nouns with stereotypical and definitional gender were contrasted in an eye-tracking study by Kreiner, Sturt, and Garrod (2008), with reflexive pronouns appearing in anaphoric or cataphoric positions (see also Van Gompel & Liversedge, 2003; and Sturt, 2003, for resolution of pronouns in cataphoric position). When reflexives were anaphoric (e.g., *Yesterday the minister/the king left London after reminding himself/herself about the letter*), definitional and stereotypical gender produced the same mismatch costs in terms of longer fixation times. With reflexives in cataphoric position, in contrast, only definitional role nouns led to mismatch costs (e.g., *After reminding himself/herself about the letter, the minister/the king immediately went to the meeting at the office*), which suggests that stereotypical cues can be outweighed by a prior specification of the referent gender.

Evidence for gender stereotype effects on anaphor resolution also comes from event-related potentials data in Osterhout, Bersick, and McLaughlin (1997). The experiment investigated the processing of stereotypically and definitionally male and female role nouns followed by a reflexive pronoun. The reflexives either matched or mismatched the gender of the role noun. A positive deflection around 600 ms after onset of the reflexive pronoun was found in the condition of mismatch between the gender of a role noun and the reflexive pronoun, with a wider amplitude for sentences containing role nouns whose gender was determined by definition, compared to stereotypical ones.

These studies on gender stereotypes in English document a gender typicality effect that emerges as a disruption in reference resolution in the condition of gender mismatch between an antecedent and a personal or reflexive pronoun. This typicality effect appears weaker than the effect generated by biological/definitional gender and can be modulated through previous context. Possible differences in the mismatch effect produced by male in comparison to female stereotypes, as well as by the two personal pronouns, were usually not analyzed. In a sentence-reading experiment with English material, Carreiras et al. (1996, exp. 1) presented role nouns with male, female, and neutral gender typicality, followed by a masculine or a feminine anaphoric pronoun. The analysis of the gender-stereotyped items showed a main effect of gender match/mismatch but no interaction with the gender stereotype of the role, which suggests that the mismatch effect was of equal size for male and female roles. In the experiment by Kennison and Trofe (2003)

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

mentioned above, the authors report data showing a gender mismatch effect for both the masculine and the feminine pronoun. Altogether, these data may suggest that in natural gender languages the mismatch effect is symmetrically triggered by the two genders. To accurately answer the question, however, further research is needed to systematically analyze possible interactions among role noun stereotype, pronoun gender, and the mismatch effect.

In natural gender languages, most role nouns convey only semantic and stereotypical cues to gender. In contrast, personal nouns in grammatical gender languages, such as Spanish or German, generally contain grammatical markings that indicate the gender of the referent. Therefore, psycholinguistic studies on gender stereotypes in grammatical gender languages have always studied the effects of gender typicality in interaction with grammatical gender.

In the self-paced reading experiment with Spanish material conducted by Carreiras et al. (1996), sentences contained a role noun followed by a pronominal anaphor. The grammatical gender of the role noun could match or mismatch its own stereotypical gender. Moreover, the stereotypical gender of the role noun could either match or mismatch a subsequent pronoun (e.g., *El carpintero/La carpintera tomó las medidas para hacer el armario. Era un encargo bastante urgente. El/Ella tenía que terminarlo en el plazo de una semana*. "The carpenter took measurements to make the cupboard. It was a quite urgent order. He/She had to finish it in the space of one week."). Results showed slower reading times on the initial region in the condition of mismatch between grammatical and stereotypical gender (e.g., *La carpintera* "the carpenter$_{\text{feminine}}$"). In the last sentence, which contained the anaphoric reference, no effect of typicality was found when referent gender was already established via morphological features of the role noun and its preceding article. This study shows that when a role noun is encountered, the gender information provided by stereotypicality is compared with, and if necessary overruled by, gender cues provided by the local morphology. Once the referent gender is signaled through grammatical cues, no typicality effect emerges in the subsequent steps of discourse comprehension.

In German, a grammatical gender language with three gender categories and fewer overt gender markings than Romance languages, the mismatch effect between antecedent and anaphor emerged asymmetrically for male and female antecedents. In an eye-tracking study on reference resolution, Irmen (2007, exp. 1) found a mismatch effect between the stereotypical gender of the antecedent and the lexical gender of the anaphor only with stereotypically male role nouns followed by a female anaphoric noun phrase ("these women"). Similarly, in an event-related potential experiment on reference resolution, Irmen, Holt, and Weisbrod (2010) detected a larger mismatch effect, in the P600 window, for sentences where male antecedents were followed by a female anaphor. In both experiments, however, all antecedents were presented in the grammatically masculine form, which may have biased readers' expectations toward a masculine anaphor.

One possibility of analyzing the effect of gender stereotypes without interference of grammatical gender lies in the use of bigender role nouns, which do not possess a definite grammatical gender and can refer to both male and female persons (Cacciari, Carreiras, & Barbolini Cionini, 1997). Irmen (2007, exp. 2) used nominalized adjectives and present participles, whose plural forms are

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

bigender forms in German, as antecedents in an eye-tracking study with an anaphor resolution paradigm. Typically male, female, and neutral role nouns were followed by the anaphoric expression *diese Männer*, "these men," or *diese Frauen*, "these women." Because of the scarcity of stereotypical bigender role nouns in German, only a small number of role nouns was employed (three typically male, three typically female, and six neutral roles). Results showed an interaction between stereotypical gender and anaphor gender, and a male bias in the resolution of the anaphor, with longer fixation times for the female anaphor "these women," regardless of the stereotypical gender of the antecedent. This suggests that grammatically unmarked role nouns in German are understood as indicating primarily male referents, whereas a group consisting exclusively of female referents is expected only after an antecedent with feminine grammatical gender.

Bigender nouns were also employed in a study on Italian by Cacciari and Padovani (2007). The authors used bigender role nouns with a neutral morphological marker (suffix *–e*) in a single word priming study. Participants were instructed to read a role noun (e.g., *insegnante*, "teacher") followed by a personal pronoun (*lui/lei*, "he/she") and to identify the gender of the pronoun, regardless of the preceding role noun. Results showed an effect of gender typicality on response times. Interestingly, an inhibitory effect was detected for typically female role nouns followed by the incongruent pronoun (e.g., *insegnante/lui*, "teacher/he") but not for typically male role nouns followed by the incongruent pronoun (e.g., *ingegnere/lei*, "engineer/she"), which may indicate an asymmetry in the processing of male and female roles.

The reviewed studies in grammatical gender languages dealt with the complex interference of gender stereotypes and grammatical gender information, showing that the two sources of gender information can compete with each other or even override one another, as in the case of the feminine suffix for stereotypically male roles. Studies employing bigender role nouns may allow a separate investigation of gender stereotype and grammatical gender. The restricted number of available items, however, represents a limitation for languages such as German, Italian, or Spanish, where there are few bigender role nouns with strong gender typicality, especially for typically female roles (cf. Irmen, 2007).

The present study aims to overcome the limitation mentioned above by using an approach that enables us to isolate the influence of gender-stereotypical cues from grammatical gender cues without restricting the range of roles that can be included in the investigation. This is achieved by replacing role nouns with role descriptions, that is, sentences describing role-typical behavior and activities. The descriptions were empirically developed to convey the contents of a role noun, but without the presence of any morphological or grammatical gender cue. This approach offers insights into the effects of gender stereotype activation during anaphor resolution in a grammatical gender language, without any interference of morphological gender markings and grammatical gender agreement. The study focuses on professional activities, because they represent a critical area where gender stereotypes play an important role (Heilman & Eagly, 2008).

The rationale of the study relies on the assumption that the anaphor is resolved through the use of stereotypical but not grammatical gender information. However, it could be argued that the job descriptions spontaneously activate their

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

corresponding role nouns, and consequently grammatical gender markings, in the reader's mind. To test this hypothesis, we conducted a reaction time priming experiment (Experiment 1). Participants were presented with typical role descriptions and had to accomplish a decision task on the semantic relatedness of a following role noun, which could be gender-typical or neutral and grammatically masculine or feminine. We postulated that if the job descriptions spontaneously activate grammatical gender, this would affect the processing of the target role nouns with matching or mismatching grammatical gender. A lack of mismatch effect between job descriptions and the grammatical gender of stereotypically neutral role nouns would suggest that the descriptions did not prime grammatical gender information.

In Experiment 2, we employed the same role descriptions, combined with a target sentence containing an anaphoric personal pronoun, which could match or mismatch the stereotypical gender of the description. We expected a gender stereotype mismatch effect on anaphor resolution for both masculine and feminine pronouns. We used the methodology of eye tracking to obtain a precise assessment of the time course of sentence processing and the localization of possible effects with high spatial resolution on the target sentence.

