LANGUAGE, COGNITION AND GENDER

EDITED BY : Alan Garnham, Jane Oakhill, Sabine Sczesny
and Lisa Von Stockhausen
PUBLISHED IN : Frontiers in Psychology

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

LANGUAGE, COGNITION AND GENDER

Topic Editors:
Alan Garnham, University of Sussex, UK
Jane Oakhill, University of Sussex, UK
Sabine Sczesny, University of Bern, Switzerland
Lisa Von Stockhausen, University of Duisburg-Essen, Germany

Gender inequality remains an issue of high relevance, and controversy, in society. Previous research shows that language contributes to gender inequality in various ways: Gender-related information is transmitted through formal and semantic features of language, such as the grammatical category of gender, through gender-related connotations of role names (e.g., manager, secretary), and through customs of denoting social groups with derogatory vs. neutral names. Both as a formal system and as a means of communication, language passively reflects culture-specific social conditions. In active use it can also be used to express and, potentially, perpetuate those conditions.

The questions addressed in the contributions to this Frontiers Special Topic include:

• how languages shape the cognitive representations of gender
• how features of languages correspond with gender equality in different societies
• how language contributes to social behaviour towards the sexes
• how gender equality can be promoted through strategies for gender-fair language use

These questions are explored both developmentally (across the life span from childhood to old age) and in adults. The contributions present work conducted across a wide range of languages, including some studies that make cross-linguistic comparisons.

Among the contributors are both cognitive and social psychologists and linguists, all with an excellent research standing. The studies employ a wide range of empirical methods: from surveys to electro-physiology. The papers in the Special Topic present a wide range of complimentary studies, which will make a substantial contribution to understanding in this important area.

Table of Contents

3. Decision making and Teachers' Attitudes

Editorial: Language, Cognition, and Gender

Alan Garnham[1]*, Jane Oakhill[1], Lisa Von Stockhausen[2] and Sabine Sczesny[3]

[1] Psychology, University of Sussex, Brighton, UK, [2] Psychology, University of Duisburg-Essen, Essen, Germany, [3] Psychology, University of Bern, Bern, Switzerland

The Editorial on the Research Topic

Language, Cognition, and Gender

Gender inequality remains a contentious issue in many societies, despite legislative, and other less formal attempts to tackle it. It is perpetuated, in part, by gender stereotyping. Previous research indicates that language contributes to gender inequality in various ways: Gender-related information is transmitted through formal and semantic features of language, such as the grammatical category of gender, through gender-related connotations of role names (e.g., *manager, secretary*), and through customs of denoting social groups with derogatory as opposed to neutral names. Both as a formal system and as a means of communication, language passively reflects culture-specific social conditions. Furthermore, language can also be used to express actively, and can potentially perpetuate, those conditions. Tackling these issues successfully depends on a proper understanding of their cognitive and societal underpinnings, but also on understanding the effects of attempted interventions. With these points in mind, the editors of this Special Topic, in collaboration with other colleagues, proposed a *Marie Curie Initial Training Network* entitled *Language, Cognition, and Gender* (ITN LCG), to address a range of questions about language and gender inequality. This project received funding from the European Commission's Seventh Framework Programme (FP7/2007-2013). ITN LCG included 10 European universities in the Czech Republic, Germany, Italy, Norway, Spain, Switzerland, and the United Kingdom, together with 12 associate partners in Germany, Italy, Switzerland, and the United Kingdom.

The research conducted within the ITN was organized into four work packages, addressing the questions of:

- how languages shape cognitive representations of gender
- how features of European languages correspond with gender equality in European societies
- how language contributes to social behavior toward the sexes
- how gender equality can be promoted through strategies for gender-fair language use.

These questions also appeared in the call for papers for this Special Topic, as it was intended that the Special Topic should showcase findings from ITN LCG together with related research.

Reflecting ITN LCG's focus on both cognitive and broader language-based and societal issues, the Special Topic has nine papers in Frontiers in Psychology, Cognition, and eight papers in Frontiers in Psychology, Language Sciences. However, it was originally thought that all papers would be referenced in both sections, so that the allocation of a paper to either Cognitive or Language Sciences is of no particular significance. Of the nine papers in the Cognition section, seven report work from ITN LCG and the other two (Garnham and Yakovlev; Garnham et al.) report related work by Garnham and colleagues, which arose out of discussions within ITN LCG, but which was carried out by students at the University of Sussex who were not funded from the

ITN LCG grant. Of the eight papers in the Language Sciences section, six report work from ITN LCG. Of the other two, one (Wolter et al.) was carried out in collaboration with members of the ITN LCG, whereas the other (Gustafsson Sendén et al.) was an independent study, closely related to the interests of ITN LCG.

In keeping with ITN LCG's multidisciplinary approach, the contributors to this Special Topic include both cognitive and social psychologists, and linguists. For the most part the contributions report original research, with a wide range of methods, from surveys to electro-physiological studies. In addition, the Special Topic includes one Review paper (Sczesny et al.) and one Hypothesis and Theory paper (Esaulova and von Stockhausen). Most of the contributions address questions about either the cognitive representation of gender or the use and effects of gender-fair language. They present a range of complementary studies, which make a substantial contribution to the understanding of these important issues.

COGNITIVE REPRESENTATION OF GENDER

The Special Topic includes papers from four ITN LCG laboratories with a strong interest in cognitive representations of gender (University of Duisburg-Essen, Basque Center on Cognition, Brain and Language, University of Modena and Reggio Emilia, University of Sussex). The main focus in these studies is on how gender stereotypes, other gender-biased content, and grammatical information combine to determine the representation of the gender of characters, in word as well as in text processing. In this work a particular notion of stereotyping is used, one based on people's estimates of the proportions of females and males filling certain roles. The experimental techniques include ERPs, eye tracking, and various reading time and reaction time paradigms, in particular one devised by Oakhill et al. (2005), in which people have to say whether two terms, for example uncle and nurse, can refer to the same person. The relation between stereotypes defined in this way, and real world ratios of females to males, is explored in the paper by Garnham et al.

Canal et al. showed different ERP signatures for the gender mismatch effect for definitional and stereotype-based information in the interpretation of English reflexives. They also included individual difference measures, which correlated with participants' performance. In another ERP study, Su et al. looked at the corresponding match/mismatch effect for Chinese reflexives following definitional or stereotypical role nouns. They argue that the Chinese equivalent of "himself" is the default reflexive, and requires less complex processes of resolution than the Chinese equivalent of "herself".

Reali et al. present an eye tracking study in which people were described as performing typically female or typically male activities without using role names. Strength of typicality was manipulated, but evidence for stereotyping was found with both strong and weak typicality, suggesting a difference between typicality and stereotyping.

Siyanova-Chanturia et al. tested Italian 3rd and 5th graders, and young and older adults on the Oakhill et al. (2005) two-word task. They found some interesting asymmetries for both male vs. female participants and for masculine vs. feminine stereotyped role names. Their most important finding, however, was that the basic stereotyping effect was seen in all age groups.

Hanulíková and Carreiras looked not at stereotyping, but at gender information conveyed by a speaker's voice. The effect of this information was contrasted with that of morphosyntactic marking on subject nouns, and the study investigated how both types of information affected subject-verb agreement. Different ERP signatures were found for the two types of agreement.

Garnham and Yakovlev report a study of the reading of short passages in Russian with either singular or plural stereotyped role nouns. The grammar of Russian has complexities not seen in languages previously investigated in studies of this kind, and Garnham and Yakovlev found complex interactions between grammatical and stereotypical information about gender. In addition, they provide a set of stereotype norms for 160 role names in Russian.

The final experimental paper in this group is Finnegan et al.'s study of the use of counter-stereotype pictures to overcome automatic stereotyping. This study used the two-word task, and looked at changes in responding before and after exposure to a set of pictures with people in either stereotypical or counter-stereotyped roles.

The Hypothesis and Theory paper by Esaulova and von Stockhausen argues that gender should be treated as a prominence feature, which influences, for example, the assignment of thematic roles. Consistent with the notions of stereotyping and discrimination, stereotypically masculine role names were more easily integrated with agent roles than were stereotypically feminine ones.

GENDER FAIR LANGUAGE

Another group of papers, including work from ITN LCG labs in Berlin (Free University), Bern, and Padua, focuses specifically on the use of gender fair language. As Sczesny et al. point out in their Review, two methods of eliminating the overuse of masculine or male-related terms are neutralization (e.g. replacing policeman with police officer) and feminization [e.g. replacing an allegedly generic masculine plural, such as German Lehrer (teacher) with a word pair Lehrerinnen und Lehrer]. These authors look at the use of these two strategies and how they can feed into future research and policy making.

An example of research of this kind is presented by Horvath et al. who show that word pairs such as Lehrerinnen und Lehrer increase female visibility in occupations, but decrease estimated salaries (though not competence/status), though for female-biased professions only. Similarly, in a study of Swiss French speaking adolescents, Vervecken et al. found that the use of word pairs reduces differences in ascription of success in occupations and ascriptions of warmth. Ascriptions of competence were not affected by language forms. Hansen et al. looked at the effect of including generic masculine or word pairs in German newspaper

reports. The linguistic forms used in the reports affected readers own use of word pairs, and led to more gender-balanced representations.

Gustafsson Sendén et al. report on a different attempt to use gender-fair language, the introduction, in Sweden in 2012, of a new gender neutral pronoun, "hen." Initial hostile attitudes to the new pronoun reduced dramatically over the following 4 years, though take up of use of the word was relatively slow. From a broader perspective, Formanowicz et al. review the effects of gender fair language on support for social initiatives in Poland and Austria. Gender fair language is a relatively new idea in Poland and had detrimental effects on support for social initiatives. In Austria, where it is better established, it had positive effects, suggesting the need for gender fair language to establish itself in a particular society before it can be effective in reducing discrimination.

DECISION MAKING AND TEACHERS' ATTITUDES

In the two final studies of the Special Topic, Fabre et al. showed that female responders were typically treated less fairly in the ultimatum game, and when female responders were treated fairly, more cognitive effort was needed. Wolter et al. showed that schoolteachers' attitudes are an important factor

affecting whether the stereotype "reading is for girls," measured by both motivation to read and competence at reading, is realized.

All in all, the papers in the Special Topic both contribute to our understanding of how language determines the representation of gender and feed in to discussion of and strategies for mitigating against negative effects of language.

AUTHOR CONTRIBUTIONS

AG, wrote draft, incorporated comments of other authors into revised version. JO, LS, SS, commented on draft. AG, JO, LS, SS, proposed and edited the Special Topic for which this is the editorial.

FUNDING

REFERENCES

Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cognit.* 3, 72–983. doi: 10.3758/BF03193206

True gender ratios and stereotype rating norms

Alan Garnham[1]*, Sam Doehren[1] and Pascal Gygax[2]

[1] *School of Psychology, University of Sussex, Brighton, UK,* [2] *Department of Psychology, University of Fribourg, Fribourg, Switzerland*

We present a study comparing, in English, perceived distributions of men and women in 422 named occupations with actual real world distributions. The first set of data was obtained from previous a large-scale norming study, whereas the second set was mostly drawn from UK governmental sources. In total, real world ratios for 290 occupations were obtained for our *perceive* vs. *real world* comparison, of which 205 were deemed to be unproblematic. The means for the two sources were similar and the correlation between them was high, suggesting that people are generally accurate at judging real gender ratios, though there were some notable exceptions. Beside this correlation, some interesting patterns emerged from the two sources, suggesting some response strategies when people complete norming studies. We discuss these patterns in terms of the way real world data might complement norming studies in determining gender stereotypicality.

Keywords: stereotypes, role names, ratings, true ratios, archival sources

Introduction

Gender stereotyping forms a cornerstone of psychology with many sub-domains researching the topic in detail, in particular, and for very different reasons, social psychology and psycholinguistics. The study of gender stereotyping in social psychology focuses on the processes that lead to stereotyping – applying a set of beliefs about the characteristics of a social category to members of that category (Greenwald and Banaji, 1995) – and the wider result of this stereotyping; see, for example, Peterson and Zurbriggen (2010) and Latrofa et al. (2012). Gender stereotyping in psycholinguistics has typically been studied as an example of inference in the comprehension of discourse and text. When an individual is described as an engineer, researchers have been interested in possible inferences about the gender of this engineer. Much of this research uses anaphor resolution as in index of stereotyping (e.g., Carreiras et al., 1996) or judgments about words that explicitly or implicitly refer to a person of a given gender (e.g., Oakhill et al., 2005; Gygax and Gabriel, 2008).

In this psycholinguistic literature, stereotyped words are often compared and contrasted with words with definitional gender, such as *king* and *queen* (e.g., Banaji and Hardin, 1996; Osterhout et al., 1997; Oakhill et al., 2005). As in the case of *king* and *queen*, these words often form morphologically unrelated pairs. Morphologically related pairs, such as actor and actress, have undergone considerable changes in usage over the past 50 years. The definitions of role names that are gendered by stereotype do not contain gender information as part of their core meaning, which defines the role itself (what a footballer does, for example, or a secretary). It therefore follows that if the effect of the gender stereotypicality of a noun, or a role name (e.g., *taxi driver*) more generally, is to be studied, the extent of the stereotyping of the noun first needs to be measured.

Because psycholinguistic studies of stereotyping look at whether, or how likely, an inference about a person's gender will be made on the basis of stereotype information, it has been considered appropriate to assess the proportions of men and women thought, by people similar to those tested in the core experiment, to fill various roles in the real world. In collecting stereotype norms, therefore, the method has been to collate a set of role names (which may be either single nouns, such as *nurse*, or phrases such as *primary school teacher*), present them to judges, and use a variant of the instruction: estimate to what extent the groups are made up of women or men. Data is typically collected on a Likert-type scale (e.g., Kennison and Trofe, 2003; Gabriel et al., 2008; Irmen and Kurovskaja, 2010; Misersky et al., 2014). These studies have often been carried out as pre-tests for a particular further study, rather than as studies in their own right. Misersky et al. (2014) pointed out that, therefore, the methods have varied enough to prevent direct comparison between studies. The study carried out by Misersky et al. (2014) used a common data collection tool, designed for the study but extensible to other languages, to collect stereotype norms in seven languages and for a large set of role names. Four hundred and twenty-two role names were chosen to be tested for English, and as many of those in the other six languages that had translations from English. Selection was based on previous norming studies (e.g., Kennison and Trofe, 2003; Gabriel et al., 2008), as well as on brainstorming sessions and trawls of dictionaries. All of the chosen terms were intended to be stereotypically applied to males or females, but not definitionally. This distinction is not always completely clear-cut, partly because of changing matters of usage. A particularly tricky case is *waiter*, which was once part of a gender marked pair *waiter/waitress,* and was the subject in the United States of a largely failed attempt to replace it with the supposedly gender neutral term *server*. The Cambridge free English dictionary (Cambridge Dictionaries Online, 2015), for example, defines a waiter as "a man whose job is to bring the food to customers at their tables in a restaurant," though other sources reflect more progressive thinking (Cambridge Dictionaries Online, 2015, under *Sexist language*). In the Misersky et al. (2014) study, respondents were free to indicate that they thought 100% of waiters were male, though the actual figure was 45%, and the true data from ONS sources suggested 75%. As in previous studies carried out by our group (e.g., Gabriel et al., 2008), an 11-point Likert type scale was used, ranging from 0% women/100% men to 100% women/0% men, in 10% steps, and participants were asked to estimate to what extent the roles presented to them were carried out by women or men. Participants were specifically asked to think of the *real* proportion of men and women in the roles (and not to base their responses on how they thought things should be). Data were collected online, and in the English sample there were 281 respondents, far more than in previous studies and hence providing reasonably accurate estimates of beliefs about the proportions of men and women filling the roles studied (see original paper for data).

Though stereotyping is often seen as a negative and prejudicial activity, it is widely accepted as a required process for simplifying a complex world via the use of schemas (Augoustinos and Walker, 1998; López-Sáez et al., 2008; Wilbourn and Kee, 2010). Within the social psychology literature, attempts have been made to determine whether stereotyping is based on outdated true gender bias (Wilbourn and Kee, 2010), or (possibly incorrect) assumptions about current female/male ratios (Lopez-Zafra and Garcia-Retamero, 2012; Mills et al., 2012). However, exact gender ratios are not usually reported, so the conclusions can be difficult to evaluate. In the psycholinguistic domain, it is sensible to assume that comprehension is driven by beliefs about male/female ratios, rather than unknown (to the comprehender) true ratios. Nevertheless, the question can be asked about the relation between assumed and true ratios. The answer to that question bears both on the interpretation of psycholinguistic findings, and also, more importantly in the present context, potential prejudice based on completely incorrect assumptions. The current study, therefore, aims to provide true gender ratios for as many of the English role names that appear in the Misersky et al. (2014) study as possible, and to compare them with the reported ratios in the Misersky et al. (2014) data set. Because of the lack of previous research on true gender ratios it is an open question how closely related the norm data and true gender ratios will be.

The main source of information about true gender ratios was, where possible, archival data collected by the UK Office of National Statistics (ONS, http://www.ons.gov.uk/). Where necessary other archival resources were used. The primary objective the current research is, therefore, to collect true gender ratios for the role names presented in Misersky et al. (2014), and to compare them with the normative data from that study.

Materials and Methods

We used archival data to collect true gender ratios for as many as possible of the 422 English role descriptions from Misersky et al. (2014), reproduced in data sheet 1 in the supplementary material.

The data were primarily collected from governmental, in particular the UK Office for National Statistics (ONS), and academic sources. In a minority of cases other sources were considered appropriate, and were used. Where no source was available, or considered to be reliable, no estimate of the true ratio was obtained.

The archive search had a number of stages, and proceeded on an item-by-item basis, rather than a source-by-source basis. An attempt was made to locate each item in each source in order. If a source failed to provide relevant data, the next source was consulted. If relevant data were found at any stage, the process ended and the next source was not searched. If the mapping between a role name in the Misersky norms and information in a source was unclear, supplementary information on governmental and academic sites was used to clarity the definition of the role name in the archival data (no definitions were provided in the normative study). The ONS Standard Occupational Classification (ONS, 2010) was the most important document in this context. On occasion more than one definition was available. In such cases, all definitions were incorporated, if possible.

The process and sources were as follows:

(1) 2011 Census, Population Estimates by single year of age and sex for Local Authorities in the UK (ONS, 2013a)

  (i) This source is a list of demographic information about age and gender of the population of the UK.

(2) Reference table EMP16 'Employment by occupation' (ONS, 2013c), in conjunction with the Standard Occupational Classification 2010 Volume 2 The coding index (ONS, 2010)

  (i) Reference table EMP16 is a list of general job roles with the numbers of people from each gender that perform that role as an occupation, both full and part time.

  (ii) The Standard Occupational Classification coding index is a detailed list of job roles and provides the four-level classification ONS uses in EMP16.

    (I) The Standard Occupational Classification (SOC) coding index was searched for the role name; all occurrences of the role name were used. This search provided a list of 'SOC' codes that were cross-referenced with EMP16 to provide the gender ratios.

    (II) If two, or more, job roles returned the same SOC code for one role name, each SOC code was only used once to estimate the gender ratio for each role name.

(3) Other UK governmental sources

  (i) A Google search was performed with the role name combined with the search terms 'gender statistics' and 'gender ratio' to find appropriate websites sources.

  (ii) Only sites with UK governmental top-level domains were accepted at this stage; for example, .gov.uk or .mod.uk.

    (I) Sports based role names were the exception to this rule; statistics obtained directly from governing bodies were accepted if UK specific statistics were provided; as was the case, for example, for the Football Association.

(4) Academic sources

  (i) Scopus and Google Scholar were searched for the role names with, and without, the addition of the phrases 'gender statistics' and 'gender ratio.'

(5) Other sources

  (i) As with 'Other UK governmental sources,' a Google search was performed with the role name combined with search terms 'gender statistics' and 'gender ratio' to find appropriate website sources.

  (ii) Each source was judged on its own merits; for example, national UK news sources and national bodies were accepted, but blogs were not.

Each ratio was assessed for quality. The first criterion for quality was recency. Ratios dated prior to 2008 (5 years prior to the work being carried out) were marked as questionable.

Only one ratio was considered questionable on these grounds. Initially recency was to be the only criterion for the quality of the ratios, as the quality of the sources was supposed to be guaranteed by the collection process. However, during the process of data collection a second set of issues became apparent in the ratios produced from the ONS employment data (stage 2, above). The process of collating the list of SOC codes from the Standard Occupational Classification (ONS, 2010) involved identifying all occurrences of the relevant role name in the list, and it produced two types of problem. First, a specific term in Misersky et al.'s (2014) list was only located in one broader category. For example, the role name 'Zoologists' was deemed to be part of the job role 'Biological scientists and biochemists,' which covers more than just 'Zoologists.' Second, a single term in Misersky et al.'s (2014) list was associated with a large number of job roles. For example the role name 'Manager' was part of 1336 job descriptions, which were associated with 121 different SOC codes. In such cases, it is not clear that Misersky et al.'s (2014) participants would have all these possibilities in mind when making their judgments. Therefore, if the job role was deemed too broad, or if it was associated with more than ten SOC codes, the resulting ratio was classified as questionable.

We found archival data on true gender ratios for 290 (out of 422) of the role names in Misersky et al.'s (2014) English list. As can be seen in **Table 1**, the vast majority of the true gender ratios were found in stage 2 of the archival search process, though many of these have been classified as questionable. In total, 86 ratios of the 290 ratios have been so classified. The stage where each questionable ratio was collected is shown in **Table 1**, and the role names with questionable ratios are flagged in data sheet 1 in the supplementary material.

The 132 role names for which no data have been found include about 20 cases where data are unlikely to be obtainable. Some roles, such as 'Executioners' no longer exist in British society, others are difficult to define or collect data for (e.g., 'Clients'), and others may be protected by considerations of security (e.g., 'Spies'). For the rest, data are in principle obtainable, though possibly from sources that would be unreliable.

Results

The mean true gender ratio of the 290 role names was 0.44 (SD = 0.17), where 1.00 would represent 100% females and, 0.00, 100% males. This mean is similar to the mean found in Misersky

TABLE 1 | List of data collection stages.

Stage	Role names	Questionable
1	17 (4.03%)	0
2	230 (54.5%)	84
3	30 (7.11%)	1
4	2 (0.47%)	1
5	11 (2.61%)	0
No data	132 (31.28%)	NA

et al. (2014) for the same role names ($M = 0.43$, SD = 0.30). The range of the ratios was 0.00 to 1.00, this compares to the Misersky et al. (2014) range of 0.15 to 0.84. Skew and kurtosis were modest, 0.49 and −0.68, respectively.

A two-tailed Pearson's correlation was calculated to investigate how the findings of Misersky et al. (2014) related to the true gender ratios collected in this study. It was found that there was a strong significant positive relationship between the two data sets ($r = 0.755$, $N = 290$, $p < 0.001$).

As many of the ratios had been highlighted as questionable during the collection process, it was decided to separate these ratios from the non-questionable data and perform a Pearson's correlation on each set separately. Removing the questionable ratios improved the correlation ($r = 0.849$, $N = 205$, $p < 0.001$). The questionable ratios also correlated significantly with the relevant judged ratios, though much less strongly ($r = 0.273$, $N = 85$, $p = 0.011$).

Figure 1 highlights the difference in the range of the ratios found in the two studies, as well as separately indicating the questionable and non-questionable ratios. Numerical values for all the ratios can be found in data sheet 1 in the supplementary material.

Discussion

The primary aim of this study was to produce true gender ratios for as many as possible of the 422 English role names for which Misersky et al. (2014) reported judgments of gender ratio. These true gender ratios were to be compared with the normative judgments reported by Misersky et al. (2014). To date there has

been relatively little study of true gender ratios, and none on the scale of the current survey.

The overall mean in this study (0.44) suggests a tendency for the role names selected to be predominantly male. This fact may be explained by the fact that majority of the true gender ratios are for occupational role names and ONS (2014) states that the majority of workers are male, with a true gender ratio of 0.47. Figures from the past would show a greater proportion of males in the UK workforce. Full details of true gender ratios for individual role names are available in data sheet 1 in the supplementary material.

The second aim of this study was to compare the true gender ratios with the normative judgment data on stereotypicality from Misersky et al. (2014). The two studies found similar means for the gender ratio across the 290 role names for which both types of data were available (current study, $M = 0.44$, SD = 0.17; previous study, $M = 0.43$, SD = 0.30). Misersky et al. (2014) attribute this male bias to stronger male stereotypes, as did a previous study that was similar in nature (Gabriel et al., 2008). This study, looking at true gender ratios, found a similar mean to Misersky et al. (2014), Rather than suggesting stronger male stereotypes, in any sense suggesting a mismatch with reality, it appears that the role names investigated refer to roles that, on average, more males than females fill. Looking at the role names, it is clear that the majority of them are occupations, or could be viewed as occupations, and, as previously mentioned, the work force in the UK is predominantly male. It would, therefore, be expected that there would be a slight male bias (ONS, 2014).

As well as finding similar means, the two studies produced data for the 290 roles names that are significantly correlated ($r = 0.755$, $p < 0.001$). This correlation improved when the
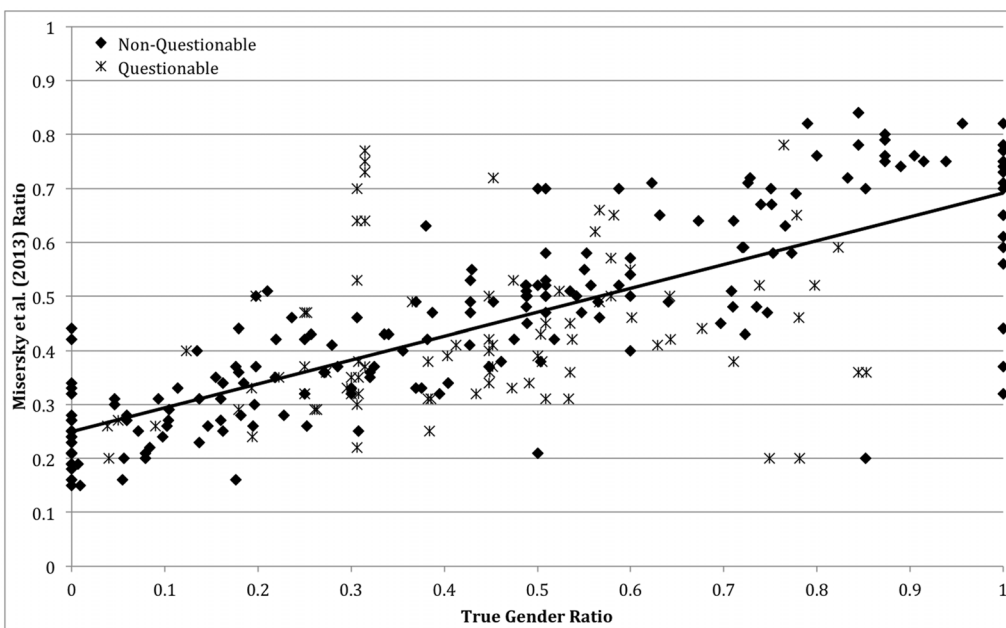

FIGURE 1 | Scatter plot of real gender ratios from current study against normative judgments from Misersky et al. (2014). The solid line is the line of best fit for all data.

ratios classified as questionable were removed from the analysis ($r = 0.849$, $N = 205$, $p < 0.001$). The two findings together, means and correlation, suggests that people are generally able to provide an accurate estimate of the true gender ratio for a role name.

Though people were generally correct in their estimates of gender ratios, there were exceptions. For a small number of role names, the discrepancy between the estimate and the true gender ratio was over 0.50. These role names were: Tailor, Barber, Probation Officer, Hunter, Archivist, Curator, and Butler. All discrepancies over 0.50 involved overestimation of the proportion of males who performed the role. This fact provides some limited support for the conclusion drawn by Gabriel et al. (2008) and Misersky et al. (2014) that male stereotypes are stronger than female stereotypes. However, except for Barber, the true gender ratios for the seven roles names in this category are considered questionable, four for having absolute values of 1.00 (see below) and the remaining two because their SOC code refers to an overly broad category.

As mentioned in the methods section, some of the ratios are classified as questionable (for our purposes) because the source provides information about a similar, but not the same, role name as the one we believe people were making judgments about. For example, in the ONS data, 'Author' was included in the broader categories "Authors, writers and translators" and "Programmers and software development professionals," with no possibility of disaggregating the data. It is unlikely that Misersky et al.'s (2014) participants had this definition of 'Author' in mind when making their judgments.

In addition, as can be seen in **Figure 1**, a number of role names (53) have ratios of 0.00 (all men) or 1.00 (all women). Two of these ratios, 'Admirals' [all men, DASA (Navy) (2013)] and 'Synchronized swimmers' (all women, Fédération Internationale de Natation, 2013), came from stage 3 of the collection process. Both of these ratios came from reliable sources and are accepted as correct. The remaining 51 of these ratios came from stage 2 of the collection process and reflect the fact that the number of workers of one gender is considered "too small for reliable estimate," and so cannot be distinguished from zero (ONS, 2013c). In EMP16 (ONS, 2013c) no information is given about what counts as too small. However, it can be inferred that the cut off for this classification occurs between 0 and 4713 people occupying the role, this number being one less than the lowest statistic that is provided for any job role. The effect on the resultant ratio varies considerably between role names. For example, 470,749 males are said to be 'Electricians,' whereas the number females is "too small for reliable estimate." In this case, even if there were 4713 females electricians, the ratio would only change from 0.00 to 0.01. 'Shoemakers,' on the other hand,

also has a 0.00 ratio, but with only 6305 males; in this case the potential change from including 4713 women is from 0.00 to 0.43.

Another issue arises from the use, by Misersky et al. (2014), of an 11-point Likert scale with 10% increments for the estimation of ratios. Participants might be reluctant to use extreme values (0% men, 0% women) when they know that some women or men do occupy certain roles. They might have been less reluctant to provide values closer to 0 or 100% on a less coarse scale, though the issue of whether sliders are preferable to radio button/Likert-type scales is a complex one (Cook et al., 2001). Another reason why participants might be reluctant to use extreme values could be that they try to produce socially desirable responses, and hence avoid extreme values, to look open minded. Although the instructions did ask participants to dissociate themselves from their view of gender equality, we cannot be sure to what extent they followed this instruction.

The true gender ratios collected as part of this study should aid future research on stereotyping. Not only do they provide a detailed catalog of true gender ratios. They also allow a distinction to be drawn between stereotyped role names that are correctly judged to be typical of one gender and those that are not. The question of why some occupations are typical of one gender still remains, but the question of why some estimates are better than others is an interesting one for future research and researchers may well want to consider their data set in terms of how big the discrepancy is between stereotype beliefs and true typicality.

One issue that neither the current, nor previous, research has addressed is the familiarity of the role names. It is reasonable to assume that the more familiar a person is with a role name, the more likely it is that they will have specific knowledge related to that role, including knowledge of true gender ratios. There are at least two different ways to incorporate questions about familiarity into research of this kind. First, the data collection tool developed for the Misersky et al. (2014) study could be augmented to collect familiarity information. Second, Blair et al. (2002) found that estimates of word frequency using Internet search methods correlate reasonably well with familiarity ratings. This second method would not be as satisfactory, as it would not provide direct estimates of familiarity. However, it could produce results more quickly, and might be preferred for that reason.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01023

References

Augoustinos, M., and Walker, I. (1998). The construction of stereotypes within social psychology: from social cognition to ideology. *Theor. Psychol.* 8, 629–652. doi: 10.1177/0959354398085003

Banaji, M. R., and Hardin, C. D. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x

Berman, G., and Dar, A. (2013). *Prison Population Statistics*. Available at: www.parliament.uk/briefing-papers/sn04334.pdf [accessed December 15, 2013].

Blair, I. V., Urland, G. R., and Ma, J. E. (2002). Using Internet search engines to estimate word frequency. *Behav. Res. Meth. Ins. C* 34, 286–290. doi: 10.3758/BF03195456

Cambridge Dictionaries Online. (2015). Available at: http://dictionary.cambridge.org/dictionary/british/waiter [accessed June 1, 2015].

Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol. A* 49, 639–663. doi: 10.1080/713755647

Cook, C., Heath, F., Thompson, R. L., and Thompson, B. (2001). Score reliability in web or internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educ. Psychol. Meas.* 61, 697–706. doi: 10.1177/00131640121971356

DASA (Navy). (2013). *Naval Service Quarterly Pocket Brief January 2013*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/280536/january_2013.pdf [accessed December 12, 2013].

Department for Education. (2010). *A Profile Of Teachers In England From The 2010 School Workforce Census*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/182407/DFE-RR151.pdf [accessed February 23, 2014].

Department for Transport. (2010). *National Travel Survey: 2010*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/8933/nts2010-02.pdf [accessed January 26, 2014].

Fédération Internationale de Natation. (2013). Synchronised Swimming Medallists and Statistics. Available at: http://www.fina.org/H2O/docs/histofina/synchro.pdf [accessed November 15, 2013].

Financial Reporting Council. (2013). *Key Facts and Trends in the Accountancy Profession*. Available at: http://www.frc.org.uk/Our-Work/Publications/Professional-Oversight/Key-Facts-and-Trends-in-the-Accountancy-Profession.aspx [accessed December 20, 2013].

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au pairs are rarely male: norms on the gender perception of role names across English, French, and German. *Behav. Res. Meth.* 40, 206–212. doi: 10.3758/BRM.40.1.206

General Teaching Council for Wales. (2012). *Annual Statistics Digest March 2012*. Available at: http://www.gtcw.org.uk/gtcw/images/stories/downloads/Annual%20Statistics%20Digest/Annual_Stats_12_E.pdf [accessed February 28, 2014].

Government Digital Service. (2014). *Ministers*. Available at: https://www.gov.uk/government/ministers [accessed January 15, 2014].

Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27. doi: 10.1037/0033-295X.102.1.4

Gygax, P., and Gabriel, U. (2008). Can a group of musicians becomposed of women? Generic interpretation of French masculine rolenames in absence and presence of feminine forms. *Swiss J. Psychol.* 67, 141–153. doi: 10.1024/1421-0185.67.3.143

Hassell, K., and Shann, P. (2003). *Pharmacy Workforce Census – February 2003*. Available at: http://www.rpharms.com/about-pharmacy-pdfs/census03.pdf [accessed February 23, 2014].

Health and Care Professions Council. (2012). *Registration statistics - 1 November 2012*. Available at: http://www.hpc-uk.org/publications/foi/index.asp?id=625 [accessed February 15, 2014].

Health and Care Professions Council. (2013). *Hospital Outpatient Activity - 2012-13* [*NS*]. Available at: http://www.hscic.gov.uk/searchcatalogue?productid=13684&q=gender&sort=Relevance&size=10&page=1#top [accessed January 26, 2014].

Health and Social Care Information Centre. (2013). *Statistics on Alcohol - England, 2013* [*NS*]. Available at: http://www.hscic.gov.uk/article/2021/Website-Search?productid=11719&q=Statistics+on+Alcohol&sort=Relevance&size=10&page=1&area=both#top [accessed January 18, 2014].

Home Office. (2013a). *Drug Misuse: Findings from the 2012 to 2013 Crime Survey for England and Wales*. Available at: https://www.gov.uk/government/publications/drug-misuse-findings-from-the-2012-to-2013-csew/drug-misuse-findings-from-the-2012-to-2013-crime-survey-for-england-and-wales [accessed January 30, 2014].

Home Office. (2013b). *Tables for "Police Workforce, England and Wales, 31 March 2013."* Available at: https://www.gov.uk/government/publications/tables-for-police-workforce-england-and-wales-31-march-2013 [accessed November 15, 2013].

Irmen, L., and Kurovskaja, J. (2010). On the semantic content of grammatical gender and its impact on the representation of human referents. *Exp. Psychol.* 57, 367–375. doi: 10.1027/1618-3169/a000044

Joseph Rowntree Foundation. (2013). *Job Seeker's Allowance Claimant Count By Age And Gender Over Time*. Available at: http://www.ons.gov.uk/ons/dcp171778_298904.pdf [accessed February 28, 2014].

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

Latrofa, M., Vaes, J., and Cadinu, M. (2012). Self-stereotyping: the central role of an ingroup threatening identity. *J. Soc. Psychol.* 152, 92–111. doi: 10.1080/00224545.2011.565382

López-Sáez, M., Morales, J. F., and Lisbona, A. (2008). Evolution of gender stereotypes in Spain: traits and roles. *Span. J. Psychol.* 11, 609–617. doi: 10.1017/S1138741600004613

Lopez-Zafra, E., and Garcia-Retamero, R. (2012). Do gender stereotypes change? The dynamic of gender stereotypes in Spain. *J Gend. Stud.* 21, 169–183. doi: 10.1080/09589236.2012.661580

Mills, M. J., Culbertson, S. S., Huffman, A. H., and Connell, A. R. (2012). Assessing gender biases: development and initial validation of the gender role stereotypes scale. *Gend. Manage. Int. J.* 27, 520–540. doi: 10.1108/17542411211279715

Ministry of Defence. (2013). *UK Armed Forces Quarterly Personnel Report: 2013*. Available at: https://www.gov.uk/government/publications/uk-armed-forces-quarterly-personnel-report-2013 [accessed February 23, 2014].

Ministry of Justice. (2011). *Statistics On Women And The Criminal Justice System 2011*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220081/statistics-women-cjs-2011-v2.pdf [accessed February 2, 2014].

Misersky, J., Gygax, P., Canal, P., Gabriel, U., Garnham, A., Braun, F., et al. (2014). Norms on the gender perception of role names in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behav. Res. Meth.* 46, 841–871. doi: 10.3758/s13428-013-0409-z

Nardi, P. M. (1988). The social world of magicians: gender and conjuring. *Sex Roles* 19, 759–770. doi: 10.1007/BF00288991

Oakhill, J. V., Garnham, A., and Reynolds, D. J. (2005). Immediate activation of stereotypical gender information. *Mem Cogn.* 33, 972–983. doi: 10.3758/BF03193206

ONS. (2010). *Standard Occupational Classification 2010 Volume 2 the Coding Index*. Available at: http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/soc2010—volume-2-of-the-coding–index/index.html [accessed September 21, 2013].

ONS. (2013a). *2011 Census, Population Estimates by Single Year of Age and Sex for Local Authorities in the United Kingdom*. Available at: http://www.ons.gov.uk/rel/census/2011-census/population-estimates-by-single-year-of-age-and-sex-for-local-authorities-in-the-united-kingdom/index.html [accessed September 20, 2013].

ONS. (2013b). *2011 Census: Lone Parent Households With Dependent Children, Local Authorities In The United Kingdom*. Available at: http://www.ons.gov.uk/rel/census/2011-census/population-estimates-by-single-year-of-age-and-sex-for-local-authorities-in-the-united-kingdom/index.html [accessed September 20, 2013].

ONS. (2013c). *EMP16 Employment by Occupation*. Available at: http://www.ons.gov.uk/ons/search/index.html?newquery=EMP16 [accessed September 21, 2013].

ONS. (2014). *The Latest on the UK Labour Market*. Available at: http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/february-2014/sty-employment.html [accessed February 24, 2014].

Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cogn.* 25, 273–285. doi: 10.3758/BF03211283

Peterson, B. E., and Zurbriggen, E. L. (2010). Gender, sexuality, and the authoritarian personality: authoritarianism and gender. *J. Pers.* 78, 1801–1826. doi: 10.1111/j.1467-6494.2010.00670.x

Ratcliffe, R. (2013). *The Gender Gap at Universities: Where are all the Men? the Guardian*. Available at: http://www.theguardian.com/education/datablog/2013/jan/29/how-many-men-and-women-are-studying-at-my-university

Royal College of Psychiatrists. (2012). *Ethnicity, Gender and Age Monitoring*. Available at: http://www.rcpsych.ac.uk/pdf/Graphs%202007-2012.pdf [accessed January 15, 2014].

Sanders, L., Sander, P., and Mercer, J. (2009). Rogue males? Approaches to study and academic performance of male psychology students. *Psychol. Teach .Rev.* 15, 3–17.

Shields, L., Hall, J., and Mamun, A. A. (2011). The "gender gap" in authorship in nursing literature. *J. R. Soc. Med.* 104, 457–464. doi: 10.1258/jrsm.2011.110015

UK Parliament. (2014a). *Frequently Asked Questions: MPs*. Available at: http://www.parliament.uk/about/faqs/house-of-commons-faqs/members-faq-page2/ [accessed February 28, 2014].

UK Parliament. (2014b). *Lords by Party, Type of Peerage and Gender*. Available at: http://www.parliament.uk/mps-lords-and-offices/lords/composition-of-the-lords/ [accessed February 28, 2014].

Wikipedia. (2014). *List Of Current Heads Of State And Government*. Available at: http://en.wikipedia.org/wiki/List_of_current_heads_of_state_and_government [accessed February 14, 2014].

Wilbourn, M. P., and Kee, D. W. (2010). Henry the nurse is a doctor too: implicitly examining children's gender stereotypes for male and female occupational roles. *Sex Roles* 60, 670–683. doi: 10.1007/s11199-010-9773-7

Womens Sport and Fitness Foundation. (2012). *Womens Football Fact Sheet Oct 2012*. Available at: http://www.thefa.com/*sim*/media/files/thefaportal/governance-docs/equality/women-and-girls/womens-football-fact-sheet-oct-2012.ashx [accessed April 2, 2014].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Beyond Gender Stereotypes in Language Comprehension: Self Sex-Role Descriptions Affect the Brain's Potentials Associated with Agreement Processing

Paolo Canal[1,2]*, Alan Garnham[1] and Jane Oakhill[1]

[1] University of Sussex, Brighton, UK, [2] Center for Neurocognition Epistemology and Theoretical Syntax, Institute for Advanced Study of Pavia, Pavia, Italy

We recorded Event-Related Potentials to investigate differences in the use of gender information during the processing of reflexive pronouns. Pronouns either matched the gender provided by role nouns (such as "king" or "engineer") or did not. We compared two types of gender information, definitional information, which is semantic in nature (a mother is female), or stereotypical (a nurse is likely to be female). When they followed definitional role-nouns, gender-mismatching pronouns elicited a P600 effect reflecting a failure in the agreement process. When instead the gender violation occurred after stereotypical role-nouns the Event Related Potential response was biphasic, being positive in parietal electrodes and negative in anterior left electrodes. The use of a correlational approach showed that those participants with more "feminine" or "expressive" self sex-role descriptions showed a P600 response for stereotype violations, suggesting that they experienced the mismatch as an agreement violation; whereas less "expressive" participants showed an Nref effect, indicating more effort spent in linking the pronouns with the possible, although less likely, counter-stereotypical referent.