The present study aims at determining the effects of gender stereotypes. Therefore, we assessed individual attitudes toward the sexes and implicit stereotypical associations, because gender stereotypical beliefs and the individual representation of social gender roles may affect participants' expectations in assigning referent gender and may modulate the disruptive effect after a mismatching referent is encountered. For this purpose, participants completed a set of questionnaires on sexism and sex role attribution, and an implicit association test for gender stereotypes, to control for possible covariation with the eye-movement data.

## EXPERIMENT 1

The goal of the first experiment was to test whether reading descriptions of a profession automatically activates the grammatical gender that corresponds to the gender typicality of the profession. The job descriptions were developed to convey the gender typicality of the job without any grammatical cues to referent gender. Even in the absence of grammatical cues in the stimulus material, it may be argued that grammatical gender is an intrinsic feature of the language and might still be activated when reading the descriptions, namely, through a spontaneous activation of the role noun corresponding to the occupation described.[2] Previous studies have shown that word recognition can be facilitated by a prime word with matching grammatical gender and inhibited by a prime with mismatching grammatical gender (about the priming effect of grammatical suffixes, see Bates, Devescovi, Hernandez, & Pizzamiglio, 1996; Cubelli, Lotto, Paolieri, Girelli, & Job, 2005). If the descriptions actually activate morphological gender cues, then target items with corresponding grammatical gender are likely to be processed faster than the same items with the opposite grammatical gender. The possible activation of grammatical gender was tested through a priming task, employing job descriptions as a prime and role nouns as a target. To control for the influence of gender typicality, the test was conducted employing gender-typical as well as gender-neutral role nouns.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

### Method

*Participants.* Thirty-two native speakers of German (16 male, 16 female, mean age = 21.9 years, *SD* = 2.2), students at the Department of Psychology at the University of Heidelberg, participated in the experiment. They received a course credit for their participation.

*Materials.* The job descriptions were empirically developed through a procedure consisting of four steps, as outlined below. Different samples of participants, all native speakers of German, contributed to the different tasks, except for Steps 2 and 3, which were carried out by the same group of participants. None of the participants of the different pretests took part in the reaction time study or the eye-tracking study.

In Step 1, a set of 77 role nouns was selected from published materials providing gender typicality ratings (Gabriel, Gygax, Sarrasin, Garnham, & Oakhill, 2008; Irmen, 2007; Kennison & Trofe, 2003). The aim was to gather a large sample of nouns describing professional roles or occupations. In the following *production task* (Step 2), 30 female and 20 male students of the Department of Psychology at the University of Heidelberg were instructed to produce two descriptions for each role noun. The role nouns were presented in the masculine singular form plus the feminine suffix (e.g., *Florist/in*, "florist$_{m/f}$"). The descriptions were to follow the basic structure verb + noun (e.g., "sells flowers"). Other words could be added after the verb and after the noun, to allow for the use of prepositions or adjectives and of separable verbs (e.g., *arbeitet in einer medizinischen Praxis*, "works in a medical surgery"; *stellt Möbel her*, "produces furniture"). Participants were requested to describe each profession as specifically as possible in two phrases, so that another person would be able to guess the role names by reading their descriptions. In a following *rating task* (Step 3), participants estimated the extent to which the occupational group denoted by each role noun consisted of women or men, with 1 = *only men*, 7 = *only women*, 4 = *same amount of women and men* (see Gabriel et al., 2008). Items were presented on a computer screen in random order for each participant. Based on the results of these ratings, role nouns were classified as typically male, typically female, or neutral (male $\leq$ 2.5, neutral = 3.5–4.5, and female $\geq$ 5.5), which yielded 21 male, 16 neutral, and 14 female role nouns. The grammatical subject of the described activity was represented by initials (e.g., "A. B. repairs cars"). The descriptions did not contain any grammatical cue to the gender of the sentence subject. In the *reverse task* (Step 4), the 51 descriptions were shown to a sample of 40 participants, who were asked to guess the role noun that corresponded to each described occupation. Only those descriptions that reached the threshold of 80% of correct responses were considered valid for the experimental material. From these, we selected 12 typically male, 12 typically female, and 12 neutral items. The same participants also rated the gender typicality of the descriptions, following the same procedure that had been used for the role noun rating. The correlation between the typicality ratings of the role nouns and those of the descriptions was solid ($r = .995$, $p < .001$). The resulting 36 descriptions were employed as experimental materials in both experiments.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

The descriptions consisted of two or three propositions and could vary from 43 to 89 characters per item, but they did not differ significantly in length between typicality groups.

### *Procedure*

Participants were presented with the typically male and female descriptions, each followed by a role noun. Their task was to decide as fast and as accurately as possible if the role noun corresponded to the preceding description by pressing two different keys on the computer keyboard. The position of the correct response key (right/left) was balanced across participants. The role noun following each description could be semantically related (corresponding to the description) or unrelated (not corresponding to the description). In addition, the role noun could appear in the grammatical gender that matched the gender typicality of the description or in the incongruent grammatical gender form, as shown in Table 1.

Semantically related role nouns were selected on the basis of the reverse task pretest (Step 4 of the material pretesting), where participants had produced role nouns corresponding to the descriptions. The semantically unrelated role nouns were randomly selected among the items with neutral typicality. The lack of semantic relatedness between these items and the descriptions was tested by having a different sample of 20 participants (native speakers of German, students of the Department of Psychology at the University of Heidelberg) rate the semantic relatedness between descriptions and role nouns on a 7-point scale (1 = *minimum*, 7 = *maximum relatedness*). Only items with mean ratings lower than 2 were considered semantically unrelated.

Each participant saw all the descriptions followed by a role noun displayed in two conditions: in one condition the noun was semantically related to the description, requiring a "yes" response; in the other condition it was semantically unrelated, requiring a "no" response to the task question ("Does the role noun correspond to the description?"). Conditions 1 (semantically related, grammatically congruent) and 4 (semantically unrelated, grammatically incongruent), as well as Conditions 2 (semantically related, grammatically incongruent) and 3 (semantically unrelated, grammatically congruent), were displayed within participants, so as not to expose participants to four repetitions of the priming description. Participants received the four conditions in equal proportion. We used E-Prime 2.0 software to present the stimuli and to record response times and accuracy.

### *Design and analysis*

If occupational descriptions automatically activate the grammatical gender of the corresponding role noun, then a response facilitation should be detected for the role nouns with corresponding grammatical gender, compared to role nouns in the opposite grammatical gender. This effect should influence both semantically related (typically male or female) and semantically unrelated (typically neutral) role nouns.

Analyses were computed on the basis of participant means across items ($F_1$) and on item means across participants ($F_2$; Clark, 1973). The $F_1$ analysis of

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

Table 1. *Experiment 1 factorial structure and results*

| Prime Description | Semantically Related Target | Mean (*SD*) | By Subjects | By Items |
|---|---|---|---|---|
| Typically male (e.g., "X repairs furniture, . . .") | 1. Tischler/carpenter$_{\text{masculine}}$ | 11.71 (167.38) | $t_{31} = -1.12$, | $t_{11} = -1.13$, |
| | 2. Tischlerin/carpenter$_{\text{feminine}}$ | 27.98 (148.97) | $p > .1$ | $p > .1$ |
| Typically female (e.g., "X sells flowers, . . .") | 1. Floristin/florist$_{\text{feminine}}$ | −23.16 (144.13) | $t_{31} = -3.95$, | $t_{11} = 3.57$, |
| | 2. Florist/florist$_{\text{masculine}}$ | 64.34 (171.63) | $p < .001$ | $p < .05$ |
| | Semantically Unrelated Target | | | |
| Typically male (e.g., "X repairs furniture, . . .") | 3. Sänger/singer$_{\text{masculine}}$ | 0.29 (151.06) | $t_{31} = -1.61$, | $t_{11} = 1.06$, |
| | 4. Sängerin/singer$_{\text{feminine}}$ | −20.27 (135.37) | $p > .1$ | $p > .1$ |
| Typically female (e.g., "X sells flowers, . . .") | 3. Sängerin/singer$_{\text{feminine}}$ | −10.51 (128.27) | $t_{31} = -0.76$, | $t_{11} = -0.49$, |
| | 4. Sänger/singer$_{\text{masculine}}$ | −21.15 (125.28) | $p > .1$ | $p > .1$ |

variance (ANOVA) was performed with Description Typicality (male, female) × Role Noun Grammatical Gender (masculine, feminine) as within-subjects factors. The $F_2$ ANOVA was performed with Description Typicality (male, female) as a between-items factor and Role Noun Grammatical Gender (masculine, feminine) as a within-items factor. Separate analyses were run for semantically related and unrelated role nouns, in order to investigate "yes" and "no" responses separately. The results of contrast comparisons based on the $F_1$ analysis are reported below. Contrast comparisons based on the $F_2$ analysis produced the same pattern of statistical significance and are reported in Table 1. Only reaction times of correct responses were included in the data analysis (96.1% of the data). Response times beyond 3 standard deviations over the mean were excluded (1.9% of the data). Response times were corrected for word length (Trueswell, Tannenhaus & Garnsey, 1994).[3]

The first group of analyses investigated response times to semantically related role nouns (only "yes" responses). Because all semantically related role nouns were typically male or typically female, this first comparison tested possible effects of grammatical gender *in addition* to those of gender typicality. In contrast, the second analysis concerned semantically unrelated role nouns (only "no" responses), which were neutral with regard to gender typicality. This analysis tested possible effects of grammatical gender without the influence of role noun typicality.