Keywords: electrophysiology of language comprehension, individual differences, social perception of gender, P600, Nref, gender stereotypes, anaphor processing

INTRODUCTION

Research has shown that readers make inferences based on information that is explicit in a text, and on readily available general knowledge, to establish a coherent representation of the text. When a character is introduced in a text, readers use different sources of information to construct an incremental model of the discourse in which the representation of the character is specified to a greater or lesser extent. This representation creates expectations about what the character is likely to do or not to do. In the present study we explored the extent to which information that is not included in the text, and psychological factors that are unrelated to reading abilities, contribute to the representation of characters mentioned in the text. In particular we investigated how different types of information about gender, based either on the semantic definition of a noun or on stereotypical information associated with it, are used to determine the gendered representation of

the text character. We also looked at whether a reader's commitment to gender-related information is modulated by individual differences in the social perception of gender.

When reading a gender-specific noun (*mother* or *father*) or a proper name (*Alan* or *Jane*) the gender of the character is incorporated into the mental model of the discourse (Chang, 1980; Garnham and Oakhill, 1985). In a natural gender language such as English, in which nouns have no grammatical gender (although pronominal forms vary depending on the gender of their referents), gender information can be conveyed through definitional or stereotypical information (Corbett, 1991). Definitional gender derives from the semantic definition of a noun: *mothers* are women. Stereotypical gender by contrast is the gender bias that is often associated with "role" nouns such as *nurse*, which refer to professions and social roles, sometimes via titles. Stereotypical information about gender is not provided by grammar or semantics but derives at least partly from individuals' world knowledge about the proportion of men and women carrying out certain jobs or holding certain social roles. Different studies (Kennison and Trofe, 2003; Gygax et al., 2008; Misersky et al., 2013) showed that the cognitive representation of a role name is often gender biased (e.g., *mechanics* are typically male). Therefore, from the simple mention of a *mechanic* people may infer that the noun's referent is a man and strongly commit to this information (see the surgeon riddle in Sanford, 1985, p. 311; but also Banaji and Hardin, 1996; Carreiras et al., 1996; Osterhout et al., 1997; Sturt, 2003; Reynolds et al., 2006; Siyanova-Chanturia et al., 2012). Research has also shown that additional disambiguating information can interact with stereotype information, and override it (Carreiras et al., 1996; Duffy and Keir, 2004).

In the present study we used the Event Related Potential (ERP) technique, to compare how reading comprehension processes differ when people activate gender information that is categorical because it is semantic in nature (a female *mother*) and when they activate probabilistic information based on a stereotypical representation (a female *nurse*). ERPs are small voltage changes in the electrical activity of the brain, recorded from the scalp, consistently triggered by an external stimulus or a cognitive event. In comparison to the methodologies used in the majority of the studies cited above (reading times, response times, and eye-movements), knowledge about the functional meaning of ERP components, i.e., the neural activity generated by a neural source when a specific operation is performed (Luck, 2005), allows researchers to test hypotheses about qualitative differences in the processes under scrutiny.

As in Osterhout et al. (1997), a gender violation paradigm was used with reflexive pronouns as probes: a male or female character was introduced in a sentence and followed by a male or female pronoun. Pronouns explicitly signal that the entity to which they refer is female or male, so that reading *she* (or *herself*) rather than *he* (or *himself*), referring to *John* will result in some kind of cognitive cost, as previously reported in the psycholinguistic literature using behavioral (e.g., Caramazza et al., 1977) and ERP measures (e.g., Osterhout and Mobley, 1995). Pronouns are one instance of coreferential anaphoric expressions, i.e., words or phrases that refer to an entity

previously introduced in the discourse (for an extensive overview of the mental processes involved in anaphor processing, see Garnham, 2001). An influential model of coreferential anaphor processing has been proposed by Garrod and Sanford (1994). Such processing involves at least two stages in which surface-form features and semantic-pragmatic factors interact in linking the anaphor to the appropriate referent, introduced by its antecedent. In the first stage (bonding), a loose attachment between the pronoun and potential antecedents is made on the basis of superficial information: this automatic process is constrained by lexical and syntactic factors. In the second stage (resolution), the link between pronoun and antecedent(s) made in the bonding process is evaluated and re-computed, if necessary taking into account the overall discourse representation. Both antecedent features (grammatical features, such as gender and number, but also accessibility of the antecedent, and discourse focus) and anaphor features (gender, number, and type) affect anaphor interpretation.

Many ERP studies have contributed to the identification of the cognitive mechanisms underlying anaphor processing (for a review see Callahan, 2008). The existing research investigates many different aspects of these mechanisms, ranging from referential ambiguity (Van Berkum et al., 1999, 2003, 2007; Nieuwland and Van Berkum, 2006), to the effects of processing repeated names (Camblin et al., 2007), from the role played by the antecedent's features (Filik et al., 2008, 2011), to the direct comparison of different types of anaphor (Streb et al., 1999, 2004). The few ERP studies that provide evidence on the basic mechanisms of pronoun processing during sentence reading in English (Osterhout and Mobley, 1995; Osterhout et al., 1997), and thus are particularly relevant for the present study, found that when no antecedent is available for the anaphor, as in the sentence *The aunt heard that *he*, gender mismatching pronouns elicit an enhanced P600 component compared with gender-matching pronouns. The P600 component is a positive deflection observed in parietal electrodes, which develops in a late time-interval. Modulations of the P600 component were initially reported for morphosyntactic agreement manipulations (for a review, see Molinaro et al., 2011). Such effects are thought to represent difficulties in a late stage of processing, reflecting sentence revision or reanalysis processes (e.g., Kaan and Swaab, 2003; Friederici, 2011), they often involve syntactic information but more recently they have been observed during the processing of non-syntactic anomalies (e.g., Kuperberg, 2007; Brouwer et al., 2012).

The relation between anaphor and antecedent can be conceived of as semantic or "loose" agreement (e.g., Corbett, 1979) as the anaphor (target) has different forms depending on the referent's (controller) semantic gender, but the domain in which referent and anaphor occur is often non-local ("unbound" personal pronouns can bind to antecedents outside the immediate clause containing them). In this study, we chose to focus on the processing of reflexive pronouns. Unlike definites, reflexives exhibit syntactically constrained behavior as they are governed by the verb, and their domain is local (Bosch, 1983; Principle A in Chomsky, 1993). When processing reflexive pronouns, rather than personal pronouns, readers should have

strong expectations of finding a suitable antecedent, since reflexives must be coreferential, and thus they must agree with the antecedent in number and gender, otherwise the sentence would be syntactically anomalous. Indeed Osterhout et al. (1997) found that reading *The queen prepared himself* elicit an enhanced P600 response that is similar to what is observed in other cases of agreement violation (e.g., Molinaro et al., 2011).

But, what if a gender mismatch occurs on the basis of stereotypical gender information (e.g., *nurse – himself*)? As Osterhout et al. (1997) argued, one might expect that the anomaly of a male playing a stereotypically female role results from the evaluation of the pragmatic plausibility of the situation, and thus could be reflected in the modulation of the N400 component, which is associated with, among other things, the processing of semantically unexpected, or anomalous words (for review see Kutas and Federmeier, 2011). In contrast, Osterhout et al. (1997) found that stereotypical gender violations elicited a "syntactic" P600 effect, which was reduced in size, compared with the definitional gender case, but still reflected a qualitatively similar response to gender violation based on the noun's semantics. This result is not necessarily surprising if we assume (i) that the activation of stereotypical gender information is the result of inference based on pragmatic knowledge that is carried out when processing the noun, but (ii) the use of this information (as with semantic gender information) can be controlled by syntactic factors when the pronoun explicitly requires evaluation of whether anaphor and antecedent are coreferential. Crucially, however, the two types of gender information differ: the gender of a *nurse* is not categorical as the gender of a *mother* is, but probabilistic. To process *he* referring to *nurse* when nurses are thought to be female in 74% of the cases (see British ratings in Misersky et al., 2013), should not be perceived as an outright agreement violation, as it would be if no possible referents were provided in the previous context. All that is necessary is to re-establish the appropriate, although less likely, reference to a male nurse. In the ERP literature on anaphor processing, the effort spent in establishing the appropriate reference when the antecedent is ambiguous, and thus difficult to link with the anaphor, has been associated with a frontal negativity dubbed the Nref effect: this negative deflection has been interpreted as reflecting the process of re-establishing the reference using information from the situation/discourse model (Van Berkum et al., 1999, 2003, 2007; Nieuwland and Van Berkum, 2006; Nieuwland, 2014).

Osterhout et al. (1997) carried out the first study using ERPs to compare the violation of gender expectations based on either semantic or stereotypical information. Their participants read sentences in which the gender of the introduced character could be semantically determined (*mother*, *king*) or stereotypically biased (*nurse*, *mechanic*). ERPs were time-locked to the presentation of reflexive pronouns that could either match the gender of the antecedent or not. The authors observed that the ERP response to both stereotypical and definitional gender violations affected the P600 component. This similarity was explained by postulating that stereotypical gender information is encoded in the grammar and thus produced "syntactic" P600 effects.

In the present study, as well as revisiting Osterhout et al.'s (1997) main results, we also use an individual differences approach. The rationale for using this approach derives from a specific result in Osterhout et al.'s (1997) study: definitional and stereotypical gender violation, but not subject-verb agreement violations, elicited larger P600 components for female participants than for male participants. One idea suggested by those authors referred to the possibility that "the amplitude of the positive shift reflects the 'strength' of stereotypic beliefs" (Osterhout et al., 1997, p. 281): to our knowledge this hypothesis has not been further tested and in the present study we will test it by exploring how differences in the social perception of gender are related to the way stereotype gender mismatch is processed. The variability in the P600 response between female and male participants suggests that considering mediating factors– instead of relying on average data and treating inter-individual variability as measurement noise – could provide a better understanding of the cognitive processes involved in a given mental operation (for a similar view, see Kanai and Rees, 2011).

In the present study we aimed to replicate Osterhout et al.'s (1997) and extend their findings by examining inter-individual variance and testing the hypothesis that the flexibility of the gender representation of a role noun might depend on the individual's social perception of gender. A person who has strongly "sexist" attitudes might be less prone to accept a reference to a female *surgeon*, compared to a less "sexist" person. Or, if a person is more sensitive to gender stereotypes, she or he could activate gender information to a greater extent and thus show more difficulty in establishing the less likely reference. To test this hypothesis we looked for covariation between the electrophysiological effects associated with processing gender mismatching pronouns and individual scores on a battery of additional measures widely used in social psychology. These tests included both implicit and explicit measures and were designed to capture individuals' perception of gender by monitoring the strength of the automatic associations between gender and career (Gender-Career Implicit Association Test – IAT, Greenwald et al., 1998), self sex-role descriptions (Bem Sex Role Inventory – BSRI, Bem, 1974), and explicit measures of sexism (Ambivalent Sexism Inventory – ASI, Glick and Fiske, 1996). Previous studies of individual differences in the ERP correlates of language processing have mostly used predictors that are specific to the language domain, such as (verbal) working memory (WM; e.g., Friederici et al., 1998; Vos and Friederici, 2003; Nieuwland and Van Berkum, 2006; Nieuwland, 2014), or, have considered the impact of proficiency in monolingual native speakers (e.g., Pakulak and Neville, 2010), or individual differences in sentence processing for second language learners (e.g., Tanner et al., 2013; Tanner and Van Hell, 2014). The present work thus explored a more indirect link between non-domain-specific factors, such as social perception of gender, and the gendered representation of role-nouns and its effect on anaphor processing. We explored the impact of these variables using Linear Mixed Models (LMMs) on single trials. This is a relatively new and promising method for ERP research (e.g., Newman et al., 2012; Payne et al., 2015).

The experimental predictions are thus the following: in a minimal sentential context with only one available antecedent, the processing of a reflexive pronoun will incur processing costs if anaphor and antecedent do not match on gender. When the gender of the character is based on the noun's semantics and is thus categorical, gender mismatch should elicit a P600 effect, because no appropriate referent is available. When the gender of the introduced character is instead based on a stereotypical representation, the link between anaphor and antecedent can in principle be made, if readers can mentally create a representation of a female mechanic. The establishment of a possible although less likely reference to a counter stereotypical representation might require additional inferential effort, and thus elicit an Nref effect. Furthermore, we expect to find individual variability in the response to mismatching pronouns in the stereotypical condition, and to capture some of this variability using the additional measures on the social perception of gender.

MATERIALS AND METHODS

Ethics Statement

The experimental work reported in this paper was approved by the University of Sussex Life Sciences and Psychology Cross-Schools Research Ethics Committee. All procedures complied with the British Psychological Society's Code of Human Research Ethics.

Participants

Thirty-four right-handed native monolingual speakers of British English (17 female), with normal or corrected to normal vision, were recruited from the population of Sussex University to participate in the study. Ages ranged from 18 to 36 (mean = 20). Participants were paid £15 for their time. Three participants were removed from the final analyses because of excessive numbers of ERP artifacts.

Additional Measures

After the ERP experiment, participants completed the battery of tests used to assess individual differences in the social perception of gender. Computerized versions of all the tests were used[1] (programmed in PsyScope).

The gender-career Implicit Association Test (IAT; Greenwald et al., 1998) was presented following the guidelines from Greenwald's website[2] and the latest scoring algorithm was used (Greenwald et al., 2003). Briefly, in the Gender-Career IAT participants respond to a series of items from four categories: two represent the "concept discrimination", i.e., men and women (five male and five female proper names) and two represent the "attribute discrimination", i.e., career and family (seven career related words and seven family related words). Participants are asked to respond quickly by pressing one key for items representing one concept and one attribute (e.g., men and career in the related condition), and another key for items from the other two categories (e.g., women and family). Participants then perform the task again with the key assignment for one of the pairs switched (so that women and career share a response, and men and family). The IAT measure derives from the differences in response latencies between these two tasks (before and after the key assignment switch).

The Bem Sex-Role Inventory (BSRI; Bem, 1974) consists of a list of 60 words or phrases, and participants are asked to rate the degree to which they believe each word describes them, using a 7-point Likert scale. Twenty trials represent desirable masculine traits (e.g., "Acts as a leader"), 20 desirable feminine traits (e.g., "Affectionate"), and 20 neutral traits. From the BSRI three indexes are obtained: Androgyny (BEM), Masculinity (BEM-M), and Femininity (BEM-F). Masculinity and Femininity are the mean scores from the masculine and the feminine items, respectively. The Androgyny score is the absolute value of the Student t test ratio between masculinity and femininity scores (scores close to 0 thus indicate an androgynous person).

The Ambivalent Sexism Inventory (ASI; Glick and Fiske, 1996) consists of 22 statements about men and women and their relationships in contemporary society. Participants rate their agreement with the statements on a 6-point scale. The ASI is organized into two subscales measuring the constructs of Hostile Sexism (HS; e.g., She usually tries to put him on a tight leash.) and Benevolent Sexism (BS; e.g., Men should be willing to sacrifice their own well being in order to provide financially for the women in their lives.). The ASI (and BS and HS) scores are the mean scores, across items, on the scales.

Stimuli

A set of 160 role nouns, including titles (e.g., *king*), states (e.g., *bachelor*), and occupations (e.g., *nurse*), was selected. The gender of half of the nouns was explicit and semantically defined (e.g., *mother*). In the other half, the gender was not explicit and could only be derived from the stereotype associated with the noun (e.g., *nurse*). The stereotypical gender of the nouns was taken from a previously collected database (Hamilton, 2006, unpublished data) in which people rated the role-nouns on an 11 point scale running from "strongly female" to "strongly male". Participants were instructed to base their ratings on how the world is and not how it ought to be. We selected the 80 most male/female biased stereotypical role-nouns (40 female, 40 male) from the norms: the average rating of the nouns selected as stereotypically female was 3.21 (ranging from 1.63 to 4.79) whereas stereotypically male nouns obtained an average rating of 9.24 (ranging from 7.29 to 10.56).

One set of 160 sentences (plus 80 fillers) containing a noun in subject position and a reflexive pronoun as object of the main verb was created. In contrast to Osterhout et al. (1997), where more than 50% of the sentences had adjectives or other pre-nominal modifiers, the role nouns in the present study were not modified and were always followed by the main verb, to make sure that additional information would not further bias the gender representation of the nouns. Sentences continued for a few words following the reflexive pronouns (average 3.4 words). Two experimental lists were created using a latin-square design

---

[1] http://psy.ck.sissa.it
[2] http://faculty.washington.edu/agg/iat_materials.htm

so that each participant was presented with each of the 160 role-nouns. Eighty sentences contained a definitionally male or female role-noun. In 40 of these sentences, the reflexive pronoun and subject agreed in number and gender, whereas in the other 40 sentences they disagreed. The other 80 target sentences contained a subject noun indicating a social role or occupation that was stereotypically male or female. The gender of the reflexive was consistent with the gender information provided by the role nouns in half of the sentences and inconsistent in the remainder (see **Table 1** for example sentences). Equal numbers (20) of male or female nouns were used in each condition. To keep the duration of the experiment below 75 min we restricted the number of filler sentences to 80, 40 of which were acceptable. Also to make the motivation of the experiment less obvious to our participants, 30 incorrect filler sentences contained pronoun-verb number agreement violations, instead of gender agreement anomalies. Ten semantic violations were then added to increase the variability in the materials. Hence, across all of the materials, 120 sentences were grammatically and semantically well formed and 120 were ill formed.

Procedure

Participants were tested individually in a dimly lit, sound attenuated room. They sat approximately 80 cm from a computer screen and were instructed to read the sentences carefully, as they would have to judge the acceptability of each sentence in terms of grammar and meaning. Each trial (presented in pseudo-randomized order) consisted of the following events: a fixation cross appeared at the center of the screen for 1000 ms, and was followed by word-by-word presentation of the sentence, with each word appearing for 350 ms at the center of the screen, followed by a 250 ms blank interval. Sentence final words were followed by a full stop. The acceptability question ("Was the sentence acceptable? Y or N") appeared after a 1000 ms blank, which followed the final word of each sentence. Participants responded by pressing one of two buttons corresponding to yes/no answers (half of the participants responded "Y" with the left hand; the other half responded "Y" with the right hand). The question remained on screen until a response was given, after which the next trial began. Words were presented in white 18-point Arial font against a black background. Throughout the trial, appropriate triggers were sent to the EEG system, through the parallel port, using Presentation software[3]. The EEG session lasted for about 1 h, and the overall experimental session (EEG set-up, EEG recording, washing, and collection of the additional measures) lasted 120 min on average.

---

[3]www.neurobs.com

EEG Recordings and Analysis

Electroencephalographic activity (EEG) was recorded from 35 Ag/AgCl electrodes (FP1, FP2, AF3, AF4, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T7, C3, CZ, C4, T8, CP5, CP1, CPZ, CP2, CP6, P7, P3, PZ, P4, P8, PO5, POZ, PO6, O1, OZ, O2) placed on the scalp using an elastic cap (Quik-Cap – Compumedics Neuroscan, Charlotte, NC, USA) following the Standard International 10–20 system. Vertical and horizontal eye movements were monitored with four electrodes, two placed beneath and above the left eye and two placed close to the left and right ocular canthi. Activity at the left and right mastoids (M1, M2) was also recorded. The EEG signal was referenced online to an electrode close to the vertex. Electrode impedance was kept below 5 kΩ at all scalp sites and mastoids, and below 15 kΩ for the eye electrodes. The EEG signal was amplified and digitized with a SynAmps2 amplifier (Compumedics Neuroscan, Charlotte, NC, USA) sampling at a rate of 250 Hz, and using a DC to 100 Hz low-pass filter during acquisition. The EEG signal was re-referenced offline to the linked mastoids, and band-pass filtered from 0.05 to 45 Hz (second order Butterworth filter). The signal was then segmented in epochs from −350 to 1100 ms around the presentations of pronouns. In this time interval, artifact rejection was carried out determining an allowed maximum voltage range of 100 μV in each epoch, and through the visual inspection of the remaining epochs. Epochs from −150 to 1100 ms relative to critical word onset were selected for ERP analysis. The artifact-free epochs were baseline corrected by subtracting the mean amplitude in the 150 ms pre-stimulus interval from the post stimulus activity. Data processing was carried out using the EEGLAB (Delorme and Makeig, 2004) and FieldTrip (Oostenveld et al., 2010) open-source toolboxes for MATLAB (MathWorks, Natick, MA, USA). Thirty-one participants were included in the analysis with an average epoch loss of 13.46%. The total rejection rate for these participants ranged from 4.37 to 28.12% of the epochs.

We performed statistical analyses[4] (using the R statistical package) in one time-window corresponding to the P600 canonical time-window, ranging from 500 to 900 ms. We used LMMs (lme4 package, Bates et al., 2015b) to account for the effects of within subjects factors and their interactions with the continuous covariates. LMMs lend themselves to ERP data (e.g., Bagiella et al., 2000; Newman et al., 2012) as they deal with non-sphericity, unbalanced experimental cells and, unlike ANCOVA, do not assume homogeneity of regression slopes across combinations of the independent variables. LMMs were used to predict the average ERP amplitude in the time window

---

[4]http://cran.r-project.org

TABLE 1 | Example of the experimental materials.

Type of noun	Agreement	Sentence	Condition
Definitional	Match	*The actress prepared herself to face the crowd.*	Definitional Match Condition
Definitional	Mismatch	*The actress prepared himself* to face the crowd.*	Definitional Mismatch Condition
Stereotypical	Match	*The architect saw himself in the mirror.*	Stereotypical Match Condition
Stereotypical	Mismatch	*The architect saw herself in the mirror.*	Stereotypical Mismatch Condition

of interest for each epoch recorded during the experiment, except those excluded by the artifact rejection procedure, and "outliers" lying outside a fixed threshold of minimum and maximum allowed amplitude ($+/-25\,\mu V$, 1.07% data loss) as the tails of the distribution departed from normality. Matrix size: 33 channels by 160 sentences by 31 participants.

Channels (except FP1 and FP2 as usuallly noisier than the rest of channels because placed close to the eye and front muscles) were organized by two topographic factors [Mediality: Left (all 12 left channels), Midline (all seven midline channels), Right (all 12 right channels); Longitude: Frontal (AF, F, and FC electrodes – 12 channels), Central (all C and CP electrodes – 10 channels), Parietal (all P, PO, and O electrodes – 11 channels)]. LMMs evaluated the effect of four within-subjects predictors (Agreement, Type of Noun, Longitude, and Mediality) and their interactions. Also the individual difference scores (after centering values on the mean of each covariate), and participants' sex (in interaction with the covariates), entered the model as fixed effects. To warrant the conservativeness of the analysis we tested a model with maximal random structure as suggested by Barr et al. (2013). However, the high number of parameters (81) that the optimizer had to estimate determined a lack of convergence, which could be reached only when models had to estimate less than 25 parameters. Therefore, the number of factors in the random effects structure was determined on the grounds of feasibility (e.g., Bates et al., 2015a). The decision about which random slopes had to be included in the random structure was also constrained by feasibility (three levels factors – Longitude – easily increased the model complexity, compared to two levels factors) and by the fact that by subject and item random slopes for Agreement or Type of Noun should be included in the random structure to provide more conservative estimation of the factors that were manipulated. Since "random slopes for subjects pertain to properties of the words, and the random slopes for word pertain to properties of the subjects" (Baayen and Milin, 2010, p. 21) we further allowed by-item random slopes of two variables (Sex and BSRI-f). The reliability of the fixed effects was evaluated by model comparison using the LMERConfenienceFunctions package (Tremblay and Ransijn, 2015), as in Newman et al. (2012). In particular, a backfitting procedure was used, which compared models of decreasing complexity using log-likelihood ratio tests. The procedure removed terms in the model that did not make significant contribution to fit, to obtain a parsimonious model. To obtain a good compromise between computation time and conservativeness, we first backfitted the fixed effect structure on a simple random structure, and then we forward fitted the more complex random structure, including Sex and BSRI-f as the two individual factors that resulted the most significant fixed effects. Analysis of variance for each fixed effect is reported (F ratios between sum of squares of the model's terms and the model's residuals from the REML estimation), and lower-bound p values were calculated using the denominator degrees of freedom obtained by subtracting the number of estimated parameters from the number of data points, although the determination of the appropriate denominator degrees of freedom for such tests is at least problematic

(e.g., Baayen et al., 2008). Main effects of topographic factors or interactions not involving the experimental factors (e.g., Mediality X Longitude or Longitude X Sex) are not reported as they can be considered irrelevant. Deviance coding was used for all categorical factors.

RESULTS

Acceptability Judgments

Participants judged sentences as acceptable as follows: gender match and mismatch to semantically defined nouns, 92.10% ($SD = 8.41\%$) and 16.58% ($SD = 9.62\%$); gender match and mismatch to stereotypical gender nouns, 94.43% ($SD = 7.26\%$) and 89.28% ($SD = 12.78\%$). To evaluate the differences in acceptability judgments we used generalized mixed-models, using a binomial distribution. The model was specified as following: Agreement and Type of Noun were treated as fixed effects, whereas the random structure was maximally specified with by-subjects random intercepts and random slopes for Agreement by Type of noun and by-item random intercepts and random slopes for Agreement only, because the manipulation of type of noun was between-items. Reliable differences emerged between gender matching and mismatching pronouns in both Definitional ($\beta = -4.51, z = -18.73, p < 0.001$) and Stereotypical conditions ($\beta = -0.70, z = -3.17, p < 0.01$), although mismatching pronouns following stereotypical role nouns are far more acceptable than mismatching pronouns in the definitional condition ($\beta = 4.40, z = 15.23, p < 0.001$).

Individual Differences

In **Table 2** the correlations between predictors from the battery of tests (BSRI, BSRI-m, BSRI-f, ASI, ASI-h, ASI-b, IAT) are reported. High correlations emerged between the scores obtained in subscales and global scores, for different tests: BSRI was correlated with the associated BSRI-m [$r = -0.59, t(29) = -3.89, p < 0.001$] and BSRI-f [$r = 0.67, t(29) = 4.93, p < 0.001$] subscales; ASI was correlated with ASI Hostile [$r = 0.85, t(29) = 8.78, p < 0.001$] and ASI Benevolent [$r = 0.81, t(29) = 7.54, p < 0.001$]. These correlations reflect collinearity between the main indexes and the subscales from which they are derived and, therefore, only BSRI and ASI subscales were further tested as predictors. Interestingly, a strong negative

TABLE 2 | Correlations between the seven measures derived from the battery of tests investigating social perception of gender.

Measure	1	2	3	4	5	6	7
(1) IAT	—						
(2) BSRI	0.49*	—					
(3) BSRI-M	0.68***	−0.59**	—				
(4) BSRI-F	−0.03	0.67***	0.16	—			
(5) ASI	0.07	−0.06	0.09	0.05	—		
(6) ASI-H	0.04	−0.19	0.02	−0.21	0.85***	—	
(7) ASI-B	0.16	0.14	0.11	0.31^	0.81***	0.41*	—

Levels of significance are indicated by ^, <0.1; *, <0.05; **, <0.01; ***, <0.001.

TABLE 3 | Definitional condition. ANOVA table for the Event Related Potential (ERP) amplitude during the time window of interest (500–900 ms).

Factor	df	SumSq	MeanSq	F	dendf	pval	Sig
Agreement	1	558.11	558.11	10.79	70412	0.001	**
Agreement:Longitude	2	3162.96	1581.48	30.58	70412	0	***
Agreement:Mediality	2	1202.89	601.44	11.63	70412	0	***
Agreement:BSRI-f	1	1.03	1.03	0.02	70412	0.8878	
Agreement:BSRI-m	1	130.71	130.71	2.53	70412	0.1119	
Agreement:ASI-b	1	5.32	5.32	0.10	70412	0.7483	
Agreement:ASI-h	1	3.61	3.61	0.07	70412	0.7916	
Agreement:Sex	1	3.41	3.41	0.07	70412	0.7972	
Agreement:Longitude:Mediality	4	480.70	120.18	2.32	70412	0.0542	
Agreement:Longitude:BSRI-f	2	866.78	433.39	8.38	70412	0.0002	***
Agreement:Longitude:IAT	2	133.20	44.40	0.86	70412	0.4618	
Agreement:Longitude:ASI-b	2	1110.81	555.40	10.74	70412	0	***
Agreement:Longitude:ASI-h	2	218.81	109.40	2.12	70412	0.1206	
Agreement:Sex:BSRI-f	1	1.43	1.43	0.03	70412	0.8681	
Agreement:Sex:BSRI-m	1	157.33	157.33	3.04	70412	0.0812	
Agreement:Sex:IAT	1	725.70	725.70	14.03	70412	0.0002	***
Agreement:Mediality:IAT	2	286.68	143.34	2.77	70412	0.0626	
Agreement:Mediality:ASI-b	2	253.94	126.97	2.45	70412	0.0859	
Agreement:Mediality:ASI-h	2	320.01	160.01	3.09	70412	0.0453	^
Agreement:Longitude:Sex:BSRI-f	2	1597.59	798.80	15.44	70412	0	***
Agreement:Longitude:Sex:BSRI-m	2	500.81	125.20	2.42	70412	0.0461	^
Agreement:Mediality:Sex:BSRI-m	2	473.45	118.36	2.29	70412	0.0574	
Agreement:Mediality:Sex:IAT	2	286.98	143.49	2.77	70412	0.0624	

*Levels of significance are indicated by ^, <0.05; *, <0.025; **, <0.005; ***, <0.0005.*

correlation between IAT and BSRI-m emerged [$r = -0.68$, $t(29) = -5.00$, $p < 0.001$]. We also tested by means of Welch two-sample t-tests whether male and female participants obtained significantly different scores on each scale: marginally significant differences due to participants sex emerged for ASI Hostile [men scored an average of 2.15 vs. an average of 1.65 for women, $t(29.35) = 1.73$, $p < 0.1$], whereas for the remaining scales no differences due to participants' sex emerged [all ts < 1].

Event related Potentials

From a visual inspection of the grand averages (**Figure 1**), time-locked to the presentation of the reflexive pronoun, the effect of gender mismatch is evident in the Definitional condition. Its broad and posterior distribution, its timing (450 ms to the end of the epoch) and the polarity of the effect are compatible with a modulation of the P600 component. In the Stereotypical condition, the effect of mismatch is less clear: there seems to be a positive deflection in posterior and right lateralized electrodes in a narrower time window (500–750 ms) that is consistent with a P600 effect. Moreover, gender mismatching pronouns also elicit a negative deflection in frontal left electrodes which temporally overlaps to the parietal Positivity in the 500 to 900 ms time-window (see also **Figure 2**). Looking at the grand averages and the difference waves, we fitted models with the following contrasts on the topographic factors. Because of the left frontal negative deflection for stereotype mismatching pronouns we coded the Mediality factor using Left as the reference level for comparisons with the

Mediality and Right levels. The Parietal level of the Longitude factor was the reference for comparisons with Frontal and Central.

Statistical analysis carried out in the P600 time window (500–900 ms) confirmed that agreement affects pronoun processing in the Stereotypical and Definitional conditions to a different extent [Agreement × Type of Noun: $F(1,141489) = 175.62$, $p < 0.001$]. The mismatch effect is larger in the Definitional condition [$M = +0.80$ μV] compared to the Stereotypical condition [$M = -0.02$ μV]. The effect of Agreement on the ERPs is focussed on more posterior locations [Agreement × Longitude $F(2,141489) = 56.12$, $p < 0.001$], and this pattern is consistent with the canonical distribution of the P600 component as being larger in Parietal with respect to Frontal electrodes [$M_{Frontal} = -0.13$ μV vs. $M_{Parietal} = +0.80$ μV, $t = 10.15$] and Central electrodes [$M = +0.33$ μV vs. $M_{Central}$, $t = 3.34$]. However, results show also that the effect has an asymmetric distribution [Agreement × Longitude × Mediality $F(4,141489) = 3.59$, $p < 0.01$]. This complex interaction (**Figure 3**; **Table 3**) reflects the fact that the effect was not different between levels of Mediality ($M_{LeftvsMidline} = -0.16$ μV, $M_{LeftvsRight} = -0.27$ μV, $M_{CentralvsRight} = +0.09$ μV) in Parietal electrodes, it was focused along the midline on central electrodes ($M_{LeftvsMidline} = -0.96$ μV, $M_{LeftvsRight} = -0.52$ μV, $M_{CentralvsRight} = +0.43$ μV), and was reduced in Frontal ($M_{LeftvsMidline} = -0.76$ μV, $M_{LeftvsRight} = -0.62$ μV, $M_{CentralvsRight} = +0.14$ μV) – and reversed in left Frontal – electrodes (**Figures 3A,B**).

FIGURE 1 | Grand Average Event Related Potential (ERP) response time-locked to the visual presentation of reflexive pronouns. The ERP recorded from 33 electrodes associated with Definitional Match condition (black solid line), Definitional Mismatch condition (red solid line), Stereotypical Match condition (black dashed line), Stereotypical Mismatch condition (green solid line) are displayed. Negative polarity is plotted upward.

Concerning the effect of participant sex and that of individual covariates, several significant three-way (nine) and four-way (eight, of which six involved participant Sex) interactions emerged. Also one five-way interaction was significant [Agreement × Noun × Longitude × Sex × BSRI-f: $F(2,141489) = 10.37, p < 0.001$]. It is worth noting that the Agreement × Noun Type interaction was further modulated by individual differences in BSRI-f [$F(1,141489) = 93.10, p < 0.001$], BSRI-m [$F(1,141489) = 48.51, p < 0.001$] and ASI-h [$F(1,141489) = 8.77, p < 0.01$] and also by an Agreement × Noun Type × Longitude × ASI-h interaction [$F(2,141489) = 15.53, p < 0.001$], but not by Sex [$F < 1$]: these interactions always have the same pattern representing a stronger modulation of the ERP effect by individual differences in the Stereotypical condition, than in the Definitional condition. To better describe this pattern of results we broke down the analysis by running two subsidiary models, on Definitional and Stereotypical role-nouns data, separately.

LMM Results on Defintional Role-Nouns

A main effect of Agreement [$F(1,70412) = 10.79, p < 0.01$] emerged. It was modulated by Longitude [$F(2,70412) = 30.58, p < 0.001$] and Mediality [$F(2,70412) = 11.63, p < 0.001$]. Planned contrasts confirmed the posterior distribution of the effect: differences between mismatching and matching pronouns in Parietal ($M = +1.23\,\mu V$) compared to Frontal ($M = +0.25\,\mu V$) electrodes were in fact consistent ($t = +7.02$), whereas the effect in Central electrodes ($M = +1.00\,\mu V$) was less pronounced with respect to that recorded in Parietal electrodes ($t = -3.19$). The interaction between Agreement and Mediality revealed reliable differences in the effect of mismatch between Left ($M = +0.51\,\mu V$) and Midline ($M = +1.11\,\mu V$) electrodes ($t = -4.52$), and less pronounced differences between Left and Right ($M = +0.87\,\mu V$) electrodes ($t = -4.04$), supporting the idea that the mismatch effect was more focused on Midline electrodes, and particularly reduced in Frontal and Left scalp sites (with

FIGURE 2 | Difference Waves (Mismatch minus Match) in Stereotypical and Definitional conditions. Difference waves were obtained by subtracting ERPs to pronouns that agreed with the gender of Stereotypical (green line) or Definitional (red line) role-nouns from those that disagreed. Negative polarity is plotted upward.
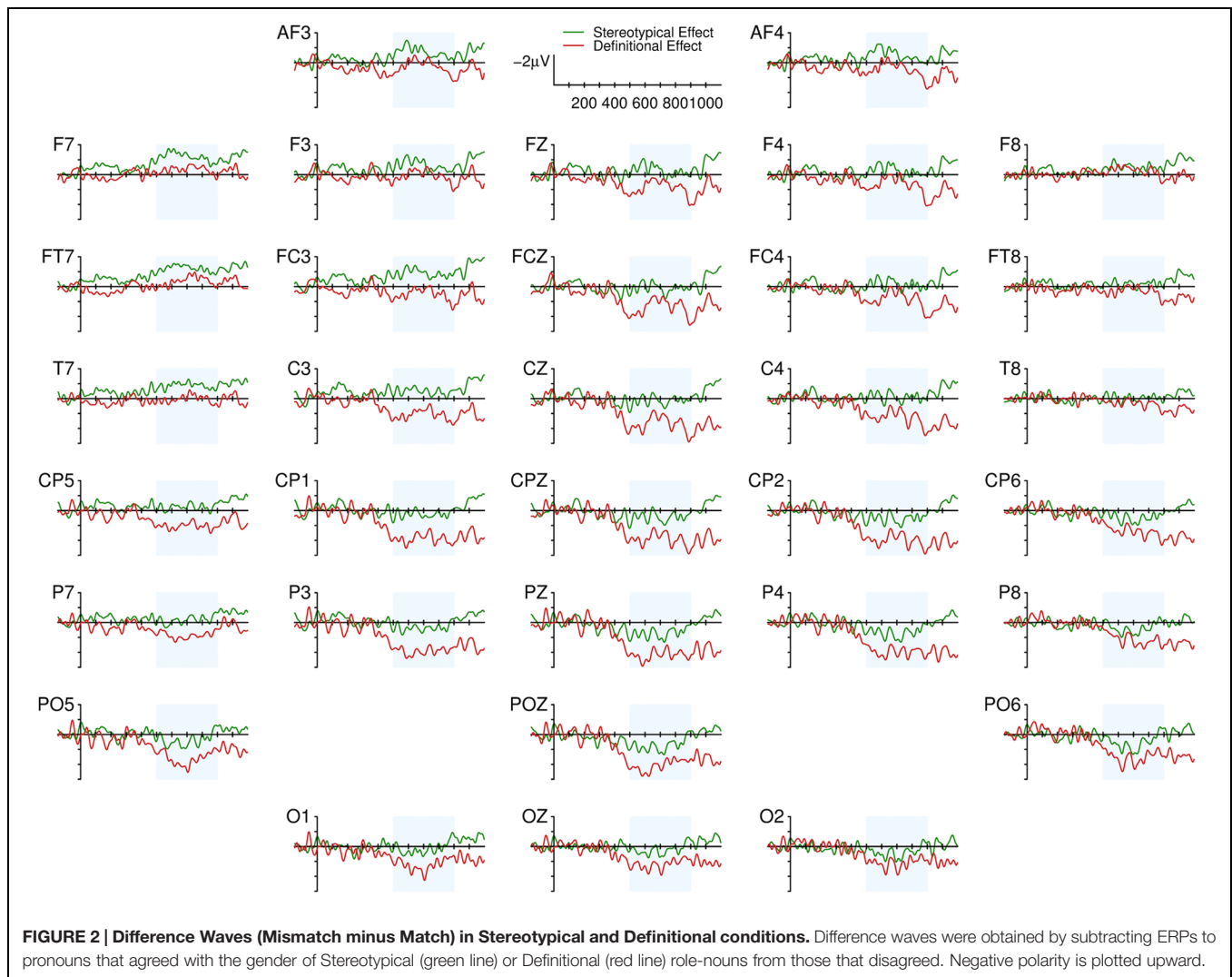
a marginally significant Agreement × Longitude × Mediality interaction).

Notably, when considering the effect of individual covariates different significant interactions emerged involving BSRI-f, ASI-b, Sex, and IAT (**Table 4**). ASI-b and BSRI-f were involved in similar interactions with Longitude and Agreement. Probably because of moderate collinearity [$r = 0.31$, $t(30) = 1.81$, $p < 0.1$] between these two measures, even though the F values for both interactions were large, the change in slope between Agreement conditions across levels of Longitude, did not consistently vary with ASI-b scores [Frontal vs. Parietal $\beta_{diff} = -0.13$, $t < 1$; Central vs. Parietal $\beta_{diff} = -0.03$, $t < 1$] but did so with BSRI-f scores [Frontal vs. Parietal $\beta_{diff} = +0.48$, $t = 3.81$; Central vs. Parietal $\beta_{diff} = +0.07$, $t < 1$]. Indeed, the effect of BSRI-f had a stronger impact in the EEG value as it was further qualified by the Agreement × Longitude × Sex × BSRI-f interaction which attested to differences between male and female participants in the BSRI-f modulation of the Agreement effect (**Figure 4**): such differences were strong comparing BSRI-f slope change associated with the Agreement effect between

Male and Female participants in Frontal – where women showed $\beta = +0.72$ and men $\beta = -0.45$ – vs. Parietal – where women showed $\beta = -0.38$ and men $\beta = -0.25$ – electrodes. Such differences were reliable in the comparison between Frontal and Parietal electrodes [$\beta = +1.36$, $t = 4.87$] but not in the comparison between Central and Parietal electrodes [$\beta = +0.39$, $t = 1.34$]. The ERP pattern as modulated by BSRI-f was thus similar for Male and Female participants in Central and Parietal electrodes (as also showed by the Agreement by Longitude by BSRI-f significant interaction). However, in Frontal electrodes the pattern was inversed. Female participants showed larger Frontal Positivity associated with an increase in BSRI-f scores, whereas Male participants showed a reduction of the Frontal portion of the P600 associated with an increase in BSRI-f scores. Moreover, participants Sex was involved in a Agreement × Sex × IAT interaction (**Figure 5**): the size of the Mismatch effect (across all scalp-sites) increased as function of IAT score ($\beta = +2.23$) for male participants and decreased ($\beta = -1.21$) for female participants ($\beta_{diff} = +3.44$, $t = 3.70$).