### *Results*

The first ANOVA concerned response times to semantically related role nouns, which required a "yes" response. Results showed a main effect of grammatical gender, $F_1$ (1, 31) = 6.02, $MSE$ = 6,741.79, $p < .05$, $F_2$ (1, 22) = 3.92, $MSE$ = 4,455.71, $p = .06$, with responses to feminine role nouns being faster, $M_{\text{masculine}} = 38.03$, $M_{\text{feminine}} = 2.41$ (means are based on $F_1$ analysis) and an interaction between description typicality and grammatical gender, reliable in both by-subjects and by-item analyses, $F_1$ (1, 31) = 19.13, $MSE$ = 4,501.16, $p < .001$, $F_2$ (1, 22) = 11.90, $p < .05$.

Following typically female descriptions, response times were shorter for the congruent feminine role noun than for the masculine one (e.g., "B. A. teaches pupils from the first to the fourth class"), and response times were shorter for the feminine role noun ("primary school teacher$_{\text{feminine}}$") than for the masculine role noun ("primary school teacher$_{\text{masculine}}$"; $M_{\text{Ff}} = -23.16$, $M_{\text{Fm}} = 64.34$), $t$ (31) = –3.95, $p < .001$. Following typically male descriptions, response times for masculine and feminine role nouns did not differ (e.g., after "A. B. develops computer software"), and no difference was found in response times for the masculine and the feminine role noun ("IT-specialist$_{\text{masculine}}$" and "IT-specialist$_{\text{feminine}}$"; $M_{\text{Mm}} = 11.71$, $M_{\text{Mf}} = 27.98$), $t$ (31) = –1.12, *ns*.

The second ANOVA was run on response times to semantically unrelated role nouns, which required a "no" response. Results revealed a marginally significant interaction between description typicality and role noun grammatical gender in the by-subjects analysis, $F_1$ (1, 31) = 2.93, $MSE$ = 2,662.11, $p = .097$, $F_2$ (1, 22) = 1.31, *ns*. Contrasts were computed to test possible effects of grammatical gender while excluding the influence of gender typicality, because all unrelated

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

role nouns were typically neutral. No significant difference was found between masculine and feminine role nouns, both after male ($M_{Mm} = 0.29$, $M_{Mf} = –20.27$), $t\,(31) = –1.61$, *ns*, and female ($M_{Ff} = –10.51$, $M_{Fm} = –21.15$), $t\,(31) = –0.76$, *ns*, descriptions.

Participants' sex did not affect the results, neither as a main effect nor in interaction with other factors in either ANOVA.

### *Discussion*

The data showed no priming effect on targets with neutral typicality, either with matching or mismatching grammatical gender. This result suggests that the role descriptions did not automatically activate the corresponding grammatical gender. With regard to gender-typical target nouns, only typically female descriptions affected response times to role nouns with matching (feminine) or mismatching (masculine) grammatical gender, with longer response times in the mismatching condition. Therefore, in this case, the hypothesis that descriptions elicit grammatical priming cannot be rejected, but only as a possible additional factor besides the gender typicality effect.

Results on gender-typical role nouns revealed an asymmetry between male and female items, with only female descriptions triggering the mismatch effect. We considered two possible interpretations of this asymmetry, a linguistic one and a sociocognitive one. The linguistic explanation is based on the asymmetry of grammatical gender use in German: the feminine form is applicable only to female referents, whereas the masculine form can be used to refer to both sexes (generic masculine). If the descriptions elicited the corresponding role nouns with morphological gender markers, this effect could have been more relevant for female descriptions, activating the feminine form, which cannot be applied to male referents. However, the mismatch effect does not occur with typically neutral targets. This suggests excluding a purely linguistic explanation. A second interpretation would be that it was easier for participants to accept both genders as fitting a typically male profession, whereas it was more complex to accept a masculine role noun as matching the description of a typically female occupation. This interpretation finds support in recent social psychology findings and will be taken up in the general discussion.

The experimental descriptions of Experiment 1 were employed in an eye-tracking experiment to test the effects of gender typicality cues on pronominal anaphor resolution.

## EXPERIMENT 2

In the second experiment, participants' eye movements were recorded during reading. Experimental sentences presented the description of a profession followed by a target sentence containing an anaphoric personal pronoun. The job descriptions did not contain any grammatical cue to the referent gender, which was revealed later on through the anaphor. The descriptions were either gender biased (male or female) or neutral, whereas the target sentence was always neutral with regard to gender typicality. Eye movements were recorded in order to measure the effect

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

of gender typicality of the role description on the resolution of the following anaphor, which either matched or mismatched the gender typicality of the job. After the eye-tracking session, participants performed an Implicit Association Test Gender–Career and completed three questionnaires on sexism and sex role attribution.

### *Method*

*Participants.* Thirty-two volunteers participated in the study (16 men, mean age = 25.1 years, *SD* = 4.4). The data of 1 participant were excluded from the analyses because of technical problems. Participants were students at the University of Heidelberg. They were all native speakers of German and had normal or corrected to normal vision. They received either course credit or money for their participation. None of them had participated in Experiment 1.

*Materials.*

EYE-TRACKING MATERIALS. Experimental materials consisted of the 36 descriptions of typically male, typically female, and neutral occupational activities that had been employed in the previous experiment, each followed by a target sentence containing a masculine or feminine anaphoric pronoun (see Example (1) and Appendix A for further information).

(1) Description:
*M. F. repariert und stellt Möbel her, arbeitet mit Holz.*
"M. F. repairs and produces pieces of furniture, works with wood."
Target sentence:
*Gewöhnlich hat er/sie ein ausreichendes Einkommen.*
"Usually he/she has a sufficient income."

The development of the job descriptions is described in detail in the previous Material section. The target sentences were constructed with a fixed linguistic structure (adverb/verb/pronoun/article/adjective/noun). The target sentences were pretested for gender neutrality by a sample of 30 participants, who read the sentences with an X in place of the pronoun. The gender typicality of the target context was rated on a 7-point Likert scale (1 = *typically male*, 7 = *typically female*). Thirty-six target sentences that lay in the neutral range between 3.5 and 4.5 points were selected and combined with the descriptions to constitute the experimental materials.

To prevent specific resolution strategies in reading the experimental target sentences, we used filler items that had a similar structure but contained a pronominal anaphor referring back to an inanimate object in the description. The filler descriptions dealt with neutral nonprofessional roles (e.g., neighbor, moviegoer). In addition, we also created fillers with a different linguistic structure, to increase variation in the linguistic features of the materials. These fillers described gender-neutral activities; the anaphoric pronoun they contained was either masculine or feminine, assigned at random and in equal proportions. Finally, we created fillers

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

that described occupations that had not shown pronounced gender typicality in the earlier ratings. As anaphor, we used the pronoun with higher cloze probability according to the typicality ratings, in order to avoid incongruity effects in the filler material (i.e., "he" for items between 2.6 and 3.4, those considered slightly male; and "she" for slightly female items with ratings between 4.6 and 5.4). Content-related questions were presented after one fourth of the sentences to ensure reading for comprehension.

IMPLICIT ASSOCIATION TEST. After the eye-tracking session, participants performed an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT is a reaction time test that measures the strength of association between two concepts. For our study, we employed the IAT "Gender–Career" (see Nosek, Banaji, & Greenwald, 2002), which measures the strength of association between the concepts of *men* and *career* and the concepts *women* and *career* as well as *women* and *family*, and *men* and *family*. Participants categorized a series of items presented on the screen as belonging to one of these four categories (*men*, *women*, *family*, or *career*). Reaction times reflected which pairs of categories were more strongly associated in each participants representation.

QUESTIONNAIRES. In the final part of the experimental session, participants completed three questionnaires: the Bem Sex Role Inventory (Bem, 1974; German version, Schneider-Düker & Kohler, 1988), the Ambivalent Sexism Scale (Glick & Fiske, 1996; German version, Eckes & Six-Materna, 1999), and the Modern Sexism Scale (Swim, Aikin, Hall, & Hunter, 1995; German version, Eckes & Six-Materna, 1998). The individual measures were collected to investigate possible covariations with the effects of gender typicality analyzed in the eye-movement measures.

The Bem Sex Role Inventory is a list of 60 typically male, typically female, and neutral personality traits. Participants marked on a 7-point scale to which extent each trait applied to themselves. Three scores were calculated on the basis of their ratings: masculinity, femininity, and androgyny scores. Masculinity and femininity scores consist of the mean self-rating on the male and female items. The androgyny score is based on the difference between masculinity and femininity scores. Masculinity and femininity scores indicate the extent to which a person regards masculine and feminine characteristics as self-descriptive. In contrast to previous instruments, the Bem Inventory considers the two scores as conceptually independent of each other, so that an individual can obtain high scores in both typically male and female traits. The androgyny score reflects the relative degrees of masculinity and femininity that individuals attribute to themselves; the closer the score is to zero, the more the participant includes both male and female traits in his or her self-description. Sex-typed individuals may be more likely to process information in terms of a gender schema (Bem, 1981), a cognitive structure that imposes expectations and meaning on the incoming information. For this reason, we expected more gender-typed participants to apply a gender-typed scheme to the experimental descriptions and to have stronger expectations in the direction of stereotype-congruent referent gender.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

The Ambivalent Sexism Inventory is composed of 22 statements for which participants mark their degree of agreement on a 6-point scale. The Inventory comprises two positively correlated components of sexism that represent opposite evaluative orientations toward women: hostile sexism, which reflects overt aversion toward women, and benevolent sexism, which reflects gender-stereotypical attitudes that are nevertheless experienced as positive by the subject and tend to elicit typically prosocial behavior (e.g., paternalistic help). Both subscales are intercorrelated and can predict the endorsement of gender stereotypes (Jost & Kay, 2005) as well as the assignment of complementary roles to men and women.

While the Ambivalent Sexism Inventory investigates interpersonal attitudes, the Modern Sexism Scale focuses on a sociopolitical level. It is composed of 10 statements, for which participants express their degree of agreement on a 6-point scale. The scale aims at capturing *modern* sexism attitudes, which, in contrast to *traditional* ones, are more indirect. Items refer to three major areas: denial of discrimination against women, antagonism toward women's demands, and resentment of special concessions for women. The modern sexism score is calculated by specifying the mean rating of all items. It has been shown that individuals with higher scores in modern sexist beliefs are more likely to overestimate the percentage of women in typically male jobs than are individuals with lower scores (Swim et al., 1995). The questionnaire was introduced to check for potential correlations between modern sexism scores and gender expectations in reference resolution.

### *Procedure*

The experiment started with the reading task, during which eye movements were recorded. Eye movements were monitored with a video-based head-mounted eye-tracker (Eyelink II, sampling rate of 250 Hz). Participants were seated 70 cm away from a computer screen, their chin resting on a chinrest during the whole experiment. Materials were presented with the software Eyetrack.[4] Reading was binocular, and participants' dominant eye was tracked. The experiment began after a calibration procedure. The presentation of sentences started with a small rectangle indicating the position of the first word of the sentence. The item appeared only when this rectangle was fixated accurately. Sentences were displayed in a monospaced 22 point Lucida Console font. After reading a sentence, participants pressed a button on a keypad to prompt the next item or a question. Two buttons of the keypad were used for answering the questions.

To familiarize participants with the task, the experiment started with four practice trials, one of which was followed by a comprehension question. Then experimental sentences and filler items were presented in random order. Items were displayed in three lines.

After the eye-tracking recording, participants performed the IAT Gender–Career. Finally, they filled out the three questionnaires on individual sexism measures and gender roles. In all, one session lasted about 45 min.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

### *Design and hypotheses*

The experimental factors were gender typicality of the role description and gender of the anaphoric pronoun, resulting in a 3 (Typicality: male, female, neutral) × 2 (Pronoun: masculine or feminine) factorial design. In the analysis by subjects, the gender typicality of the description and the grammatical gender of the pronoun served as within-subjects factors. In the analysis by items, description typicality served as a between-items factor and pronoun gender as a within-items factor.

The description of a professional activity in the priming sentence was assumed to activate the cognitive representation of the corresponding referent gender. When this representation did not match the referent gender expressed by the pronoun, a longer processing time should be required to integrate the conflicting information, that is, to resolve the pronoun. We therefore predicted that incongruence between the typical gender of the description and the grammatical gender of the pronoun would result in longer fixation times on the target sentence compared to the congruent condition. In the case of prime sentences describing a neutral context, no difference was expected between the target sentence with a masculine and the one with a feminine pronoun.

### *Results*

#### *Eye-tracking data.*

DATA ANALYSIS. In order to determine the effects of gender typicality on pronoun resolution we analyzed fixation times and regression patterns on the target sentence, which was presented in the third line of each item. Table 2 provides an example of an experimental item, consisting of a description of the occupation and a subsequent target sentence with the anaphoric reference. The example shows the segmentation of the target sentence into five regions. The region of interest, where the effect was expected, was the anaphor region including the pronoun ("he" or "she") plus the following indefinite article. The article was included in the region because the monosyllabic pronoun alone would constitute a very small area that could frequently have been skipped. The other analyzed regions were the verb region preceding the pronoun, as a possible launching region for saccades skipping the pronoun, and the adjective of the noun phrase following the pronoun region, as a possible spillover region.

Following Rayner, Sereno, Morris, Schmauder, and Clifton (1989) and the current practice in eye-tracking research (cf. Breen & Clifton, 2011), we removed fixations below 70 ms and above 600 ms before analyzing the data (3.2% of the data). Analyses were computed on the basis of participant means across items ($F_1$) and on item means across participants ($F_2$; Clark, 1973). Because the regions of interest differed in length across items, analyses were based on residual fixation times that had been corrected for length.[5] In order to reflect the process of understanding from early to late stages, results are reported for the following eye-tracking measures: first fixation time, first pass time, regression path time, total time, and probabilities of regressions into a region. First fixation time represents the duration of the first fixation in a given region. First pass time reflects the time

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

Table 2. *Example sentences and factorial structure of Experiment 2*

| | |
|---|---|
| Prime | |
| Male role description | *M. F. repariert und stellt Möbel her, arbeitet mit Holz.*<br>M. F. repairs and produces pieces of furniture, works with wood. |
| Female role description | *K. P. verkauft Blumen, bindet Sträuße in einem Geschäft.*<br>K. P. sells flowers, makes up bouquets in a shop. |
| Neutral role description | *F. H. spielt beruflich ein Instrument in einem Orchester.*<br>F. H. plays an instrument professionally in an orchestra. |
| Target | |
| Anaphoric reference | *Gewöhnlich - hat - er / sie ein - ausreichendes - Einkommen.*<br>[Usually - has - he / she a - sufficient - income.] |

*Note:* The regions of analysis in the target sentence are delimited by a dash. The German word order is preserved in the target sentence translation and enclosed in brackets.

from first entering a region of interest from the left until leaving it either to the right (i.e., moving forward in the sentence) or to the left. Regression path time is the time from first entering a region until leaving it to the right, including the time for regressions from this region. Total time is the total amount of time spent in a certain region, including rereading but not including regressions from this region (cf. Boland, 2004; Sturt, 2003). In general, longer fixation times and a higher probability of regressions indicate comparatively greater difficulty in processing the respective region.

Means of fixation times and probabilities of regressions on the pronoun and spillover region are summarized in Table 3; details of the statistical tests are given in Table 4 and Table 5. An interaction between type of description and pronoun gender occurred consistently in both $F_1$ and $F_2$ analyses in an early (first fixation time) and a late (total time) measure, and was localized on the region of interest (pronoun region), which is described in detail below; no effect occurred consistently in both analyses outside the pronoun region, and no main effect occurred consistently in both analyses, in any region. Pairwise contrast analyses on the pronoun region were conducted across typicality and across pronoun. Unless otherwise specified, $F_2$ contrast analyses replicated the result pattern obtained in $F_1$ analyses.

FIRST FIXATION TIME. On the pronoun region, first fixations revealed an interaction between typicality and pronoun, reliable in $F_1$ and $F_2$ analyses. Contrast analyses showed that after a typically female description, mean fixation times were longer for masculine than for feminine pronouns, at a marginal level in $F_1$ ($M_{\mathrm{Fm}} = 10.88$, $M_{\mathrm{Ff}} = -0.95$), $t\,(30) = 1.91$, $SEM = 6.18$, $p = .06$, and reliably in $F_2$ (see Table 5 for details of the by-items contrasts). No effect was found after a male description ($M_{\mathrm{Mm}} = -1.28$, $M_{\mathrm{Mf}} = -2.23$), $t\,(30) = 0.18$, *ns*. After neutral