FIGURE 3 | Scalpmaps representing the distribution of agreement effect (Mismatch minus Match) in Stereotypical and Definitional conditions. The ERP differences (from 500 to 900 ms) of two types of contrasts is shown: **(A,B)** represent Stereotypical and Definitional mismatch effects, respectively. **(C,D)** represent the distribution of the effect obtained by subtracting the ERPs to reflexives in sentences that were rated as "acceptable" from those in sentences rated as "unacceptable", in the Stereotypical and Definitional conditions, respectively. The bottom scalp-map **(E)** represents ERP differences between Stereotypical gender mismatching and Stereotypical gender matching sentences for all sentences that were rated as "acceptable".

LMM Results on Stereotypical Role-Nouns

The effect of Agreement was modulated by Longitude $[F(2,70038) = 25.86, p < 0.001]$ and by Mediality $[F(2,70038) = 13.38, p < 0.001]$. The mismatch effect differed $[+0.92 \ \mu V, t = 7.20]$ between Frontal $(M = -0.54 \ \mu V)$ and Parietal electrodes $(M = +0.38 \ \mu V)$ and also between Parietal and Central $(M = +0.05 \ \mu V)$ electrodes $[+0.44 \ \mu V, t = 3.19]$. Electrodes over the Left hemisphere $(M = -0.48 \ \mu V)$ showed a different gender mismatch effect from both Midline $(M = +0.17 \ \mu V, t = -4.95)$ and Right $(M = +0.09 \ \mu V, t = -4.05)$ lateralized electrodes. These results confirm that that the gender mismatch effect in the stereotypical condition is associated with a Frontal, and Left negativity overlapping with a Parietal positivity.

The analysis revealed two reliable interactions between Agreement and the individual covariates in the Stereotypical condition. One involved Agreement and BSRI-f (**Figure 6**) and was explained by more positive slopes of Mismatch compared to Match condition $(\beta = +0.80, t = 2.73)$ across scalp locations. The crossed slopes suggest that the overall null effect of Agreement is masked by the summation of negative and positive ERP responses to stereotype mismatch. Furthermore, the interaction between

Agreement, Longitude, and ASI-h (**Figure 7**) showed slope differences for the agreement effect across levels of longitude: comparing Frontal locations where the slope change was large and positive $(\beta = +0.68)$ to Parietal electrodes where this change was reduced and negative $(\beta = -0.17)$ revealed strong differences $(\beta\text{diff} = +0.86, t = 5.48)$ which also emerged in the comparison between Central $(\beta = +0.27)$, and Parietal electrodes $(\beta\text{diff} = +0.41, t = 2.76)$: less ASI-h scores were associated with a larger Frontal Negativity and larger Posterior Positivity, whereas more ASI-h participants showed a more positive Fronto-Central Positivity.

DISCUSSION

In the present experiment we investigated the ERP correlates of anaphor processing when the establishment of reference involves the evaluation of gender information. We presented participants with short sentences in which an antecedent was introduced and we recorded the ERPs to the presentation of reflexive pronouns occurring after the verb. We assumed that when a definitional role-noun (*mother*, *father*) is presented, readers access categorical information about the gender of the text character: if the form of the following pronoun is not consistent with the gender of the noun, no available referent can be found thus making the sentence unacceptable. After a stereotypically male or female character is introduced, readers also access information about the gender of the character and create a consistent representation of the discourse. However, upon reading the following pronoun it is possible to find a referent even when stereotype gender and pronoun gender are inconsistent: the counter-stereotypical referent might not be readily available, but because stereotype information is probabilistic and not categorical, it should be possible to search and find it. When nouns have definitional gender and the anaphor cannot be bound to the only available antecedent, a clear P600 effect was found. When a noun conveys gender information through the stereotypical representation associated with it, the ERP correlate of stereotype gender mismatch is biphasic, as showing a negative effect in Frontal Left electrodes and a positive effect in Parietal electrodes. Inspecting individual variability in the ERP response we showed how the biphasic pattern can be explained by the fact that grand averages reflect the summation of two different types of ERP responses: below, we argue that these effects reflect neural activity of Nref and P600 components. Different predictors (BSRI-f, ASI-h, IAT, and Sex) had effects on the ERP response. The different ERP response to gender mismatch in the Definitional and Stereotypical conditions suggest that gender information about stereotypes is not the same information conveyed by definitional gender role-nouns.

The positive part of the effects can be safely interpreted as part of the P600 component. The P600 effect to definitional gender mismatching pronouns confirms previous findings (Osterhout and Mobley, 1995; Osterhout et al., 1997; Hammer et al., 2008; and the literature on morphosyntactic Agreement, see Molinaro et al., 2011). Consistently with grammaticality judgments that fall

TABLE 4 | Sterotypical condition. ANOVA table for the Event Related Potential (ERP) amplitude during the time window of interest (500–900 ms).

Factor	df	SumSq	MeanSq	F	dendf	pval	Sig
Agreement	1	15.36	15.36	0.30	71038	0.5859	
Agreement:Longitude	2	2678.87	1339.44	25.87	71038	0	***
Agreement:Mediality	2	1385.03	692.51	13.38	71038	0	***
Agreement:BSRI-f	1	496.52	496.52	9.59	71038	0.002	**
Agreement:BSRI-m	1	6.62	6.62	0.13	71038	0.7206	
Agreement:ASI-b	1	9.68	9.68	0.19	71038	0.6655	
Agreement:ASI-h	1	40.69	40.69	0.79	71038	0.3753	
Agreement:Sex	1	13.51	13.51	0.26	71038	0.6094	
Agreement:Longitude:Mediality	4	289.66	72.41	1.40	71038	0.2315	
Agreement:Longitude:BSRI-f	2	120.76	60.38	1.17	71038	0.3116	
Agreement:Longitude:IAT	2	178.64	59.55	1.15	71038	0.3272	
Agreement:Longitude:ASI-b	2	15.71	7.85	0.15	71038	0.8592	
Agreement:Longitude:ASI-h	2	1590.88	795.44	15.36	71038	0	***
Agreement:Sex:BSRI-f	1	95.22	95.22	1.84	71038	0.1751	
Agreement:Sex:BSRI-m	1	0.26	0.26	0.01	71038	0.9436	
Agreement:Sex:IAT	1	32.78	32.78	0.63	71038	0.4262	
Agreement:Mediality:IAT	2	376.47	188.24	3.64	71038	0.0264	^
Agreement:Mediality:ASI-b	2	6.25	3.12	0.06	71038	0.9414	
Agreement:Mediality:ASI-h	2	112.48	56.24	1.09	71038	0.3375	
Agreement:Longitude:Sex:BSRI-f	2	84.47	42.24	0.82	71038	0.4423	
Agreement:Longitude:Sex:BSRI-m	2	114.78	28.69	0.55	71038	0.6959	
Agreement:Mediality:Sex:BSRI-m	2	389.27	97.32	1.88	71038	0.1109	
Agreement:Mediality:Sex:IAT	2	93.47	46.73	0.90	71038	0.4055	

*Levels of significance are indicated by ^, <0.05; *, <0.025; **, <0.005; ***, <0.0005.*

very low (16%) for definitional gender mismatching pronouns, participants fail to find an appropriate referent for the pronoun. In contrast to what might happen in the processing of free or unbound pronouns (e.g., *he, she*) a reflexive pronoun cannot link to a yet unmentioned antecedent, and therefore the observed P600 effect can be taken to reflect a genuine "failure" in linking anaphor and antecedent (consistent with Osterhout and Mobley, 1995).

The biphasic pattern associated with Stereotypical gender mismatching pronouns replicates previous findings only partially: the positive part of the effect is consistent with Osterhout et al. (1997), whereas the overlapping negativity is not. The interpretation of the Left Anterior effect that is visible in the Grand Averages (**Figures 1** and **2**) elicited by stereotype gender mismatching pronouns is not straightforward since different language related ERP components, reflecting the activity of different neural mechanisms, have been described as occurring in Frontal or Left Frontal locations: the focal/morphosyntactic LAN (e.g., Friederici, 2002, 2011), the sustained LAN (e.g., King and Kutas, 1995; Fiebach et al., 2002), and the Nref effect (e.g., Van Berkum et al., 1999). The effect observed here is compatible with modulation of either type of LAN component, but only because the polarity of the effect is negative, and the distribution of the effect is left anterior when looking at the grand averages. In contrast, the timing of the effect and the functional interpretation of the focal LAN do not fit with the effect we observed and the hypothesized undergoing cognitive processes. Focal LANs are usually seen between 300 and 500 ms (i.e., they are not sustained), whereas our effect was

sustained (it onsets before 500 ms and is lasts until 900 ms). But the strongest reason to believe that the observed effect is not a LAN effect is that current accounts of the functional meaning of the LAN component (e.g., Friederici, 2011; Molinaro et al., 2011) proposed that LAN should be observed when morphological cues of target and controller in the agreement process are both transparent, and conflicting. In our study, the pronoun's form provides a transparent morphological gender cue, but for the vast majority of the sentences no gender related morphological cues are provided by English nouns (although a few Definitional nouns, such as *actress* or *mistress* convey morphologically transparent – and female – cues). Therefore, one crucial condition for eliciting "morphosyntactic" LAN effects is not met. And even if we assumed that LAN is triggered by disconfirmed syntactic predictions (as is also proposed in Molinaro et al., 2011) it is clear that syntactically driven expectations should be much stronger when gender information is categorical. On this hypothesis, we should have observed a more negative LAN in the definitional condition. In fact, it was more pronounced in the stereotypical condition. Therefore the negative effect observed here cannot be interpreted as a focal LAN.

Alternatively, the effect may look more similar to the "sustained" LAN which has been found in the processing of long distance syntactic dependencies (e.g., King and Kutas, 1995; Fiebach et al., 2002) and has been associated with working memory costs for holding open gaps in the syntactic representation of the sentence. But this functional interpretation also does not fit, because if any gap has been opened (at the
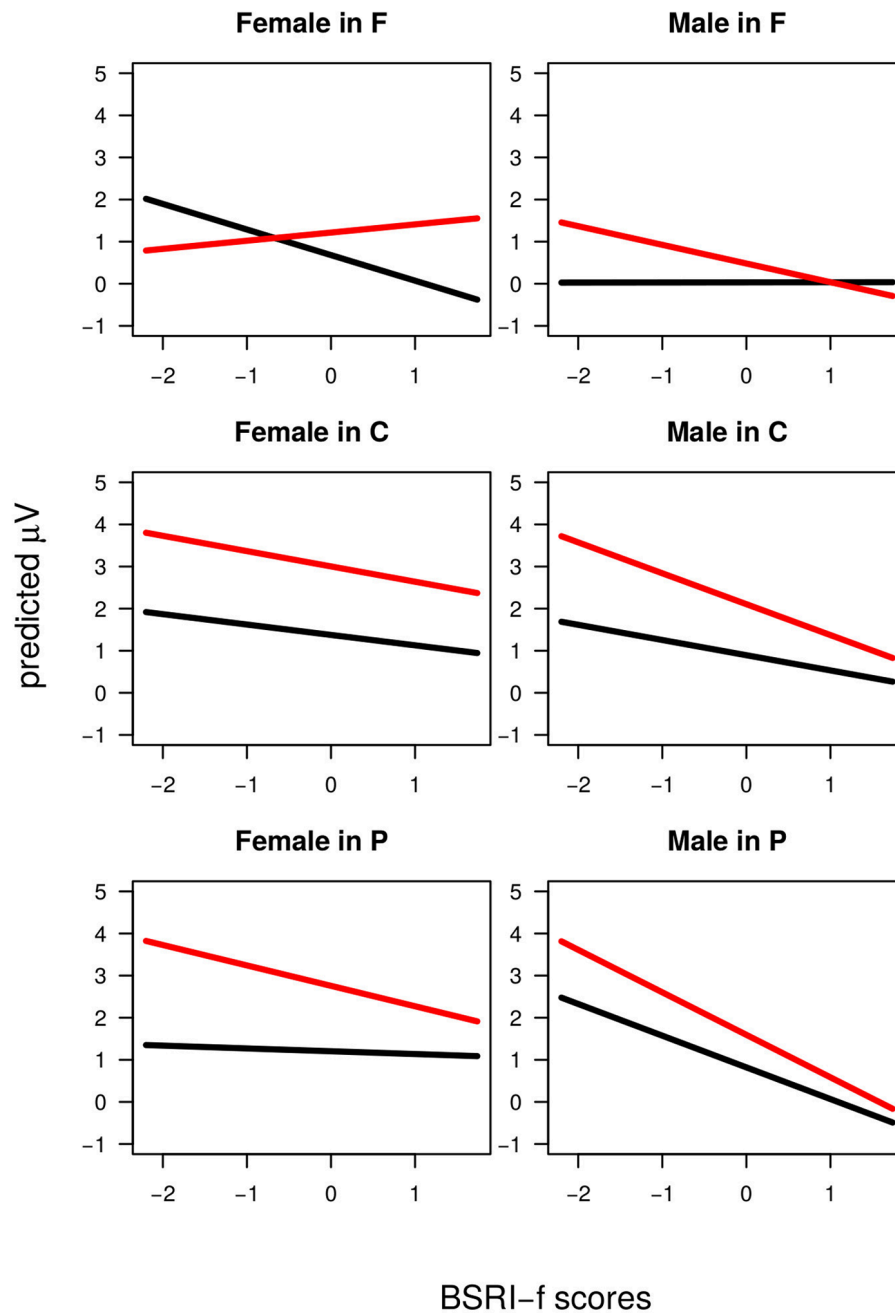
FIGURE 4 | Definitional condition. The Agreement × Longitude × Sex × BSRI-f interaction. The *x*-axis is BSRI-f scores, and the *y*-axis is amplitude. The solid black line is the Agreement Match condition and the solid red line is the Agreement Mismatch condition. Top row represents electrodes in Frontal scalp-locations, middle row represents Central scalp-locations and bottom row represents EEG in Parietal scalp-locations. In the left column results for female participants are displayed. In the right column results for male participants are displayed.

Verb) it should be "filled" when processing the reflexive. Our preferred interpretation is, therefore, that the Left Anterior effect is an Nref effect (Van Berkum et al., 1999, 2003, 2007; Nieuwland and Van Berkum, 2006; Nieuwland, 2014). Nref effects have been reported in cases in which two or more antecedents are equally plausible referents for an anaphor (Nieuwland and Van Berkum, 2006; Nieuwland et al., 2007), or when a mismatch occurs between the only available antecedent and an unbounded pronoun, that can be linked to an as yet unmentioned, unknown referent associated with the discourse (Nieuwland, 2014). Based on these findings, the Nref effect has been taken to reflect the search for additional information to link anaphor and antecedent. In the present experiment we manipulated the relation between anaphor and antecedent and it is likely that,

FIGURE 5 | Definitional condition. The Agreement × Sex × IAT interaction. The x-axis is IAT scores, and the y-axis is amplitude. The solid black line is the Agreement Match condition and the solid red line is the Agreement Mismatch condition. The top figure shows results for female participants from all scalp-locations. The bottom figure shows results for male participants from all scalp-locations.

FIGURE 6 | Stereotypical condition. The Agreement × BSRI-f interaction. The x-axis is BSRI-f scores, and the y-axis is amplitude. The solid gray line is the Agreement Match condition and the solid green line is the Agreement Mismatch condition.

when processing Stereotypical gender mismatching pronouns, participants might need to look for additional information to realize that antecedent and pronoun are coreferential, even though a *mechanic* is more often *male* than *female*. Stereotypical gender information is a probabilistic bias that guides the assignment of a male/female feature to a role-noun, but does not determine the antecedent gender categorically. Consistently with this idea, the acceptability ratings for stereotype gender mismatch passages are very high (89%) showing that (at least at the end of the sentence) pronouns and antecedent are judged as coreferential, although the corresponding sentences were still perceived as less well formed than stereotype matching sentences (94%). The distribution of the effect we observed may seem at odds with the canonical distribution of the Nref effect that tends to be bilateral, but a few examples of more left lateralized Nrefs have been reported (Experiments 1 and 2 in Nieuwland, 2014; Figure 1 in Nieuwland and Van Berkum, 2008). Moreover, although the Grand Averages show a frontal left distribution, the effects of the covariates (see below) often interacted with the agreement pattern and the Longitudinal rather than the Mediality dimension. Therefore we believe that the particular distribution
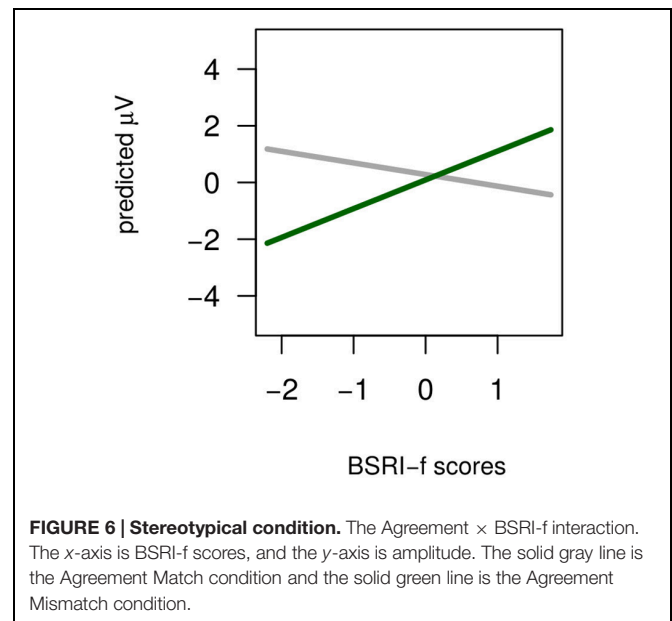
of the effect is due to the summation of two types of "late" ERP responses: a broad anterior Nref and a posterior P600.

Interesting insights derive from our investigation of individual differences. Differences between male and female participants were reported by Osterhout et al. (1997): gender violations (both stereotypical and definitional) elicited larger P600 responses for female subjects than for male subjects. We also found differences between Male and Female participants but they emerged only in relation to individual covariates and, somewhat unexpectedly, following Definitional rather than Stereotypical role nouns: an increase in IAT scores was associated with larger P600 effects to the processing of definitional gender mismatching pronouns for male participants only; furthermore, an increase in BSRI-f scores for male participants was related to smaller positive effects to definitional gender mismatch in both Frontal Central and Parietal electrodes, whereas female participants showed a similar pattern only in Parietal and Central electrodes, while in Frontal electrodes the effect was reversed, with larger Frontal positive effect for higher BSRI-f scores. The lack of strong asymmetries between Female and Male participants could be due to the use of the individual covariates that might have captured the EEG variance better then a dichotomous variable such as participants' sex. To review the effects of covariates that affected the EEG amplitude independently of participant sex, we notice that BSRI-f and ASI-h were the most relevant. BSRI-f appeared to modulate agreement following both Definitional and Stereotypical rolenouns: when role-nouns gender information was stereotypical, participants that described themselves as less feminine showed also a larger negative response to mismatching pronouns, but when the gender was semantically defined the low BSRI-f participants showed larger P600 effects in Central and Parietal electrodes; on the other hand, male participants with more "Feminine Traits" showed a reduced size of the P600 effect across all scalp sites but an increased Frontal P600 if participants were female.

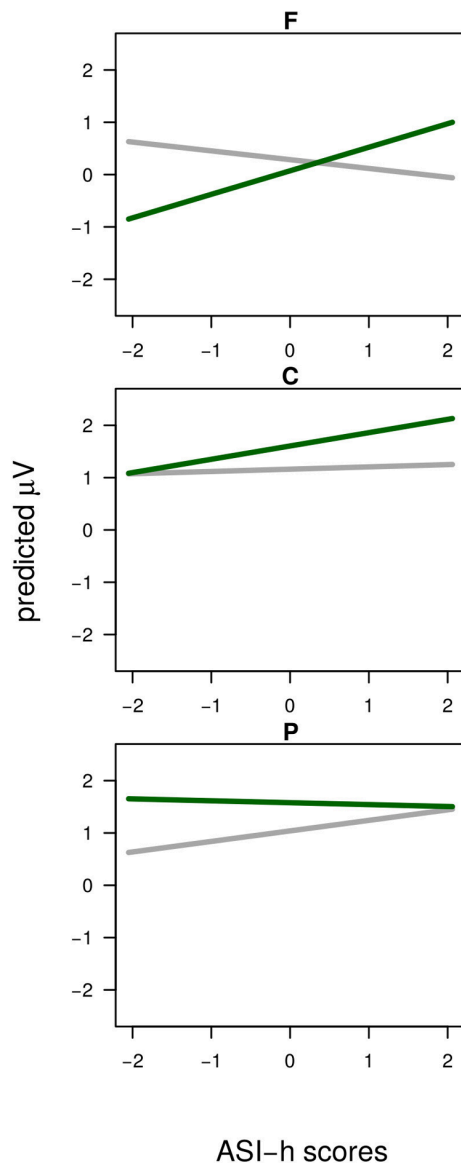**FIGURE 7 | Stereotypical condition.** The Agreement × Longitude × ASI-h interaction. The x-axis is ASI-h scores, and the y-axis is amplitude. The solid gray line is the Agreement Match condition and the solid green line is the Agreement Mismatch condition. Top row represents electrodes in Frontal scalp-locations, middle row represents Central scalp-locations and bottom row represents EEG in Parietal scalp-locations.

We believe that the finding that both BSRI-f and ASI-h were associated with the size of the Negative effect in Frontal electrodes (mainly with Stereotypical role nouns, but to some extent also with Definitional role nouns), and that these interactions did not involve differences across levels of Mediality, support the idea that the observed negativity is not strongly lateralized and thus the ERP pattern can be described as the temporal overlap of a frontally distributed Nref with the P600 effect in Parietal electrodes (that in the Grand Averages shows a more left-lateralized distribution).

Differences between the present study and that of Osterhout et al. (1997) may partly explain the differences between their results and ours. Was the British Brightonian sample more liberal than the American Seattle sample in 1997? Do differences in stereotype bias exist between countries (Misersky et al., 2013)? Is today's society less biased than 15 years ago? And if so, was it the efforts of governments that helped to reduce the gender gap? Clearly these questions cannot be easily answered from a psycholinguistic perspective, which instead suggests alternative hypotheses. One is that the linguistic materials were slightly more biased in Osterhout et al. (1997), because of the use of adjectives or other modifiers, which might have induced stronger commitment to probabilistic gender information, either because of further gender biasing in the modifiers themselves or because the presence of modifying information encouraged a more highly specified representation of the person. Another possibility is that because Osterhout et al. (1997) had lower spatial density in the EEG recording (13 electrodes in total), they might have missed the effect over frontal left electrodes revealing the biphasic pattern. In both Osterhout and Mobley (1995) and Osterhout et al. (1997) some hints of a frontal negativity can be seen by inspecting their figures. In the penultimate paragraph of their study, Osterhout et al. (1997, p. 282) acknowledge the unexpected nature of their findings: "Anomalies involving social categories that are not marked in the grammar (e.g., race) should not elicit the P600 effect but might elicit the N400 effect associated with semantic/pragmatic aspects of language". On the basis of the present results we believe that their idea that social categories should not elicit the same response as the response for grammatically encoded linguistic features was correct but the prediction of a N400 component effect was disconfirmed by their and the present study's results: stereotype gender mismatch did not elicit an N400 but rather a Nref effect as an index of inferencing about the most suitable referent of the discourse (Van Berkum, 2009).

If the mechanisms underlying the P600 in reflexive pronouns processing can be taken to reflect a failure to link the anaphor with the antecedent, when processing sentences with stereotypical gender role nouns, the P600 effect suggests that participants behave as if sometimes the link between gender inconsistent pronouns and antecedents cannot be established, whereas participants with lower scores in the BSRI-f or ASI-h that show an Nref effect suggest that less Feminine or less explicitly sexist participants may have actively searched for an appropriate although less likely antecedent. The modulation of the size of the Nref and P600 components may be linked to the strength of the stereotype bias that participants use to create the gendered representation of the text characters. The study of individual variation in the ERP response was fruitful because it allowed us to distinguish two ways in which co-reference can be evaluated when gender information is not categorical. One way is to use stereotype information as a categorical feature, perceiving the mismatch as an agreement violation (at least initially – by the end of each sentence most ratings turn out to be "acceptable"). The other way is to consider it as indicating a case of possible referential ambiguity, which requires additional processing effort to search for the possible although less likely

referent. We believe that the complex pattern of interactions between individual measures of sexism and the way the anaphoric relation is evaluated is an interesting finding, because it suggests that language processing depends on participants' characteristics that are unrelated to language competence. However it is not straightforward to explain the observed relation between personality traits and anaphor resolution. For instance, BSRI-f traits are termed "expressive" in the literature critiquing the BSRI (Payne, 1985; Choi and Fuqua, 2003). Participants who described themselves as not having the traits regarded as desirable when attributed to women in 1974 (less "Affectionate", less "Cheerful", less "Childlike", less "Compassionate", less "Does not use harsh language") appear to have been more actively engaged in trying to resolve the loose agreement between anaphor and antecedent when it involved stereotypical representations. Those who, on the other hand, had a more "expressive" self-representation, were either more sensitive to stereotypical information, or less prone to search for a counter-stereotypical representation of a role-noun.

Overall, the present study suggests that cognition can be better described when accounting for individual variation and, importantly, that variation in a linguistic task can be predicted also on the basis of personality factors that are largely independent of linguistic competence: likewise, researchers in the framework of cognitive psychology may benefit from investigating the effect of non-domain-specific factors that may not seem obviously relevant. The reported evidence is consistent with the view that language comprehension is influenced by the larger (non-linguistic) context of individuals' experience and personal beliefs, which likely plays a role in generating the mental representation of the text, of a communicative interaction, or more generally of the situation model.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Baayen, R. H., and Milin, P. (2010). Analyzing Reaction Times. *Int. J. Psychol. Res.* 3, 12–28.

Bagiella, E., Sloan, R. P., and Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology* 37, 13–20. doi: 10.1111/1469-8986.3710013

Banaji, M. R., and Hardin, C. D. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). *Parsimonious Mixed Models. arXiv Preprint arXiv:1506.04967*.

Bates, D. Maechler, M. Bolker, B., and Walker S. (2015b). *lme4: Linear Mixed-Effects Models Using Eigen and S4. R Package Version 1.1-9*. Available at: https://cran.r-project.org/web/packages/lme4/index.html

Bem, S. L. (1974). The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* 42, 155–162. doi: 10.1037/h0036215

Bosch, P. (1983). *Agreement and Anaphora: A Study of the Role of Pronouns in Syntax and Discourse*. London: Academic Press.

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055

Callahan, S. M. (2008). Processing anaphoric constructions: insights from electrophysiological studies. *J. Neurolinguistics* 21, 231–266. doi: 10.1016/j.jneuroling.2007.10.002

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., and Swaab, T. Y. (2007). Processing new and repeated names: effects of coreference on repetition priming with speech and fast RSVP. *Brain Res.* 1146, 172–184. doi: 10.1016/j.brainres.2006.07.033

Caramazza, A., Grober, E., Garvey, C., and Yates, J. (1977). Comprehension of anaphoric pronouns. *J. Verbal Learn. Verbal Behav.* 16, 601–609. doi: 10.1016/S0022-5371(77)80022-4

Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol. A* 49, 639–663. doi: 10.1080/713755647

Chang, F. R. (1980). Active memory processes in visual sentence comprehension: clause effects and pronominal reference. *Mem. Cogn.* 8, 58–64. doi: 10.3758/BF03197552

Choi, N., and Fuqua, D. R. (2003). The structure of the bem sex role inventory: a summary report of 23 validation studies. *Educ. Psychol. Meas.* 63, 872–887. doi: 10.1177/0013164403258235

Chomsky, N. (1993). *Lectures on Government and Binding: The Pisa Lectures*. Berlin: Walter de Gruyter.

Corbett, G. G. (1979). The agreement hierarchy. *J. Linguist.* 15, 203–224. doi: 10.1017/S0022226700016352

Corbett, G. G. (1991). *Gender. Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University press.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cogn.* 32, 551–559. doi: 10.3758/BF03195846

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *J. Mem. Lang.* 47, 250–272. doi: 10.1016/S0749-596X(02)00004-9

Filik, R., Leuthold, H., Moxey, L. M., and Sanford, A. J. (2011). Anaphoric reference to quantified antecedents: an event-related brain potential study. *Neuropsychologia* 49, 3786–3794. doi: 10.1016/j.neuropsychologia.2011.09.043

Filik, R., Sanford, A. J., and Leuthold, H. (2008). Processing pronouns without antecedents: evidence from event-related brain potentials. *J. Cogn. Neurosci.* 20, 1315–1326. doi: 10.1162/jocn.2008.20090

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 68–84. doi: 10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D., Steinhauer, K., Mecklinger, A., and Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biol. Psychol.* 47, 193–221. doi: 10.1016/S0301-0511(97)00033-1

Garnham, A. (2001). *Mental Models and the Interpretation of Anaphora.* Hove: Psychology Press.

Garnham, A., and Oakhill, J. (1985). On-line resolution of anaphoric pronouns: effects of inference making and verb semantics. *Br. J. Psychol.* 76, 385–393. doi: 10.1111/j.2044-8295.1985.tb01961.x

Garrod, S., and Sanford, A. J. (1994). "Resolving sentences in a discourse context: how discourse representation affects language understanding," in *Handbook of Psycholinguistics,* ed. M. Gernsbacher (New York, NY: Academic Press), 675–698.

Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *J. Pers. Soc. Psychol.* 70, 491–512. doi: 10.1037/0022-3514.70.3.491

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1498. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216. doi: 10.1037/0022-3514.85.2.197

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., and Garnham, A. (2008). Generically intended, but specifically interpreted: when beauticians, musicians, and mechanics are all men. *Lang. Cogn. Proc.* 23, 464–485. doi: 10.1080/01690960701702035

Hammer, A., Jansma, B. M., Lamers, M., and Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Res.* 1230, 177–191. doi: 10.1016/j.brainres.2008.07.004

Kaan, E., and Swaab, T. (2003). Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* 15, 98–110. doi: 10.1162/089892903321107855

Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Neurosci.* 12, 231–242. doi: 10.1038/nrn3000

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

King, J. W., and Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *J. Cogn. Neurosci.* 7, 376–395. doi: 10.1162/jocn.1995.7.3.376

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., et al. (2013). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behav. Res. Methods* 46, 841–871. doi: 10.3758/s13428-013-0409-z

Molinaro, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., and Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *J. Cogn. Neurosci.* 24, 1205–1223. doi: 10.1162/jocn_a_00143

Nieuwland, M. S. (2014). 'Who's he?' Event-related brain potentials and unbound pronouns. *J. Mem. Lang.* 76, 1–28. doi: 10.1016/j.jml.2014.06.002

Nieuwland, M. S., Otten, M., and Van Berkum, J. J. A. (2007). Who are You talking about? Tracking discourse-level referential processing with event-related brain potentials. *J. Cogn. Neurosci.* 19, 228–236. doi: 10.1162/jocn.2007.19.2.228

Nieuwland, M. S., and Van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: evidence from ERPs. *Brain Res.* 1118, 155–167. doi: 10.1016/j.brainres.2006.08.022

Nieuwland, M. S., and Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: evidence from ERPs. *Brain Lang.* 106, 119–131. doi: 10.1016/j.bandl.2008.05.001

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2010). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:9. doi: 10.1155/2011/156869

Osterhout, L., Bersick, M., and Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cogn.* 25, 273–285. doi: 10.3758/BF03211283

Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033

Pakulak, E., and Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *J. Cogn. Neurosci.* 22, 2728–2744. doi: 10.1162/jocn.2009.21393

Payne, B. R., Lee, C. L., and Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: evidence from single-word event-related brain potentials. *Psychophysilogy* 52, 1456–1469. doi: 10.1111/psyp.12515

Payne, F. D. (1985). "Review of the bem sex-role inventory," in *The Ninth Mental Measurements Yearbook,* ed. J. V. Mitchell Lincoln, NE: University of Nebraska Press, 178–179.

Reynolds, D. J., Garnham, A., and Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Q. J. Exp. Psychol. A* 59, 886–903. doi: 10.1080/02724980543000088

Sanford, A. J. (1985). *Cognition and Cognitive Psychology.* London: Weidenfeld and Nicolson.

Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE* 7:e48712. doi: 10.1371/journal.pone.0048712

Streb, J., Hennighausen, E., and Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *J. Psychol. Res.* 33, 175–201. doi: 10.1023/B:JOPR.0000027961.12577.d8

Streb, J., Rösler, F., and Hennighausen, E. (1999). Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain Lang.* 70, 273–286. doi: 10.1006/brln.1999.2177

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3

Tanner, D., Mclaughlin, J., Herschensohn, J., and Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Biling. Lang. Cogn.* 16, 367–382. doi: 10.1017/S1366728912000302

Tanner, D., and Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56, 289–301. doi: 10.1016/j.neuropsychologia.2014.02.002

Tremblay, A., and Ransijn, J. (2015). *LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models.* R Package Version 2.10. Available at: http://CRAN.R-project.org/package=LMERConvenience Functions

Van Berkum, J. J. A. (2009). "The neuropragmatics of 'simple' utterance comprehension: an ERP review," in *Semantics, and Pragmatics: From Experiment to Theory,* eds U. Sauerland and K. Yatsushiro (Basingstoke: Palgrave Macmillan), 276–316.

Van Berkum, J. J. A., Brown, C. M., and Hagoort, P. (1999). Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182. doi: 10.1006/jmla.1999.2641

Van Berkum, J. J. A., Brown, C. M., Hagoort, P., and Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology* 40, 235–248. doi: 10.1111/1469-8986.00025

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. doi: 10.1016/j.brainres.2006.06.091

Vos, S. H., and Friederici, A. D. (2003). Intersentential syntactic context effects on comprehension: the role of working memory. *Cogn. Brain Res.* 16, 111–122. doi: 10.1016/S0926-6410(02)00226-4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer Stefanie Nickels and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

When "He" Can Also Be "She": An ERP Study of Reflexive Pronoun Resolution in Written Mandarin Chinese

Jui-Ju Su [1]*[†], Nicola Molinaro [1, 2][†], Margaret Gillon-Dowens [3], Pei-Shu Tsai [4], Denise H. Wu [5, 6] and Manuel Carreiras [1, 2, 7]

[1] Basque Center on Cognition, Brain, and Language, San Sebastián, Spain, [2] Ikerbasque, Basque Foundation for Science, Bilbao, Spain, [3] Neuroscience of Language Laboratory, The University of Nottingham Ningbo China, Ningbo, China, [4] Graduate Institute of Translation and Interpretation, National Changhua University of Education, Changhua, Taiwan, [5] Institute of Cognitive Neuroscience, National Central University, Taoyuan, Taiwan, [6] Joint Research Center for Language and Human Complexity, The Chinese University of Hong Kong, Hong Kong, Hong Kong, [7] Department of Philology, University of Basque Country, San Sebastián, Spain

The gender information in written Chinese third person pronouns is not symmetrically encoded: the character for "he" (他, with semantic radical 人, meaning human) is used as a default referring to every individual, while the character for "she" (她, with semantic radical 女, meaning woman) indicates females only. This critical feature could result in different patterns of processing of gender information in text, but this is an issue that has seldom been addressed in psycholinguistics. In Chinese, the written forms of the reflexive pronouns are composed of a pronoun plus the reflexive "自己/self" (他自己/himself and 她自己/herself). The present study focuses on how such gender specificity interacts with the gender type of an antecedent, whether definitional (proper name) or stereotypical (stereotypical role noun) during reflexive pronoun resolution. In this event-related potential (ERP) study, gender congruity between a reflexive pronoun and its antecedent was studied by manipulating the gender type of antecedents and the gender specificity of reflexive pronouns (default: 他自己/himself vs. specific: 她自己/herself). Results included a P200 "attention related" congruity effect for 他自己/himself and a P600 "integration difficulty" congruity effect for 她自己/herself. Reflexive pronoun specificity independently affected the P200 and N400 components. These results highlight the role of 他自己/himself as a default applicable to both genders and indicate that only the processing of 她自己/herself supports a two-stage model for anaphor resolution. While both reflexive pronouns are evaluated at the bonding stage, the processing of the gender-specific reflexive pronoun is completed in the resolution stage.

Keywords: ERPs, reflexive pronoun resolution, type of gender information, gender specificity, Mandarin Chinese

INTRODUCTION

Anaphor resolution involves linking a given anaphor to a previously-mentioned antecedent in a sentence context, while interpreting both as related to the same discourse-level entity. Most studies related to the effects of gender information on anaphor resolution have proposed that (morpho)-syntactic rules constrain anaphor resolution (Carreiras et al., 1993, 1996; Garnham et al., 1995; Osterhout et al., 1997; Kennison and Trofe, 2003; Sturt, 2003; Duffy and Keir, 2004; Kreiner et al., 2009; Esaulova et al., 2014). From a theoretical perspective, anaphoric processing is considered to involve two stages of processing (Garrod and Sanford, 1994; Garrod and Terras, 2000). A bonding stage occurs first, for the purpose of searching for a best-fit referent among possible candidates (i.e., the antecedents) based on information related to gender, number, and syntactic rules. A resolution stage subsequently accomplishes the interpretation of the anaphor, taking into account world knowledge, and contextual information.

There is debate, however, about the degree to which lexical-semantic and syntactic cues are employed to resolve the anaphor. Carreiras et al. (1996) studied what linguistic information constrains pronoun resolution processing, by employing stereotypical role nouns as antecedents. They measured English and Spanish speakers' self-paced reading times and found cross-linguistic differences for the different morphological gender marking systems in English and Spanish. According to these authors, as soon as the mismatch of stereotypical gender is detected by the participants (i.e., on the pronouns encountered after stereotypical role nouns in English or on stereotypical role nouns at the beginning of a sentence in Spanish), it immediately influences processing, resulting in different patterns during pronoun resolution. Eye-tracking studies using stereotypical role nouns as antecedents (Sturt, 2003, in English; Esaulova et al., 2014, in German) have supported a two-stage model of anaphor resolution and these authors interpreted their findings as evidence for the syntactic constraints employed in resolving the link between an anaphor and its antecedent at the initial stage. Esaulova et al. (2014) concluded: "*anaphor resolution […] seems to depend above all on the rules of grammatical agreement in the context of overlapping gender cues*" (p. 798).

Osterhout et al. (1997) carried out an ERP study in English, in which the gender type of the antecedent (definitional vs. stereotypical) and the antecedent-reflexive pronoun gender congruity were manipulated (see also Kreiner et al., 2009 for an ERP study and Kreiner et al., 2008 for an eye-tracking study). While definitional role nouns had a definite gender (e.g., mother/father), gender in stereotypical role nouns was inferred based on world knowledge (i.e., a role noun that could refer to two genders but is biased toward one, e.g., electrician is male-biased and beautician is female-biased). The results showed similar P600 effects for antecedent-reflexive pronoun gender mismatches in conditions of both definitional (*mother—himself*) and stereotypical (*nurse—himself*) role nouns. These authors concluded that gender information was grammatically encoded even for stereotypical role nouns. Based on their interpretation,

it would follow that anaphor resolution correlates with a single ERP component, the P600, and engages a process that is syntactic in nature rather than semantic/pragmatic. However, Nieuwland and Van Berkum (2006) found an N400 effect for pronouns with antecedents of the same gender as compared to those with antecedents of different gender in a Dutch study and attributed the results to differences in the contextual bias that would modulate the N400 effect (used to index semantic/context related processing) during anaphor resolution. Results obtained from German studies are heterogeneous. Schmitt et al. (2002) investigated ERP responses to pronouns related to biological (definitional) and grammatical gender entities. The authors claimed that anaphor resolution is basically syntactically driven (P600 effect found) but can interact with semantic information in the N400 time interval. Irmen et al. (2010) focused on the link between antecedent stereotypical gender and anaphor lexical-semantic gender (these men/women/people) in German. They reported an N400 stereotypical gender effect and a P600 effect for anaphor mismatch with the antecedent's stereotypical gender. These authors interpreted their findings as supporting the two-stage model: while stereotypical gender information is collected in the bonding stage, the resolution stage represents integration, driven by either lexical-semantic mismatch, or syntactic violation on the anaphors.

Results from Chinese ERP studies also reveal somewhat different patterns. Qiu et al. (2012) manipulated the distance and gender congruity between antecedent and pronoun in Chinese sentences. N400 and P600 mismatch effects were found respectively for short and long distance manipulations. These authors claimed that the processing of gender information in Chinese pronoun resolution is more "semantics-based" when the pronoun is closer to the antecedent and this representation decays as the distance increases. Xu et al. (2013) (Experiment 1, singular antecedent) also manipulated gender congruity between antecedent and pronoun across clause boundaries (i.e., long-distance dependency). Only P600 effects were reported for gender mismatches. However, the authors did not interpret this P600 effect as reflecting purely syntactic processing, but either semantics-based processing -computation of the semantic relationship between antecedent and anaphor- or general integration difficulty resulting from conflicts on gender during Chinese anaphor resolution. It is important to note, however, that in these two Chinese studies, the gender specificity of the pronouns (i.e., 他/he and 她/she) was considered to be symmetrical, as is the case in morphological gender languages. However, co-reference processing between the anaphor and the antecedent could in fact differ in Chinese, due to the asymmetry of gender specificity encoded in the Chinese characters for these pronouns.

As can be seen from the research reported above, the processing of gender information has been investigated by employing the co-indexation structure of anaphor resolution mainly in languages in which the morpho-syntactic gender marking on anaphor[1] is "symmetrically" expressed, i.e.,

---

[1]The term anaphor here refers to all kinds of anaphors in general that could refer back to an antecedent in previous text.