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

Table 3. *Means (standard deviations) of residual fixation times and probabilities of regressions, differentiated for region and experimental factor*

| | Experimental Factors | | Eye-Tracking Measures | | | | |
|---|---|---|---|---|---|---|---|
| Region | Description Typicality | Pronoun | First Fix. Time | First Pass Time | Regression Path | Total Fix. Time | Regressions Into Region |
| Pronoun | Male | Masculine | −1.28 (32.48) | −10.43 (83.50) | −15.90 (101.67) | −36.81 (124.99) | 18.28 (21.24) |
| | | Feminine | −2.24 (29.10) | −7.74 (87.57) | 2.95 (111.97) | 13.60 (123.07) | 24.19 (25.03) |
| | Female | Masculine | 10.89 (33.47) | 23.50 (83.67) | 26.72 (127.50) | 23.99 (105.46) | 25.27 (22.31) |
| | | Feminine | −0.96 (29.95) | −6.26 (75.11) | 2.19 (113.63) | −7.14 (112.43) | 17.20 (17.99) |
| | Neutral | Masculine | −5.36 (30.67) | −3.12 (83.07) | 19.62 (135.75) | 11.09 (121.75) | 19.89 (22.12) |
| | | Feminine | 4.40 (33.79) | −6.97 (69.34) | −11.79 (92.63) | −4.60 (103.54) | 24.19 (24.28) |
| Spillover | Male | Masculine | 2.05 (47.36) | 0.49 (78.67) | 7.72 (318.77) | −23.24 (87.63) | 13.98 (12.98) |
| | | Feminine | 2.26 (34.91) | 9.24 (92.79) | −7.18 (195.11) | 19.29 (140.05) | 9.68 (13.45) |
| | Female | Masculine | 2.52 (38.20) | −10.44 (69.61) | −32.42 (227.11) | −17.45 (77.41) | 13.44 (13.89) |
| | | Feminine | −4.17 (38.75) | −11.34 (78.37) | −33.75 (226.57) | −6.98 (103.62) | 11.29 (13.87) |
| | Neutral | Masculine | 0.49 (46.05) | 4.77 (78.45) | 7.37 (195.16) | 8.80 (79.39) | 12.37 (12.15) |
| | | Feminine | 2.96 (42.04) | 8.05 (73.67) | 70.19 (257.53) | 20.89 (96.12) | 12.90 (17.06) |

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

Table 4. *Results of Experiment 2 statistical analyses of variance, differentiated for eye-tracking measures and regions of analysis*

| Measure | Region of Analysis | Effect | *F*1 | *F*2 |
|---|---|---|---|---|
| First fix. time | Pronoun | Typicality | $F = 2.026$, $p = .141$ | $F = 1.943$, $p = .159$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F_{2,60} = 3.879$, $MSE = 466.06$, $p = .026$ | $F_{2,33} = 3.526$, $MSE = 232.34$, $p = .041$ |
| | Spillover | Typicality | $F < 1$ | $F < 1$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F < 1$ | $F < 1$ |
| First pass time | Pronoun | Typicality | $F = 2.193$, $p = .120$ | $F_{2,33} = 2.746$, $MSE = 1219.25$, $p = .079$ |
| | | Pronoun | $F = 2.607$, $p = .117$ | $F = 1.378$, $p = .249$ |
| | | Typicality × Pronoun | $F_{2,60} = 2.68$, $MSE = 1709.95$, $p = .078$ | $F = 1.256$, $p = .298$ |
| | Spillover | Typicality | $F_{2,60} = 2.566$, $MSE = 2213.24$, $p = .085$ | $F < 1$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F < 1$ | $F < 1$ |
| Regression path | Pronoun | Typicality | $F = 1.162$, $p = .320$ | $F < 1$ |
| | | Pronoun | $F < 1$ | $F = 1.062$, $p = .310$ |
| | | Typicality × Pronoun | $F_{2,60} = 3.126$, $MSE = 3681.28$, $p = .051$ | $F = 1.658$, $p = .206$ |
| | Spillover | Typicality | $F_{2,60} = 3.532$, $MSE = 22700.63$, $p = .035$ | $F < 1$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F < 1$ | $F < 1$ |
| Total time | Pronoun | Typicality | $F = 1.293$, $p = .282$ | $F = 1.014$, $p = .374$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F_{2,60} = 6.925$, $MSE = 4199.17$, $p = .002$ | $F_{2,33} = 5.880$, $MSE = 1904.26$, $p = .007$ |
| | Spillover | Typicality | $F = 1.782$, $p = .177$ | $F < 1$ |
| | | Pronoun | $F = 2.780$, $p = .106$ | $F_{2,33} = 4.518$, $MSE = 2011.32$, $p = .041$ |
| | | Typicality × Pronoun | $F < 1$ | $F = 1.113$, $p = .341$ |
| Regressions into region | Pronoun | Typicality | $F < 1$ | $F < 1$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F_{2,60} = 3.006$, $MSE = 301.57$, $p = .057$ | $F_{2,33} = 3.017$, $MSE = 119.37$, $p = .063$ |
| | Spillover | Typicality | $F < 1$ | $F < 1$ |
| | | Pronoun | $F < 1$ | $F < 1$ |
| | | Typicality × Pronoun | $F < 1$ | $F < 1$ |

Table 5. *Results of Experiment 2 statistical analyses (t test), differentiated for eye-tracking measures, on the pronoun region*

| Measure | Comparisons | By Subjects | By Items |
|---|---|---|---|
| First fix. time | Male role, he/she | $t_{30} = 0.178, p = .860$ | $t_{11} = 0.167, p = .871$ |
| | Female role, he/she | $t_{30} = 1.915, p = .065$ | $t_{11} = 3.141, p = .009$ |
| | Neutral role, he/ she | $t_{30} = -1.905, p = .066$ | $t_{11} = -1.646, p = .128$ |
| | Male/female, he | $t_{30} = -2.437, p = .021$ | $t_{11} = -2.219, p = .048$ |
| | Male/female, she | $t_{30} = -0.231, p = .819$ | $t_{11} = -0.280, p = .785$ |
| First pass time | Male role, he/she | $t_{30} = -0.266\ p = .792$ | $t_{11} = -0.244, p = .812$ |
| | Female role, he/she | $t_{30} = 2.720, p = .011$ | $t_{11} = 1.924, p = .081$ |
| | Neutral role, he/ she | $t_{30} = 0.350, p = .729$ | $t_{11} = 0.598, p = .562$ |
| | Male/female, he | $t_{30} = -3.285, p = .003$ | $t_{11} = -2.383, p = .036$ |
| | Male/female, she | $t_{30} = -0.134, p = .894$ | $t_{11} = -0.072, p = .944$ |
| Regression path | Male role, he/she | $t_{30} = -1.243, p = .224$ | $t_{11} = -1.379, p = .195$ |
| | Female role, he/she | $t_{30} = 1.370, p = .181$ | $t_{11} = 1.341, p = .207$ |
| | Neutral role, he/ she | $t_{30} = 1.593, p = .122$ | $t_{11} = 1.110, p = .291$ |
| | Male/female, he | $t_{30} = -2.730, p = .011$ | $t_{11} = -1.922, p = .081$ |
| | Male/female, she | $t_{30} = 0.040, p = .968$ | $t_{11} = 0.024, p = .981$ |
| Total time | Male role, he/she | $t_{30} = -3.099, p = .004$ | $t_{11} = -3.705, p = .003$ |
| | Female role, he/she | $t_{30} = 1.993, p = .055$ | $t_{11} = 1.564, p = .146$ |
| | Neutral role, he/ she | $t_{30} = 0.844, p = .405$ | $t_{11} = 0.976, p = .350$ |
| | Male/female, he | $t_{30} = -4.091, p < .001$ | $t_{11} = -3.318, p = .007$ |
| | Male/female, she | $t_{30} = 0.999, p = .326$ | $t_{11} = 0.869, p = .404$ |
| Regressions into region | Male role, he/she | $t_{30} = -1.134, p = .266$ | $t_{11} = -1.803, p = .099$ |
| | Female role, he/she | $t_{30} = 2.540, p = .016$ | $t_{11} = 1.378, p = .195$ |
| | Neutral role, he/ she | $t_{30} = -1.052, p = .301$ | $t_{11} = -1.199, p = .256$ |
| | Male/female, he | $t_{30} = -1.748, p = .091$ | $t_{11} = -1.556, p = .148$ |
| | Male/female, she | $t_{30} = 1.686, p = .102$ | $t_{11} = 1.836, p = .093$ |

descriptions, masculine pronouns tended to be fixated shorter than feminine ones ($M_{Nm} = -5.36, M_{Nf} = 4.40$), $t\,(30) = -1.90$, $SEM = 5.12$, $p = .07$. The tendency became not significant in the by-items analysis. This first grouping compared the effects of the different gender typicalities on resolving the pronoun. To analyze the impact of the pronoun gender, a second grouping of contrasts was based on the anaphor gender. This contrast revealed that the mismatch effect occurred only with the masculine pronoun, which was fixated shorter after congruent than incongruent typicality ($M_{Mm} = -1.28$, $M_{Fm} = 10.88$), $t\,(30) = -2.44$, $SEM = 4.99$, $p = .02$, whereas no effect was found when comparing the feminine pronoun after male and female typicality ($M_{Mf} = -2.23$, $M_{Ff} = -0.95$), $t\,(30) = 0.23$, *ns*.

FIRST PASS TIME. First pass time on the pronoun region showed a marginally significant interaction between typicality and pronoun. Contrast analyses across typicality showed that after a typically female description, mean fixation times were longer for masculine than for feminine pronouns, ($M_{Fm} = 23.50$, $M_{Ff} = -6.25$), $t\,(30) = 2.72$, $SEM = 10.09$, $p = .01$. No effect was found after a male description ($M_{Mm} = -10.43$, $M_{Mf} = -7.74$), $t\,(30) = -0.26$, *ns*, and after neutral

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

descriptions ($M_{Nm} = –3.11$, $M_{Nf} = –6.97$), $t$ (30) = 0.35, *ns*. Contrast analyses across pronouns revealed that the mismatch effect was statistically significant when the anaphor was a masculine pronoun, which was fixated shorter after congruent than incongruent typicality ($M_{Mm} = –10.43$, $M_{Fm} = 23.50$), $t$ (30) = –3.28, $SEM$ = 10.33, $p$ = .003, whereas no effect was found with the feminine pronoun after male and female typicality ($M_{Mf} = –7.74$, $M_{Ff} = –6.25$), $t$ (30) = –0.13, *ns*.

REGRESSION PATH TIME. A significant interaction between typicality and pronoun emerged in $F_1$ analysis on the pronoun region. Contrast analyses across typicality showed no significant effect. Contrast analyses across pronouns showed that the mismatch effect occurred only with the masculine pronoun, which resulted in shorter fixations after congruent than incongruent typicality, reliably in the by-subjects analysis ($M_{Mm} = –15.90$, $M_{Fm} = 26.72$), $t$ (30) = –2.73, $SEM$ = 15.61, $p$ = .01, and at a marginal level in the by-items analysis. No effect was found when comparing the feminine pronoun after male and female typicality.

TOTAL TIME. The expected interaction between typicality and pronoun occurred on the pronoun region. Contrast analyses showed that after a typically female description, mean fixation times were longer for masculine than for feminine pronouns in the by-subjects analysis ($M_{Fm} = 23.99$, $M_{Ff} = −7.14$), $t$ (30) = 1.99, $SEM$ = 15.62, $p$ = .05. This difference was not significant in the by-items analysis. After a typically male description, the incongruent anaphor was fixated longer ($M_{Mm} = –36.81$, $M_{Mf} = 13.60$), $t$ (30) = − 3.09, $SEM$ = 16.26, $p$ = .004. No effect occurred after neutral descriptions ($M_{Nm} = 11.09$ vs. $M_{Nf} = –4.60$), $t$ (30) = 0.84, *ns.* In contrast analyses across pronouns, the mismatch effect occurred again only with the masculine pronoun, which was fixated shorter after congruent than incongruent typicality ($M_{Mm} = –36.80$, $M_{Fm} = 23.99$), $t$ (30) = –2.44, $SEM$ = 14.86, $p < .001$, whereas no effect was found when comparing the feminine pronoun after male and female typicality ($M_{Mf} = 13.60$, $M_{Ff} = –7.12$), $t$ (30) = 0.99, *ns*.

REGRESSIONS INTO A REGION. The expected interaction between typicality and pronoun was found as a tendency on the pronoun region in $F_1$ and $F_2$ analyses. Contrast analyses across typicality showed that after a typically female description, mean regression probabilities were higher for masculine than for feminine pronouns ($M_{Fm} = 25.67$, $M_{Ff} = 17.20$), $t$ (30) = 2.54, $SEM$ = 3.17, $p$ = .02. This difference was not significant in the by-items analysis. No effect was found after a male description ($M_{Mm} = 18.28$ vs. $M_{Mf} = 24.19$), $t$ (30) = –1.13, *ns*, and after neutral descriptions ($M_{Nm} = 19.89$ vs. $M_{Nf} = 24.19$), $t$ (30) = –1.05, *ns*. Contrast analyses across pronouns showed no significant result for this measure.

Participants' sex did not affect eye movements as a main effect and did not cause any systematic interaction effects with other ANOVA factors.[6]

*Relating eye movements to individual measures.*

EYE MOVEMENTS AND GENDER TYPICALITY RATINGS. In order to investigate whether eye movements reflect not only congruity or incongruity with gender

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

expectations but also, in a finer-grained manner, the degree of violation of an expected typicality, we ran a by-item linear regression analysis with typicality ratings as a predictor of eye movements. The typicality ratings of the descriptions had been collected in the pretesting phase. The ratings were given on a Likert scale with 1 as the *typically male* and 7 as the *typically female* pole. The ratings were correlated to fixation durations and proportion of regressions for each item on the pronoun region. Correlational analyses were conducted separately for eye movement data on items in the masculine and feminine anaphor condition. The linear regression revealed that the typicality ratings predicted eye movements on items presenting the masculine pronoun, in first fixations ($\beta = 0.34$, $p = .044$), first pass ($\beta = 0.34$, $p = .041$), and total time ($\beta = 0.47$, $p = .007$).[7] This means that lower ratings (closer to the typically male pole) produced shorter fixations on the target region containing the pronoun "he," and higher ratings (closer to the typically female pole) led to longer fixations on the corresponding items presenting the pronoun "he." The correlation was not symmetrical for the same items in the feminine pronoun condition. No significant correlation emerged between ratings and eye-movement data on items containing the pronoun "she" (maximum coefficient $\beta = -0.29$, $p = .082$, in regressions into the pronoun region; the negative coefficient indicates that lower ratings, corresponding to male items, where fixated longer, and higher ratings, corresponding to female items, were fixated shorter, when presenting the feminine pronoun). The results indicate that eye movements on the pronoun region following a gender-typical description reflected the degree of gender typicality revealed in explicit ratings of the corresponding role nouns, but only when the typical descriptions were related to a masculine referent.

EYE MOVEMENTS AND IAT. The IAT index was calculated for each participant according to the scoring algorithm proposed by Greenwald, Nosek, and Banaji (2003). This index reflects the difference, in terms of reaction times and accuracy, between the congruent and incongruent blocks of an IAT. In the congruent block, experimental categories are associated according to the traditional stereotypical representation (*Men* combined with *Career* and *Women* with *Family*), whereas the opposite coupling is presented in the incongruent block (*Men* + *Family* and *Women* + *Career*). A positive IAT index represents a stronger implicit association between the concepts in the stereotypical association. A negative IAT index represents a stronger implicit association between the concepts in the counterstereotypical association.

The IAT index showed that 29 participants out of 31 had a positive index, which indicates a stronger implicit association between the concepts of *Men* and *Career*, and between *Women* and *Family*. Two participants had a negative score, indicating the counterstereotypical tendency (stronger association between *Men* and *Family*, and *Women* and *Career*). For our sample, the mean IAT index (0.59, $SD = 0.39$) was higher than the mean index reported by Nosek et al. (0.39, $SD = 0.36$), which was averaged on a sample of 83.084 Gender–Career IATs collected on a publicly available website between 2002 and 2006 (Nosek et al., 2007). We analyzed possible covariation between the IAT index and eye-movement measures. As outlined above, the IAT index results from the subtraction of reaction times for the congruent block from reaction times for the incongruent block. For our study,

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

we calculated an eye-movement score following the same logic. Specifically, we subtracted fixation times or proportions of regressions on the pronoun in the congruent condition (i.e., description of typically male role/masculine pronoun; description of typically female role/feminine pronoun) from fixation times or proportion of regressions in the incongruent condition. As before, the pronoun region was selected as the most representative region of eye-movement effects. The analyses showed that the IAT index did not correlate with eye-movement measures (maximum correlation coefficient: $r = .22$, $p > .1$).

EYE MOVEMENTS AND QUESTIONNAIRES. The average questionnaire scores in our sample were close (within 1 *SD*) to the norms reported for the Ambivalent Sexism Inventory and the Bem Sex Role Inventory, German versions, respectively. The Modern Sexism Scale scores were higher in our sample (within 2 *SD*) than the norms of 1998. We investigated possible covariations between explicit individual measures and eye movements. The eye-movement effect was calculated with the same procedure as described for the IAT. The Bem Sex Role Inventory showed a weak positive correlation between the masculinity scale and the proportion of regression into the pronoun region ($r = .30$, $p = .09$). The two sexism questionnaires showed no reliable correlation with the eye-tracking measures (maximum correlation coefficient: $r = –.19$, $p > .1$).[8]

### *Discussion*

The eye-movement results showed a mismatch effect in the condition of incongruence between gender typicality of the description and the referential gender revealed by the anaphoric pronoun. In contrast to earlier studies on grammatical gender languages, the antecedent completely lacked morphological gender cues in the present experiment. Still, the descriptions of gender-stereotypical professional roles activated a representation of the referent gender, as indicated by the disruption in resolving an incongruent pronoun. The mismatch effect occurred on the pronoun region, including the pronoun itself plus a spillover word, in correspondence with previous findings in natural gender languages (Duffy & Keir, 2004; Sturt, 2003). Specifically, fixation times and proportions of regressions increased when the anaphor disagreed with the gender typicality of the occupation described in the previous sentence. This mismatch effect was observed reliably or as a tendency in very early, middle, and late stages of sentence processing, which suggests that the integration of gender-stereotypical cues and pronoun gender took place as soon as the incongruent pronoun was encountered and also affected later wrap-up processes.