"*he/himself*" is used specifically for male antecedents and "*she/herself*" exclusively refers to female antecedents. However, in a language without inflectional morphological gender markings, such as Chinese[2], the gender specificity is not symmetrically encoded in the written forms of the pronouns. In the spoken language, the third person singular pronoun is pronounced the same, /tā/, for both genders and the gender of a pronoun is inferred based on the context. Gender distinction is thus made only in written Chinese. Although the characters for the male and female pronouns share the same phonological component "也," they differ in their semantic radicals[3]. According to the web-based Dictionary of Chinese Character Variants[4] established by the Ministry of Education in Taiwan, the character 他/*he*, which contains the semantic radical "人" (/rén/*human*) is the "third person pronoun, refers to a third person" and the character 她/*she*, pronounced as /tā/, containing the semantic radical "女" (/nǔ/*woman*) is the "female third person pronoun[5]." When reading Chinese, therefore, during anaphor resolution it could be the case that the relevance of the gender information provided by an antecedent may differ depending on the extent to which gender information is specifically presented or not in the anaphoric pronoun. The processing of the anaphor may also differ when the antecedent's gender is either definitional, with a clear gender, or stereotypical, where gender can only be inferred. This critical feature (i.e., asymmetry of gender specificity in Chinese pronouns) has rarely been tested in previous Chinese studies and could result in distinct patterns of processing of gender information.

We illustrate below why the gender specificity encoded in the pronouns is not likely to be symmetrical, from three different perspectives: the historical background of the characters, the difference in their semantic radicals, and the usage of 他/*he* as a default. First, from a historical point of view, 她/*she* was only recently proposed as the third person singular female pronoun by a linguist, Liu in 1921 for the convenience of translation from western languages (Ling, 1989; Chang, 2007; Hua, 2012)[6]. Previous to this, the default 他/*he* was used in written Chinese.

The use of the character 她/*she* was not generally accepted at first and even now its appearance and necessity remains controversial (Moser, 1997; Chang, 2007; Wang, 2010; Hua, 2012). In recent years, due to campaigns for gender equality, the two pronouns have been gradually differentiated but the use of 她/*she* is still not compulsory for female antecedents (Peng, 2009).

Secondly, the semantic radicals encoded in the two pronouns play a critical role in bringing out the gender specificity. In Chinese, the radical 人/rén/ means *human* and the radical 女/nǔ/ means *woman* and these different semantic radicals make the characters orthographically distinct from each other. Increasing evidence has shown that Chinese speakers rely very much on sub-lexical units -semantic radicals and phonological components- during text comprehension (Perfetti and Zhang, 1991; Feldman and Siok, 1999; Ho et al., 2003; Liu et al., 2003; Ding et al., 2004; Lee et al., 2007; Hsu et al., 2009; Tsang and Chen, 2009). A study carried out by Cherng et al. (2009), which explored whether Chinese script reflects negative attitudes toward women (whether characters containing the semantic radical for "woman" have a more negative valence), found no evidence of this in Chinese speakers' perception of gender-based characters. They reported no negative attitudes toward characters containing the semantic radical for "woman" (the meaning conveyed by the characters was rated by participants as positive, negative or neutral). However, while characters containing the radical 子/zǐ/*son* and radical 女/nǔ/*woman* were rated as positive, characters with the radical 人/rén/*human* were rated as neutral. These results may imply differences in the mental representations of the gendered semantic radicals, especially when they appear in pronouns denoting different gender specificity.

Third, from an empirical point of view, the asymmetry on gender specificity presented by the pronouns is reflected in the usage of 他/*he* as a default. Wu and Liang (2008) analyzed 150 news items taken from the Academia Sinica Balanced Corpus (ASBC), for a rule-based corpus analysis of Chinese pronominal anaphor resolution. The results showed an error rate of 0.21 for gender mismatches between 他/*he* and female antecedents. The authors attributed this relatively high rate of mismatch to the use of 他/*he* as default in Chinese written text. Different learning sequences of the two pronouns at school also contribute to the tendency to use 他/*he* as a default. A textbook analysis carried out by Huang and Luh (2012) in Taiwan reported that while children learn 他/*he* in the first year of elementary school, 她/*she* is learnt in the second year. In some textbook articles, 他/*he* is used to refer to female antecedents before and even after the pronoun 她/*she* is learnt. Word frequencies of the pronouns reported by Academia Sinica (Word List with Accumulated Word Frequency in Sinica Corpus[7]) correspond to such usage trends in reality (see **Table 1**).

---

[2]According to Packard (2000) on Chinese word formation, the basic orthographic unit in Chinese written words is the character and generally one character represents one morpheme. From this point of view, Chinese has morphological gender markings, which are represented by individual characters/morphemes instead of inflections on a root. The term "morphological markings/cues," however, usually refers to inflectional morphology.

[3]Typically, a written Chinese character represents one morpheme and one syllable and is composed of two parts: a semantic radical indicating the meaning and a phonological component that provides information about the pronunciation.

[4]Dictionary of Chinese Character Variants (2000). Available online at: http://dict.variants.moe.edu.tw/eng.htm (Accessed April 30, 2015).

[5]The plural is formed by adding another character, 們/men/, which is a collective marker, to either of these pronouns to form 他們/tā men/*they* (male) or 她們/tā men/*they*(female) (Li and Shi, 2000).

[6]The character 她/*she* was recorded as first appearing in 玉篇 (Article of Jade, a manuscript written in the Yuan Dynasty), but the pronunciation and meaning is identical to 姐/jiě/ meaning elder sister or ladies (Ling, 1989; Hua, 2012; Dictionary of Chinese character variants established by the Ministry of Education in Taiwan). The Yuan Dynasty was from 1271 A.D.-1368 A.D.

[7]Word List with Accumulated Word Frequency in Sinica Corpus (2005). Available online at: http://elearning.ling.sinica.edu.tw/eng_teaching_index.html (Accessed April 30, 2015).

TABLE 1 | Word frequencies for the two Chinese pronouns extracted from Word List with cumulated Word Frequency in Sinica Corpus, Academia Sinica, Taiwan.

	他/*he*	她/*she*
Cumulative Word Frequency for Modern Chinese words (based on the corpus size of 5 million words)	29,938	10,755
Cumulative Word Frequency for Pre-modern Chinese Corpus	37,259	2
Cumulative Word Frequency for Old Chinese Corpus	36	Word not found

Data from these three perspectives thus clearly indicate that gender specificity in the characters for Chinese third person pronouns is non-symmetrical. Investigating this asymmetry can shed light not only on the processing of gender information in written Chinese but also on the general processing of pronouns in text.

In the present study, we investigate how the asymmetry of gender specificity interacts with antecedent noun type in which the biological gender is differently inferred during anaphor resolution in written Chinese. The experiment had a 2 × 2 × 2 design, with three factors manipulated: antecedent gender type (definitional vs. stereotypical), reflexive pronoun gender specificity (default vs. specific), and gender congruity between the reflexive pronoun and antecedent (congruent vs. incongruent).

First, antecedent gender type was manipulated. For definitional gender, because most female definitional role nouns in Chinese carry a 女/nǚ/*woman* radical in the same position as 她/*she*, (such as 媽/mā/*mother*, 嬸/shěn/*aunt*, and 姐/jiě/*elder sister*, Tang, 1988), proper names were used, to limit priming effects due to the presence of the same semantic radical in antecedent and reflexive pronouns (see Feldman and Siok, 1999; Ding et al., 2004). So, proper names such as 達賴喇嘛/*Dalai Lama* or 呂秀蓮/*Annette Lu* were used as antecedents for definitional gender. Stereotypical role nouns (e.g., 拳擊手/quán jī shǒu/*boxer*, male-biased; or 啦啦隊長/lâ lâ duì zhǎng/*cheerleader*, female-biased) were used as antecedents for the stereotypical gender condition.

Second, gender specificity (default vs. specific) was manipulated on reflexive pronouns instead of personal pronouns. In Chinese, the use of pronouns is not constrained by the local structure (Principle B, Chomsky, 1981: pronouns cannot co-refer with antecedents in a local clause). A reflexive term, 自己/zì jǐ/*self* is allowed to appear after a pronoun to make it a reflexive pronoun co-referential to the subject in the same clause (Principle A, Chomsky, 1981), (see Li and Thompson, 1981). For instance, in (1a) the 她/*she* could refer to the teacher or another female. In (1b), the reflexive, 自己/zì jǐ/*self*, helps to make the pronoun unambiguously co-referential to the previously mentioned animate antecedent (i.e., Mary) in the same clause (i.e., local binding of 自己/zì jǐ/*self*, see Jäger et al., 2015). Therefore, to avoid any confusion in co-reference between an anaphor and its antecedent, the reflexive 自己/zì jǐ/*self* was added after the third person pronouns to form the third person

reflexive pronouns (他自己/tā zì jǐ/ *himself*, default; and 她自己/tā zì jǐ/ *herself*, specific).

(1) a. 這個老師 覺得 Mary 不喜歡 *她*。
/zhè gè/ /lǎo shī/ /jiào dé/ /Mary/ /bú xi huān/ /tā/
(This teacher thinks that Mary doesn't like **her**.)

b. 這個老師 覺得 Mary 不喜歡 *她自己*。
/zhè gè/ /lǎo shī/ /jiào dé/ /Mary/ /bú xi huān/ /tā zì jǐ/
(This teacher thinks that Mary doesn't like **herself**.)

Third, we manipulated gender congruity between a reflexive pronoun and its antecedent (congruent vs. incongruent). It should be noted that when 他自己/*himself* appears after a female/female-biased antecedent, the sentence might still be acceptable in Chinese because 他自己/*himself* can be used as a general term referring to both genders, even though we consider this as a mismatch in the data analysis. An ERP mismatch paradigm was employed.

ERPs (Event-Related Potentials) are scalp recordings of electrical brain activity time-locked to a stimulus event. Compared to other neuroimaging techniques, they offer very good temporal detail of brain activity. This makes the ERP technique one of the best measures for disentangling the temporally incremental neural processes typically assumed by cognitive models. In ERP studies related to anaphor resolution, three major correlates have been discussed (see Callahan, 2008): the LAN (Left Anterior Negativity), the N400, and the P600. The LAN is a negative-going wave mostly observed at left anterior scalp electrodes from around 250–500 ms post target word onset. The LAN is related to automatic parsing involving morphosyntax and rule-based decomposition processes (Hahne and Friederici, 2002; Barber and Carreiras, 2005; Molinaro et al., 2008a,b, 2011; Hagoort, 2009). The N400 is a negative-going wave peaking around 400 ms after the onset of the target word, mainly recorded from electrodes in the centroparietal scalp regions (Kutas and Hillyard, 1980; Molinaro et al., 2010). The N400 is thought to represent context-dependent, lexico-semantic processing of a given stimulus. Its amplitude can be modulated depending on the lexical properties of single words and, at the sentence level, the anticipation/contextual semantic fit of a word with the previous context (Kutas and Federmeier, 2011, for a review). The P600 effect is a positive-going wave observed ∼500–700 ms after target word onset, with centroparietal scalp distribution. This component was initially reported as correlating with (morpho-) syntactic violations (Osterhout and Holcomb, 1992). Recently, however, the interpretation of the P600 effect has been extended

to reflect more general (also semantic) integration difficulties during sentence processing (Münte et al., 1998; Kaan et al., 2000; Kaan and Swaab, 2003; Carreiras et al., 2004; van Herten et al., 2005; Van de Meerendonk et al., 2009; Brouwer et al., 2012; Molinaro et al., 2012). In addition to these three components, the P200 component reflects a wide range of attention-related feature analysis including color, orientation and size of a feature (Luck and Hillyard, 1994). It is reported in studies related to word frequency (Dambacher et al., 2006), syllable frequency/structure (Barber et al., 2004; Carreiras et al., 2005), and Chinese character recognition (graphic, semantic, or phonological) (Liu et al., 2003; Lee et al., 2007; Hsu et al., 2009; Yum et al., 2014). Critically, the P200 component has recently been reported in Chinese discourse inference (Hung and Schumacher, 2012, 2014) and is considered to be related to a certain expectation driven by the context. It is reported as distributed mostly in the anterior region; the more positive the P200 amplitude, the less familiar, lower frequency and less expected the target word.

Based on previous ERP studies and the two-stage model of anaphor resolution (Garrod and Sanford, 1994), interactions among antecedent gender explicitness, reflexive pronoun gender specificity, and gender congruity are thus expected mainly in the N400 or P600 time intervals. Since the pronoun 他/*he* could be considered as a default pronoun (that can refer to both male and female antecedents) (Wu and Liang, 2008) and supported by the word frequencies of both pronouns in the Sinica Corpus, two hypotheses are possible about how the default pronoun is perceived for gender. First, if the 他/*he* is recognized as male-biased, a gender mismatch N400 or P600 effect is predicted for 他自己/*himself*. Second, if the 他/*he* is understood as equally applicable to both genders, no gender mismatch effects are expected for 他自己/*himself* following female or female-biased antecedents (e.g., 呂秀蓮/*Annette Lu* or 啦啦隊長/*cheerleader*). Whether or not the second hypothesis results in a mismatch effect is considered a key result in the present study, to evaluate participants' sensitivity to the gender asymmetry of the reflexive pronouns. On the other hand, stronger gender mismatch effects are expected for the female specific reflexive pronoun, 她自己/*herself*, because this contrast (male/male-biased antecedent vs. specific female reflexive) could lead to strong gender incongruity. This contrast is also critical in evaluating how the specific reflexive pronoun is processed depending on the gender type of the antecedent. Both N400 and P600 effects are expected in this "pure" mismatch gender contrast, reflecting semantic/pragmatic processing costs (Osterhout et al., 1997; Callahan, 2008; Tsai et al., 2009; Irmen et al., 2010; Molinaro et al., 2012; Qiu et al., 2012; Xu et al., 2013). For the relatively less familiar and lower frequency orthographic form (i.e., the specific reflexive pronoun 她自己/*herself*), the attention-related P200 effect (Luck and Hillyard, 1994; Liu et al., 2003; Hung and Schumacher, 2012, 2014; Lee et al., 2012) and frequency-related N400 effect (Kutas and Federmeier, 2011) are also expected. Specifically, we are interested in the time course of the ERP effects, to see when and how the linguistic

sources of gender information denoted by the reflexive pronouns interact with antecedent gender type and gender congruity during reflexive pronoun resolution.

MATERIALS AND METHODS

Participants

Forty native Chinese speakers (20 males, mean age: 21.8, aged 20–36 years) were recruited from the National Central University, Taiwan and were paid for their participation. They were healthy, right-handed, with normal or corrected-to-normal vision and reported no neurological or psychiatric history. The design and execution of the experiment conformed to the ethical regulations of the Institute of Cognitive Neuroscience at National Central University in Taiwan, which are equivalent to international standards. Informed consent was obtained from each participant.

Materials

Proper names of eighty celebrities (40 males) were selected from the news, based on Google search from July to September 2011. The names chosen had a number of occurrences larger than 300,000. Eighty sentences related to the 40 male and 40 female celebrities were used. Half of the 40 sentences containing the male/female proper names had a congruent third person reflexive pronoun and the other half had an incongruent one. These sentences were the experimental materials for the condition of antecedents with definitional gender.

The stereotypical role nouns were selected based on the results of a questionnaire containing 348 generic role nouns (e.g., 拳擊手/*quán jī shǒu*/*boxer*, 學生/*xué shēng*/*student*, or 啦啦隊長/*lā lā duì zhǎng*/*cheerleader*). Participants were asked to mark on an 11-point Likert scale from 0 to 100% with the percentage of 10 as the scale interval. The point of 0%-male/100%-female was always on the left and 100%-male/0%-female was always on the right. Fifty-seven college students (14 males, mean age: 21.58) from Tsing Hua University and Sun Yat-Sen University filled in this questionnaire. The forty most male-biased role nouns and 40 most female-biased ones were used as antecedents with stereotypical gender in 80 sentences. Half of the male-biased/female-biased antecedents were associated with a congruent third person reflexive pronoun and the other half with an incongruent one in sentences. The sentence structure for all materials was constructed as short-distance [i.e., the antecedent and the reflexive pronoun were in the same clause. See (2) and (3)]. At the beginning of each sentence, a segment denoting a time, location, or circumstance appeared. This was followed by a clause with S+V+O structure. The subject (i.e., antecedent) was presented by means of a proper name or stereotypical role noun. The object (i.e., anaphor) was the reflexive pronoun referring back to the subject. The target word was always the reflexive pronoun in each sentence, located in the fourth, fifth, or sixth position and never appeared AT THE END of the sentence.

(2) 記者會上，呂秀蓮 表示 她自己 會 繼續 支持 反核。
/jì zhě huì shàng/, /Annette Lu/ /biǎo shì/ /tā zì jǐ/ /huì/ /jì xù/ /zhī chí/ /fǎn hé/

(In the press conference, Annette Lu expressed **herself** about continuing to support the anti-nuclear movement.)

(3) 在訪談中，達賴喇嘛 表示 *他自己* 不會 繼續 擔任 西藏的 政教領袖。
/zài fǎng tán zhōng/, /Dalai Lama/ /biǎo shì/ /tā zí jǐ/ /bú huì/ /ji xù/ /dān rèn/ /xī cáng de/ /zhèng jiāo ling xiù/
(In the interview, the Dalai Lama expressed **himself** about not continuing to serve as the political religious leader in Tibet.)

An additional 80 filler sentences were created. The manipulations on critical words were focused on whether their semantic meaning could fit into the sentence or not (40 semantic match vs. 40 semantic mismatch). Data from these 80 sentences were not included in the data analysis. In total, 240 sentences were employed in this study. Thirty percent of the sentences were accompanied by comprehension yes/no questions (i.e., 72 questions) to evaluate participants' understanding of the sentences. The questions were related to the description of the main character and never related to any gender information of our interest. List 1 contained all the 240 sentences mentioned above. For counter-balancing purposes, a second list was created. List 2 contained the same 240 sentences as in List 1 but with all the target words presenting the opposite manipulation (see **Table 2** for examples of materials) (those items used were listed as Supplementary Material available online).

Procedure

All the stimuli were presented in white letters on a black background. Each trial began with a fixation point "+" at the center of the screen for 500 ms, followed by a blank screen for 400 ms. Each word was presented for 400 ms followed by a blank screen for 400 ms. According to some previous studies (Ye et al., 2007; Jiang and Zhou, 2009), 400 ms (word) + 400 ms (blank) word presentation is natural and comfortable for Chinese readers. A variable inter-trial time interval (from 1700 to 3000 ms) appeared after each sentence.

Participants were comfortably seated in a sound-attenuated cubicle and were instructed to read each sentence silently and carefully. Their task was to answer yes/no comprehension questions by pressing one of the pre-designated buttons ("J" for "Yes" and "F" for "No"). A practice session with 12 trials was conducted before the main experiment. The main experiment was arranged in six blocks with five breaks. Each block contained 40 sentences. The 240 sentences were randomly presented, differently for each participant.

Data Recoding and Analysis

Continuous EEG data (SynAmps2, NeuroScan) were acquired from 32 active electrodes mounted in a 66-channel Quick Cap. Electrodes were positioned according to the 10–20 system. The impedance was kept below 5Ω in each electrode. The sampling rate (A/D) was 500 Hz. The on-line reference electrode was set to be the left mastoid (M1) and we also recorded the right mastoid (M2). The signals were amplified with a bandpass of 0.05–100 Hz. The ground electrode was set between FPZ and FZ. HEOGs were placed at the outer canthi of the eyes and VEOGs were placed above and below the left eye in a bipolar montage.

The EEG raw data were re-referenced to the average activity of M1 and M2. The signal was bandpass filtered between 0.1 and 30 Hz. Epochs of interest were from −100 ms before the onset of the target word to 1000 ms after stimulus onset. Baseline correction was set from −100 ms to the onset of the target words. Trials with artifacts, such as eye blinks or saccades, or with activity exceeding ±120 μV were rejected. As a result, 5.6% of the trials were removed due to artifact rejection.

Except for trials with artifacts, all the correctly and incorrectly judged trials were included for statistical analysis. Data analysis focused on the mean voltage of each electrode within a time interval of interest after the onset of the target words in each participant. Based on the findings from previous ERP studies, four components, P200, LAN, N400, and P600 are used to index the processing correlates. The time intervals were chosen based on visual inspection of the averaged wave patterns.

Repeated measures ANOVAs were separately employed for electrodes in the midline region (anterior: average activity of Fz and FCz; central: Cz and CPz; and posterior: Pz and Oz) and in the lateral scalp (left anterior: average activity of Fp1, F3, F7, and FT7; left central: FC3, C3, CP3, and T7; left posterior: TP7, P3, P7, and O1; right anterior: Fp2, F4, F8, and FT8; right central: FC4, C5, CP4, and T8; and right posterior: TP8, P4, P8, and O2). A four-way repeated measure ANOVA was employed for the midline region considering *antecedent gender type* (definitional vs. stereotypical), *reflexive pronoun gender specificity* (default vs. specific), *gender congruity* (congruent vs. incongruent), plus the *latitude* topographical factor (anterior, central and posterior). For electrodes in the lateral scalp, a five-way repeated measures ANOVA was used: *latitude* and *lateral scalp* (left vs. right) were the topographical factors added to the three main factors. The Greenhouse-Geisser corrected *p*-value was used if the degree of freedom was larger than one. For interactions among the experimental factors and/or topographic factors, planned paired *t*-tests (with FDR adjusted *p*-value) were carried out mainly focused on the comparison of gender congruity respectively in each two levels of the main factors and/or separately in the topographic region to look for the location of the effect. Non-significant effects obtained from the planned paired *t*-tests following significant interactions are not reported in the data analysis.

RESULTS

Comprehension Questions

The average of participants' accuracy in the comprehension questions was 93%, ranging from 85 to 99%, showing that participants understood very well the sentences they read.

ERPs on the Reflexive Pronouns

Figure 1 reports the grand average of the ERPs elicited by the two reflexive pronouns, taking into account gender congruity. Based on visual inspection of the overall ERP results (and supporting evidence in the literature), the time interval for the analysis of the P200 component was set as 150–250 ms after the onset of stimuli, that for the N400 component was 250-600, and that for the P600 component was 600–800. The repeated measures ANOVAs on

TABLE 2 | Example sentences used in the experiment.

	Default reflexive pronoun, 他自己/*himself*	Specific reflexive pronoun, 她自己/*herself*
ANTECEDENT WITH DEFINITIONAL GENDER		
Gender congruent	在訪談中，達賴喇嘛 表示 *他自己* 不會 繼續 擔任 西藏的 政教領袖。(20 sentences) (*In the interview, the **Dalai Lama** expressed **himself** about not continuing to serve as the political religious leader in Tibet.*)	記者會上，呂秀蓮 表示 *她自己* 會繼續 支持 反核。(20 sentences) (*In the press conference, **Annette Lu** expressed **herself** about continuing to support the anti-nuclear movement.*)
Gender incongruent	*記者會上，呂秀蓮 表示 *他自己* 會繼續 支持 反核。 (20 sentences) (*In the press conference, **Annette Lu** expressed **himself** about continuing to support the anti-nuclear movement.*)	*在訪談中，達賴喇嘛 表示 *她自己* 不會 繼續 擔任 西藏的 政教領袖。(20 sentences) (*In the interview, the **Dalai Lama** expressed **herself** about not continuing to serve as the political religious leader in Tibet.*)
ANTECEDENT WITH STEREOTYPICAL GENDER		
Gender congruent	比賽前，那個 拳擊手 認為 *他自己* 很有實力 一定可以 贏得 金牌。(20 sentences) (*Before the race, **that boxer** considered **himself** very competent for winning the gold medal.*)	表演後，那個 啦啦隊長 評論 *她自己* 覺得 表現 不錯。(20 sentences) (*After the performance, **the cheerleader** commented on **herself** for doing not badly.*)
Gender incongruent	*表演後，那個 啦啦隊長 評論 *他自己* 覺得 表現 不錯。 (20 sentences) (*After the performance, **the cheerleader** commented on **himself** for doing not badly.*)	*比賽前，那個 拳擊手 認為 *她自己* 很有實力 一定可以 贏得 金牌。(20 sentences) (*Before the race, **that boxer** considered **herself** very competent for winning the gold medal.*)
Fillers	**Semantic match**	**Semantic mismatch**
	每天早上 我 搭 公車 去學校 上課。(40 sentences) (*Every morning, I take **the bus** to school.*)	*每天早上 我 搭 *手術* 去學校 上課。(40 sentences) (*Every morning, I take **the surgery** to school*)

The sentence which is marked with an asterisk (*) is taken as antecedent-reflexive pronoun gender incongruent or semantic anomaly during data analysis.

the earlier time intervals did not show any statistically reliable effects (i.e., baseline correction: −100 to 0 ms; 0 to 150 ms).

ERPs on the Reflexive Pronouns at the P200 Time Interval: 150–250 ms

The repeated measures ANOVA on the amplitude of the evoked activity for electrodes in the midline region showed an interaction among reflexive pronoun specificity, gender congruity, and latitude [$F_{(2, 78)} = 4.43$, $p = 0.031$]. The effect was found to be located in the mid-anterior region for 他自己/*himself* [congruent: 0.35 μV; incongruent: −0.33 μV; $t_{(39)} = 2.02$, $p = 0.050$] and no such effect was found for 她自己/*herself* (see **Figure 1**).

The repeated measures ANOVA for electrodes in the lateral scalp regions showed an interaction between *reflexive pronoun specificity* and *latitude* [$F_{(2, 78)} = 9.34$, $p = 0.002$] and an interaction among *reflexive pronoun specificity*, *gender congruity* and *latitude* [$F_{(2, 78)} = 6.18$, $p = 0.010$]. For the former interaction, the paired *t*-tests showed significant differences in the anterior region in which 她自己/*herself* was more positive than 他自己/*himself* [default: −0.49 μV; specific: −0.002 μV; $t_{(39)} = −2.75$, $p = 0.009$]. However, the second interaction did not reveal any relevant effects.

In this early time interval, a P200 *gender congruity* effect for the default reflexive pronoun (他自己/*himself*) in the mid-anterior region was observed (congruent > incongruent) (see
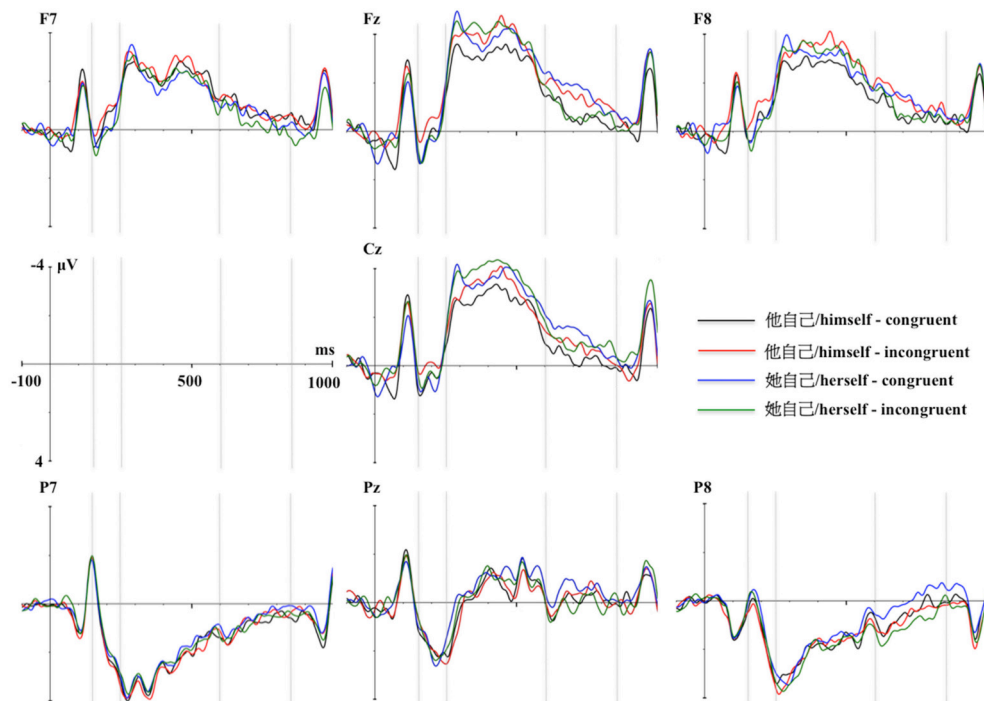
FIGURE 1 | Overall averaged brain activities of reflexive pronoun gender specificity (general vs. specific) by gender congruity (congruent vs. incongruent) in the representative electrodes.
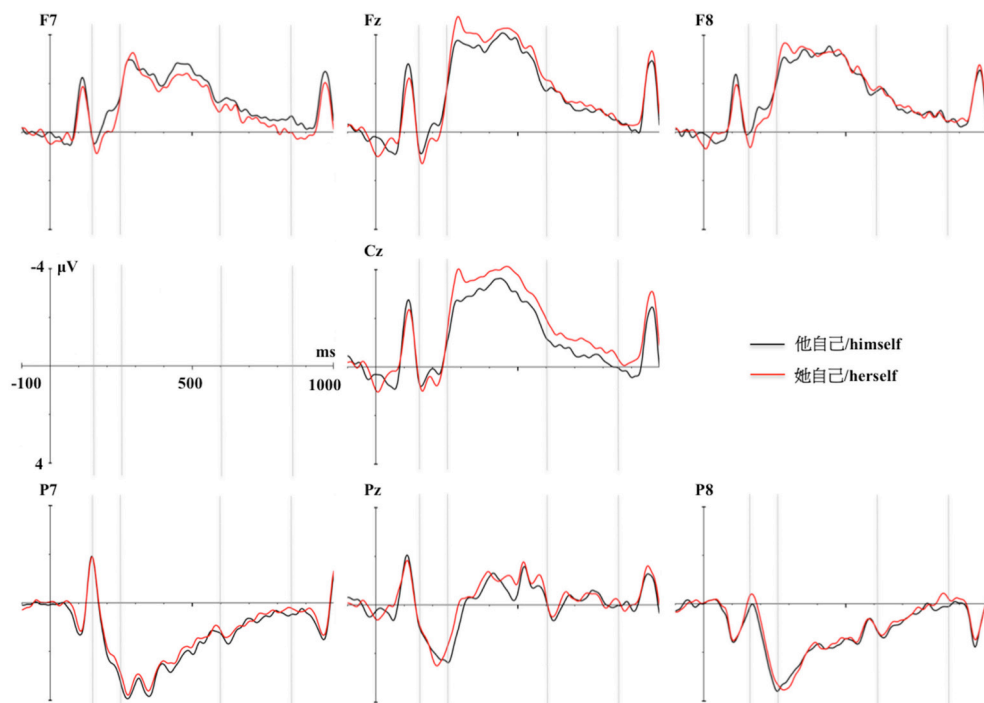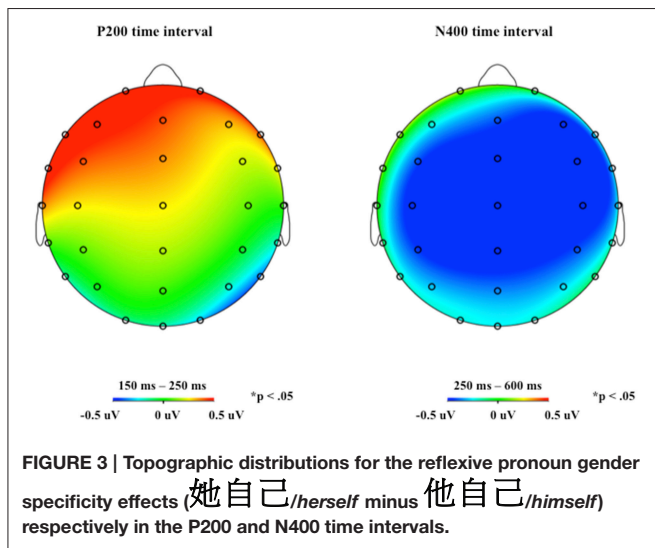

FIGURE 2 | The averaged brain activities separately presented by reflexive pronoun gender specificity (default: 他自己/himself vs. specific: 她自己/herself) in the representative electrodes.

P200 time interval

N400 time interval

150 ms – 250 ms   *p < .05
-0.5 uV   0 uV   0.5 uV

250 ms – 600 ms   *p < .05
-0.5 uV   0 uV   0.5 uV

FIGURE 3 | Topographic distributions for the reflexive pronoun gender specificity effects (她自己/*herself* minus 他自己/*himself*) respectively in the P200 and N400 time intervals.

Figures 1, **4**) and a P200 *reflexive pronoun gender specificity* effect (specific > default) emerged in the lateral anterior region (see **Figures 2**, **3**).

ERPs on the Reflexive Pronouns at the N400 Time Interval: 250–600 ms

The repeated measures ANOVA on the amplitude of the evoked activity showed the following pattern: For electrodes in the midline region, a main effect of *reflexive pronoun specificity* revealed that 她自己/*herself* was significantly more negative than 他自己/*himself* [default: $-1.56\,\mu V$; specific: $-2.04\,\mu V$; $F_{(1, 39)} = 7.33$, $p = 0.010$] and an interaction between *gender congruity* and *latitude* [$F_{(2, 78)} = 4.55$, $p = 0.028$] were found. However, this interaction did not reveal any relevant effects in the paired comparisons.

For the electrodes in the lateral scalp, the ANOVA showed two interactions. The first one showed that the *antecedent gender type* interacted with *latitude*, [$F_{(2, 78)} = 3.81$, $p = 0.035$]. However, the paired comparisons did not show any significant differences. The second one was an interaction among *reflexive pronoun specificity*, *latitude,* and *lateral scalp* [$F_{(2, 78)} = 4.19$, $p = 0.019$]. The paired *t*-tests showed that 她自己/*herself* was significantly more negative than 他自己/*himself* in the left central region [default: $-1.37\,\mu V$; specific: $-1.72\,\mu V$; $t_{(39)} = 2.18$, $p = 0.035$] and no relevant effects were found in other lateral regions.

Considering the whole experimental design, no interaction between the factors of interest emerged in the N400 time interval. Only an independent N400 *reflexive pronoun specificity* main effect was found in which the specific reflexive pronoun (她自己/*herself*) elicited more negative waveforms than the default (他自己/*himself*) (see **Figures 2**, **3**).

ERPs on the Reflexive Pronouns at the P600 Time Interval (600–800 ms)

The repeated measures ANOVA for electrodes in the midline region revealed three interactions. First, *antecedent gender type*

was found to interact with *gender congruity* [$F_{(1, 39)} = 4.53$, $p = 0.040$], but the paired *t*-tests did not show any significant effects. Second, the *reflexive pronoun specificity* interacted with *gender congruity* [$F_{(1, 39)} = 6.21$, $p = 0.017$]. The planned paired *t*-tests showed that the incongruent 她自己/*herself* elicited more positive amplitude as compared to the congruent one [congruent: $-1.02\,\mu V$; incongruent: $-0.38\,\mu V$; $t_{(39)} = -2.10$, $p = 0.042$], but no such difference was found for 他自己/*himself* [congruent: $-0.22\,\mu V$; incongruent: $-0.61\,\mu V$; $t_{(39)} = 1.26$, $p = 0.215$]. Third, an interaction among *reflexive pronoun specificity*, *gender congruity*, and *latitude* [$F_{(2, 78)} = 5.08$, $p = 0.015$] emerged. The paired *t*-tests showed only a significant difference for incongruent 她自己/*herself* as compared to the congruent 她自己/*herself* in the anterior region [congruent: $-1.77\,\mu V$; incongruent: $-0.85\,\mu V$; $t_{(39)} = -2.39$, $p = 0.022$].

The repeated measures ANOVA for electrodes in the lateral scalp showed two interactions. The first one showed that the *reflexive pronoun specificity* interacted with *gender congruity* [$F_{(1, 39)} = 8.84$, $p = 0.005$]. The paired *t*-tests showed that the incongruent 她自己/*herself* elicited more positive amplitude as compared to the congruent one [congruent: $-0.67\,\mu V$; incongruent: $-0.12\,\mu V$; $t_{(39)} = -2.56$, $p = 0.015$], but no such effects emerged for 他自己/*himself* [congruent: $-0.21\,\mu V$; incongruent: $-0.46\,\mu V$; $t_{(39)} = 1.31$, $p = 0.198$]. The second interaction was among *reflexive pronoun specificity*, *gender congruity,* and *lateral scalp* [$F_{(1, 39)} = 6.77$, $p = 0.013$]. The paired *t*-tests revealed that the incongruent 她自己/*herself* was more positive than the congruent 她自己/*herself* in the right hemisphere [congruent: $-0.86\,\mu V$; incongruent: $-0.07\,\mu V$; $t_{(39)} = -3.26$, $p = 0.002$].

The most relevant finding in this P600 time interval was the interaction between *reflexive pronoun gender specificity* and *gender congruity*. While the amplitude for incongruent 她自己/*herself* was more positive than that for congruent 她自己/*herself*, no such effects emerged for 他自己/*himself* (see **Figures 1**, **4**).

DISCUSSION

The present study evaluates how the asymmetry of gender specificity encoded in characters for Chinese reflexive pronouns influences the processing of gender information during reflexive pronoun resolution in Chinese text. The results showed: First, two independent effects for 她自己/*herself*; a P200 effect (她自己/*herself* is more positive than 他自己/*himself*) and a N400 effect N400 effect (她自己/*herself* is more negative than 他自己/*himself*), supporting the assumption of 他自己/*himself* as the default pronoun. Second, a P200 gender congruity effect for 他自己/*himself* (congruent is more positive than incongruent) and a P600 gender congruity effect for 她自己/*herself* (incongruent is more positive than congruent)
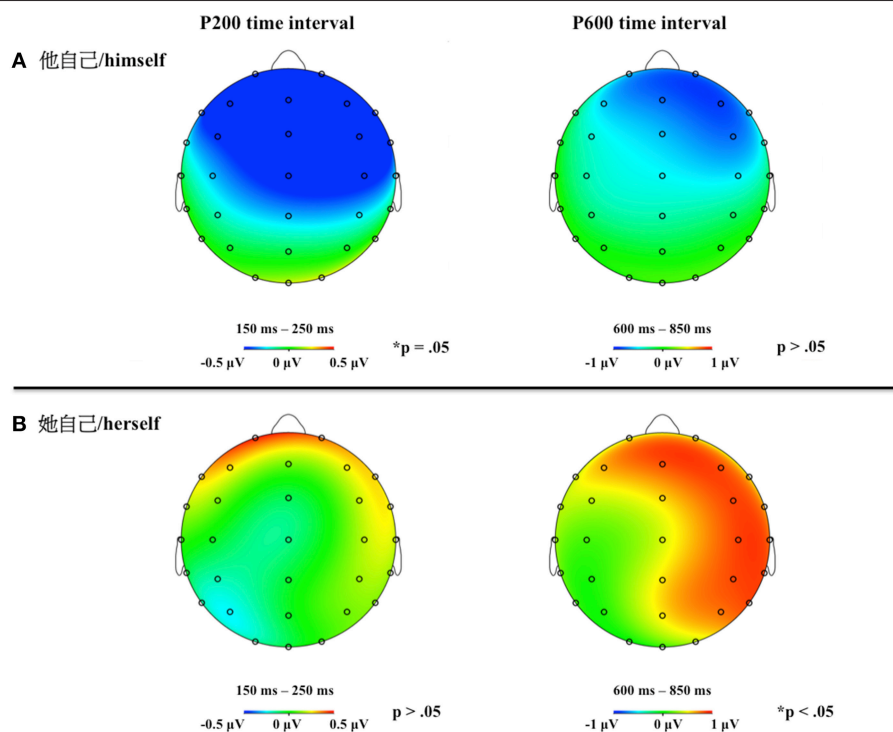
FIGURE 4 | Topographic distribution of the gender congruity effect (incongruent minus congruent) respectively for the two reflexive pronouns in the P200 and P600 time intervals. Panel **(A)** represents effects for the general reflexive pronoun, 他自己/*himself*, and panel **(B)** represents effects for the specific reflexive pronoun, 她自己/*herself*.

also emerged. The dissociation between 他自己/*himself* and 她自己/*herself* in the two time intervals provides relevant evidence of the asymmetry of gender specificity for anaphor resolution and suggests the distinct time courses involved in the processing of gender information during reflexive pronoun resolution. Most importantly, such dissociation further clarifies that the default pronoun 他/*he* is perceived as equally applicable to both genders instead of being a male-biased pronoun because no gender congruity effects emerged in the semantic-related N400 or the semantic/integration-related P600 time intervals.

The Processing of Chinese Third Person Reflexive Pronouns

As predicted, the less familiar and less frequent reflexive pronoun, 她自己/*herself* elicited more positive amplitude in the P200 time interval and more negative amplitude in the N400 time interval. The results thus reveal that the semantic radicals are processed at an early perceptual phase during Chinese orthographic recognition and lexical access is faster and easier for the default pronoun 他自己/*himself*, as compared to the specific 她自己/*herself*.

A straightforward explanation for the P200 and N400 specificity results could be that these arise because of frequency effects, due to the large difference in the word frequencies between the two characters (see **Table 1** and **Figure 2**)

(Dambacher et al., 2006; Kutas and Federmeier, 2011). According to Dambacher et al. (2006), the P200 could index processing differences resulting from word frequency comparisons (low frequency words elicit more positive amplitude) and the "*[…] lexical access was presumably completed for high-frequency words while low-frequency words were still being processed*" (p. 96). Meanwhile, N400 word frequency effects were reported for low frequency words as compared to high ones (Van Petten and Kutas, 1990; Dambacher et al., 2006), even in the context of word repetition in sentences (Van Petten et al., 1991). Although word frequency could explain the P200 and N400 effects reported here, the mechanism of how the two reflexive pronouns are processed and the dissociation of the two reflexive pronouns in the P200 and P600 time intervals are not clarified by this explanation, taking into account the co-reference between a reflexive pronoun and its antecedent within the structure of anaphor resolution.