Furthermore, the data revealed an asymmetry in the processing of the pronouns. The masculine pronoun triggered the mismatch effect, being fixated longer after a typically female than after a typically male description in early, intermediate, and late measures, whereas the mismatch effect for the female anaphor emerged only in the comparison across typicality in the final wrap-up stage. Thus, female referents were generally perceived as more compatible with both male and female contexts, whereas male referents suited male but not female occupational roles. An asymmetry in the same direction is also reported by Cacciari and Padovani

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

(2007) in the aforementioned priming study with bigender role nouns, where the mismatch effect was found only with the masculine pronoun after typically female role nouns ("teacher"–"he") but not with feminine pronouns after male roles ("engineer"–"she"). A possible explanation of these findings could lie in the fact that during the last decades women in industrialized societies have begun to enter typically male professions, whereas men do not seem to enter typically female professional areas to an equal degree (Cacciari & Padovani, 2007; Diekman & Eagly, 2000).

The individual attitude measures applied in the present study (sexism questionnaires and Gender Role Attribution Inventory) showed no reliable correlation with the eye-tracking data. Thus, the highly automatized processes of language comprehension may not recruit attitudes or stereotypical self-representations but rather seems to be based on typical distributions of men and women in different professional fields, as the high correlation between eye-tracking data and typicality ratings suggests.

Likewise, no correlation was found between eye movements and the IAT. This lack of correlation can also be due to the fact that the IAT and the eye-tracking items measured two theoretically different constructs: the IAT tested the strength of a specific job-related stereotypical association, namely, the association between gender and career, whereas the eye-tracking sentences focused on the cognitive link between referent gender and occupational activities, which were not necessarily associated with the concept of career, even in the case of male professions (e.g., plumber or janitor; see Appendix A).

## GENERAL DISCUSSION

Our investigation has shown the influence of stereotypical gender information on personal pronoun anaphor resolution during sentence reading. In contrast to natural gender languages such as English, the effect of gender typicality in grammatical gender languages is generally confounded with information coming from grammatical gender cues, which usually indicate the gender of the referent. The present study intended to overcome this constraint by replacing role nouns with equivalent descriptions of an agent performing a professional activity. These descriptions carried purely conceptual gender information (morphological gender cues were completely avoided) and served as primes for the target sentences that contained a pronominal anaphor. Eye-movement results revealed a mismatch effect of the stereotypical gender of the description, which emerged as soon as the anaphor region was entered and persisted in later stages of sentence processing. The structure of the paradigm does not allow us to determine if stereotypical expectations are activated during reading of the descriptions or when the anaphor is met. However, the fact that the effect is recorded in the earliest measure (first fixation time) and localized on the pronoun region with no spillover on the following region may suggest that the stereotypical gender information could have been activated before encountering the pronoun.

When comparing the effects for the pronouns *er*, "he," and *sie*, "she," the mismatch effect was observed consistently across measures only when the referent was a man, as indicated by the masculine pronoun. Results suggest that in initial stages

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

of processing, female referents suited both typically male and typically female occupational roles, whereas male referents were perceived as suiting typically male but not typically female occupations. This imbalance cannot be ascribed to different degrees of typicality in the materials, because role nouns were controlled for degrees of typicality. A source of ambiguity could lie in the German pronoun *sie*, which is used both for the third-person singular feminine and the third-person plural (without gender distinction). However, because a third-person singular verb form was presented before the anaphor, we would exclude the hypothesis of a plural (and thus generic) interpretation of the feminine pronoun. An asymmetrical pattern in the same direction was found as well in the reaction time experiment. After a typically female description, participants responded more slowly to a semantically related masculine than to a semantically related feminine role noun. No such difference occurred after typically male descriptions.

Taken together, the results may be interpreted as an indication that, in the absence of grammatical cues, gender roles are interpreted more flexibly for female than for male referents. A disruptive effect was found when male referents were to be integrated into a counterstereotypical occupational context, whereas less effort seemed to be required to match female referents with both gender contexts, especially in the initial stages of sentence processing. This perspective is compatible with social cognition findings that female roles have changed in the direction of incorporating formerly male attributes, whereas stereotypically male roles have changed to a lesser extent (Diekman & Eagly, 2000).

Another possible interpretation of the results would lie in postulating that the descriptions actually carry grammatical information because they would spontaneously activate the corresponding role noun with its grammatical gender in the reader. Female descriptions, even if grammatically gender free in their overt linguistic form, would thus activate in readers the corresponding role noun and its feminine suffix (*–in*), which constrains the possible referent gender. Male descriptions, in contrast, would activate masculine grammatical gender, which can be interpreted as generic in German (Duden, 1995). The first experiment, however, suggests that the descriptions do not activate a grammatical gender marking, as indicated by the lack of grammatical gender priming with typically neutral target stimuli. However, a priming effect was detected when stereotypical role nouns served as targets. Therefore, it seems to be possible that grammatical gender, even when not overtly present in the stimulus material, may still constitute an additional factor that can enhance the stereotypicality effect in grammatical gender languages. This is compatible with the fact that the asymmetry between male and female typicality has been reported, to our knowledge, only in studies on grammatical gender languages (German and Italian).

We found no reliable correlation between eye movements and measures of individual attitudes toward the sexes and sex role attribution. This finding is in line with the literature on correlation between explicit and implicit measures, which reports a generally weak correlation between self-reports and indirect measures especially for socially sensitive topics (Hoffman, Gawronsky, Gschwendner, Le, & Schmitt, 2005). The lack of correlation between the explicit individual measures and the eye-tracking data points to the importance of integrating the assessment of gender stereotypes with data from different methodologies, including indirect ones

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

such as eye-movement behavior. A nonstereotypical gender attitude may still fail to prevent stereotypes from affecting highly automatized cognitive processes. The IAT Gender–Career as well showed low correlation with the eye-tracking data. The strength of stereotypical associations between the concepts of *men* and *career*, and *women* and *family* did not covary with the mismatch effect observed in the eye-tracking data for an occupational description and a counterstereotypical referent. As an implicit measure of gender-stereotypical associations, the IAT was expected to correlate more consistently with the indirect measure of gender-stereotypical association offered by the eye-movement paradigm. However, the two measures focused on two different aspects of gender stereotypes in professions: while the IAT focused on career-related aspects, the eye-tracking experiment covered a wider range of professional activities. By contrast, a reliable covariation was found between the eye-tracking data and explicit gender typicality ratings, which therefore appeared to be a valid predictor of the stereotypicality effect in eye movements. The correlation between eye movements and explicit ratings was obtained with items that were either strongly stereotyped or clearly defined as gender unbiased. It would be interesting to explore whether this by-item correlation between implicit and explicit measures is also valid for roles that do not strictly belong to the male, female, or neutral category, but lie in between the usual rating cutoffs. This would be the case, for example, with professions whose current gender distributions contradict the traditional gender stereotype. For instance, *physician* has traditionally been a male role, but the increasing number of women entering medical universities may influence explicit typicality judgments, which are based on the perceived proportion of men and women in the field. In such cases of discrepancy, a highly automatized measure such as eye movements might tend to reflect more accurately the established gender stereotype, whereas typicality ratings might be more sensitive to recent changes in the distribution rates of men and women observed in a given professional area.

The present research suggests that gender-stereotypical information is activated in early stages of sentence processing and integrated with other gender cues available in the text to build the cognitive representation of the referent gender. This process can be interpreted in the framework of the scenario mapping and focus theory proposed by Sanford and Garrod (1998). According to the model, discourse comprehension relies on mapping specific text units into a world-knowledge scenario activated from long-term memory. In our study, the scenario was prompted by the gender-typical descriptions, which preactivated a representation of the referent, whereas the pronoun in the target sentence defined the referent gender. In case of a conflict between the implicit focus of the scenario and the explicit focus of the pronoun, as in the case of gender-incongruent anaphors, the initial cognitive representation of the referent requires correcting. This correction process becomes manifest as time cost, which was precisely reflected in our eye-tracking data through longer fixation times on the critical referent region.

To conclude, we presented a new paradigm that assessed the influence of gender-stereotypical cues on reference resolution in a grammatical gender language while avoiding the interference of morphological markers of grammatical gender. In a next step, these results should be systematically contrasted with data from comparable materials in a language without grammatical gender. Theoretically, the

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

results should be overlapping. If differences should emerge in this comparison, this might suggest an automatic activation of grammatical gender even in the absence of morphological cues when the discourse is processed in a grammatical gender environment. This would inform a cross-linguistic model of how diverse gender cues affect referent resolution in different grammatical systems. Implications of a possible automatic activation of grammatical gender, even in the absence of morphological gender cues, should be taken into account in the development of strategies for language use aiming at a balanced representation of gender.