Instead, the attention-related feature analysis (P200) and semantic expectation/predictability (N400) viewpoints may well explain the processing mechanism and the dissociation of the two reflexive pronouns during anaphor resolution. The P200 gender specificity effect could be interpreted as an attention-related mapping cost of the high similarity of graphic form in the two reflexive pronouns during Chinese character recognition (他自己/*himself* vs. 她自己/*herself*) (Luck and Hillyard, 1994; Liu et al., 2003). Liu et al. (2003) manipulated graphic similarity between a prime, 凉/liáng/*cool*, and target,

惊/jīng/startled, in a pronunciation task and reported a P200 effect for the change in semantic radical. Following the attention-related feature analysis interpretation of Luck and Hillyard (1994), the authors concluded that the P200 effect is related to orthographic and phonological processing. This may possibly be the case in the 他自己/himself-她自己/herself contrast at the word level. As a default, the graphic form of 他自己/himself may be more familiar to participants and so make it easier to recognize/retrieve. In contrast, when 她自己/herself is encountered, participants need more effort to process the relatively less familiar graphic as compared to the default. This effect could be interpreted as basically driven by the semantic radicals encoded in the two reflexive pronouns.

Hung and Schumacher (2012, 2014) reported similar P200 effects in studies comparing the topicality (the amplitude for novel-topic was more positive than topic-shift, that was more positive than topic-continuity) and topic-worthiness (new topic was more positive than given topic) effects in Chinese discourse processing. The authors interpreted the P200 effect as reflecting early perceptual processing costs during discourse inference, taking into account the topicality or topic-worthiness. According to Hung and Schumacher (2014), the P200 "[...] is likely to be a neural response to the involvement of selective attention that facilitates perceptual processing of an item that fulfilled contextually-induced expectation" (p. 43). Due to the fact that similar sentence structures were employed in the present experiment, the default reflexive pronoun may be considered as a given topicality expected after every antecedent mentioned in the previous sentence fragment. The 他自己/himself as a default may account for this context-induced expectation of discourse inference. It could facilitate the perceptual processing and result in a neural brain response to its antecedent because it already fulfills the expectation driven by the contextual information. The more positive amplitude observed for 她自己/herself would accordingly reflect the processing of an unexpected item detected by the brain during early perceptual processing and result in a P200 effect, thus in line with Hung and Schumacher (2012, 2014).

On the other hand, the fact that the N400 effect (她自己/herself is more negative than 他自己/himself) could reflect the differential semantic expectation/predictability between the default (referring to any human) and the specific (referring to females only) reflexive pronouns. Since 他自己/himself is applicable to every mentioned antecedent, less negative amplitude is expected, either due to its all-inclusive semantic meaning (Hagoort et al., 2004; Lau et al., 2008; Rabovsky and McRae, 2014) or its more accessible orthographic form (see Delong et al., 2005: the expected article "a" or "an" in English, cf. Kutas and Federmeier, 2011).

Anaphor Resolution in a Language without Inflectional Morphological Gender Markings

Taking into account the gender incongruity effects reported here, in terms of reflexive pronoun resolution, the present findings are consistent with Osterhout et al. (1997) in two respects. First, antecedent gender type (definitional as opposed to stereotypical gender) has no differential effect on the processing of gender information in anaphoric reflexive pronouns. Second, the processing cost for mismatches on gender emerges in the P600 time interval during reflexive pronoun resolution. According to Osterhout et al. (1997), although there is variability between definitional and stereotypical gender, as long as the gender information is activated (either male or female), the co-reference of this gender information during anaphor resolution should not vary by antecedent gender type. Thus, due to the syntactic constraints for definitional role nouns, and similar mismatch P600 effects found for both definitional and stereotypical role nouns, the authors concluded: gender information is " [...] encoded within grammar" (p. 282) and results in syntactic processing. Following Osterhout and colleagues, Kreiner et al. (2008, Experiment 1) further argued that the gender representation for such inference differed by noun types (i.e., definitional: lexical and stereotypical: pragmatic). Even so, Kreiner et al. (2008) reported no differences on the reflexive pronouns between sentences with definitional roles and those with stereotypical role nouns. Accordingly, our finding that no significant effects related to antecedent gender type emerged in any time interval very likely corresponds to what these authors described about the assignment of gender information regardless of antecedent noun type (definition or stereotypical) or how gender is represented by the antecedent (lexical or pragmatic). As long as the antecedent's gender is assigned, the activated gender is taken for the following anaphor resolution. However, in the present study, the P600 gender congruity effect found for 她自己/herself cannot be interpreted as reflecting syntactic processing, but is, rather, semantically-driven. As no morpho-syntactic gender agreement is required for Chinese anaphor resolution, the sentence remains grammatical when the character 他 is replaced by 她. Only the meaning-based semantic radicals can account for the co-reference difficulty between antecedent and anaphor. This result is in line with Xu et al.'s (2013) proposal that the P600 effect in Chinese anaphor resolution reflects an integration difficulty based on semantic anomaly instead of reflecting processing difficulties with syntactic violations. This semantics-based interpretation of the P600 effect has already been offered by studies that did not involve any extra syntactic manipulations in their experiments (Münte et al., 1998; Kaan et al., 2000; Kaan and Swaab, 2003; Kolk et al., 2003; Kim and Osterhout, 2005; van Herten et al., 2005; Callahan, 2008; Van de Meerendonk et al., 2009; Molinaro et al., 2012). Thus, while findings in morphological gender languages (e.g., Osterhout et al., 1997; Kreiner et al., 2008, 2009) have shown the importance of syntactic gender agreement between antecedents and pronouns, the case of Chinese seems to be different. Reflexive pronoun resolution in written Chinese needs to rely on the semantic information unequally encoded in the orthographic forms of the characters, given that one orthographic form codes for a generic gender while the other codes specifically for a feminine representation. The generic gender encoded by the default pronoun 他/he is confirmed

here to be a gender-neutral pronoun (appropriate to both genders) in line with the rating study that words with the radical 人/rén/*human* are rated as neutral (Cherng et al., 2009). Thus, mechanisms of co-reference of gender information between antecedents and reflexive pronouns in this study are modulated by the information carried by the semantic radicals denoting the different gender specificities of the pronouns.

The new finding here related to anaphor resolution is the P200 gender congruity effect found for the default reflexive pronoun 他自己/*himself*. This mid-anteriorly distributed P200 effect (congruent conditions elicited more positive amplitude as compared to incongruent ones) is different from the P200 reflexive pronoun gender specificity effect observed in the lateralized electrodes (她自己/*herself* is more positive than 他自己/*himself*, discussed in Section The Processing of Chinese Third Person Reflexive Pronouns), because the former effect is related to the gender of antecedents at sentence level. Thus, it seems that the conceptual processing of gender (the former P200) and the perceptual processing of character recognition (the latter P200) are processed in a similar time interval but are independent and associated with distinct scalp regions. Consistent with the attention-related mapping cost and context-induced expectation accounts, because of the similar sentence structures in the experimental stimuli, participants could expect the appearance of reflexive pronouns after encountering a proper name or a stereotypical role noun. If the antecedent is female, it is possible that participants are prepared for both reflexive pronouns, as both are applicable to a female antecedent. During the time interval in which the default (他自己/*himself*) is recognized, the evaluation of gender congruity is also easily accomplished because of its wide range of applicability. If the antecedent is male however, within this experimental context (because of the number of incongruent items of this nature presented), both the perceptual and conceptual processing systems might require more cognitive resources for encountering either 他自己/*himself* or 她自己/*herself*. The amplitude difference between the two contrasts (himself-congruent is more positive than himself-incongruent) might reflect such attention-related cost, related to prior contextual information (i.e., antecedent's gender here) deployed for processing the reflexive pronoun (Luck and Hillyard, 1994; Blanchet et al., 2007). In this case, the P200 is also sensitive to contextual information. On the other hand, when the specific reflexive (她自己/*herself*) is encountered, its recognition and retrieval are more complicated (as discussed in Section The Processing of Chinese Third Person Reflexive Pronouns) as compared to the default. It is possible that the word recognition is accomplished in the P200 time interval (i.e., the P200 effect for 她自己/*herself*) and the evaluation of gender congruity is delayed and resolved in a later time interval (i.e., the P600 effect).

Taking the main findings together, it is clear that 他自己/*himself* as a default with neutral gender is a critical feature in the perceptual and conceptual processing of gender

information during Chinese reflexive pronoun resolution. Both types of processing rely on the semantic radicals encoded in the characters, suggesting the essential importance of the gender-based radicals to decoding of gender specificity. In terms of gender congruity effects, only when there is a clear mismatch (i.e., male/male-biased antecedent followed by the specific female reflexive), do our results support a two-stage model of anaphor resolution (Garrod and Sanford, 1994). As a whole, given the more familiar orthographic form (perceptual) and applicability to both genders (conceptual) of 他自己/*himself*, this pronoun may serve as a baseline during anaphor resolution. In addition, the ERP pattern for Chinese reflexive pronoun resolution confirms distinct time courses of processing for the two reflexive pronouns. While the default, 他自己/*himself*, is processed mainly at the early perceptual stage of character recognition and gender evaluation, the processing of the specific reflexive pronoun, 她自己/*herself*, lasts from the early perceptual stage (bonding stage, possibly including the N400 time interval) to the late integration stage (resolution stage). It is possible that there is more than one way to resolve anaphors. One is a two-stage model when the gender of an anaphor is specific and the mismatch is definite. The other is a one-stage processing model in which a default anaphor is eligible for every mentioned antecedent, as is the case of 他自己/*himself* reported in the present study. This does not mean that these results are specific to languages with default and specific distinctions on pronouns, as in Chinese. Instead, in addition to the well-established two-stage model for anaphor resolution (when pronouns have a specific gender), the one-stage model extends the description of anaphor resolution to when a pronoun can be used for both genders (in other words, a genderless pronoun). According to Siewierska (2013), genderless pronouns are used in 67% (254 out of 378) of world languages (such as Finnish, Turkish, Thai, Indonesian, Vietnamese or Maori … etc.). The results observed here may thus be relevant to the 67% languages with genderless pronouns.

Conclusion

The non-symmetrical gender specificity of the Chinese characters for third person reflexive pronouns was studied during anaphor resolution. Independent P200 and N400 gender specificity effects confirmed processing differences resulting from the different gender specificity of reflexive pronouns (encoded in their semantic radicals) and also suggested the functional role of 他自己/*himself* as a default during anaphor resolution. During reflexive pronoun resolution, the two types of gender specificity interact with gender congruity respectively in the P200 (他自己/*himself*) and P600 (她自己/*herself*) time intervals. These results provide further evidence in support of the two-stage model of anaphor resolution only when there is an unambiguous mismatch between the antecedent and anaphor. The ERP patterns of the two reflexive pronouns also highlight the distinct time courses of anaphor resolution resulting from the two types of gender specificity. Overall, the findings in the present study demonstrate the importance of taking into account the asymmetry of gender specificity in Chinese third person reflexive

pronouns, as well as confirming 他自己/*himself* as the default applicable to both genders.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00151

REFERENCES

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Barber, H., Vergara, M., and Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: evidence from ERPs. *Neuroreport* 15, 545–548. doi: 10.1097/00001756-200403010-00032

Blanchet, S., Gagnon, G., and Bastien, C. (2007). Event-related potential of dynamic neural mechanisms of semantic organizational strategies in verbal learning. *Brain Res.* 1170, 59–70. doi: 10.1016/j.brainres.2007.07.024

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055

Callahan, S. M. (2008). Processing anaphoric constructions: insights from electrophysiological studies. *J. Neurolinguist.* 21, 231–266. doi: 10.1016/j.neuroling.2007.10.002

Carreiras, M., Garnham, A., and Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns: evidence from Spanish. *Eur. J. Cogn. Psychol.* 5, 93–116. doi: 10.1080/09541449308406516

Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol.* 49, 639–663. doi: 10.1080/713755647

Carreiras, M., Salillas, E., and Barber, H. (2004). Event related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cogn. Brain Res.* 20, 98–105. doi: 10.1016/j.cogbrainres.2004.01.009

Carreiras, M., Vergara, M., and Barber, H. (2005). Early ERP effects of syllabic processing during visual word recognition. *J. Cogn. Neurosci.* 17, 1803–1817. doi: 10.1162/089892905774589217

Chang, B.-M. (2007). 《"她"從哪裡來？現代女性指稱的源流考釋》(Where does "她" come from? *An enquiry into the origin of the modern feminine pronoun*),《歷史月刊》(*Historical Monthly*), 237, 123–127.

Cherng, R.-J., Chang, C.-L., and Chen, J.-Y. (2009). A new look at gender inequality in Chinese: a study of Chinese speakers' perception of gender-based characters. *Sex Roles* 61, 427–433. doi: 10.1007/s11199-009-9639-z

Chomsky, N. (1981). *Lectures on Government and Binding.* Dordrecht; Holland: Foris Publications.

Dambacher, M., Kliegl, R., Hofmann, M., and Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Res.* 1084, 89–103. doi: 10.1016/j.brainres.2006.02.010

Delong, K. A., Urback, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

Ding, G., Peng, D., and Taft, M. (2004). The nature of the mental representation of radicals in Chinese: a priming study. *J. Exp. Psychol. Learn.* 30, 530–539. doi: 10.1037/0278-7393.30.2.530

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cognit.* 32, 551–559. doi: 10.3758/BF03195846

Esaulova, Y., Reali, C., and von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Lang. Cogn. Neurosci.* 29, 781–803. doi: 10.1080/01690965.2013.794295

Feldman, L. B., and Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chiense characters. *J. Mem. Lang.* 40, 559–576. doi: 10.1006/jmla.1998.2629

Garnham, A., Oakhill, J., Ehrlich, M. F., and Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: new evidence from Spanish and French. *J. Mem. Lang.* 34, 41–62. doi: 10.1006/jmla.1995.1003

Garrod, S., and Sanford, A. J. (1994). "Resolving sentences in a discourse context: how discourse representation affects language understanding," in *Handbook of Psycholinguistics,* ed M. A. Gerbsbacher (San Diego, CA: Academic Press), 675–698.

Garrod, S., and Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: bonding and resolution. *J. Mem. Lang.* 42, 526–544. doi: 10.1006/jmla.1999.2694

Hagoort, P. (2009). "Reflections on the neurobiology of syntax," in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmáry (Cambridge, MA: MIT Press), 279–296.

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441. doi: 10.1126/science.1095455

Hahne, A., and Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cogn. Brain Res.* 13, 339–356. doi: 10.1016/S0926-6410(01)00127-6

Ho, S.-H., Ng, T.-T., and Ng, W.-K. (2003). A "radical" approach to reading development in Chinese: the role of semantic radicals and phonetic radicals. *J. Lit. Res.* 35, 849–878. doi: 10.1207/s15548430jlr3503_3

Hsu, C.-H., Tsai, J.-L., Lee, C.-Y., and Tzeng, O. J. L. (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: an event-related potential study. *Brain Lang.* 108, 56–66. doi: 10.1016/j.bandl.2008.09.002

Hua, X.-H. (2012). 《近三十年"她"字研究綜述》(An integrated discussion on the research of 她 nearly thirty years), 《現代語文：語文研究版》(Modern Chinese). 8, 6–8.

Huang, C.-H., and Luh, W.-M. (2012). Analysis of third person pronouns with gender semantic features in elementary school Mandarin textbooks. *J. Textbook Res.* 5, 85–113.

Hung, Y.-C., and Schumacher, P. B. (2012). Topicality matters: position- specific demands on Chinese discourse processing. *Neurosci. Lett.* 511, 59–64. doi: 10.1016/j.neulet.2012.01.013

Hung, Y.-C., and Schumacher, P. B. (2014). Animacy matters: ERP evidence for the multi-dimensionality of topic-worthiness in Chinese. *Brain Res.* 1555, 36–47. doi: 10.1016/j.brainres.2014.01.046

Irmen, L., Holt, D. V., and Weisbrod, M. (2010). Effects of role typicality on processing person information in German: evidence from an ERP study. *Brain Res.* 1353, 133–144. doi: 10.1016/j.brainres.2010.07.018

Jäger, L., Engelmann, F., and Vasishth, S. (2015). Retrieval interference in reflexive processing: experimental evidence from Mandarin, and computational modeling. *Front. Psychol.* 6:617. doi: 10.3389/fpsyg.2015.00617

Jiang, X., and Zhou,. X. (2009). Processing different levels of syntactic hierarchy: an ERP study on Chinese. *Neuropsychologia* 47, 1282–1293. doi: 10.1016/j.neuropsychologia.2009.01.013

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Proc.* 15, 159–201. doi: 10.1080/016909600386084

Kaan, E., and Swaab, T. Y. (2003). Repair, revision, complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* 15, 98–110. doi: 10.1162/089892903321107855

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psychol. Res.* 32, 355–377. doi: 10.1023/A:1023599719948

Kim, A., and Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *J. Mem. Lang.* 52, 205–225. doi: 10.1016/j.jml.2004.10.002

Kolk, H. H. J., Chwilla, D. J., van Herten, M., and Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain Lang.* 85, 1–36. doi: 10.1016/S0093-934X(02)00548-5

Kreiner, H., Mohr, S., Kessler, K., and Garrod, S. (2009). "Can context affect gender processing: evidence from ERP about the differences between definitional and stereotypical gender," in *Brain Talk: Discourse with and in the Brain*, eds K. Alter, M. Horne, M. Lindgren, M. Roll, and J. von Koss Torkildsen (Lund: Lunds Universitet), 107–119.

Kreiner, H., Sturt, P., and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007.09.003

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the Event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 1–27. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lee, C.-Y., Liu, Y.-N., and Tsai, J.-L. (2012). The time course of contextual effects on visual word recognition. *Front. Psychol.* 3:285. doi: 10.3389/fpsyg.2012.00285

Lee, C.-Y., Tsai, J.-L., Chan, W.-H., Hsu, C.-H., Hung, D. L., and Tzeng, O. J. L. (2007). Temporal dynamics of the consistency effect in reading Chinese: an event-related potentials study. *Neuroreport* 18, 147–151. doi: 10.1097/WNR.0b013e328010d4e4

Li, C. N., and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar.* Berkeley: University of California Press.

Li, Y. H. A., and Shi, Y. Z. (2000). 《漢語量詞系統的建立與複數標記"們"的發展》(The story of "men"). *Contemp. Linguist.* 2, 27–36.

Ling, Y.-Z. (1989). 《"她"字的創造歷史》(The history of the character '她'), 《語言教學與研究》(Language Teaching and Research). 4, 139–151.

Liu, Y., Perfetti, C. A., and Hart, L. (2003). ERP evidence for the time course of graphic, phonological, and semantic information in Chinese meaning

and pronunciation decisions. *J. Exp. Psych. Learn.* 29, 1231–1247. doi: 10.1037/0278-7393.29.6.1231

Luck, S., and Hillyard, S. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308. doi: 10.1111/j.1469-8986.1994.tb02218.x

Molinaro, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019

Molinaro, N., Carreiras, M., and Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *Neuroimage* 59, 3488–3501. doi: 10.1016/j.neuroimage.2011.11.009

Molinaro, N., Conrad, M., Barber, H., and Carreiras, M. (2010). On the functional nature of the N400: contrasting effects related to visual word recognition and contextual semantic integration. *Cogn. Neurosci.* 1, 1–7. doi: 10.1080/17588920903373952

Molinaro, N., Kim, A., Vespignani, F., and Job, R. (2008a). Anaphoric agreement violation: an ERP analysis of its integration. *Cognition* 106, 963–974. doi: 10.1016/j.cognition.2007.03.006

Molinaro, N., Vespignani, F., and Job, R. (2008b). A deeper reanalysis of a superficial feature: an ERP study on agreement violations. *Brain Res.* 1228, 161–176. doi: 10.1016/j.brainres.2008.06.064

Moser, D. (1997, January). Covert sexism in Mandarin Chinese. *Sino-Platonic Papers*, No. 74, 1–23.

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., and Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia* 36, 217–226. doi: 10.1016/S0028-3932(97)00119-X

Nieuwland, M. S., and Van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: evidence from ERPs. *Brain Res.* 1118, 155–167. doi: 10.1016/j.brainres.2006.08.022

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cognit.* 25, 273–285. doi: 10.3758/BF03211283

Packard, J. L. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach.* Cambridge, UK: Cambridge University Press.

Peng, J.-J. (2009). 《新詞"她 X"探析》(Discussion on the new word 'She X'), 《經濟研究導刊》(Economic Research Guide). 1, 235–236.

Perfetti, C. A., and Zhang, S. (1991). Phonetic processes in reading Chinese words. *Exp. Psychol. Learn.* 17, 633–643. doi: 10.1037/0278-7393.17.4.633

Qiu, J., Swaab, T. Y., Chen, H.-C., and Wang, S. (2012). The role of gender information in pronoun resolution: evidence from Chinese. *PLoS ONE* 7:e36156. doi: 10.1371/journal.pone.0036156

Rabovsky, M., and McRae, K. (2014). Simulating the N400 ERP component as semantic network error: insights from a feature-based connectionist attractor model of word meaning. *Cognition* 132, 68–89. doi: 10.1016/j.cognition.2014.03.010

Schmitt, B. M., Lamers, M., and Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cogn. Brain Res.* 14, 333–346. doi: 10.1016/S0926-6410(02)00136-2

Siewierska, A. (2013). "Gender distinctions in independent personal pronouns," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: http://wals.info/chapter/44 (Accessed August 06, 2015).

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3

Tang, T.-C. (1988). 《國語詞彙的『重男輕女』現象》 (Male chauvinism in the Chinese lexical structure)," in 《漢語詞法句法論集》(*Studies on Chinese Morphology and Syntax*), eds T.-C. Tang (Taipei: Student Book Co., Ltd.), 59–65.

Tsai, P.-S., Yu, B. H.-Y., Lee, C.-Y., Tzeng, O. J.-L., Hung, D. L., and Wu, D. H. (2009). An event-related potential study of the concreteness

effect between Chinese nouns and verbs. *Brain Res.* 1253, 149–160. doi: 10.1016/j.brainres.2008.10.080

Tsang, Y.-K., and Chen, H.-C. (2009). Do position-general radicals have a role to play in processing Chinese characters? *Lang. Cogn. Process.* 24, 947–966. doi: 10.1080/01690960802154615

Van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., and Vissers, C., Th,.W. M. (2009). Monitoring in language perception. *Lang. Linguist. Compass* 3, 1211–1224. doi: 10.1111/j.1749-818X.2009.00163.x

van Herten, M., Kolk, H. H. J., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Brain Res. Cogn. Brain Res.* 22, 241–255. doi: 10.1016/j.cogbrainres.2004.09.002

Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cognit.* 18, 380–393. doi: 10.3758/BF03197127

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., and McIsaac, H. (1991). Fractionating the word repetition effect with Event-Related Potentials. *J. Cogn. Neurosci.* 3, 131–150. doi: 10.1162/jocn.1991.3.2.131

Wang, Y.-Y. (2010). 《從 'ta' 看 '他' 與 '她'》(From 'ta' to look at '他' and '她'),《大眾文藝》(Da Zhong Wen Yi). 6, 78.

Wu, D.-S., and Liang, T. (2008). Chinese pronominal anaphora resolution using lexical knowledge and entropy-based weight. *J. Am. Soc. Inf. Sci. Tec.* 59, 2138–2145. doi: 10.1002/asi.20922

Xu, X.-D., Jiang, X.-M., and Zhou, X.-L. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain Cogn.* 81, 223–236. doi: 10.1016/j.bandc.2012.11.002

Ye, Z., Zhan, W., and Zhou, X. (2007). The semantic processing of syntactic structure in sentence comprehension: an ERP study. *Brain Res.* 1142, 135–145. doi: 10.1016/j.brainres.2007.01.030

Yum, Y.-N., Law, S.-P., Su, I.-F., Lau, D. K.-Y., and Mo, K.-N. (2014). An ERP study of effects of regularity and consistency in delayed naming and lexicality judgment in a logographic writing system. *Front. Psychol.* 5:315. doi: 10.3389/fpsyg.2014.00315

Role descriptions induce gender mismatch effects in eye movements during reading

Chiara Reali[1]*, Yulia Esaulova[1], Anton Öttl[2] and Lisa von Stockhausen[1]

[1] *Department of Psychology, University of Duisburg-Essen, Essen, Germany,* [2] *Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway*

The present eye-tracking study investigates the effect of gender typicality on the resolution of anaphoric personal pronouns in English. Participants read descriptions of a person performing a typically male, typically female or gender-neutral occupational activity. The description was followed by an anaphoric reference (*he* or *she*) which revealed the referent's gender. The first experiment presented roles which were highly typical for men (e.g., blacksmith) or for women (e.g., beautician), the second experiment presented role descriptions with a moderate degree of gender typicality (e.g., psychologist, lawyer). Results revealed a gender mismatch effect in early and late measures in the first experiment and in early stages in the second experiment. Moreover, eye-movement data for highly typical roles correlated with explicit typicality ratings. The results are discussed from a cross-linguistic perspective, comparing natural gender languages and grammatical gender languages. An interpretation of the cognitive representation of typicality beliefs is proposed.

Keywords: gender typicality, gender stereotypes, eye-tracking, sentence reading, anaphor resolution

INTRODUCTION

In talking about human beings, gender information can be transmitted in different ways, e.g., via grammatical gender cues and gender-typical lexemes. Grammatical gender is marked, for example, in morphological elements which may express the gender of the referent such as the suffix *-in* in German (e.g., *Lehrer-in*, teacher$_{feminine}$). The gender typicality of lexemes results from the likelihood of personal nouns to refer to men or women. Thus, the noun *nurse* has female typicality and *surgeon* male typicality, because of their likelihood to be associated with a female or a male referent respectively, as shown in typicality ratings (cf. Kennison and Trofe, 2003). The purpose of the present paper is to analyze the effect of gender typicality on the resolution of a pronominal anaphor when gender typicality is conveyed by a description of a role rather than a role noun antecedent. Namely, we investigate a socio-psychological concept, expectations about gender roles, with the help of a psycholinguistic tool, the paradigm of anaphor resolution during sentence reading. Our approach makes use of verbal descriptions and allows for comparing a natural gender language with a grammatical gender language, as will be outlined in detail below. The present study deals with English, a language which does not possess a grammatical gender system ("natural gender language," see Hellinger and Bußmann, 2001). Since most professional roles lie in the range of moderate stereotypicality, we explore both the effect of roles with high and moderate degrees of gender typicality. Previous studies, however, mainly focused on the gender typicality effect of strongly stereotyped roles; thus, in a reading time study employing

role nouns, Kennison and Trofe (2003) presented gender-typical roles as antecedents and personal pronouns as anaphors. The gender mismatch condition (e.g., *The executive…She…*) prompted longer reading times in the spillover region following the pronoun compared to the matching condition. The results indicated that the role nouns triggered gender-typical representations of the referent which either agreed or disagreed with the following pronominal anaphor.

Garnham et al. (2002) conducted a reading study employing both role nouns and short expressions referring to gender typical habits or biological characteristics (e.g., wearing a bikini; giving birth). The study shows that a mismatch between the two pieces of information produced longer reading times, even when the presentation order of the two pieces of information was reversed, suggesting that gender inferences were made elaboratively and not only when the inference was necessary for the coherent interpretation of the text.

In a reaction time study, Oakhill et al. (2005) asked participants to judge if pairs composed of gender stereotypical and gender definitional role nouns (e.g., *surgeon-sister*) could apply to the same person. Results showed that the activation of stereotypical information was automatic and difficult to suppress, even with instructions encouraging participants to explicitly reconsider the stereotypical representations of the roles.

Pyykkönen et al. (2010) explored the effect of gender stereotypes on spoken language processing in Finnish, a language which also does not possess a grammatical gender system, by means of the visual-world paradigm. Participants heard stories presenting a gender typical role noun, in association with pictures of male or female characters. Results showed an activation of gender stereotypes triggered by the spoken role nouns, even if this activation was not needed to establish greater discourse coherence.

Most psycholinguistic studies investigating gender typicality effects on anaphor resolution in English (e.g., for eye-tracking methodology Sturt, 2003; Duffy and Keir, 2004; Kreiner et al., 2008; for ERP methodology, Osterhout and Mobley, 1995; Osterhout et al., 1997) used reflexive pronouns (*himself/herself*) to reveal referential gender. The results of these studies document a consistent mismatch effect on the anaphor region or the subsequent region, caused by conflicts between the gender typicality of role noun antecedents and the following anaphors.

To summarize the main findings of studies on natural gender languages, one can state that incongruence between the gender typicality of the antecedent role nouns and the anaphor gender triggers a slowdown in resolution, for both personal and reflexive pronouns.

In grammatical gender languages, in contrast to natural gender languages, role nouns carry additional grammatical gender cues, which also affect the representation of referential gender. As a consequence, the effect of grammatical gender and gender typicality usually appear in interaction, and the specific contribution of the different factors can be difficult to disentangle.

Esaulova et al. (2014), for example, analyzed anaphor resolution after role nouns carrying both grammatical gender cues and gender typicality in an eye-tracking study on German, (e.g., *Oft hatte der Elektriker/die Elektrikerin gute Einfälle, regelmäßig plante er/sie neue Projekte.* "Often had the electrician$_{masculine/feminine}$ good ideas, regularly planned he/she new projects."). In the condition of a mismatch between grammatical gender and gender typicality of the role noun results showed a mismatch effect not only on the anaphor region but also on the role noun region. The antecedent contained grammatical gender markings (either masculine or feminine ones), therefore the effect of the noun's gender typicality on anaphor resolution resulted from a combined processing of grammatical gender cues and typicality (see also Gygax et al., 2008; Irmen and Schumann, 2011).

A series of experiments conducted by Jäger et al. (2015), analyzed the online processing of reflexives in German and pronominal possessives in Swedish, by means of self-paced reading and eye-tracking methodology. The study focused on grammatical gender, conveyed through gender markings on role nouns (in German) or proper names (in Swedish). Materials presented an antecedent and a distractor, which could match or mismatch in gender (masculine/feminine). In contrast to previous studies, the results of these experiments showed no evidence for an online similarity-interference effect triggered by a gender overlap between the competitor role nouns. Only offline response accuracy to the comprehension questions in the self-paced reading experiment showed that the similarity-interference might have produced misretrievals of the distractors. These results suggest that the previously reported interference effects in reflexive processing may arise at the stage of retrieval rather than at the encoding stage.

The interplay of grammatical gender and gender typicality was further explored in a reading study on another grammatical gender language (Italian): Cacciari et al. (2011) investigated the resolution of personal pronouns in interaction with gender typicality. In the first part of each item, gender typicality was established through a context which described a typically male, female or neutral setting, for example "During the last Grand Prix of Formula One a terrible car accident provoked a crash close to the stands" (typically male context), or "Within the couple, scenes of jealousy were frequent but this time they came to blows and they got close to tragedy" (typically female context). In the second part of the item an epicene (a noun with a defined grammatical gender, but which can refer to both a male or female referent, e.g., *vittima,* male or female victim$_{feminine}$) or a bigender role noun (a noun which can function both as a feminine and a masculine noun, e.g., *assistente,* assistant) was introduced as antecedent for an anaphoric pronoun. The anaphor could match or mismatch the typical context and/or the grammatical gender of the epicene. Results showed that for bigender role nouns, which did not present a defined grammatical gender, the influence of gender typicality was essential to trigger the mismatch effect; however, when the antecedent was an epicene the grammatical gender of the role noun, even though purely formal, affected the resolution of the anaphor and interfered with the typicality effect.

The reviewed literature shows that role nouns can represent a useful tool to convey and investigate gender typicality. However, role nouns can preclude a direct comparison of natural gender languages and grammatical gender languages,

because in grammatical gender languages personal role nouns are usually marked for grammatical gender and therefore carry an additional cue to referential gender, whereas in natural gender languages most role nouns are not morphologically marked. This causes different processes in the resolution of anaphors with role noun antecedents, for in grammatical gender languages readers are presented both with grammatical information and information from gender typicality, while natural gender languages mostly present only cues from gender typicality. The complex interaction between grammatical cues and gender typicality represents a challenge in investigating effects of gender typicality, since the grammatical gender of role nouns may compete with gender typicality cues in the representation of referent gender. To overcome this issue, the present study employs a paradigm which replaces role nouns with corresponding role descriptions, in order to convey the gender typicality of a role without presenting the role noun itself. In a study by Reali et al. (2015), a description-based paradigm was developed to study the effect of gender typicality on anaphor resolution in a grammatical gender language, while excluding grammatical cues of the antecedents. This research raised a further research question, namely a cross-linguistic comparison of cognitive processes occurring in a "naturalized" grammatical gender language (i.e., a grammatical gender language without grammatical gender cues) and those in a natural gender language. Even in the absence of grammatical gender cues in the materials, speakers of a grammatical gender language may process gender typicality cues differently from speakers of a language without grammatical gender. Evidence from studies with bilinguals suggests that readers may activate different cognitive representations of referent gender according to the language of the task they are engaged in, shifting gender representations when switching from a natural gender language to a grammatical gender language and vice versa (see Sato et al., 2013). Starting from these considerations, the present study analyzes the processing of gender typicality in a natural gender language and compares the resolution process with previous studies conducted on a grammatical gender language (cf. Reali et al., 2015).

Another research question concerns the degree of gender typicality of the items. Earlier studies employing the anaphor resolution paradigm usually relied on highly typical roles and thus excluded the majority of social and professional roles, which do not occupy extreme positions on the gender typicality scale. Therefore, the second experiment of the present paper focuses on effects triggered by roles with lower degrees of gender typicality and examines if role descriptions with moderate degrees of gender typicality are able to elicit expectations in the referent gender representation, thus producing a disruption in the reading process when the mismatching pronoun is encountered.

The present research employs the methodology of eye-tracking, which provides high spatial and temporal resolution in mapping the process of anaphor resolution during reading.

EXPERIMENT 1

The aim of Experiment 1 was to analyze the effect of gender typicality on pronominal anaphor resolution with a description-based paradigm. Specifically, the paradigm employed descriptions of gender-typical occupational roles instead of role nouns to convey gender typicality. The absence of role nouns allows us to compare the processing of gender typicality cues in natural gender and grammatical gender languages.

Method
Participants
Thirty-one students (17 women and 14 men) from the University of Sussex, UK, participated in the study. Participants were English native speakers, with normal or corrected-to-normal vision (mean age = 21 years, SD = 3.9). They received monetary compensation or course credit for their participation. Ethical approval for the study was granted by the University of Sussex's Research Ethics Committee and all participants provided written informed consent before taking part in the study.

Design and Hypothesis
The experiment was designed to test the interaction between the gender typicality of the occupational role (*typicality*: male, female, or neutral) and the gender of the anaphoric reference (*pronoun*: masculine or feminine). In accord with the German study (Reali et al., 2015) and earlier research using gender-typical role nouns, we expected a mismatch between gender-typical role description and anaphor gender to evoke longer fixation times and more frequent regressions compared to the matching and neutral conditions.

Materials
Materials were created to provide gender-typical information associated with different occupational activities without employing role nouns. The experimental sentences are based on the material of a study which had been conducted in German (Reali et al., 2015). In this previous study, a list of roles had been first selected from published collections of role nouns gender typicality ratings for different languages (Kennison and Trofe, 2003; Irmen, 2007; Gabriel et al., 2008). Then participants (30 women, 20 men, mean age = 23.1, SD = 4.1, students from the University of Heidelberg, Germany) estimated to which extent a specific professional role (e.g., primary school teacher) was held by men and/or women, using a 7-point scale with anchor points 1 = only men, 7 = only women, and 4 = same amount of women and men. Items (N = 77) were categorized as follows: male: $\leq$ 2.5, neutral: 3.5–4.5, female: $\geq$ 5.5. The same sample provided, through a written computer-based production task, a description of each role, on which the experimental items were based. These descriptions were then presented, in a paper-based questionnaire, to a new participant sample (N = 40, students from the University of Heidelberg), which had to guess the role nouns corresponding to the descriptions. This sub-test had the goal to check the correspondence between the role representation conveyed by the descriptions and the corresponding role nouns. Descriptions presenting less than 80% description-noun correspondence were discarded. This selection yielded 12 female, 12 male, and 12 neutral descriptions, to constitute the final material of 36 experimental items for the

eye-tracking study. The last participant sample also rated the typicality of the final descriptions, which presented a strong correlation with the role noun rating ($r = 0.995$, $p < 0.001$). The differences between the three typicality conditions, calculated on the description typicality ratings ($M_{\text{male}} = 1.87$, $SD = 0.42$, $M_{\text{female}} = 5.98$, $SD = 0.37$, $M_{\text{neutral}} = 4.17$, $SD = 0.37$) were statistically significant, male–female: $t_{(22)} = -30.23$, $p < 0.001$; male–neutral: $t_{(22)} = -20.24$, $p < 0.001$; female–neutral: $t_{(22)} = -18.99$, $p < 0.001$. The pre-test procedure was fully conducted at the University of Heidelberg, Germany (see Reali et al., 2015). The resulting experimental material was translated and adapted to be employed for the present eye-tracking study.

Each experimental sentence consisted of a first part which described an occupation ("context"), and a second part containing a pronominal anaphor ("target sentence"). The personal pronoun ("he"/"she") referred back to the person presented in the previous context, which had been introduced with initials, as in examples (1) (male typicality), and (2) (female typicality):

(1) K. L. installs power lines and cables, checks electricity voltage.
    In this field he/she has a lot of experience.
(2) L. K. teaches at a primary school, instructs children in reading.
    At work he/she wears thick glasses.

The gender neutrality of the target sentences had been ensured through a rating pre-test. In order to keep the anaphoric pronoun in a comparable position across items, all target sentences had a fixed linguistic structure, with the anaphor positioned between an initial adverbial expression and the verb.

In addition to the experimental sentences we presented 50 filler sentences containing descriptions of non-professional roles (e.g., moviegoer) and anaphoric expressions referring back to an inanimate object, to avoid drawing attention to the gender topic. Finally, we presented 24 content-related questions (e.g., "Is the lab coat green?") in order to promote attentive reading, leading to a total number of 110 trials (including experimental items, fillers and questions).

Procedure

Eye movements were monitored with a video-based head mounted eye-tracker (Eyelink II, sampling rate of 250 Hz, average accuracy 0.5°). Materials were presented with the software Eyetrack[1] on a 21-inch CRT computer screen, with an active screen size of 40×30 centimeters and a resolution of 1024×768 pixels. Participants were seated 70 cm away from the screen, at which distance 3 characters subtended approximately 1° of visual arc. A chinrest was used to minimize head movements. Reading was binocular but only the dominant eye was tracked. The dominant eye was determined through the Miles test[2].

---

[1] 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 Massachusetts, UMass at Amherst, USA).
[2] Participants extended both arms and created an opening with their hands, through which they fixated a point on the wall. Then they slowly moved their hands

The experiment began after a calibration procedure which was performed on a nine-point grid.

The presentation of sentences started with a small rectangle indicating the position of the first word of the sentence. The item appeared when the rectangle was fixated accurately. Whenever, the fixation on the rectangle was judged as inaccurate, re-calibration was carried out.

To familiarize participants with the task, the experiment started with four practice trials, one of which was followed by a comprehension question. Then the experimental sentences and filler items were presented. Sentences were displayed in a monospaced 22-point Lucida Console font, in black characters on a light gray background and consisted of three lines, presenting a maximum number of 49 characters each. The first two lines contained the role description; the third line presented the target sentence with the anaphoric reference. Experimental items were presented in randomized order across participants. After reading an item, 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 comprehension questions.

As a follow-up procedure, participants completed a questionnaire asking for gender typicality ratings, on a 7-point Likert scale, concerning the job descriptions that were presented in the eye-tracking session. The experiment lasted in total approximately 30–45 min.

Results
Data Analysis

In order to investigate the effect of the priming context on the target sentence, we analyzed fixation times and regression patterns on different regions of the target sentences. The target sentence was divided into four regions of analysis: adverb region, anaphor region, spillover region, and final region. The segmentation into regions of analysis is shown in **Table 1**.

In order to reflect the processing of the text from early to late stages, data were analyzed for the following eye-tracking measures: first fixation time, first pass time, regression path time, total time, and probabilities of regressions into and out of a region. First fixation time is the duration of the first fixation in a given region. First pass time is the time 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

---

TABLE 1 | Experiment 1 factorial structure and regions of analyses (delimited by a dash).

Context	Male role description	C. R. repairs and produces furniture, works with wood.
	Female role description	K. P. sells flowers, makes up bouquets in a shop.
	Neutral role description	F. H. plays an instrument professionally in an orchestra.
Target	Anaphoric reference	Usually - he/she has - a sufficient - income.
		Reg.1 Reg.2 Reg.3 Reg.4

---

toward their eyes, while fixating the point through the opening. At a close distance, in order to continue to fixate the point, the opening was drawn either in front of the left or the right eye, according to ocular dominance.

path 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 re-reading, but not including regressions from this region. Regressions into and out of a region, respectively, consist of the proportion of backward movements into a specific region, or leaving the region to the left after a first pass fixation of the region (cf. Sturt, 2003; Boland, 2004). In general, longer fixation times and a higher probability of regressions are indicative of greater difficulty in processing the respective region.

Initial stages of data analysis were carried out using the software EyeDoctor and EyeDry provided by the Department of Psychology at the University of Massachusetts Amherst. Short fixations (below 70 ms) were merged with neighboring fixations within three characters. Following Reali et al. (2015), we removed fixations below 70 ms and above 600 ms, as they can be assumed to be not representative of regular information acquisition during reading (4.1% of the data). The remaining data have been logarithmically transformed to meet the normality assumption for the following analyses. No significant difference emerged in the distribution of missing data across typicality conditions for all regions and fixation duration measures [$M_{male} = 74.00$; $M_{female} = 74.19$; $M_{neutral} = 69.06$, $F_{(2, 45)} = 0.86$, ns]. Analyses were based on linear mixed-effect modeling, implemented by the *lmer* function from the lme4 package (Bates et al., 2014) in R (R Core Team, 2012, version 2.15.2). We included in our models participants and items as random effects (see Baayen et al., 2008). As fixed effects for our models we selected the experimental factors that were assumed to influence the target sentence processing: *gender typicality* of the priming sentence (male, female, or neutral) and *pronoun* of the target sentence (masculine, feminine). In addition, we included *region length* (number of characters for each region of analysis) in all fixation duration measures (i.e., excluding regression measures), and *participant gender*, as fixed effects, since these factors could affect the reading processes, Model<- lmer [fixation_time ~ typicality * pronoun * participant_gender * region_length + (1 |participants) + (1 |items)].

To systematically detect the best fitting model for each measure and region, we employed the *step* function available in lmerTest package (Kuznetsova et al., 2013), which was developed with the purpose of automatizing and standardizing the model building process. Starting from a fully specified model, *step* performs a backward elimination of both random and fixed effects that are not warranted by the data by conducting iterative model comparisons. The function is based on likelihood ratio tests and step-wise removal of non-significant fixed effect terms. Significant effects of pronoun, typicality and their interaction were further explored through contrast analyses. Pairwise comparisons tested each typicality condition followed by masculine and feminine pronouns (male–he vs. male–she; female–he vs. female–she; neutral–he vs. neutral–she).

Eye-tracking Results

The final models for each measure and region (including all significant random effects, fixed effects, and interactions) are reported in Supplementary Material (Table S1). Means and

standard deviations of fixation duration time and percentages of regressions are reported in **Table 2**[3]. Details on statistical results are reported in **Table 3**. We report below eye-tracking measures presenting statistically significant fixed effects of *typicality*, *pronoun*, and *typicality*pronoun* ($p < 0.05$), and corresponding significant or marginally significant ($p < 0.1$) results of contrast analyses, separated for measure.

First pass time

The first reliable interaction effect between *typicality* and *pronoun* was detected in first pass time on the region immediately following the pronoun (spillover)[4]. Contrast analyses revealed that the effect was statistically significant only when the priming sentence was female, with congruent trials being read faster, $M_{femaleHE} = 302$, $M_{femaleSHE} = 263$, $t_{(948)} = 2.55$, $p = 0.01$; $M_{maleHE} = 257$, $M_{maleSHE} = 269$, ns; $M_{neutralHE} = 269$, $M_{neutralSHE} = 288$, ns.

Regression path time

A main effect of *pronoun* appeared on the pronoun region and on the spillover. Contrast analyses showed that the feminine pronoun condition was read faster, $M_{HE} = 295$, $M_{SHE} = 269$, $t_{(514)} = 2.35$, $p = 0.002$ (pronoun region); $M_{HE} = 457$, $M_{SHE} = 407$, $t_{(941)} = 2.14$, $p = 0.03$ (spillover region).

Regressions out of a region

The interaction between *typicality* and *pronoun* emerged in the proportion of regressions out of the last region of the target sentence. Contrast analyses showed a significant effect for the neutral condition, presenting less regressions in association with a masculine as compared to a feminine pronoun, $M_{neutralHE} = 8.1$, $M_{neutralSHE} = 13.2$, $t_{(947)} = -2.26$, $p = 0.02$; $M_{maleHE} = 8.9$, $M_{maleSHE} = 11.7$, ns; $M_{femaleHE} = 14.8$, $M_{femaleSHE} = 11.2$, ns.

Total fixation time

The interaction between *typicality* and *pronoun* emerged on the spillover region. Pairwise comparisons revealed a significant effect for the female condition, but not for the male and neutral conditions, with shorter fixation time on congruent trials as compared to incongruent ones, $M_{femaleSHE} = 380$, $M_{femaleHE} = 427$, $t_{(998)} = 2.14$, $p = 0.03$; $M_{maleHE} = 363$, $M_{maleSHE} = 355$, ns.; $M_{neutralHE} = 437$, $M_{neutralSHE} = 437$, ns. Furthermore, a main effect of *participant gender* emerged on the pronoun region. Contrasts revealed a tendency for female participants to read faster, $M_{men} = 355$, $M_{women} = 316$, $t_{(30)} = 1.86$, $p = 0.073$.

Gender Typicality Ratings and Eye Movements

Typicality, ratings for Experiment 1 are reported in Supplementary Material (Table S2). Typicality ratings were based on the data collected in a previous study (see Materials section),

---

[3]Estimates obtained from the fitted models represent the model's prediction and take the crossed random effects into consideration. Therefore, values reported in the text may differ from the aggregated means reported in the tables.

[4]In first fixation time and first pass time, the first region of the target sentence was discarded from the analysis because of high percentage of missing values (33.5%) in comparison to the average skipping rate (17.4%). The high skipping rate of the first region may be explained by the fact that this region is represented by a short temporal adverb (e.g., "Today") which may be easily skipped in early reading stages.

TABLE 2 | Means (standard deviations) of fixation duration time (ms) and percentages of regressions for Experiment 1.

Region	Typ.	Pron.	Reg. 1		Reg. 2		Reg. 3		Reg. 4	
FF	Male	He	202	(59)	191	(66)	205	(79)	216	(94)
		She	214	(82)	192	(67)	204	(66)	230	(106)
	Fem.	He	209	(68)	198	(75)	215	(87)	232	(111)
		She	205	(76)	184	(66)	207	(73)	237	(108)
	Neutr.	He	197	(61)	198	(62)	211	(81)	233	(111)
		She	196	(63)	188	(64)	217	(80)	224	(105)
FP	Male	He	245	(85)	254	(140)	313	(182)	340	(282)
		She	253	(98)	272	(176)	328	(203)	334	(242)
	Fem.	He	246	(108)	270	(165)	348	(203)	340	(253)
		She	233	(84)	266	(159)	307	(204)	339	(236)
	Neutr.	He	235	(82)	248	(114)	316	(244)	295	(192)
		She	226	(68)	255	(137)	327	(216)	322	(236)
RP	Male	He	265	(164)	372	(312)	585	(681)	950	(776)
		She	290	(187)	388	(335)	538	(496)	1047	(869)
	Fem.	He	246	(108)	369	(270)	563	(422)	1096	(877)
		She	246	(126)	347	(191)	496	(380)	1093	(969)
	Neutr.	He	243	(110)	325	(232)	680	(618)	901	(828)
		She	243	(121)	306	(202)	629	(719)	973	(815)
TT	Male	He	275	(139)	384	(239)	456	(294)	427	(322)
		She	295	(170)	406	(255)	439	(267)	412	(270)
	Fem.	He	275	(146)	416	(224)	497	(299)	466	(330)
		she	261	(130)	389	(227)	459	(297)	428	(281)
	Neutr.	He	279	(139)	389	(207)	512	(390)	371	(245)
		She	264	(118)	395	(260)	501	(319)	393	(267)
RI	Male	He	28	(45)	30	(46)	22	(42)	–	–
		She	22	(41)	32	(47)	22	(42)	–	–
	Fem.	He	26	(44)	35	(48)	26	(44)	–	–
		She	30	(46)	30	(46)	22	(42)	–	–
	Neutr.	He	22	(42)	44	(50)	20	(40)	–	–
		She	21	(41)	42	(49)	20	(40)	–	–
RO	Male	He	2	(15)	21	(41)	30	(46)	47	(50)
		She	4	(19)	17	(38)	25	(43)	53	(50)
	Fem.	He	0	(0)	19	(40)	32	(47)	59	(49)
		She	2	(13)	19	(40)	30	(46)	52	(50)
	Neutr.	He	1	(10)	13	(34)	42	(49)	45	(50)
		She	2	(13)	9	(29)	35	(48)	56	(50)

FF, first fixation time; FP, first pass time; RP, regression path; TT, total time; RI, regressions into the region; RO, regressions out of the region.

from a sample which did not participate in the eye-tracking experiment. In order to investigate if eye movements reflected the extent of gender expectations, we conducted a by-item linear regression analysis with typicality ratings as predictors of eye movements. We selected the regions of analysis where the gender mismatch effect emerged. Since pairwise comparisons revealed an asymmetry between the male and female condition, we conducted separate analyses for the two anaphoric pronouns. Results revealed that typicality ratings predicted first pass fixation times after a masculine anaphor ($\beta = 0.35$, $p < 0.05$).

TABLE 3 | Statistical results for Experiment 1.

	First fixation time		First pass time		Total time	
	(DF) F-value	Pr (>F)	(DF) F-value	Pr (>F)	(DF) F-value	Pr (>F)
FIRST REGION						
Pronoun	(1690) 0.332	0.564	(1668) 0.072	0.789	(1875) 0.524	0.469
Typicality	(2691) 5.655	0.003*	(242) 1.557	0.223	(232) 1.120	0.339
Pron. * Typ.	(2697) 0.406	0.666	(2674) 1.662	0.190	(2873) 1.031	0.357
PRONOUN REGION						
Pronoun	(1892) 2.842	0.092	(1883) 0.522	0.470	(11,008) 1.134	0.287
Typicality	(2888) 0.349	0.706	(2883) 0.131	0.877	(235) 0.165	0.848
Pron. * Typ.	(2886) 1.571	0.208	(2883) 0.435	0.647	(21,016) 2.003	0.136
SPILLOVER REGION						
Pronoun	(1958) 0.022	0.883	(1948) 0.055	0.816	(11,011) 1.265	0.261
Typicality	(232) 0.521	0.599	(232) 0.551	0.582	(231) 0.143	0.867
Pron. * Typ.	(2955) 0.578	0.561	(2948) 4.442	0.012*	(21,003) 3.015	0.049*
FINAL REGION						
Pronoun	(1795) 0.324	0.569	(1761) 0.521	0.471	(1773) 0.008	0.928
Typicality	(2799) 0.596	0.551	(231) 0.130	0.879	(232) 0.255	0.776
Pron. * Typ.	(2793) 0.469	0.626	(2755) 0.197	0.821	(2765) 0.167	0.846

	Regression path		Regressions in		Regressions out	
FIRST REGION						
Pronoun	(1678) 0.046	0.830	(11,043) 0.282	0.595	(11,082) 2.714	0.100
Typicality	(231) 2.418	0.105	(233) 0.939	0.401	(21,083) 2.876	0.057
Pron. * Typ.	(2677) 0.628	0.534	(21,043) 1.176	0.308	(21,077) 0.222	0.801
PRONOUN REGION						
Pronoun	(1886) 7.491	0.006*	(11,048) 1.092	0.296	(11,042) 1.646	0.199
Typicality	(233) 0.781	0.466	(233) 2.705	0.082	(233) 2.184	0.128
Pron. * Typ.	(2855) 0.360	0.698	(21,045) 0.752	0.472	(21,042) 0.148	0.862
SPILLOVER REGION						
Pronoun	(1941) 4.594	0.032*	(11,050) 0.206	0.650	(11,049) 3.713	0.054
Typicality	(232) 1.055.	0.358	(233) 0.266	0.768	(233) 1.180	0.320
Pron. * Typ.	(2938) 0.805	0.447	(21,042) 0.321	0.726	(21,046) 0.216	0.806
FINAL REGION						
Pronoun	(1762) 0.486	0.486	–	–	(11,047) 1.608	0.205
Typicality	(2757) 1.514	0.221	–	–	(233) 0.392	0.679
Pron. * Typ.	(2755) 0.324	0.723	–	–	(21,047) 3.363	0.035*

*Significance codes: *p < 0.05.*

As the scale for typicality ratings presented the poles 1 = male, and 7 = female, the β coefficient showed a direct correlation in the condition of the masculine pronoun, with lower ratings predicting shorter fixations after the pronoun *he*. This result indicates that fixation time on a region where the mismatch effect emerged corresponded to the degree of gender typicality expressed in the explicit typicality ratings of the respective items.

Follow-up Typicality Ratings

Follow-up typicality ratings were collected from participants immediately after completing the eye-tracking experiment. The follow-up ratings showed a high correlation with the pre-test ratings ($r = 0.966$, $p < 0.001$). However, male and female typicality turned out to be more skewed toward neutrality, so that typically male and particularly typically female occupations received less extreme ratings as compared to the pretest ratings, M_{male}, pretest = 1.87, M_{male}, follow-up = 2.32, $t_{(22)} = 2.88$, $p = 0.009$; M_{female}, pretest = 5.98, M_{female}, follow-up = 5.20, $t_{(22)} = 4.20$, $p < 0.001$; $M_{neutral}$, pretest = 4.04, $M_{neutral}$, follow-up = 4.16, $t_{(22)} = 0.85$, *ns*.

Discussion

The study analyzed the effect of gender typicality cues on the resolution of a pronominal anaphor. As antecedents, the commonly used role nouns were replaced with role descriptions which contained only gender typicality cues to referent gender. The experiment was conducted in English, a language which does not possess a grammatical gender system.

A main effect of pronoun emerged in regression path on the pronoun and spillover region, with the feminine pronoun

receiving shorter fixation time than the masculine pronoun. This effect may suggest a general greater difficulty to integrate a male as compared to a female referent. However, it should be noted that this effect is limited to this time measure, therefore representing an isolated finding rather than a systematic pattern.

The interaction between gender typicality of the description and pronoun gender is in the focus of the study and emerged in measures representing different stages of processing. Results showed that a mismatch effect between the two factors occurred reliably in a measure of early processing on the region following the anaphoric pronoun. Moreover, this interaction was detected consistently in a measure of intermediate stage of processing (i.e., when participants regressed from the last region at the end of the target sentence to re-check the previously read sentence) and in one measure of late processing, namely the total amount of time spent on the pronoun spillover region. Furthermore, correlational analyses with gender typicality ratings showed that the typicality degree of the different items predicted the mismatch effect revealed by early fixation times, confirming the validity of the description paradigm as a tool to investigate gender typicality.

The location of the early mismatch effect is consistent with data from reading studies in English which employed role nouns as antecedents and personal pronouns as anaphors (Kennison and Trofe, 2003). The effect appears to be delayed in location and time in regard to studies employing reflexive pronouns to trigger the mismatch (e.g., Sturt, 2003). However, the effect cannot be compared directly because of relevant differences in sentence structure and paradigms used in the studies.

The present data can now be compared to a parallel study on German, where grammatical gender cues were avoided in the materials (Reali et al., 2015). Interestingly, in the German study the mismatch effect occurred earlier (in first fixations), on the pronoun region. Furthermore, in the German experiment the mismatch effect surfaced in two further measures (regressions in and total time) on the pronoun region itself. A possible explanation of the difference to the present findings concerns the presence or absence of grammatical gender in the two languages. The description-based paradigm served to keep the texts free of morphological gender cues in both languages. However, the processing of gender typicality cues may activate grammatical gender in the language with a grammatical gender system and thus cognitively facilitate the assignment of referent gender in the direction suggested by gender typicality. This would explain why the reference resolution process appears to be faster in the grammatical gender language. Previous eye-tracking studies using plural role nouns as antecedents also may support the interpretation that grammatical gender cues make gender typicality cues more salient and speed up the eventual gender mismatch effect. For example, in an eye-tracking experiment with German material, Irmen (2007) employed a noun phrase as anaphor ("these men/these women"). When antecedents were masculine generics, the typicality mismatch effect appeared on the first word of the anaphoric phrase itself in first pass reading ("these"). In contrast, when the antecedents had the form of gender-unmarked role nouns (e.g., *Alleinerziehende*, single parents) the typicality mismatch effect fully emerged only in later measures on the spillover region.

A further point of discussion is the asymmetry for the male and female condition, revealed in the pairwise comparisons of the mismatch effect. Specifically, gender mismatch was reliable only for the female condition, which produced an impairment in the sentence processing when followed by a masculine pronoun. This asymmetry was reliable in early and later stages of processing, on the target sentence spillover. The asymmetry effect may be interpreted as indicative of readers' difficulty to integrate a male referent with the representation of a typically female occupation; in contrast, reconciling a female referent with a typically male professional role apparently required less cognitive effort. Moreover, regressions launched from the last region show that the neutral condition may be integrated more easily with a masculine rather than a feminine anaphoric pronoun. This finding may represent a wrap-up effect emerging at the end of the sentence, after all the available information presented in the text had been collected. In this case, it may reflect a generally easier integration for the masculine as compared to the feminine referent when no specific gender cue is available, as in the case of neutral context.

Finally, follow-up typicality ratings, collected immediately after the eye-tracking session, showed less extreme ratings as compared to the pre-test ratings, for the male and particularly for the female condition. This finding is surprising since it was the female typicality that triggered the significant mismatch effect. In other words, participants found it particularly difficult to associate the representation of a male referent to a female occupation in the online measure, while the explicit ratings show that the female roles were judged as partially suitable also for men. We believe that participants may have been primed with counter-stereotypical representations of the roles through the recent exposure to the eye-tracking stimuli. While the present experiment was not designed to determine such a priming effect, it is plausible to suspect such an effect after a task where participants had to perform the cognitive task to integrate a stereotypical gender context with the gender incongruent referent. As shown by the eye movement data, this task may have been particularly surprising and consequently more salient for the female condition, thus priming later, on the offline ratings, a more equal representation of the gender distribution in the typical occupational roles.

EXPERIMENT 2

Experiment 1 investigated the effect of typicality with the help of highly gender-typical items. However, the selection of such items excluded occupational roles in the range between gender-typical and neutral (see the Materials section for details). Therefore, the second experiment examines the following research question: Do occupational roles which are judged as slightly typical—but not as gender-neutral—affect the process of anaphor resolution? In other words, do readers develop a probabilistic cognitive expectation of referent gender when reading a description of roles with low gender typicality, such as *psychologist* or *lawyer*, which were rated as only *slightly* female and *slightly* male in the off-line measures?

Method

Participants

Twenty-nine students (17 women and 12 men) from the University of Sussex, UK, participated in the study. Participants were native English speakers, with normal or corrected-to-normal vision (mean age = 21 years, SD = 2.4). None of them had participated in Experiment 1. They received monetary compensation or course credit for their participation. All participants provided written informed consent before taking part in the study.

Design and Hypothesis

The experiment was designed to test the interaction between the gender typicality of the occupational role (*typicality*: slightly male, slightly female, or neutral) and the gender of the anaphoric reference (*pronoun*: masculine or feminine). If stimuli with moderate degrees of gender typicality can elicit expectations on the referent gender, then a disruption in the reading process would emerge when the mismatching pronoun is presented. This disruption would result in longer fixation times and higher probabilities of regressions. No effect is expected with neutral priming stimuli.

Materials

Item structure was identical to the one used in Experiment 1. In Experiment 2, the priming context was constituted of slightly male, slightly female, or neutral occupational roles. The selection of the roles was based on the role noun pretest (see Materials section, Experiment 1). We selected items with role noun typicality ratings between 2.5 and 3.5 (slightly male), 4.5 and 5.5 (slightly female) and 3.5 and 4.5 (neutral) on a 7-point Likert scale for gender typicality, where 1 represented the pole of male and 7 the pole of female typicality ($M_{s.male}$ = 2.99, SD = 0.16, $M_{s.female}$ = 4.98, SD = 0.31, $M_{neutral}$ = 4.04, SD = 0.14). (3) and (4) are examples of a slightly male (3) and a slightly female (4) experimental item:

(3) C. H. earned a degree in law after many years of study. Nowadays he/she does mostly paperwork.
(4) H. C. receives calls from many customers at the call-center. Regularly he/she takes short breaks.

Participants were presented with 12 slightly male, 12 slightly female, and 12 neutral role descriptions. In addition, we randomly presented 50 filler sentences (the same items as in Experiment 1), and 24 content-related questions to promote attentive reading.

Procedure and Analysis

The experimental procedure with eye-tracking recordings and the analyses were identical to those in Experiment 1. No significant difference emerged in the distribution of missing data across typicality conditions for all regions and fixation duration measures [$M_{s.male}$ = 42.00; $M_{s.female}$ = 35.00; $M_{neutral}$ = 46.88, $F_{(2, 45)}$ = 1.01, *ns*]. The mixed-effect models included participants and items as random effects. As fixed effects we included *typicality* (slightly male, slightly female, neutral), *pronoun* (masculine, feminine), *region length*

(in fixation duration measures) and *participant gender*, Model<-lmer(fixation_time ~ typicality * pronoun * participant_gender * region_length + (1 |participants) + (1 |items).

Results

Eye-tracking Results

The final models for each measure and region (including all significant random effects, fixed effects, and interactions) are reported in Supplementary Material (Table S1). Means and standard deviations of fixation duration time and percentages of regressions are reported in **Table 4**. Details on statistical results are reported in **Table 5**. We report below eye-tracking measures presenting statistically significant fixed effects of *typicality*, *pronoun*, and *typicality*pronoun* ($p < 0.05$), and corresponding significant or marginally significant ($p < 0.1$) results of contrast analyses, separated for measure. Contrast analyses tested each typicality condition followed by the masculine and feminine pronoun (slightly male–he vs. slightly male–she; slightly female–he vs. slightly female–she; neutral–he vs. neutral–she).

First fixation time

A main effect of typicality emerged on the second region of the target sentence. Pairwise comparisons between all the factor levels showed no reliable difference, $M_{s.male}$ = 191, $M_{s.female}$ = 186, $M_{neutral}$ = 186, *ns*.

First pass time

The interaction between typicality and pronoun emerged on the pronoun region. Pairwise comparisons, however, showed no significant effect, $M_{s.maleHE}$ = 234, $M_{s.maleSHE}$ = 245, *ns*; $M_{s.femaleHE}$ = 240, $M_{s.femaleSHE}$ = 257, *ns*; $M_{neutralHE}$ = 251, $M_{neutralSHE}$ = 257, *ns*.

Regressions into a region

The interaction between *typicality* and *pronoun* emerged in regressions in the first region of the target sentence. Contrast analyses showed a significant effect for the female priming condition, where the congruent trials presented fewer regressions as compared to the incongruent ones, $M_{s.femaleSHE}$ = 1.6, $M_{s.femaleHE}$ = 2.5, $t_{(978)}$ = 2.48, p = 0.01. The effect was also significant for the male condition, with congruent trials presenting fewer regressions as compared to the incongruent ones, $M_{s.maleHE}$ = 2.4, $M_{s.maleSHE}$ = 3.5, $t_{(978)}$ = −2.14, p = 0.03. No effect was found for the neutral priming condition, $M_{neutralHE}$ = 2.1, $M_{neutralSHE}$ = 2.3, *ns*.

Regressions out

Regressions out of the last region showed a main effect of typicality. Pairwise comparisons revealed a smaller proportion of regressions for the neutral condition as compared to the slightly male condition, $M_{s.male}$ = 14.1, $M_{neutral}$ = 7.2, $t_{(33)}$ = −2.58, p = 0.01, as well as a tendency for the neutral condition to present fewer regressions as compared to the slightly female condition, $M_{s.female}$ = 11.2 $M_{neutral}$ = 7.2, $t_{(33)}$ = −1.75, p = 0.09. Probability of regressions did not differ for female and male conditions, $M_{s.female}$ = 11.2, $M_{s.male}$ = 14.1, *ns*.

TABLE 4 | Means (standard deviations) of fixation duration time (ms) and percentages of regressions for Experiment 2.

Region	Typ.	Pron.	Reg. 1		Reg. 2		Reg. 3		Reg. 4	
FF	Male	He	208	(74)	195	(58)	207	(70)	240	(104)
		She	211	(74)	195	(57)	202	(67)	232	(94)
	Fem.	He	209	(78)	198	(58)	222	(79)	230	(95)
		She	216	(80)	202	(67)	220	(77)	222	(84)
	Neutr.	He	207	(71)	211	(78)	212	(84)	219	(207)
		She	196	(58)	196	(69)	218	(79)	217	(196)
FP	Male	He	237	(87)	249	(117)	300	(142)	347	(218)
		She	238	(91)	269	(128)	289	(154)	352	(234)
	Fem.	He	254	(107)	292	(141)	331	(148)	339	(254)
		She	251	(104)	315	(168)	325	(138)	369	(251)
	Neutr.	He	250	(107)	278	(140)	336	(228)	310	(250)
		She	240	(87)	266	(140)	327	(166)	347	(240)
RP	Male	He	263	(246)	326	(230)	492	(365)	988	(719)
		She	256	(153)	354	(247)	484	(430)	976	(790)
	Fem.	He	267	(129)	357	(242)	536	(497)	912	(672)
		She	261	(140)	368	(245)	538	(439)	896	(646)
	Neutr.	He	280	(218)	318	(227)	687	(640)	796	(632)
		She	261	(147	331	(235)	638	(560)	916	(826)
TT	Male	He	294	(152	413	(232)	448	(238)	423	(245)
		She	323	(213	427	(292)	455	(266)	438	(276)
	Fem.	He	316	(190	454	(276)	450	(233)	415	(250)
		She	282	(142	450	(264)	447	(204)	436	(309)
	Neutr.	He	303	(145	425	(239)	495	(306)	359	(277)
		She	305	(170	419	(245)	485	(276)	397	(285)
RI	Male	He	19	(39)	37	(49)	24	(43)	–	–
		She	28	(45)	35	(48)	25	(44)	–	–
	Fem.	He	20	(40)	28	(45)	20	(40)	–	–
		She	10	(31)	29	(46)	20	(40)	–	–
	Neutr.	He	17	(37)	32	(47)	16	(37)	–	–
		She	18	(39)	36	(48)	17	(38)	–	–
RO	Male	He	1	(11)	13	(34)	30	(46)	57	(50)
		She	2	(13)	14	(35)	25	(44)	57	(50)
	Fem.	He	2	(13)	10	(31)	25	(44)	52	(49)
		She	2	(13)	7	(25)	27	(45)	53	(50)
	Neutr.	He	3	(17)	6	(23)	36	(48)	43	(50)
		She	3	(17)	10	(31)	36	(48)	43	(50)

FF, first fixation time; FP, first pass time; RP, regression path; TT, total time; RI, regressions into the region; RO, regressions out of the region.

Total fixation time

A main effect of participant gender emerged on the pronoun region. Contrasts revealed no significant difference, $M_{men} = 363$, $M_{women} = 355$, *ns.*

Gender Typicality Ratings

Typicality ratings for Experiment 2 are reported in Supplementary Material (Table S3). Follow-up typicality ratings correlated with the pretest ratings of the role nouns

TABLE 5 | Statistical results for Experiment 2.

	First fixation time		First pass time		Total time	
	(DF) F-value	Pr (>F)	(DF) F-value	Pr (>F)	(DF) F-value	Pr (>F)
FIRST REGION						
Pronoun	(1861) 0.026	0.871	(1831) 0.225	0.635	(1895) 0.103	0.748
Typicality	(2857) 1.430	0.240	(239) 1.234	0.302	(238) 1.589	0.217
Pron. * Typ.	(2855) 1.315	0.269	(2828) 0.065	0.937	(2899) 0.054	0.948
PRONOUN REGION						
Pronoun	(1903) 2.399	0.122	(1878) 0.171	0.679	(1844) 2.970	0.085
Typicality	(2905) 6.839	0.001**	(2330) 0.486	0.620	(232) 1.550	0.228
Pron. * Typ.	(2898) 0.545	0.580	(227) 3.872	0.021*	(2923) 0.371	0.690
SPILLOVER REGION						
Pronoun	(1918) 0.009	0.923	(1761) 0.749	0.387	(1940) 0.001	0.981
Typicality	(232) 2.127	0.136	(232) 0.239	0.788	(230) 3.050	0.062
Pron. * Typ.	(2913) 0.968	0.380	(2760) 0.367	0.693	(2933) 0.106	0.899
FINAL REGION						
Pronoun	(1812) 0.655	0.418	(1761) 0.749	0.387	(1781) 1.500	0.221
Typicality	(2814) 1.725	0.179	(232) 0.239	0.789	(233) 0.928	0.405
Pron. * Typ.	(2808) 0.040	0.961	(2760) 0.367	0.692	(2780) 1.080	0.339

	Regression path		Regressions in		Regressions out	
FIRST REGION						
Pronoun	(1834) 0.171	0.680	(1978) 0.004	0.952	(10) 0.048	0.826
Typicality	(229) 0.165	0.848	(233) 1.628	0.212	(20) 1.014	0.363
Pron. * Typ.	(230) 0.038	0.963	(2978) 5.466	0.004*	(20) 0.048	0.952
PRONOUN REGION						
Pronoun	(1812) 0.024	0.877	(1980) 0.097	0.756	(1980) 0.211	0.646
Typicality	(233) 0.440	0.648	(233) 1.221	0.308	(233) 2.014	0.150
Pron. * Typ.	(2515) 0.324	0.723	(2975) 0.437	0.646	(2978) 1.757	0.173
SPILLOVER REGION						
Pronoun	(1903) 0.348	0.556	(1980) 0.049	0.824	(1978) 0.190	0.663
Typicality	(232) 1.772	0.186	(233) 1.670	0.204	(233) 1.682	0.202
Pron. * Typ.	(2900) 0.744	0.475	(2975) 0.053	0.948	(2976) 0.681	0.506
FINAL REGION						
Pronoun	(1767) 0.002	0.968	–	–	(1978) 0.037	0.847
Typicality	(2769) 2.562	0.078	–	–	(233) 3.461	0.043*
Pron. * Typ.	(2757) 0.379	0.684	–	–	(2975) 0.048	0.953

Significance codes: ""p < 0.05; "***"p < 0.001.*

($r = 0.827$, $p < 0.001$). As a whole, follow-up typicality ratings did not differ from pre-test ratings, $M_{\text{pretest}} = 4.0$, $M_{\text{follow-up}} = 4.1$, $t_{(70)} = 0.325$, ns. When analyzed separately, male and female typicality turned out to be more skewed toward neutrality in the ratings collected after the eye-tracking experiment, $M_{\text{s.male, pretest}} = 2.99$, $M_{\text{s.male, follow-up}} = 3.34$, $t_{(22)} = -2.86$, $p = 0.009$; $M_{\text{s.female, pretest}} = 4.98$, $M_{\text{s.female, follow-up}} = 4.68$, $t_{(22)} = 2.20$, $p = 0.039$; $M_{\text{neutral, pretest}} = 4.04$, $M_{\text{neutral, follow-up}} = 4.16$, $t_{(22)} = 1.07$, ns.

The mismatch effect found in eye movements did not correlate with explicit typicality ratings (β's ≤ 0.07).

Discussion

Experiment 2 documents an effect of slightly gender-typical roles on the resolution of mismatching anaphoric personal pronouns, manifest in an early to intermediate stage of sentence processing. As in Experiment 1, gender typicality cues were conveyed through sentences describing a professional activity. In this experiment the occupations had been rated as only slightly typical for men or women, or as neutral. Still, slightly typical contexts were able to trigger the mismatch effect, as opposed to neutral priming trials. When description typicality and pronoun gender mismatched, readers regressed to the beginning of the target sentence, in order to re-check information and eventually resolve the gender conflict. The description-paradigm proved to be sensitive, showing that low degrees of typicality may evoke an impairment in the resolution process, and may thus be considered an adequate tool for investigating gender typicality, even when typical gender cues are too subtle to be categorized as "stereotypical."

Differently from Experiment 1, in Experiment 2 the mismatch effect emerged in relation to both gender priming contexts. This may be explained by the fact that the second experiment presented slightly typical contexts, which may not produce a specific difficulty for the integration of the two gender conditions, as in the case of the integration of male referents in highly stereotypical roles. In other words, in the second study both gender priming conditions produced a reading impairment, as opposed to the neutral priming condition, in which integration with the pronoun did not prove problematic.

GENERAL DISCUSSION

The study presented a paradigm to investigate the effect of gender typicality on pronominal anaphor resolution without relying on role nouns as antecedents. Gender typicality was prompted through descriptions of occupational roles. Results showed that gender typicality was conveyed effectively, that it affected the process of anaphor resolution in both a condition of high (Experiment 1) and low (Experiment 2) degree of the priming gender context. Incongruence between gender typicality of the description and pronoun gender produced a mismatch cost, which was mainly located on the pronoun region and its immediate spillover for fixation duration measures, and at the beginning and ending of the target sentence for the regression measures. While in Experiment 1 the explicit ratings could predict eye movements, no correlation was found in Experiment 2.

Taken together, these results offer insight into the representational format of gender typicality beliefs. First, the results suggest that the cognitive process of correcting for and integrating the initial mismatching gender representation exhibited a different time course in the two experiments: a more complex repair strategy involving early and late stages of processing was applied in the case of highly typical items, whereas less typical items only affected an early to intermediate stage of sentence processing.

Second, the results suggest that the effect of gender typicality can have two different cognitive sources: gender typicality and gender stereotypes. Gender typicality refers to the cognitive representation of the proportion of men and women in certain occupational roles and can be measured through explicit ratings. Gender stereotypes are cognitive representations which associate an occupational role with a specific gender and may be implicit, i.e., may not be directly measurable through typicality ratings, but can be captured with indirect methods such as eye movements during reading. The cognitive dissociation between these two factors is evident in the results of Experiment 2, where items possessed a low degree of gender typicality. Based on explicit ratings, the roles (e.g., manager, politician) were not classified as gender-typical, but they still triggered a mismatch effect in the eye-tracking measures, due to an automatic association of the professional role with a gender stereotype. Therefore, we can conclude that the concept of gender typicality could actually be split into two cognitive components: an explicit one, which can be recorded through classical typicality ratings

and corresponds to beliefs on the distribution of men and women in a specific field, and an automatic one, which is revealed with indirect methods and is stored in readers' long-term memory together with the semantics of the respective role.

Furthermore, a cross-linguistic comparison with studies on grammatical gender languages suggests that the presence or absence of a grammatical gender system in the investigated language may play a key role in the processing of gender typicality cues, even when morphological/grammatical gender cues are not present in the text, but only cognitively available to the reader. More specifically, we argue that a grammatical gender system may make gender typicality cues more salient in comparison to a natural gender language. This is, however, open to debate [cf. Irmen and Rossberg, 2004; Gygax et al., 2008, on the relation between gender typicality and grammatical gender]. In a study employing a picture categorization paradigm in Italian and Spanish, Cubelli et al. (2011) show that grammatical gender is automatically activated, even if its retrieval is not required to accomplish the task. This consideration may suggest that gender information is already available in the cognitive representation of a reader possessing a grammatical gender system—even when no morphological markings are required for comprehension or presented in the stimuli—and trigger a faster processing of the gender mismatch.

Finally, a cross-linguistic comparison of the present study with grammatical gender language studies reveals a similar finding on the asymmetrical distribution of the gender mismatch effect, which had been previously reported only in studies on languages with a grammatical gender system (in Italian, Cacciari and Padovani, 2007; in German, Irmen et al., 2010). Specifically, pairwise contrasts in Experiment 1 revealed a significant effect in the condition of the masculine pronoun related to the incongruent female context, but no effect on the feminine pronoun related to the incongruent male context. In a study with event related potentials, Siyanova-Chanturia et al. (2012) document an N400-like effect for the masculine pronoun only, preceded by an incongruent typically female role noun (e.g., *insegnante-lui*). The N400 is assumed to represent a violation in semantic expectations, which is also at the basis of the gender mismatch asymmetry effect in eye movements. Our findings in English supports the cross-linguistic evidence that gender stereotypes may affect the processing of masculine and feminine anaphors differently. Socio-psychological theories on expectations related to gender roles may be required to explain this effect, as it may not only be due to the features of a particular gender system. However, further comparative studies and replications are necessary to determine the exact role of the gender system of a reader's language on the interpretation of gender-typical cues and its interaction with the process of anaphor resolution.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg. 2015.01607

REFERENCES

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint* arXiv:1406.5823.

Boland, J. E. (2004). "Linking eye movements to sentence comprehension in reading and listening," in *The Online Study of Sentence Comprehension,* eds M. Carreiras and C. Clifton, Jr (New York, NY: Psychology Press), 51–76.

Cacciari, C., Corradini, P., Padovani, R., and Carreiras, M. (2011). Pronoun resolution in Italian: the role of grammatical gender and context. *J. Cogn. Psychol.* 23, 416–434. doi: 10.1080/20445911.2011.526599

Cacciari, C., and Padovani, R. (2007). Further evidence on gender stereotype priming in language: semantic facilitation and inhibition on Italian role nouns. *Appl. Psycholinguist.* 28, 277–293. doi: 10.1017/S0142716407070142

Cubelli, R., Paolieri, D., Lotto, L., and Job, R. (2011). The effect of grammatical gender on object categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 449–460. doi: 10.1037/a0021965

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye-movements and comprehension processes when text conflicts with world knowledge. *Mem. Cogn.* 32, 551–559. doi: 10.3758/BF03195846

Esaulova, Y., Reali, C., and von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Lang. Cogn. Neurosci.* 29, 781–803. doi: 10.1080/01690965.2013.794295

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au pairs are rarely male: norms on the gender perception of role names across English, French, and German. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206

Garnham, A., Oakhill, J., and Reynolds, D. (2002). Are inferences from stereotyped role names to characters'gender made elaboratively? *Mem. Cogn.* 30, 439–446. doi: 10.3758/BF03194944

Gygax, P., Gabriel, U., Sarrasin, O., Garnham, A, and Oakhill, J. (2008). Generically intended, but specifically interpreted: when beauticians, musicians, and mechanics are all men. *Lang. Cogn. Neurosci.* 23, 464–485. doi: 10.1080/01690960701702035

Hellinger, M., and Bußmann, H. (2001). *Gender Across Languages: The Linguistic Representation of Women and Men,* Vol. 1. Amsterdam: Benjamins.

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res* 36, 431–456. doi: 10.1007/s10936-007-9053-z

Irmen, L., Holt, D.V., and Weisbrod, M. (2010). Effects of role typicality on processing person information in German: evidence from an ERP study. *Brain Res.* 1353, 133–144. doi: 10.1016/j.brainres.2010.07.018

Irmen, L., and Rossberg, N. (2004). Gender markedness of language: the impact of grammatical and nonlinguistic information on the mental representation of person information. *J. Lang. Soc. Psychol.* 23, 272–307. doi: 10.1177/0261927X04266810

Irmen, L., and Schumann, E. (2011). Processing grammatical gender of role nouns: further evidence from eye movements. *J. Cogn. Psychol.* 23, 998–1014. doi: 10.1080/20445911.2011.596824

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., and Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Front. Psychol.* 6:506. doi: 10.3389/fpsyg.2015.00506

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

Kreiner, H., Sturt, P., and Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007. 09.003

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2013). *LmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (Lmer Objects of lme4 Package).* R package version 2.0-0. Available online at: http://CRAN. Rproject.org/package=lmerTest/

Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cogn.* 33, 972–983. doi: 10.3758/BF03193206

Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cogn.* 25, 273–285. doi: 10.3758/BF03211283

Osterhout, L., and Mobley, L. (1995). Event-related brain Potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033

Pyykkönen, P., Hyönä, J., and van Gompel, R. P. (2010). Activating gender stereotypes during online spoken language processing. *Exp. Psychol.* 57, 126–133. doi: 10.1027/1618-3169/a000016

Reali, C., Esaulova, Y., and von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: evidence from eye movements. *Appl. Psycholinguist.* 36, 977–1006. doi: 10.1017/S01427164140 00010

R Core Team (2012). *R: A Language and Environment for Statistical Computing.* Vienna: The R Foundation. Available online at: http://www.R-project.org/

Sato, S., Gygax, P., and Gabriel, U. (2013). Gender inferences: grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism* 16, 792–807. doi: 10.1017/S1366728912000739

Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE* 7:e48712. doi: 10.1371/journal.pone. 0048712

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3

Gender stereotypes across the ages: On-line processing in school-age children, young and older adults

Anna Siyanova-Chanturia[1]*, Paul Warren[1], Francesca Pesciarelli[2] and Cristina Cacciari[2]

[1] School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, New Zealand,
[2] Department of Biomedical, Metabolic and Neurological Sciences, University of Modena and Reggio Emilia, Modena, Italy

Most research to date on implicit gender stereotyping has been conducted with one age group – young adults. The mechanisms that underlie the on-line processing of stereotypical information in other age groups have received very little attention. This is the first study to investigate real time processing of gender stereotypes at different age levels. We investigated the activation of gender stereotypes in Italian in four groups of participants: third- and fifth-graders, young and older adults. Participants heard a noun that was stereotypically associated with masculine (*preside* "headmaster") or feminine roles (*badante* "social care worker"), followed by a male (*padre* "father") or female kinship term (*madre* "mother"). The task was to decide if the two words – the role noun and the kinship term – could describe the same person. Across all age groups, participants were significantly faster to respond, and significantly more likely to press 'yes,' when the gender of the target was congruent with the stereotypical gender use of the preceding prime. These findings suggest that information about the stereotypical gender associated with a role noun is incorporated into the mental representation of this word and is activated as soon as the word is heard. In addition, our results show differences between male and female participants of the various age groups, and between male- and female-oriented stereotypes, pointing to important gender asymmetries.

Keywords: gender stereotypes, on-line language processing, implicit measure, children, young adults, older adults

Introduction

Gender stereotyping, for better or worse, occurs frequently in everyday life. We seem to readily attribute masculine gender to doctors, surgeons, and politicians, and feminine gender to nurses, school teachers, and secretaries. When our personally held beliefs are compromised in one way or another, we feel obliged to provide additional information, as suggested by terms like <u>male nurse</u> or <u>female soldier</u>. This occurs even when other clues already point to the gender of the referent, as in the following example: *Military rules ban pregnant servicewomen from front-line duties, though last year another <u>female</u> British soldier gave birth two weeks after returning from her six-months deployment to Afghanistan* [BBC News, 24th March, 2013, emphasis added].