## APPENDIX A

The following are examples of experimental items (corresponding role nouns are in parentheses). German word order is preserved in the English translation of the target sentences (brackets). The complete list of items and relative ratings is available on request.

### *Typically male roles*

1. (Mechaniker/in) J. P. repariert Autos und Motoren, überprüft Bremsen in einer Werkstatt. / Bald braucht er einen erholsamen Urlaub.
1. (Mechanic) J. P. repairs cars and engines, checks brakes in a workshop. [Soon needs he a relaxing vacation.]
2. (Elektriker/in) K. L. verlegt Stromleitungen und Kabel, überprüft die Spannung. / Auf dem Gebiet hat er große Erfahrung.
2. (Electrician) K. L. installs power lines and cables, checks electric voltage. [In this field has he a lot of experience.]
3. (Hausmeister/in) L. T. verwaltet ein Gebäude, erledigt kleine Reparaturen, hat alle Schlüssel. / Nächsten Monat macht er einen kurzen Urlaub.
3. (Janitor) L. T. takes care of a building, carries out small repairs, keeps all the keys. [Next month has he a short holiday.]
4. (Informatiker/in) P. K. entwickelt Computerprogramme, überwacht Computersysteme. / Bei der Arbeit trägt er eine dicke Brille.
4. (IT specialist) P. K. develops computer programs, monitors computer systems. [At work wears he thick glasses.]

### *Typically female roles*

1. (Florist/in) K. P. verkauft Blumen, bindet Sträuße in einem Geschäft. / Eigentlich hat er ein großes Angebot.
1. (Florist) K. P. sells flowers, makes up bouquets in a shop. [Actually has he a wide offer of products.]
2. (Sekretär/in) L. K. vereinbart Termine, erledigt die Korrespondenz in einem Büro. / Außerdem kann er eine fremde Sprache.
2. (Secretary) L. K. makes appointments, deals with the correspondence in an office. [In addition speaks he a foreign language.]
3. (Geburtshelfer/in) M. C. unterstützt bei der Entbindung, arbeitet im Krankenhaus. / Regelmäßig hat er einen langen Arbeitstag.
3. (Obstetrician) M. C. assists in childbirth, works at a hospital. [Regularly has he a long working day.]

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

4. (Kosmetiker/in) P. J. schminkt Gesichter, zupft Augenbrauen und entfernt Haare. / Oftmals gibt er eine nützliche Empfehlung.
4. (Beautician) P. J. does clients' make up, plucks eyebrows and removes hair. [Often gives he a useful suggestion.]

### *Typically neutral roles*

1. (Schauspieler/in) K. W. verkörpert verschiedene Rollen im Theater oder in Filmen. / Eigentlich hat er eine angenehme Stimme.
1. (Actor) K. W. plays different roles on the stage or in films. [Actually has he a pleasant voice.]
2. (Künstler/in) J. W. besitzt Kreativität, malt Bilder und baut Skulpturen. / Seit Jahren hat er ein eigenes Atelier.
2. (Artist) J. W. is creative, paints and makes sculptures. [Since many years has he a personal studio.]
3. (Musiker/in) F. H. spielt beruflich ein Instrument, spielt in einem Orchester. / Zweifellos hat er ein gutes Gehör.
3. (Musician) F. H. plays an instrument professionally in an orchestra. [Undoubtedly has he a discriminatory ear.]
4. (Apotheker/in) S. L. verkauft Medikamente, hat Pharmazie studiert. / Im Dienst trägt er einen weißen Kittel.
4. (Pharmacist) S. L. sells medicine, studied pharmacy. [On duty wears he a white lab coat.]

## ACKNOWLEDGMENTS

This research was supported by the European Community's Seventh Framework Programme (FP7/2007–2013) under Grant 237907. It was conducted at the Department of Cognitive and Theoretical Psychology at the University of Heidelberg. We thank Daniel Holt for his help in the production experiment and Friederike Braun for copy editing the manuscript.

## NOTES

1. One study with a single-word priming paradigm is reported due to its relevance to the discussion. For studies with auditory material see, for example, Lattner and Friederici (2003), Most, Verbeek Sorber, and Cunningham (2005), Pyykkönen, Hyönä, and van Gompel (2010).
2. We thank Pascal Gygax and Pirita Pyykkönen-Klauck for helpful comments and suggestions on how to address the issue.
3. Length correction was computed by subtracting predicted response times on the basis of a linear regression equation relating word length to response time, from the original response times measured for that word.
4. We are grateful to Chuck Clifton for making the software available on the web page http://www.psych.umass.edu/eyelab/ (eye-tracking lab of the University of Massachussets at Amherst).

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

5. Length correction was computed by subtracting the fixation times predicted for a particular region on the basis of a linear regression equation relating length to fixation time, from the original fixation times measured for that region.
6. One two-way interaction of participants' sex and target pronoun emerged in regressions into the pronoun region, where female participants regressed more often to the feminine pronoun than did male participants, $F(1, 29) = 4.94$, $p = .034$.
7. Excluding neutral items from the analyses, the standardized coefficients β are enhanced (first fixations: $\beta = 0.41$, $p = .047$; first pass: $\beta = 0.36$, $p = .081$; and total time: $\beta = 0.57$, $p = .004$).
8. Correlations between individual IAT results and questionnaire scores were also analyzed; no reliable correlation was found (maximum correlation coefficient: $r = -17$, $p > 1$).

## REFERENCES

Bates, H., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *58*, 992–1004.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*, 155–162.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*, 354–364.

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton Jr. (Eds.), *The online study of sentence comprehension* (pp. 51–76). New York: Psychology Press.

Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, *64*, 153–170.

Bußmann, H., & Hellinger, M. (2003). German—Engendering female visibility in German. In M. Hellinger & H. Bußmann (Eds.), *Gender across languages: The linguistic representation of women and men* (Vol. *3*, pp. 141–174). Amsterdam: Benjamins.

Cacciari, C., Carreiras, M., & Barbolini Cionini, C. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, *37*, 517–532.

Cacciari, C., & Cubelli, R. (2003). The neuropsychology of grammatical gender: An introduction. *Cortex*, *39*, 377–382.

Cacciari, C., & Padovani, R. (2007). Further evidence on gender stereotype priming in language: Semantic facilitation and inhibition on Italian role nouns. *Applied Psycholinguistics*, *28*, 277–293.

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. *Quarterly Journal of Experimental Psychology*, *49A*, 639–663.

Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Corbett, G. G. (1991). *Gender.* Cambridge: Cambridge University Press.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture-word interference paradigm. *Journal of Memory and Language*, *53*, 42–59.

Dieckman, A. B., & Eagly, A. H. (2000) Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*, 1171–1188.

Duden, Volume 4 (1995). *Die Grammatik* [Grammar]. Mannheim: Dudenverlag.

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

Eckes, T., & Six-Materna, I. (1998). Leugnung von Diskriminierung: Eine Skala zur Erfassung des modernen Sexismus [Denial of discrimination: A scale measuring modern sexism]. *Zeitschrift für Sozialpsychologie*, *29*, 224–238.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

Eckes, T., & Six-Materna, I. (1999). Hostilität und Benevolenz: Eine Skala zur Erfassung des ambivalenten Sexismus [Hostility and benevolence: A scale for assessing ambivalent sexism]. *Zeitschrift für Sozialpsychologie*, *30*, 211–228.

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, *40*, 206–212.

Glick, P., & Fiske, S. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

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*, 464–485.

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In S. L. Gaertner & J. F. Dovidio (Eds), *Prejudice, discrimination, and racism* (p. 133). New York: Academic Press.

Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 393–398.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). Meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.

International Labour Organization of the United Nations. (2000). *Employment for detailed occupational groups by sex, Germany*. Retrieved from http://laborsta.ilo.org

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*, 431–456.

Irmen, L., Holt, D. V., & Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research*, *1353*, 133–144.

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498–509.

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *32*, 355–378.

Kreiner, H., Sturt, P., & Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*, 239–261.

Lattner, S., & Friederici, A. (2003). Talker's voice and gender stereotype in human auditory sentence processing—Evidence from event-related brain potentials. *Neuroscience Letters*, *339*, 191–194.

Most, S. B., Verbeck Sorber, A., & Cunningham, J. G. (2005). Auditory Stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology*, *43*, 287–294.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*, 273–285.

Pyykkönen, P., Hyönä, J., & van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye tracking. *Experimental Psychology*, *57*, 126–133.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*, 121–149.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, *26*, 159–190.

Schneider-Düker, M., & Kohler, A. (1988). Assessment of sex roles: Results of a German version of the Bem Sex-Role Inventory. *Diagnostica*, *34*, 256–270.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social communication: Frontiers of Social Psychology* (pp. 163–187). New York: Psychology Press.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*, 542–562.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*, 199–214.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Van Gompel, R. P. G., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 128–139.

https://doi.org/10.1017/S0142716414000010 Published online by Cambridge University Press