In contexts where there is no explicit information about the gender associated with a specific occupation (*doctor, nurse*), personal trait (*aggressive, nurturing*), or activity (*mending, laundering*), we rely on our beliefs and background knowledge to infer – sometimes erroneously – the more likely gender. A wealth of psycholinguistic studies has looked at the activation of

stereotypical gender information during language processing. Specifically, it has been widely documented that when language users encounter stereotypically incongruent information (male nurse or female doctor), their processing slows down (Banaji and Hardin, 1996; Carreiras et al., 1996; Garnham et al., 2002; Duffy and Keir, 2004; Oakhill et al., 2005; Cacciari and Padovani, 2007; Kreiner et al., 2008; Pyykkönen et al., 2010; Siyanova-Chanturia et al., 2012). These studies have shown that stereotypical gender information is incorporated into the mental representation of the role noun in question (doctors/surgeons/politicians are assumed to be males, while nurses/teachers/secretaries are assumed to be females), and that gender activation occurs at the time a role noun is encoded (Oakhill et al., 2005; Siyanova-Chanturia et al., 2012). These and other studies have used a range of methodologies, paradigms, and tasks to investigate moment-by-moment processing of stereotypical gender information, predominantly in young adults. As we will see in the following review of previous research, the processing of stereotypical gender violations in other age groups – such as children and older adults – remains poorly understood. There is also relatively little data that indicate whether the stereotypicality effects vary with the sex of the participants or with the gender indicated by the linguistic items involved.

Young Adults

The bulk of the research on the processing of gender stereotypes has focused on young adults, and has shown that linguistic information congruent with stereotypes is processed more rapidly than incongruent information. An early study (Banaji and Hardin, 1996) measured response times in judging the grammatical gender of personal pronouns (he, she) that followed prime words that were gender-biased either definitionally (mother, father) or because of stereotyped use (nurse, doctor). Responses were slower when there was a mismatch between the gender of the prime and that of the pronoun, especially for definitional terms (mother followed by he). Participants also responded significantly more quickly to targets that matched their own gender. Banaji and Hardin's (1996) paradigm has since been used in other behavioral studies and in studies measuring event-related brain potentials (ERPs; Cacciari and Padovani, 2007; Siyanova-Chanturia et al., 2012). Interestingly, Cacciari and Padovani (2007) found a stereotype incongruency effect with masculine pronouns (secretary-he) but not with feminine pronouns (engineer-she).

In a study employing a similar paradigm, and one on which the current study is based, Oakhill et al. (2005) asked participants to read word pairs in which a stereotypically male or female role noun (engineer or secretary, respectively) was followed by a kinship term that was either congruent (engineer – brother) or incongruent (engineer – sister), and to decide for each pair whether they could be used to refer to the same person. Participants responded more rapidly to congruent than incongruent word pairs, even when they were explicitly instructed to suppress their gender stereotypes.

In an eye-tracking study (Duffy and Keir, 2004), test sentences contained masculine and feminine role nouns as antecedents to stereotypically congruent or incongruent reflexive pronouns

(The electrician taught himself/herself …). Test sentences were preceded either by a discourse context specifying the sex of the referent or by a sex-neutral context. In the neutral contexts, automatic activation of gender stereotypical information encoded in the role nouns resulted in higher processing costs and longer fixation times when the test sentences contained incongruent pronouns (for similar results, see Irmen, 2007; Pyykkönen et al., 2010; also see Esaulova et al., 2014). However, when the preceding context signaled that the character's sex matched the reflexive pronoun, the incongruency effect disappeared. Kreiner et al. (2008) similarly found congruency effects on fixation times in anaphoric sentences (where the reflexive follows the noun to which it refers: Yesterday the minister left London after reminding himself/herself about the letter, where the reflexive pronoun refers to a preceding noun) but not in cataphoric sentences (where the reflexive precedes its noun: After reminding himself/herself, the minister immediately went to the meeting at the office).

ERP studies have shown that the brain response to gender stereotype violations in language might be indexed by two different components, the N400 and the P600. The N400 is a negative-going deflection peaking around 400 ms. after stimulus onset that has traditionally been shown to reflect semantic and world knowledge violations (for an overview, see Kutas and Federmeier, 2011). The P600 is a slow positive shift emerging 500–900 ms. after stimulus onset, traditionally associated with syntactic violations, but also linked to semantic anomaly (Osterhout and Holcomb, 1992, 1995; Kuperberg et al., 2003; Kim and Osterhout, 2005; Bornkessel-Schlesewsky and Schlesewsky, 2008). White et al. (2009) presented participants with a gender category (men/women) followed by a word stereotypically associated with males (aggressive) or females (nurturing). Participants judged whether or not the two words matched, according to their beliefs about gender stereotypes. Stereotypically incongruent combinations (men/nurturing, women/aggressive) elicited a larger N400 than congruent ones. In their ERP study, Siyanova-Chanturia et al. (2012) used Banaji and Hardin's (1996) paradigm described above with native speakers of Italian. Participants judged the grammatical gender of a personal pronoun (equivalent to English he, she) following either a definitionally gendered noun (mother, father) or a gender stereotyped role noun (teacher, driver). After definitionally gendered nouns, incongruent pronouns (mother/he or father/she) resulted in a N400 effect, but after stereotypically gendered nouns, this effect was found only with male targets (teacher/he but not driver/she), suggesting that participants were more accepting of female drivers than male teachers.

In an earlier ERP study of gender stereotypes, Osterhout et al. (1997) observed larger P600s when the stereotypical gender of an antecedent role noun was incongruent with the gender of a reflexive pronoun (doctor – herself) than when it was congruent. They found a stronger effect for female than for male participants, suggesting that females have stronger gender stereotypes. Finally, Irmen et al. (2010) conducted a study in German, in which participants read statements involving occupations (florists, pilots) followed by masculine, feminine, or neutral anaphoric noun phrases (these men/women/people).

While the occupations are stereotypically associated with males or females, all of the nouns representing them had masculine grammatical gender. When the anaphors were semantically incongruent with their antecedents, feminine anaphors produced more positive P600 responses than masculine anaphors. Irmen et al. (2010) suggested this was because the masculine anaphors were congruent at least with the masculine grammatical gender of the female antecedent noun, and that this eased integration, compared with the feminine anaphors.

A range of behavioral, eye-tracking, and ERP studies have thus shown stereotypical gender effects in experiments with young adults. These studies suggested that information about stereotypical gender – denoting an occupation or a personal characteristic – is incorporated into the reader's representation of a word, and that this information is difficult to suppress during on-line language processing. A few of these studies have also reported asymmetries that depend on the gender indicated by the words involved and on the sex of the participant, with stronger incongruency effects reported for combinations of female stereotypes with masculine pronouns and from female participants.

Older Adults

Most of the stereotype research with older populations has focused on racial rather than gender stereotypes, with older adults frequently found to be more prejudiced than younger adults. Following Devine (1989), it has been widely hypothesized firstly that what sets apart prejudiced and non-prejudiced individuals is the extent to which they are able to suppress stereotyped behavior, and secondly that this ability diminishes with age. For instance, von Hippel et al. (2000) found that younger (18–25 years of age) but not older (65–95) adults were able to ignore racial stereotypes when rating the intelligence of two characters presented as African American and Caucasian. Similarly, Gonsalkorale et al. (2009) found that older adults showed greater implicit bias because of their inability – relative to young adults – "to regulate automatically activated associations" (p. 412), and Radvansky et al. (2010) found that older adults (60–88) drew on and maintained racial stereotypic references to a much greater extent than younger adults (18–25).

To the best of our knowledge, only one study has investigated the processing of gender stereotypical information in older adults. In a self-paced reading study (Radvansky et al., 2009), younger (18–22) and older (60–87) adults read a series of short stories (adapted from Duffy and Keir, 2004). Critical sentences contained sequences such as *The babysitter/plumber found herself/himself* … in which the reflexive pronoun was either congruent or incongruent with the gender stereotyped occupation of the character. Both young and older adults showed an effect of congruency on reading time. But unlike the findings in studies on racial prejudice, both groups of adults were found to be capable of suppressing gender stereotypes when counter-stereotypic information was provided in the preceding context. There is no indication as to whether this varied with participant sex or with the gender of the stereotyped items.

So while research on racial stereotypes suggests that older adults may be less able to suppress the activation of stereotypical information than younger adults, the small amount of relevant research suggests that this may not be the case with gender stereotypes.

Children

Gender stereotyping in children has received increasing attention in recent years, with a particular focus on the development of stereotype behavior during childhood. Hill and Flom (2007) found sensitivity to gender stereotypes at 24 months but not at 18 months, using a preferential looking paradigm in which children watched male and female actors performing masculine and feminine stereotypical activities. An earlier study (Poulin-Dubois et al., 2002) used a generalized imitation paradigm in which children selected a male and a female doll to imitate masculine and feminine stereotypical activities. They found that 24-month-old girls, but not boys, were sensitive to the violation of gender stereotypical activities.

Research with children has also addressed the question of *stereotypical gender asymmetry*, that is, whether gender stereotyping is less restrictive for female than for male stereotypes, as predicted, for instance, by Social Role theory (Eagly and Steffen, 1984; Diekman and Eagly, 2000; Eagly et al., 2000). Wilbourn and Kee (2010) asked 8- and 9-year-old children to create sentences that paired male and female proper names with stereotypically masculine and feminine occupations. The results showed that children were less likely to think of males engaging in traditional feminine activities (*Henry-nurse*) than the other way around (*Mary-doctor*). As noted above, a similar asymmetry has recently been found in young adults (Siyanova-Chanturia et al., 2012, see also Cacciari and Padovani, 2007; Irmen et al., 2010; Reali et al., 2014).

Banse et al. (2010) considered both stereotype knowledge and stereotype flexibility in groups of 5-, 8-, and 11-year-old children. Stereotype knowledge is reflected in automatic stereotyping that occurs independently of whether the individual considers the stereotypes to be accurate or not, while stereotype flexibility involves a recognition that stereotypes can be wrong (see also Signorella et al., 1993; Trautner et al., 2005). The children were assessed on how they associated gender-stereotyped common objects (*iron, hammer*) with men and women, and gender-stereotyped toys (*doll, truck*) with boys and girls. The results showed that gender stereotype knowledge for toys was at ceiling as early as 5 years of age, and for common objects reached ceiling levels by 11 (while already very high at five). Stereotype flexibility, that is, the realization that stereotypes are not immutable, showed a considerable increase from the age of 5–11, and, unlike stereotype knowledge, was higher at all ages for common objects than for toys. Differences between girls and boys and between female- and male-related stereotypes were reported neither for stereotype knowledge nor for stereotype flexibility.

Few studies in this area tap into the processes underlying the moment-by-moment comprehension of gender stereotypes. A notable exception is Most et al. (2007), who used an auditory Stroop paradigm in which young adults and third-graders (∼9 years old) categorized the sex of voices that pronounced male and female proper names, or stereotypically male (*football*), female (*makeup*), or neutral (*paper*) words. Both children and

adults were slower when the voice's sex was incongruent either with the gendered stereotype of the spoken word (*makeup* spoken with a male voice; *football* spoken with a female voice) or with the gender of the proper name (*Cindy* spoken with a male voice; *Jason* spoken with a female voice). This suggests that implicit gender associations are already present in 8- to 9-year-old children. Unfortunately, it is not clear how the gender stereotypicality of the heterogeneous types of target words (nouns, adjectives, verbs, names of activities, objects, professions, concrete, and abstract words) was established. Nor does there seem to be any control of the lexical properties that are known to affect the time it takes to decode a word stimulus (such as word frequency, length, etc.).

The studies cited above provide a sketch of the development of gender stereotypical behavior in children. Automatic stereotyping is evident from an early age and firmly in place by about 11 years. At the same time, children show evidence from 5 to 11 that they are increasingly able to override their stereotype behaviors. Stronger sensitivity to stereotype violations has been reported for very young girls than for boys, and there is some evidence that children are more sensitive to gender incongruencies in which stereotypically feminine roles are paired with male persons.

The Present Study

The majority of the studies on gender stereotyping conducted with children and older adults employed explicit off-line measures such as questionnaires, off-line reading, and judgment and classification tasks. Although such measures usefully elucidate social beliefs and attitudes, they do not provide information on the underlying moment-by-moment processes that can be revealed by real-time measures such as reaction times, eye-tracking, and ERPs – methodologies that have so far been mostly used with young adults. In addition, previous studies have had little to say either on the possible differences between the sexes in terms of behavior with gender stereotypes, or on the possibility that female- and male-gendered language may be responded to differently.

The present study therefore aimed to use the same real-time measure to assess gender stereotype behavior with a range of ages, namely third- and fifth-graders (∼8 and 10 years of age), young adults (mean age of 24), and older adults (mean age of 77). It also aimed to assess differences between female and male participants and between female- and male-gendered stereotypes. To achieve this, we adapted Oakhill et al.'s (2005) paradigm outlined earlier. Participants had to decide whether two words – a gender-biased occupational role and a kinship term – could describe the same person. The two terms formed either a stereotypically congruent pair (*engineer – brother*) or an incongruent pair (*secretary – father*). Our adaptation of the paradigm was that we used auditory rather than visual presentation of the stimuli, since this seemed better suited for testing participants with different reading abilities. Response choices ('yes'/'no') and decision times for those choices were collected.

Our predictions are that all age groups will show sensitivity to the violation of gender stereotypical information, but that the

extent of this sensitivity will be age-dependent. In particular, we predict that adults will show greater stereotype flexibility and be better able than children to suppress gender stereotypes and therefore to accept the incongruent role-kinship pairs as possibly referring to the same person. The evidence from racial stereotypes indicates that older adults are less well able to suppress stereotypes than younger adults, while Radvansky et al. (2009) suggest that this may not be the case for gender stereotypes. It remains an empirical question, therefore, whether the results for older adults will show the same or lowered rates of suppression of gender stereotypes compared with those for young adults. We will look to the extent and speed of acceptance that the incongruent pairs may refer to the same person as a measure of this. Within the two groups of children, we predict that the change in stereotype flexibility demonstrated by Banse et al. (2010) for children between the ages of 5 and 11 will be reflected in stronger and more rapid acceptance of incongruent pairs by our fifth-graders than by our third-graders.

As far as the gender of the tested words is concerned, we note that the research reviewed above showed asymmetries both for 8- and 9-year-old children (Wilbourn and Kee, 2010) and for young adults (Siyanova-Chanturia et al., 2012). We predict that for our data, these groups will be more likely to accept the combination of male roles with female kinship terms (*engineer-sister*) than vice versa (*secretary-brother*). We have no reason not to expect the same of our older adults.

With regard to sex differences between our participants, we predict – on the basis of the study with very young children by Poulin-Dubois et al. (2002) – that young girls will show greater stereotype flexibility than young boys, and will therefore be more likely to accept incongruent pairs. There is little direct evidence cited above that addresses this issue in adults, but our prediction is that by adulthood, male participants will show similar degrees of stereotype flexibility as females.

Method

Participants

Our young adult group comprised 28 students at the University of Modena and Reggio Emilia (13 females, mean age: 24.1, range: 20–30, *SD:* 4.3) who participated in the experiment for course credit or a small gift (equivalent of €10).

Our group of older adults was made up of 30 cognitively preserved older adults (14 females, mean age: 77.4, range: 72–82, *SD:* 2.5) with homogenous educational and socio-economic backgrounds. They all achieved a Mini-Mental State Evaluation score (MMSE, Folstein et al., 1975) equal to, or higher than 26 (*M* = 28.2, *SD:* 1.7, range: 26–30) and had at least 10 years of formal education. They did not receive a gift for their participation.

Our two groups of children consisted of 43 third-graders (20 females, mean age: 8.5, range: 7.9–9.5, *SD:* 0.4) and 42 fifth-graders (17 females, mean age: 10.4, range: 9.7–11.2, *SD:* 0.3) from the same school in the province of Modena, Emilia Romagna (Italy). They received a small gift (equivalent of €3) for their participation. The use of these two age groups was based on our

review of earlier studies which suggests that these groups fall in a period of development where stereotype flexibility is increasing rapidly (Signorella et al., 1993; Trautner et al., 2005; Banse et al., 2010). We decided not to test children younger than third grade because of the task demands of a paradigm that requires high accuracy and speed.

All participants were residents in the province of Modena, Emilia Romagna (Italy). They were informed of their rights and gave written informed consent for participation in the study (for children, this consent was granted by their parents), according to the Declaration of Helsinki, and in line with the ethical requirements of the University of Modena and Reggio Emilia.

Materials

Material selection followed two stages of norming, with adults and with children. In all cases rating scales were used, with the scale poles reversed for half of the participants. None of the participants used in the norming studies also took part in the main experiment. An initial set of 260 Italian words (nouns, past participles, and adjectives), morphologically unmarked for gender and specifying occupations, roles and individual characteristics, was presented in two questionnaires (each containing 130 words) to 40 students (20 females). Participants rated the extent to which each word was associated with men, women, or both, using a seven-point scale. From this initial set, 60 words were selected that were rated as highly male-oriented (30 words) or female-oriented (30 words). A further set of 40 participants subsequently rated the valence (positive, negative, or neutral connotations) of the 60 selected words.

To ensure the 60 selected words were familiar to third- and fifth-graders and had gender associations from the children's perspective, they were included in additional questionnaires presented to 133 children (half third-graders and half fifth-graders; half females). Participants selected from three options, indicating that the words could be used: (1) only for men, (2) for both men and women, (3) only for women. There was a fourth option – 'I don't know' – in case the word was not known to the participant; this option always appeared last. The questionnaire also included filler items morphologically marked for gender (*amico* "male friend"). If a child performed poorly on such items, then their data were excluded from the norming procedure.

On the basis of this norming, we selected nine words that received the highest ratings of male-oriented stereotypicality (*preside* "headmaster"), and nine words that received similarly high ratings of female-oriented stereotypicality (*badante* "social care worker") in the adult rating task. All 18 selected words were known to both third- and fifth-graders. The male- and female-oriented words did not significantly differ in stereotypicality for either adults or children, nor in their valence. The words in the two groups were also comparable in terms of frequency (*Repubblica* corpus, Baroni et al., 2004), length (number of characters), and in the durations of the recorded tokens used in the experiment (see below). Norming and lexical statistics are summarized in **Table 1**.

The selected items (see Appendix) used one of three nominal endings not associated with a specific grammatical gender. Each

TABLE 1 | Mean log frequency, length, stereotypicality, valence, and millisecond duration of target stimuli.

	Male stereotype	Female stereotype	*p*
Log frequency	3.1 (2.4–3.7) *0.4*	2.3 (0.0–3.7) *1.2*	=0.11
Length (characters)	8.8 (4.0–13.0) *2.8*	9.3 (7.0–12.0) *1.5*	=0.60
Adult stereotypicality	2.6 (2.0–3.5) *0.5*	2.9 (1.3–3.8) *0.8*	=0.17
Child stereotypicality	1.7 (1.6–1.8) *0.1*	1.8 (1.4–1.9) *0.2*	=0.09
Valence	4.5 (3.8–5.8) *0.6*	4.9 (4.1–5.5) *0.5*	=0.20
Duration (ms)	779 (526–1032) *175.3*	826 (659–972) *107*	=0.22

Range is indicated in parentheses and standard deviation in italics below.

of the groups of male- and female-oriented words contained five words ending in *-ista*, three in *-e*, and one in a consonant.

Following Oakhill et al. (2005), each of the 18 role nouns was paired with each of six paired kinship terms: *sorella* "sister," *fratello* "brother," *madre* "mother," *padre* "father," *moglie* "wife," *marito* "husband," resulting in three stereotypically congruent and three stereotypically incongruent word pairs for each role noun. The words in each kinship pair were comparable in terms of their lexical characteristics (see **Table 2**).

The tokens of all words used in the experiment were created using ALFa Reader 3 voice synthesizer software. We used speech production software rather than a human voice to make the recording as neutral as possible (free of regional accents, personal traits, etc.). Two native speakers of Italian judged the recordings to be natural and to have native-like prosody.

Procedure

Participants were seated comfortably in a silent room. In each trial, a fixation point (+) appeared in the center of a computer screen for 1500 ms. followed by a blank screen for 350 ms. Participants then heard the prime (role noun) and 250 ms. later the target (kinship term), and decided whether the two words could describe the same person. They were instructed to listen carefully to both words and to press the 'yes' or 'no' button on a button box as quickly and accurately as possible (button positions were reversed for half of the participants). The subsequent trial began after the response. To ensure that the gender of the voice

TABLE 2 | Mean log frequency, length, and millisecond duration of the six kinship terms used in the experiment.

	madre *padre*	*sorella* *fratello*	*moglie* *marito*
Log frequency	4.6 4.8	4.0 4.3	4.4 4.7
Length (characters)	5 5	7 8	6 6
Duration (ms)	440 435	668 680	616 587

did not bias participants' response (for such an effect, see Most et al., 2007; Van Berkum et al., 2008), half of the participants listened to the words pronounced with a male-synthesized voice and half with a female-synthesized voice.

The experiment comprised six blocks of 36 trials (216 word pairs in total). Trials were pseudorandomised in each block, but each block contained equal numbers of stereotypically congruent and incongruent pairs, and of feminine and masculine role nouns. Each role noun occurred only once in each block. In addition to the 18 test pairs, each block contained 18 filler pairs (half congruent, half incongruent), whose primes were role nouns morphologically marked for gender (*amico* "male friend," *ragazza* "girl"). These filler pairs provided a measure of performance accuracy in the task (see below).

The experimental session was preceded by a practice block of 20 trials (half congruent and half incongruent word pairs of the same type as the fillers). After each block, participants were invited to take a short break.

Analysis and Results

A total of 16 participants (11% of the original 143) were excluded for one or more of the following reasons: they exceeded the 25% error rate threshold on the fillers ($N = 3$), they were non-native speakers of Italian ($N = 4$), they did not follow the instructions ($N = 3$), they were identified as having hearing problems or learning difficulties ($N = 3$) or as not being naïve to the nature of the experiment ($N = 1$), or because of equipment failure ($N = 2$). One further participant was excluded on the basis of having exceptionally long response times (a mean response time more than 2.5 standard deviation from the mean for their age group). Hence the analyses were conducted on 34 third-graders (17 females), 39 fifth-graders (17 females), 26 young adults (13 females), and 27 older adults (14 females). The mean error rate on the fillers for the retained participants was 9.7% for third-graders, 7.8% for fifth-graders, 3.9% for young adults, and 7.9% for older adults. The distribution of response times across both experimental and filler items was examined within each age group, and cut-off times determined for the group. A total of 2.06% of responses were removed.

Mixed effects models were computed over responses to the experimental items for response choice (logistic regression) and response times (linear regression), using the *lme4* package in R (Bates et al., 2015). The *afex* package (Singmann et al., 2015) was used to determine Chi-square values and significance levels for relevant factors. For the response choice analysis, the dependent variable was the selection of 'yes' or 'no' (prime and target could or could not describe the same person). In the analysis of response times, since these did not follow a normal distribution, the effect of a range of transformations was tested, and the inverse square root function [transformed RT = 1/sqrt(RT)] selected as the best fit to a normal distribution.

Following model comparison, the random effects structure for both response choice and response time analyses included random intercepts for participant, prime (the stereotyped role word) and target (the kinship term), and random slopes by

participants across the sequence of blocks in the experiment. The fixed effects were Participant Sex, Age Group (third grade, fifth grade, young adult, or older adult), Block[1], Target Gender (female or male kinship term), and Congruence (the target word formed a congruent or incongruent pair with the prime word).

Response Choice Analysis

To test our predictions that there will be age-dependent sensitivity to the violation of gender stereotypical information, and that there will be asymmetries in the acceptability of incongruent items depending on the gender of the items, we ran a model including as predictors Congruence, Age Group, and Target Gender, as well as Block[2]. **Figure 1** presents a summary of the proportions of 'yes' responses by Congruence, Age Group, and Target Gender.

The statistical model confirmed simple effects of Block (participants increasingly respond with 'yes' across blocks: $\chi^2 = 63.90$, df: 1, $p < 0.0001$), Congruence (the proportion of 'yes' responses was higher for congruent pairs: $\chi^2 = 1166.22$, df: 1, $p < 0.0001$), Target Gender (more 'yes' responses after pairs with male targets, $\chi^2 = 15.18$, df: 1, $p < 0.0001$) and Age Group ($\chi^2 = 69.60$, df: 3, $p < 0.0001$). The overall effect of Age Group reflects a very high level of 'yes' responses for young adults, with lower levels for old adults, then fifth-graders and, finally, third-graders. As well as demonstrating an overall effect, Congruence was involved in a number of interactions. Therefore, we subsequently modeled congruent and incongruent conditions separately, with Block, Target Gender, and Age Group as predictors.

Congruent Pairs

The analysis of congruent pairs revealed simple effects of Block ($\chi^2 = 15.14$, df: 1, $p < 0.0001$), Age Group ($\chi^2 = 68.04$, df: 3, $p < 0.0001$), and Target Gender ($\chi^2 = 7.59$, df: 1, $p < 0.01$), and a significant interaction of Target Gender and Age Group ($\chi^2 = 25.36$, df: 3, $p < 0.0001$). As can be seen from the upper panel of **Figure 1**, the interaction of Target Gender and Age Group reflects the fact that there were considerably more 'yes' responses to male targets than to female targets in congruent pairs for third graders, with smaller Target Gender differences in the same direction for older adults and fifth graders, and virtually no difference for young adults.

Subsequent analysis of each age group in the congruent condition showed no effects for young adults; this is hardly

---

[1]Block was included because of widely-reported speeding-up and slowing-down effects across response time experiments which can add variance to response data. The block effects are not important in the context of our predictions above and will not be discussed in detail. However, it is interesting to note that in the response time analyses adults typically sped up across the experiment but the children slowed down. Importantly, additional analyses showed that these age-related block effects did not interact with the experimental variables in this study, suggesting that our participant groups did not change their performance strategically in response to patterns they may have noticed in the stimuli.

[2]This initial analysis did not also include Participant Sex, because models including this as a predictor failed to converge. This is because the female young adults produced 100% 'yes' responses to subsets of data, a situation which produces inflated standard error in logistic regression. Since our prediction concerning Participant Sex relates primarily to children rather than adults, this factor will be reintroduced in subsequent analyses of each age group.
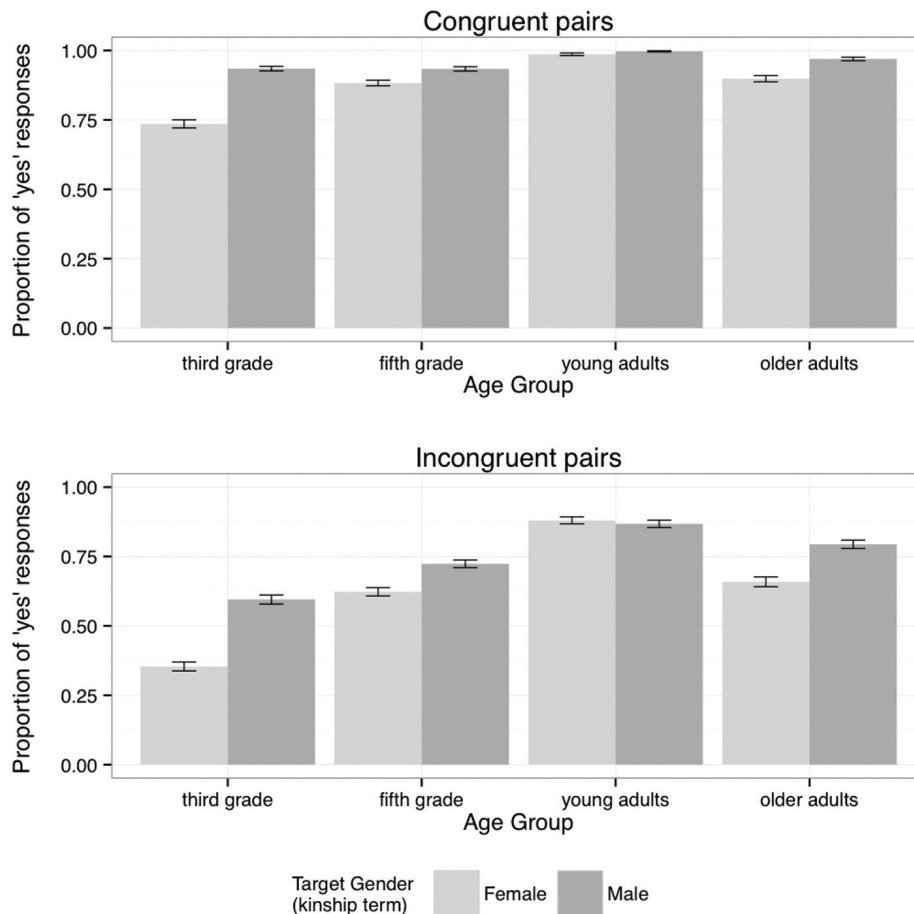
FIGURE 1 | Proportions of 'yes' responses for congruent and incongruent prime-target pairs, by Age Group and Target Gender (mean and standard error).

surprising given the ceiling-level performance that is visible in the top panel of **Figure 1**. For the other three groups we were able to introduce Participant Sex into the models (see Footnote 2). For each group there was a significant interaction of Target Gender and Participant Sex (older adults: $\chi^2 = 17.28$, df: 1, $p < 0.0001$; third-graders: $\chi^2 = 16.31$, df: 1, $p < 0.0001$; fifth-graders: $\chi^2 = 4.18$, df: 1, $p < 0.05$). In each case, there was a larger Target Gender difference for male participants than for female participants for both of these groups. In addition, males gave fewer 'yes' responses than females when the target was female, but more when it was male (see **Table 3**).

Incongruent Pairs

As with the congruent pairs, the analysis of the incongruent pairs revealed a significant interaction of Target Gender and Age Group ($\chi^2 = 35.60$, df: 3, $p < 0.0001$). In addition, there were significant simple effects of Block ($\chi^2 = 44.13$, df: 1, $p < 0.0001$), Age Group ($\chi^2 = 49.79$, df: 3, $p < 0.0001$), and Target Gender ($\chi^2 = 4.46$, df: 1, $p < 0.05$). The lower panel of **Figure 1** shows that the interaction of Target Gender and Age Group is similar to that found for the congruent pairs, but is more strongly marked.

The largest difference for Target Gender is for the third graders, followed by older adults, then fifth graders and finally young adults, who have a very small difference in the opposite direction. As before, each of these age groups was subsequently analyzed in separate models, which included Participant Sex. The young adult data showed no effects of Participant Sex or Target Gender. The older adults showed a significant effect for Target Gender ($\chi^2 = 4.34$, df: 1, $p < 0.05$), and this effect was marginally significant for the fifth-graders ($\chi^2 = 3.37$, df: 1, $p = 0.07$).

TABLE 3 | Proportion of 'yes' responses by Age Group, Participant Sex and Target Gender, congruent pairs.

	Female participants		Male participants	
	Female targets	Male targets	Female targets	Male targets
Third grade	0.79	0.92	0.68	0.95
Fifth grade	0.89	0.91	0.88	0.95
Young adults	0.99	1.00	0.98	0.99
Older adults	0.91	0.95	0.89	0.99

A major difference in the case of the third graders is that there was a significant interaction of Target Gender and Participant Sex ($\chi^2 = 29.37$, df: 1, $p < 0.0001$), as well as a significant simple effect for Target Gender ($\chi^2 = 9.61$, df: 1, $p < 0.005$). **Table 4** shows that, although both third-grade boys and girls gave more 'yes' responses to incongruent items that had a male kinship term as the target, this difference was over twice as large for boys as for girls.

Response Choice Data: Summary

Overall, congruent pairs led to higher levels of 'yes' responses (responding that the prime and target could describe the same person) than incongruent pairs. This is true of all age groups, supporting our prediction that all groups will show sensitivity to stereotype violation. Importantly, though, this effect varies across age groups, and is most marked with the younger children, indicating that they have the lowest level of stereotype flexibility.

The incongruent pairs showed strong age-related effects, as well as Target Gender effects, with male targets following female role nouns receiving higher proportions of 'yes' responses than female targets, particularly from the younger children. This finding is contrary to our prediction that combinations of male roles with female kinship terms will be more acceptable than vice versa. Note though that a similar Target Gender difference was identified in the congruent condition. We also found interactions of Participant Sex and Target Gender for congruent pairs for all age groups except for the young adults, but only for the third-graders for the incongruent pairs. We will return to these findings in the Discussion section.

Response Times

The dependent variable in this analysis was the set of transformed response times (using the inverse square root transformation). For clarity, however, the graphs below present the untransformed mean response times. Two sets of analyses were carried out, one for the 'yes' responses and one for the 'no' responses. The second of these included only responses to incongruent pairs, because the low numbers of 'no' responses to congruent pairs in some combinations of predictors made it difficult to obtain reliable regression models (see Footnote 2).

'Yes' Responses

Our initial analysis included the predictors Congruence, Age Group, Participant Sex, Target Gender, and Block. This revealed

a significant three-way interaction of Age Group, Participant Sex, and Target Gender ($\chi^2 = 22.25$, df: 3, $p < 0.0001$), significant two-way interactions of Participant Sex and Target Gender ($\chi^2 = 39.96$, df: 1, $p < 0.0001$) and Age Group and Congruence ($\chi^2 = 9.72$, df: 3, $p < 0.05$), and simple effects of Block ($\chi^2 = 19.87$, df: 1, $p < 0.0001$), Congruence ($\chi^2 = 255.63$, df: 1, $p < 0.0001$), and Age Group ($\chi^2 = 51.62$, df: 3, $p < 0.0001$).

Figure 2 illustrates the three-way interaction between Age Group, Participant Sex, and Target Gender – the different age groups clearly show different effects of the interaction of Participant Sex and Target Gender. The overall effect of Age Group is also obvious in this figure. **Figure 3** presents the interaction of Age Group and Congruence and shows how the Congruence effect differs in size but not in direction across the groups. That is, all groups more readily accept stereotype-matching pairs than incongruent pairs, with this effect stronger for the children and smallest for the young adults. To explore the two interaction effects involving Age Group, separate analyses were carried out for each group.

For the young adults, the only significant effects were for Block ($\chi^2 = 24.20$, df: 1, $p < 0.0001$) and Congruence ($\chi^2 = 27.18$, df: 1, $p < 0.0001$). These young adults were faster in accepting congruent pairs than in accepting incongruent pairs, and their responses sped-up across the experiment. The older adult group similarly demonstrated significant effects for Block ($\chi^2 = 30.56$, df: 1, $p < 0.0001$) and Congruence ($\chi^2 = 90.35$, df: 1, $p < 0.0001$), but also a significant interaction of Participant Sex and Target Gender ($\chi^2 = 10.90$, df: 1, $p < 0.001$). This interaction arises because the male participants responded more quickly to male than to female targets, while the female participants showed no difference (see **Figure 2**).

The older of the two groups of children showed a significant interaction between Participant Sex and Target Gender ($\chi^2 = 18.29$, df: 1, $p < 0.0001$) and a simple effect of Congruence ($\chi^2 = 114.44$, df: 1, $p < 0.0001$), with no other effects. The interaction of Participant Sex and Target Gender has the same pattern as reported above for the older adults. The third grade participants also showed this significant interaction of Participant Sex and Target Gender ($\chi^2 = 29.80$, df: 1, $p < 0.0001$), as well as simple effects of Block ($\chi^2 = 4.25$, df: 1, $p < 0.05$) and Congruence ($\chi^2 = 65.54$, df: 1, $p < 0.0001$). The Block effect reflects a slowing-down as the experiment progressed (see Footnote 1). As with the older adults and fifth graders, the interaction of Participant Sex and Target Gender arises because male participants responded faster to male than to female targets, while females responded equally fast to both.

'No' Responses

As noted above, the analysis of 'no' responses included only the incongruent pairs. This analysis showed significant interactions of Target Gender with Age Group ($\chi^2 = 8.53$, df: 3, $p < 0.05$) and with Participant Sex ($\chi^2 = 4.79$, df: 3, $p < 0.05$), a significant interaction of Age Group with Block ($\chi^2 = 22.47$, df: 3, $p < 0.0001$), and an overall simple effect of Age Group ($\chi^2 = 11.25$, df: 3, $p < 0.01$).

TABLE 4 | Proportion of 'yes' responses by Age Group, Participant Sex and Target Gender, incongruent pairs.

	Female participants		Male participants	
	Female targets	Male targets	Female targets	Male targets
Third grade	0.37	0.51	0.33	0.67
Fifth grade	0.61	0.68	0.63	0.76
Young adults	0.92	0.89	0.84	0.85
Older adults	0.65	0.77	0.67	0.82
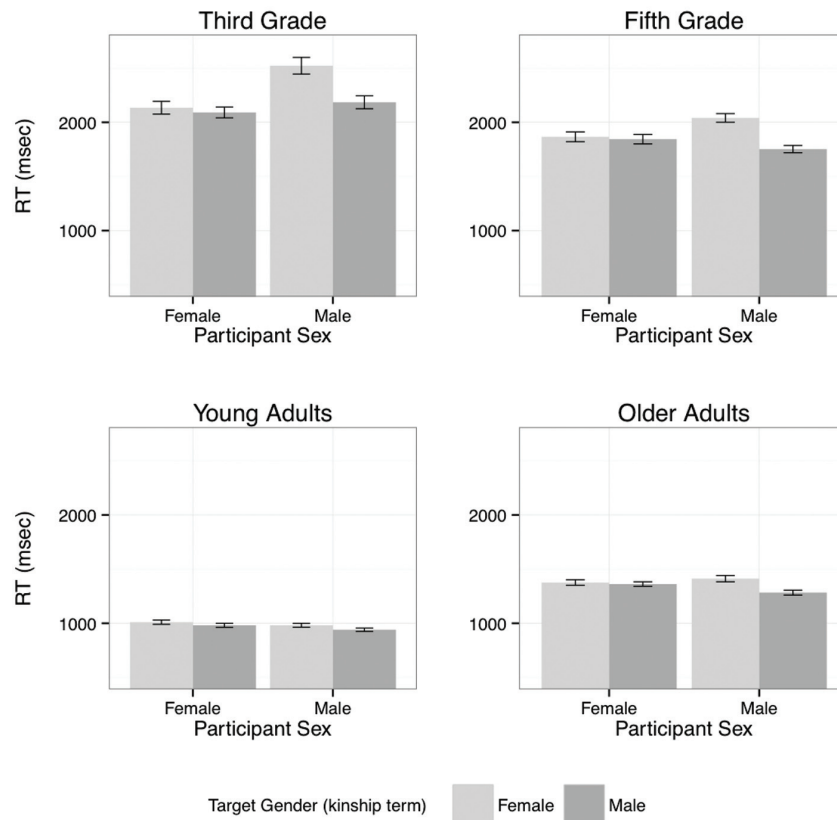
FIGURE 2 | Response times for 'yes' responses by Age Group, Participant Sex, and Target Gender (mean and standard error).
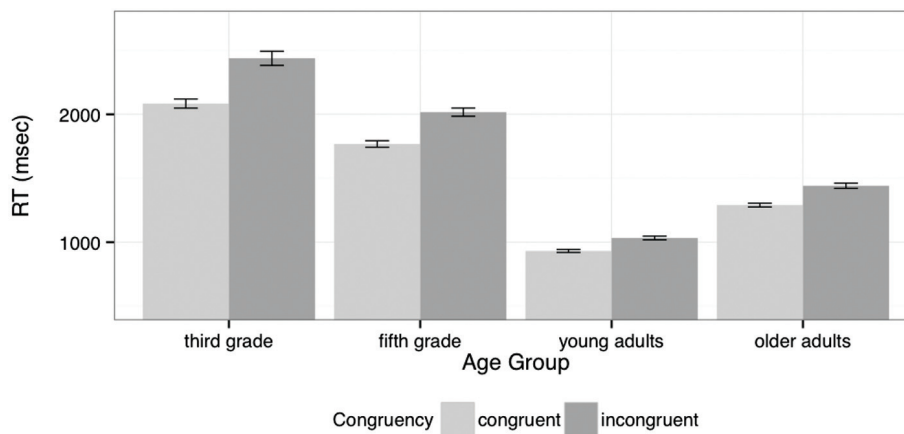
FIGURE 3 | Response times for 'yes' responses by Age Group and Congruence (mean and standard error).

The interactions of Target Gender with Age Group and Participant Sex are illustrated in the left and right panels of **Figure 4** respectively. In the left panel we see that the fifth grade children gave faster 'no' responses to incongruent pairs that involved a male target following a female prime, while the adult groups and the third grade children showed no such difference. This pattern was confirmed in further analyses for

each age group: the young and older adults and the third graders showed no effect of Target Gender (all $ps > 0.7$), whereas the difference was significant for the fifth graders ($\chi^2 = 4.17$, df: 1, $p < 0.05$). The interaction of Target Gender with Participant Sex shown in the right panel is one of degree rather than of direction (contrast the interaction effects for these variables in the analysis of 'yes' response

FIGURE 4 | Response times for 'no' responses to incongruent items (mean and standard error). The left panel shows the interaction of Target Gender and Age Group and the right panel shows the interaction of Target Gender and Participant Sex.

times) – both male and female participants gave faster 'no' responses to incongruent pairs involving a male target (following a female prime), but this difference was larger for the male participants.

Response Time Data: Summary

Overall, young adults responded fastest, followed by older adults, fifth graders and lastly third graders. All groups were faster in accepting congruent than incongruent pairs. The groups differed from one another in the relative effects on 'yes' response times of the 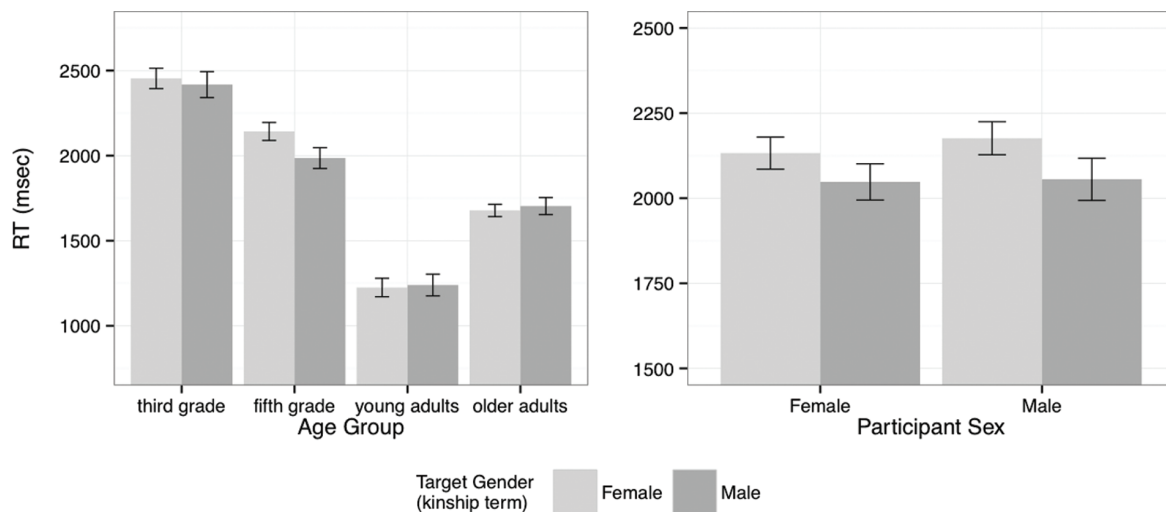interaction of Participant Sex and Target Gender. This interaction arose because while there was little difference in the response times of female participants that depended on the gender of the kinship term used as the target noun, responses from male participants were faster to male targets than to female targets (reflecting the increased level of 'yes' responses after these targets noted above). This difference was strongest for the third graders, followed by the fifth graders and the older adults, but the effect was not significant for the young adults. The interaction of Participant Sex and Target Gender in the 'no' responses to incongruent pairs shows a similar pattern, that is, a stronger difference between responses to male and female targets from the male participants.

Discussion

In a timed decision task, Italian third- and fifth-graders, young adults, and older adults were required to decide as quickly as possible if two auditorily presented words – a masculine or feminine stereotypical word combined with either a male or a female kinship term – could be used to describe the same person. Participants across all age groups were significantly more likely to respond 'yes' and to do so more rapidly when the kinship term

was preceded by a stereotypically congruent than incongruent role noun. These results provide evidence that language users of various ages – school-age children, young and older adults – are biased by gender stereotypes when making judgments about the likely identity of people fulfilling certain roles. In addition, the higher processing cost of responding to incongruent pairings of roles and kinship terms is reflected in the response time differences – across all age groups – between incongruent and congruent conditions.

Our results are in line with those of Oakhill et al. (2005), as well as those of a range of studies using a variety of behavioral, eye-tracking and ERP techniques, predominantly with young adults (Carreiras et al., 1996; Garnham et al., 2002; Duffy and Keir, 2004; Cacciari and Padovani, 2007; Most et al., 2007; Kreiner et al., 2008; Pyykkönen et al., 2010; Siyanova-Chanturia et al., 2012). Importantly, our study extends the evidence-base for automatic gender stereotype effects to children and older adults, and highlights the contribution that on-line measures can make to the assessment of gender stereotyping across ages. In contrast to more traditional off-line measures, which have been widely used in studies with children and older adults, on-line measures are based largely on automatic processes that are believed to be free of strategic responses.

In addition to finding a general effect of gender stereotypes across ages, we also discovered a number of important differences in the processing of stereotypical gender (in)congruencies in children and adults. As noted in our introduction, Radvansky et al. (2010) suggested that older adults might not show the drop-off in stereotype flexibility (operationalised as an inability to suppress gender stereotypes) that has been reported for racial stereotypes. However, taking the proportion of 'yes' responses to incongruent pairs as one measure and the speed with which such 'yes' responses are made as another, we see that older adults, like the two groups of children, show lowered levels of stereotype flexibility, that is, of being able to identify that

a stereotype can be wrong. In line with Devine (1989), von Hippel et al. (2000) and Gonsalkorale et al. (2009), we take our results to support the idea that older individuals are less likely to suppress their prejudiced behavior and are less able to regulate automatically activated associations when compared to younger adults. Our results thus appear to go against those reported in Radvansky et al. (2009) who found no reliable differences between young and older adults' ability to discount gender stereotypical bias. It should be noted, however, that in Radvansky et al. (2009), counter-stereotypic information was *explicitly* provided to the participants, while in our study, no such information was present.

Within our two groups of children, we find results that are compatible with developmental stages of stereotype acquisition, in particular, with an increase in stereotype flexibility between the ages of 5 and 11 (Signorella et al., 1993; Trautner et al., 2005; Banse et al., 2010). Our 8-year-olds were found to be less flexible, less likely to press 'yes' following a stereotypically incongruent word pair, and slower in doing so than our 10-year-olds. The latter made 'yes' choices at a level comparable to that of the older adults, although the children's responses were slower (as were their 'yes' responses to congruent pairs).

Another interesting set of effects pertains to the gender of the target (kinship) word. This is the asymmetry in the processing of incongruent pairs, predominantly in the data from our children and older adults, which favored the pairing of female roles with male kinship terms. This asymmetry is in the opposite direction to that predicted on the basis of previous results with children (Wilbourn and Kee, 2010) and young adults (Cacciari and Padovani, 2007; Siyanova-Chanturia et al., 2012; Reali et al., 2014). Note, however, that we further found that male targets in the congruent condition also received more and faster 'yes' responses than female targets for these groups. In other words, this asymmetry affects more than just the processing of incongruent pairs. In addition, we found participant sex differences in the decision choice and response times of the same three groups across both congruent and incongruent conditions, in interaction with these target gender effects. First, we found larger differences between the proportions of 'yes' responses to male and female targets for the male children and older adults than for their female counterparts, with the males providing fewer 'yes' responses than the females after female kinship terms, but more 'yes' responses than the females after male kinship terms. Second, these groups differed in how quickly female and male participants pressed 'yes' following female and male kinship terms. While female participants' response times did not differ with the gender of the kinship term, male participants' responses to male targets were faster than to female targets. In line with these findings, the analysis of the 'no' responses to incongruent pairs further suggested a bigger difference in responses to male and female targets for male than female participants.

How can we explain such gender asymmetries? We interpret the fact that male children and older adults responded more quickly to male kinship items, and their tendency, when

compared to female participants, to prefer male kinship terms, as a reflection of the use of the social category "male" as the standard – or unmarked normative group – against which other categories are judged. According to social psychologists, one group (males) can become more "normative" than another (females), being the unmarked normative group (Hegarty and Pratto, 2001). For example, Miller et al. (1991) showed that when asked to think of a prototypical voter, most people think of a male voter exemplar. Researchers have argued that such "androcentrism" is common (Bem, 1993; Hegarty and Pratto, 2001), and that attitudes, beliefs, and stereotypes are more influenced by male exemplars than female ones (Eagly and Kite, 1987). It seems that social "androcentrism" affects male and female children and older adults differently, in that females, being members of the marked normative group, may be more sensitive and able to correct for the bias than males, being members of the unmarked normative group. Interestingly, no such effect was observed for our young adults, implying that age plays an important role in one's ability to correct for the "unmarked group effect" and to inhibit stereotypical representations.

In addition, the gender asymmetry reflected in interactions of Target Gender and Participant Sex appears to be consistent with the claim of Miller et al. (2009) that boys generally have stronger stereotypical biases than girls, especially in the domain of activities. According to Miller et al. (2009), girls tend to confirm less strictly than boys to gender-role stereotypes. Interestingly, Miller et al. (2009) also maintain that gender stereotypes are differentially accessible when children think about males and females. These authors, as well as others (Higgins and King, 1981; Higgins, 1996), define accessibility as the readiness with which a construct is retrieved from memory. Our findings suggest that female children are equally fast to access and accept male and female constructs (kinship terms), while males more rapidly access male constructs than female ones.

In summary, our findings support the view according to which information about the stereotypical gender associated with occupations is incorporated into the representation of words denoting these occupations and is activated as soon as such a word is encountered. Importantly, the present study has gone beyond young adults to unveil the mechanisms of on-line processing of gender stereotypical information, as well as notable gender asymmetries associated with such processing, in two under-researched age groups – school-age children and cognitively preserved older adults.

Acknowledgments

Supplementary Material

References

Banaji, M., and Hardin, C. D. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x

Banse, R., Gawronski, B., Rebetez, C., Gutt, H., and Morton, B. (2010). The development of spontaneous gender stereotyping in childhood: relations to stereotype knowledge and stereotype flexibility. *Dev. Sci.* 13, 298–306. doi: 10.1111/j.1467-7687.2009.00880.x

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., et al. (2004). "Introducing the "la Repubblica" corpus: a large, annotated, TEI(XML)-compliant corpus of newspaper Italian," in *Proceedings of LREC* 2004, Lisbon: European Language Resources Association.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). *lme4: Linear Mixed-Effects Models using Eigen and S4. R Package Version 1.1-8.* Available at: http://CRAN.R-project.org/package=lme4

BBC News (2013). Available at: http://www.bbc.co.uk/news/uk-19657646 [accessed March 24, 2013].

Bem, S. L. (1993). *The Lenses of Gender*. New Haven, CT: Yale University Press.

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Res. Rev.* 59, 55–73. doi: 10.1016/j.brainresrev.2008.05.003

Cacciari, C., and Padovani, R. (2007). Further evidence of gender stereotype priming in language: semantic facilitation and inhibition in Italian role nouns. *Appl. Psycholinguist.* 28, 277–293. doi: 10.1017/S0142716407070142

Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol.* 49A, 639–663. doi: 10.1080/713755647

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18. doi: 10.1037/0022-3514.56.1.5

Diekman, A. B., and Eagly, A. H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Pers. Soc. Psychol. Bull.* 26, 1171–1188. doi: 10.1177/0146167200262001

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cognit.* 32, 551–559. doi: 10.3758/BF03195846

Eagly, A., and Kite, M. (1987). Are stereotypes of nationalities applied to both men and women? *J. Pers. Soc. Psychol.* 53, 457–462. doi: 10.1037/0022-3514.53.3.451

Eagly, A. H., and Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *J. Pers. Soc. Psychol.* 46, 735–754. doi: 10.1037/0022-3514.46.4.735

Eagly, A. H., Wood, W., and Diekman, A. B. (2000). "Social role theory of gender differences and similarities: a current appraisal," in *The Developmental Social Psychology of Gender*, eds T. Eckes and H. M. Trautner (Mahwah: Lawrence Erlbaum), 123–174.

Esaulova, Y., Reali, C., and von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Lang. Cogn. Neurosci.* 29, 781–803. doi: 10.1080/01690965.2013.794295

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Garnham, A., Oakhill, J., and Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Mem. Cognit.* 30, 439–446. doi: 10.3758/BF03194944

Gonsalkorale, K., Sherman, J., and Klauer, K. C. (2009). Aging and prejudice: diminished regulation of automatic race bias among older adults. *J. Exp. Soc. Psychol.* 45, 410–414. doi: 10.1016/j.jesp.2008.11.004

Hegarty, P., and Pratto, F. (2001). The effects of social category norms and stereotypes on explanations for intergroup differences. *J. Pers. Soc. Psychol.* 80, 723–735. doi: 10.1037/0022-3514.80.5.723

Higgins, E. T. (1996). "Knowledge activation: accessibility, applicability, and salience," in *Social Psychology: Handbook of Basic Principles*, eds E. T. Higgins and A. W. Kruglanski (New York, NY: Guilford), 133–168.

Higgins, E. T., and King, G. (1981). "Accessibility of social constructs: information processing consequences of individual and contextual variability," in *Personality, Cognition and Social Interaction*, eds N. Cantor and J. F. Kihlstrom (Hillsdale, MI: Erlbaum), 69–121.

Hill, S., and Flom, R. (2007). 18- and 24-month-olds' discrimination of gender-consistent and inconsistent activities. *Infant Behav. Dev.* 30, 168–173. doi: 10.1016/j.infbeh.2006.08.003

Irmen, L. (2007). What's in a (role) name? Formal and semantic aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456.

Irmen, L., Holt, D. V., and Weisbrod, M. (2010). Effects of role typicality on processing person information in German: evidence from an ERP study. *Brain Res.* 1353, 133–144. doi: 10.1016/j.brainres.2010.07.018

Kim, A., and Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *J. Mem. Lang.* 52, 205–225. doi: 10.1016/j.jml.2004.10.002

Kreiner, H., Sturt, P., and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007.09.003

Kuperberg, G. R., Sitnikova, T., Caplan, D., and Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cogn. Brain Res.* 17, 117–129. doi: 10.1016/S0926-6410(03)00086-7

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Miller, C., Lurye, L., Zosuls, K., and Ruble, D. (2009). Accessibility of gender stereotypes domains: developmental and gender differences in children. *Sex Roles* 60, 870–881. doi: 10.1007/s11199-009-9584-x

Miller, D. T., Taylor, B., and Buck, M. L. (1991). Gender gaps: who needs to be explained? *J. Pers. Soc. Psychol.* 61, 5–12. doi: 10.1037/0022-3514.61.1.5

Most, S., Sorber, A. V., and Cunningham, J. (2007). Auditory Stroop reveals implicit gender associations in adults and children. *J. Exp. Soc. Psychol.* 43, 287–294. doi: 10.1016/j.jesp.2006.02.002

Oakhill, J. V., Garnham, A., and Reynolds, D. J. (2005). Immediate activation of stereotypical gender information. *Mem. Cognit.* 33, 972–983. doi: 10.3758/BF03193206

Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cognit.* 25, 273–285. doi: 10.3758/BF03211283

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Osterhout, L., and Holcomb, P. J. (1995). "Event-related brain potentials and language comprehension," in *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*, eds M. D. Rugg and M. G. H. Coles (Oxford: Oxford University Press).

Poulin-Dubois, D., Serbin, L., Eichstedt, J., and Sen, M. (2002). Men don't put on make-up: toddlers' knowledge of the gender stereotyping of household activities. *Soc. Dev.* 11, 166–181. doi: 10.1111/1467-9507.00193

Pyykkönen, P., Hyönä, J., and Van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: evidence from visual world eye tracking. *Exp. Psychol.* 57, 126–133. doi: 10.1027/1618-3169/a000016

Radvansky, G., Copeland, D., and von Hippel, W. (2010). Stereotype activation, inhibition, and aging. *J. Exp. Soc. Psychol.* 46, 51–60. doi: 10.1016/j.jesp.2009.09.010

Radvansky, G., Lynchard, N., and von Hippel, W. (2009). Aging and stereotype suppression. *Aging Neuropsychol. Cogn.* 16, 22–32. doi: 10.1080/13825580802187200

Reali, C., Esaulova, Y., and von Stockhausen, L. (2014). Isolating stereotypical gender in a grammatical gender language: evidence from eye movements. *Appl. Psycholinguist.* 36, 977–1006. doi: 10.1017/S0142716414000010

Signorella, M. L., Bigler, R. S., and Liben, L. S. (1993). Developmental differences in children's gender schemata about others: a meta-analytic review. *Dev. Rev.* 12, 147–183. doi: 10.1006/drev.1993.1007

Singmann, H., Bolker, B., Westfall, J., Højsgaard, S., Fox, J., Lawrence, M. A., et al. (2015). *afex: Analysis of Factorial Experiments, vsn. 0.13-145.* Available at: https://cran.r-project.org/web/packages/afex/index.html

Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE* 7:e48712. doi: 10.1371/journal.pone.0048712

Trautner, N. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., and Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: developmental or differential? *Infant Child Dev.* 14, 365–381. doi: 10.1002/icd.399

Van Berkum, J. J. A., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054

von Hippel, W., Silver, L., and Lynch, M. (2000). Stereotyping against your will: the role of inhibitory ability in stereotyping and prejudice among the elderly. *Pers. Soc. Psychol. Bull.* 26, 523–532. doi: 10.1177/014616720026700

White, K. R., Crites, S. L., Taylor, J. H., and Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Soc. Cogn. Affect. Neurosci.* 4, 191–198. doi: 10.1093/scan/nsp004

Wilbourn, M. P., and Kee, D. W. (2010). Henry the nurse is a doctor too: implicitly examining children's gender stereotype flexibility for male and female occupational roles. *Sex Roles* 62, 670–683. doi: 10.1007/s11199-010-9773-7

Electrophysiology of subject-verb agreement mediated by speakers' gender

Adriana Hanulíková[1]* and Manuel Carreiras[2,3,4]

[1] *Department of German Linguistics, University of Freiburg, Freiburg, Germany,* [2] *BCBL – Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain,* [3] *IKERBASQUE, Basque Foundation for Science, Bilbao, Spain,* [4] *Departamento de Lengua Vasca y Comunicación, University of the Basque Country, Donostia, Spain*

An important property of speech is that it explicitly conveys features of a speaker's identity such as age or gender. This event-related potential (ERP) study examined the effects of social information provided by a speaker's gender, i.e., the conceptual representation of gender, on subject–verb agreement. Despite numerous studies on agreement, little is known about syntactic computations generated by speaker characteristics extracted from the acoustic signal. Slovak is well suited to investigate this issue because it is a morphologically rich language in which agreement involves features for number, case, and gender. Grammaticality of a sentence can be evaluated by checking a speaker's gender as conveyed by his/her voice. We examined how conceptual information about speaker gender, which is not syntactic but rather social and pragmatic in nature, is interpreted for the computation of agreement patterns. ERP responses to verbs disagreeing with the speaker's gender (e.g., a sentence including a masculine verbal inflection spoken by a female person 'the neighbors were upset because I *stole$_{MASC}$ plums') elicited a larger early posterior negativity compared to correct sentences. When the agreement was purely syntactic and did not depend on the speaker's gender, a disagreement between a formally marked subject and the verb inflection (e.g., the woman$_{FEM}$ *stole$_{MASC}$ plums) resulted in a larger P600 preceded by a larger anterior negativity compared to the control sentences. This result is in line with proposals according to which the recruitment of non-syntactic information such as the gender of the speaker results in N400-like effects, while formally marked syntactic features lead to structural integration as reflected in a LAN/P600 complex.

Keywords: subject–verb agreement, speaker's gender, social language processing, speaker identity, ERP, P600, N400

Introduction

An important aspect of language comprehension is that listeners are able to efficiently establish the relation between words in an utterance and to extract meaning in just the right way. In order to capture syntactic dependencies between words and their features, listeners have to keep track of surface-level agreement between the form of one linguistic unit such as the noun *cat* and another unit such as the verb *scratches*. In English, the utterance *the cat scratches* reflects a standard use of number feature agreement between the subject and the verb whereas the sentence *the cat scratch*

does not. Examining the way in which listeners respond to the standard use of agreement features provides insight into how relations between words are computed and how computational problems related to the rules of a given language or a variety are solved. Although languages vary greatly in how they reflect dependencies between words (see e.g., Corbett, 1991), previous research has shown that these dependencies are utilized and facilitate language processing in general (e.g., MacWhinney et al., 1984; Boelte and Connine, 2004; for a review see Friederici and Jacobsen, 1999).

Successful comprehension also entails the encoding of extra-linguistic information such as speaker-related characteristics. The processing of the non-standard utterance *the cat scratch* produced by a small child may not be hindered if a listener is able to use extra-linguistic information about the child's incomplete mastery of the standard verb singular agreement. In such a case, anticipating a non-standard use of syntactic dependencies could potentially facilitate rather than hinder the overall processing effort (e.g., Hanulíková et al., 2012). Anticipations can result from a listener's experience with certain speakers and their language use and can help in the interpretation and prediction of upcoming events across various modalities (e.g., Bar, 2007; Lau et al., 2014). But do listeners use speaker-specific characteristics for the computation of grammatical agreement features? Despite numerous studies on agreement, little is known about syntactic computations generated by speaker characteristics extracted from spoken language.

The question regarding how linguistic processing and speaker characteristics interact in real time has increasingly attracted research interest in cognitive neuroscience; in particular questions such as which neural mechanisms are involved and what is the time-course of speaker integration (e.g., Lattner and Friederici, 2003; Van Berkum et al., 2008; Scharinger et al., 2011; Hanulíková et al., 2012; Bornkessel-Schlesewsky et al., 2013). Such research suggests that speaker-related attributes such as gender, social group affiliation or age modulate speech perception, lexico-semantic processing and the processing of stereotypical knowledge, albeit the exact time-course varies across different linguistic levels. In contrast, little is known about how social information conveyed by a speaker's voice affects syntactic processing (e.g., Hanulíková et al., 2012).

In many languages, syntactic relations between words usually include grammatical features such as person, number and gender but also features that go beyond the sentence given and include pragmatic aspects such as features of the speaker or the addressee. Slovak (a West-Slavonic language), for example, has a rich agreement paradigm that marks multiple properties simultaneously (Corbett, 1991), involving features for number, case, and gender. Each Slovak noun bears one of three grammatical genders (masculine, feminine, or neuter) and requires agreement with determiners, attributives, predicate adjectives, verb participles, and – in the past tense – with finite verbs (Badecker and Kuminiak, 2007, p. 82). The past tense is formed with the auxiliary "be" and the so-called *l*-participle (used as active but not as passive participle) that agrees with the subject of the clause in number and gender (e.g., Migdalski, 2006). In the third person no auxiliary is present, the past tense

is expressed by the *l*-participle alone. It is called *l*-participle to reflect the fact that *l* is present in all suffixes of the participle (−*l* for masculine singular, −*la* for feminine singular, −*lo* for neuter singular). For example, the Slovak female past verb form *išla* ('went$_{FEM}$') in an utterance such as *ja som išla* ('I went$_{FEM}$') agrees with the biological gender of the female speaker (the personal pronoun *ja* 'I' is unmarked for gender and is often omitted due to Slovak being a pro-drop language). The correctness of the verb can be evaluated based on the conceptual (i.e., biological) gender of the speaker as conveyed by the speaker's voice. The utterance *ja som išla* produced by a male voice would clearly be considered ungrammatical due to the mismatch between the female participle form (*išla*) and the speaker's male gender.

Such speaker-related agreement features are found in many world languages. While within the Indo-European languages verb agreement with gender is common mainly in the Slavic subgroup (but partly also present in the French orthography), gender agreement on predicate adjectives can be seen in several languages (e.g., in Spanish *yo estaba cansada* 'I was tired$_{FEM}$'). Speaker-related gender agreement features can be considered pragmatic, because they are grammaticalized – encoded in the syntactic structure of a given language – and mark a relation between language and context (see Levinson, 1983, p. 9). Such pragmatic aspects (i.e., speaker-related physical/social information) of grammatical agreement processes have rarely been investigated. The exact nature of pragmatic (speaker-related) agreement processes compared to syntactic (speaker-independent) agreement processes remains unclear. In the present event-related potential (ERP) study, therefore, we examined pragmatic aspects of subject–verb-gender agreement by using electroencephalography (EEG) that allows for the examination of brain activity as speech unfolds over time without an additional interfering task (e.g., Hagoort and Brown, 2000; Van Berkum et al., 2008; Hanulíková et al., 2012). More specifically, we examined the nature and time course of the integration of a speaker's voice during computation of grammatical gender agreement, and how such agreement processes compare to agreement computations that depend on the formal grammatical gender of the subject and are independent of a speaker's voice.

ERP Studies on Grammatical Gender Agreement

Communication in languages with a rich inflectional system requires comprehenders to keep track of agreement features between words. Numerous ERP-studies have demonstrated that the human brain shows distinct responses to expected as opposed to unexpected use of dependencies between words during sentence processing across many languages (for a review see Molinaro et al., 2011; Steinhauer and Drury, 2012). An important and well-studied grammatical category and agreement feature across languages is gender. Gender is usually considered an inherent feature of nouns and can be either assigned based on the meaning or the form of the noun or be an arbitrary formal feature (e.g., Corbett, 2006). Gender has been central to numerous studies examining how

listeners store gender information in the brain and how it affects language production and comprehension in general and agreement processes in particular. It has been shown that listeners are sensitive to the correct use of gender, as is reflected in several studies showing that expected (congruent) gender is easier to process than unexpected (incongruent) gender (e.g., Grosjean et al., 1994; Bates et al., 1996; Boelte and Connine, 2004; for a review, see Friederici and Jacobsen, 1999). Most studies involving gender agreement effects examined grammatical gender agreement between nouns and determiners or nouns and adjectives (e.g., Münte and Heinze, 1994; Hagoort and Brown, 1999; Gunter et al., 2000; Barber and Carreiras, 2005; Van Berkum et al., 2005; Martin et al., 2012, 2014; see also Molinaro et al., 2011 for a review), while studies examining the processing of subject–verb agreement usually focus on features such as number (e.g., Kutas and Hillyard, 1983; Hagoort et al., 1993; Osterhout and Mobley, 1995; Kaan, 2002; De Vincenzi et al., 2003; Roehm et al., 2005; Silva-Pereyra and Carreiras, 2007; Zawiszewski et al., 2015) and person (e.g., Münte and Heinze, 1994; Hinojosa et al., 2003; Silva-Pereyra and Carreiras, 2007; Mancini et al., 2011; Shen et al., 2013; Zawiszewski et al., 2015). These studies have shown specific components responsive to incorrect agreement patterns such as an anterior negativity that is often left-lateralized (LAN) and peaks between 300 and 500 ms after the mismatch onset and/or a late posterior positivity (P600) peaking around 600 ms after the mismatch onset. While there still is ongoing debate about the functional significance of linguistically relevant ERP components, researchers frequently link LAN to an index of early syntactic processing (e.g., Friederici, 2002; Molinaro et al., 2011; Batterink and Neville, 2013) and a 'failure to bind' (Hagoort, 2003). The P600 on the other hand is typically associated with a later stage of processing and has been observed in response to various syntactic violations (e.g., Gouvea et al., 2010). It is assumed to index processes of syntactic integration, reanalysis, or recovery from well-formedness conflicts (Osterhout and Holcomb, 1992; Hagoort et al., 1993; Friederici, 1995, 2002; Bornkessel-Schlesewsky and Schlesewsky, 2009; for a review see Kutas and Federmeier, 2007) and may reflect controlled and strategic processes (Gunter et al., 1997; Coulson et al., 1998; Hahne and Friederici, 1999) or the competition between several syntactic unification alternatives (Hagoort, 2003).

Only a few prior studies have directly examined the processing of subject–verb agreement including grammatical gender features, i.e., cases in which verbs agree with the subject of the clause in gender. Deutsch and Bentin (2001) examined subject-predicate-gender agreement in Hebrew, in which the subject must agree with the predicate with regard to gender and number (or person for future, past, and imperative verb forms). ERP responses to predicates that were congruent or incongruent in gender with an animate (e.g., *boy*) or an inanimate (e.g., *diamond*) subject were recorded. Unlike many previous studies that report a P600 to subject–verb agreement, Deutsch and Bentin (2001) observed a larger modulation of the N400 to incongruent relative to congruent predicates, which was more pronounced in the animate than in the inanimate

condition. Furthermore, an early left anterior negativity (eLAN) was observed but only in the singular animate condition. The eLAN is an early ERP component frequently linked to phrase structure violations (e.g., Friederici, 2002). The N400 is one of the most studied ERP components often seen during semantic processing. A consistent finding across studies on the N400 effect is that its amplitude is negatively correlated with the fit of a word in the (semantic) context. The N400 has frequently been interpreted as reflecting conceptual/semantic integration or a cognitive cost associated with word recognition, often linked to predictive processing (e.g., Kutas and Hillyard, 1980; Brown and Hagoort, 1993; Kutas and Van Petten, 1994; Federmeier, 2007). The N400 result for incongruent predicates in Deutsch and Bentin's (2001) study was somewhat surprising and led to a discussion about the exact roles of formal gender and conceptual gender in agreement computations. Following Deutsch and Bentin (2001), it is the salient semantic information of an animate noun that usually functions as the thematic role of an agent and may lead to a more prominent N400 effect for the animate subject in contrast to the inanimate subject (see also Mancini et al., 2011 for similar N400-like effects in person mismatches in subject-verb agreement in Spanish).

Similar to Hebrew, Hindi future tense verbs agree in person, number, and gender with the subject of a sentence. In contrast to the Hebrew results, however, Nevins et al. (2007) observed a P600 to Hindi verb-gender agreement incongruencies relative to congruencies and no LAN or N400 effects. The discrepancy between the outcomes of the two studies could be explained by the fact that, while many languages use gender features, the extent to which gender information is used during syntactic processing may be language specific. Moreover, whether a LAN is observed may depend on specific linguistic properties, as well as on the methodology applied in a specific ERP study (for a critical discussion of the methodology used in the Hebrew study, see Molinaro et al., 2011; for a discussion of the LAN component, see Molinaro et al., 2014; Tanner, 2014). Taken together, the majority of syntactic agreement studies have observed that agreement violations lead to a P600 response or to a LAN followed by a P600. A similar pattern of results should be observed for Slovak subject–verb-gender incongruencies.

Integration of Speaker Information in Language Processing

Phonetic and voice information are extracted from the speech signal early and in parallel (e.g., Knösche et al., 2002). Voice perception studies have shown that listeners automatically extract speaker-related information such as gender, age or estimates of body size (e.g., Mullennix et al., 1995; van Dommelen and Moxness, 1995; Braun and Cerrato, 1999; Cerrato et al., 2000). An important question is whether and when in time this speaker-related information is integrated during language processing. Word, sentence and discourse processing studies suggest that listeners anticipate what might be said and use their world knowledge or stereotype-driven inferences about a speaker during linguistic processing, but the exact timing of speaker integration differs across studies (e.g., Lattner and Friederici,

2003; Van Berkum et al., 2008; Scharinger et al., 2011; Hanulíková et al., 2012; Bornkessel-Schlesewsky et al., 2013).

Electroencephalography-studies show that conflicts with inferences about what a given speaker may say lead to qualitatively distinct ERPs (Van Berkum et al., 2008). Van Berkum et al. (2008) presented participants with utterances that were either consistent or inconsistent with a speakers' age, gender, or socioeconomic status (e.g., the biologically implausible utterance produced in a male voice *I might be pregnant because I feel sick*). Van Berkum et al. (2008) found that inconsistencies between the speaker's identity and the meaning of an utterance elicited a larger N400 compared to speaker consistency (e.g., hearing a woman producing the word *pregnant* in the above utterance). This modulation of the N400 effect suggests that listeners use speaker-related attributes in the earliest stages of meaning construction. In contrast to this finding, Lattner and Friederici (2003) suggest that the neural integration of speaker at the semantic level occurs relatively late. In their study, stereotype-driven inferences about a speaker in self-referent utterances such as *I like to wear lipstick* produced by a male speaker resulted in a P600 effect relative to the same utterance produced by a woman. Lattner and Friederici (2003) suggest that their result supports the idea that the P600 reflects a 're-integration of semantic meaning and stereotypical beliefs' (Osterhout et al., 1997).

The distinct time-course patterns in these two studies could be attributed to the type of semantic/pragmatic context established by stereotypically driven inferences based on speaker characteristics. While Lattner and Friederici's (2003) study measured the effect of speaker gender on sentence final stereotypical nouns (e.g., *lipstick, skirt, soccer*), Van Berkum et al.'s (2008) study was less restricted to the use of gender stereotypical role nouns and varied speakers' gender, age, and accent (e.g., *I drink some <u>wine</u> before I go to sleep* in a child voice; *My favorite book is the <u>fairy tale</u> Sleeping Beauty* in an adult voice). Taken together, these studies suggest that violations of stereotypical role nouns as in Lattner and Friederici (2003) are likely to elicit a P600 (e.g., Osterhout et al., 1997), while the semantic-pragmatic incongruity as in Van Berkum et al. (2008) is more likely to elicit an N400 (e.g., Irmen et al., 2010). Since the pragmatic agreement examined in the present study relies on semantic-pragmatic congruity between the conceptual gender of the speaker and the predicate verb, it would be plausible to expect that pragmatic agreement involves the evaluation of speaker characteristics and reflects integration difficulties at the conceptual rather than purely syntactic level.

The Present Study

The majority of studies on syntactic processing that employ grammatical agreement were conducted in the visual modality whereas studies that manipulate speaker characteristics in the domain of auditory processing usually do not examine syntactic processing (for a review, see Kutas and Federmeier, 2007). The present study fills this gap by examining agreement computations between verbs and a speaker's gender in Slovak. We compared Slovak listeners' ERP responses to Slovak past verb forms (a) agreeing or disagreeing with the conceptual gender of the speaker (first person singular; hence pragmatic agreement) and (b) agreeing or disagreeing with the grammatical gender of the animate subject (third person singular; hence syntactic agreement).

In line with previous research on gender agreement conflicts, we expected that incongruencies between the grammatical gender of an animate subject (e.g., *žena* 'woman$_{FEM}$') and a predicate (e.g., *išiel* 'went$_{MASC}$') would result in a P600 and possibly a LAN relative to the congruent predicate (e.g., *išla* 'went$_{FEM}$'). If the computation of the pragmatic agreement resembles the syntactic agreement, similar gender incongruency effects should be observed for the first and third person agreement features. There are, however, alternative accounts for the integration of speaker information during computations of verb agreement in the pragmatic condition. Following Nevins et al. (2007), the computation of concord might take place in a bottom-up fashion during the syntactic build-up of a sentence. Agreement processing starts once gender features are identified upon hearing a verb. This triggers a search for the subject (personal pronouns are, however, not marked for gender and are omitted due to Slovak being a pro-drop language) to check for matching gender features. This checking process may be independent of the semantic-pragmatic information (for a discussion, see Mancini et al., 2013) that is provided by a speaker's voice. Under this assumption, pragmatic violations should not elicit any mismatch effects because the personal pronoun is unmarked for gender and no mismatch of the verb will be encountered (the verbal inflection is incorrect only if the pragmatic information about the speaker is considered and integrated in the syntactic build-up of the utterance).

Given prior research on the impact of speaker characteristics on linguistic processing, it seems unlikely to expect that pragmatic information is not used in the checking process. We therefore consider two possible outcomes. Following Nevins et al. (2007), it could be that the processing of agreement features that must be matched with the speaker's gender can start before hearing the verb in a top–down fashion. Under this assumption, listeners would not wait until the presentation of the verb to initiate the agreement processing. Rather, listeners check whether the verb matches the speaker's features that have been predictively built (Nevins et al., 2007). Since speaker information spreads across the entire utterance, listeners would quickly encounter a mismatch upon hearing the verb. If this mismatch is perceived as syntactic in nature, a P600 and possibly a LAN would be expected. Alternatively, in line with studies on the integration of speaker information, the mismatch could be perceived as pragmatically implausible. Since speaker characteristics convey social and pragmatic information whose violation have been shown to elicit an N400 effect (Van Berkum et al., 2008), it would be plausible to expect an N400 effect to violations of the pragmatic agreement. Such a result would also be in line with proposals according to which the recruitment of non-syntactic information about a person leads to conceptual/semantic integration reflected in an N400-like effect (e.g., Deutsch and Bentin, 2001; Mancini et al., 2013), while formally marked syntactic features lead to structural integration as reflected in a LAN/P600 complex (e.g., Molinaro et al., 2011).

Materials and Methods

Participants

Thirty-two native speakers of Slovak with no neurological or psychiatric disorders and no reported hearing problems volunteered to participate. They were all students (16 female, all right handed, mean age = 21, range = 18–24) at the Comenius University in Bratislava. All students grew up speaking Slovak only, and 27 of the students indicated communicative competence in at least one foreign language (the majority in English and German). Students received financial compensation for their participation; informed consent was obtained from all participants.

Materials

The stimuli consisted of 240 sentences all of which contained a main clause followed by a subordinate clause. Each subordinate clause contained a past verb form that agreed in grammatical gender with the third person animate subject (e.g., *lebo svokra kradla* 'because the mother-in-law_FEM stole_FEM') or with the conceptual gender of the speaker in the first person (*lebo som kradla* 'because I stole_FEM'). Each sentence was recorded in eight versions (see **Table 1** and the Supplementary Material for example sentences). A male speaker and a female speaker spoke a correct and an incorrect version of each sentence. The resulting 1920 sentences were distributed over eight experimental lists with one of the eight versions of each sentence occurring in only one experimental list. Within one experimental list, the number of correct and incorrect sentences was equally spread across conditions and voices. An additional set of four practice sentences with the same type of agreement patterns was recorded. The critical verbs at which the agreement violation became apparent were always embedded in a subordinate clause, at least two syllables before the end of the entire utterance. The critical verbs were between two to five syllables long. Verbs with the feminine inflections ended with the inflectional morpheme −*la* while the masculine inflections ended with −*l*. The mean logarithmic critical word form frequency per million was 0.76 (SD 0.76) for masculine verb forms and 0.43 (SD 0.76) for feminine verb forms (Slovenský Národný Korpus, 2009). Except for nine verbs (most of them with a stereotypically female connotation

such as *to cook, to clean, to paint nails*), the masculine verb forms were more frequent that the feminine verb forms. This is not surprising because Slovak (and many other languages) use generic masculine nouns to refer to male beings, as well as to beings of unspecified sex (e.g., *pracovník* 'worker_MASC,' *pracovníci* 'workers_MASC'), while the female nouns refer only to female beings (e.g., *pracovníčka* 'worker_FEM,' *pracovníčky* 'workers_FEM'). This pattern of usage is then reflected in the frequency distribution of the inflected verb forms as well as nouns. All nouns in the subject position in the third person utterances were balanced for gender (half were male) and referred to professions or social groups (e.g., *translator, professor, teacher, member, tourist, friend*) or relatives (e.g., *mother-in-law, father, bride, brother, niece, cousin*). The grammatical gender of the subject always corresponded to the biological gender (neuter nouns such as *dievča* 'girl' were not used). The mean logarithmic word form frequency of the masculine nouns was 0.68 (SD 0.72) and of the feminine nouns −0.047 (SD 0.77) (Slovenský Národný Korpus, 2009).

Sentences were spoken by a 31 year-old male speaker and a 33 year-old female speaker. The speakers were siblings and grew up speaking a standard variety of Slovak. Their voices clearly indicated their biological gender as determined by ratings from 8 additional participants (mean age 28; 6 women), none of whom took part in the EEG study. On a scale from 1 to 5 (with 1 meaning clearly male voice and 5 meaning clearly female voice), the male speaker had an average of 1 and the female speaker had an average of 5. There was no ambiguity with respect to the gender of the speakers given their voice characteristics.

Both speakers received a complete list of all sentences, each sentence with its correct and incorrect version. They read the sentences at a natural speech rate. To minimize possible differences in the speech rate and intonation across the male and the female speakers, and across the conditions, each sentence was first produced by one speaker and immediately repeated by the second speaker (as in Hanulíková et al., 2012). Utterances that differed in prosody or speech rate were repeated by both speakers in both the correct and incorrect versions. Correct and incorrect versions of each utterance were produced in pairs to keep them as comparable as possible across conditions. In sentences spoken by the female speaker, the mean duration of the critical verbs was 474 ms (SD 104) and the mean duration of the whole sentence was 3578 ms (SD 725). The mean duration of the critical verbs spoken by the male speaker was 473 ms (SD 112) and the mean duration of the whole sentence was 3594 ms (SD 729). There were no significant differences in duration between the male and female speakers for either sentence duration or word duration (all p's > 0.4). All sentences were adjusted in Praat to have comparable amplitude.

Procedure

After the completion of an informed consent form, participants were seated in a comfortable armchair in front of a computer in a quiet room. They were told that they would listen to a male speaker and a female speaker talking about their lives. The 240 utterances were presented over loudspeakers situated next to the computer. Participants were asked to carefully listen

---

TABLE 1 | Sentences with subject–verb-gender agreement with English translation.

First person: pragmatic agreement (*n* = 60 correct and 60 incorrect)

Female speaker: Susedia sa nahnevali, lebo som kradla/*kradol slivky

Male speaker: Susedia sa nahnevali, lebo som *kradla/kradol slivky

(neighbors themselves upset because am stole_FEM/stole_MASC plums)

'neighbors were upset because I stole plums'

Third person: syntactic agreement (*n* = 60 correct and 60 incorrect)

Female speaker: Susedia sa nahnevali, lebo svokra kradla/*kradol slivky

Male speaker: Susedia sa nahnevali, lebo svokra kradla/*kradol slivky

(neighbors themselves upset because mother-in-law_FEM stole_FEM/stole_MASC Plums)

'neighbors were upset because the mother-in-law stole plums'

Critical words are underlined. Asterisk indicates an incorrect verbal inflection in a given context.

for comprehension in order to answer comprehension questions that would follow some of the utterances. These questions (24 yes/no questions, half of which required a "yes" response) were included to ensure that participants were paying attention. To keep the task as natural as possible, and to keep the study comparable to previous task-less studies (e.g., Hagoort and Brown, 2000; Van Berkum et al., 2008; Hanulíková et al., 2012), no further grammaticality judgment or acceptability task was used. Participant performance of the comprehension questions was very high (mean percentage correct 98%, SD 4.46, range 83.3–100%). After the presentation of each utterance, a cross appeared in the middle of the screen to indicate that participants could blink or move. Participants were given button-press control over the initiation of the next trial, which started with a silence of 1000 ms followed by the utterance. The experiment consisted of six blocks and five short breaks. After the EEG study, participants were asked to complete the Edinburgh handedness test (to control for variation in lateralization of brain functions), a language-background questionnaire and comprehensibility ratings for the male and the female speakers. The ratings revealed that both speakers were equally well comprehensible. On a scale from 1 to 5 (with 1 meaning well comprehensible and 5 not comprehensible), both speakers had an average of 1.34.

EEG Recording

Electroencephalography was recorded from 27 Ag/AgCl electrodes (impedance was kept below 5 kΩ) at standard locations (Fz, Cz, Pz, Fp1/2, F3/4, F7/8, FC1/2, FC5/6, C3/4, T7/8, CP1/2, CP5/6, P3/4, P7/8, O1/2). Two additional mastoid electrodes (placed on the left mastoid A1 and on the right mastoid A2) and four additional electrooculogram electrodes (placed above and below each eye) for eye movement and blink artifacts recordings were used. All recordings were referenced to the left mastoid during online recording, amplified with BrainAmp DC amplifiers (0.016–100 Hz band pass, digitized at 250 Hz), and re-referenced offline to the mastoid average. EEG segments ranging from 200 ms before to 1200 after critical word onset were extracted and baseline corrected to a 200-ms pre-onset baseline. All segments with potentials above $\pm 75\ \mu V$ were rejected as artifacts (average segment loss 14%, range 13–15%, no differences between conditions). The segments were averaged per participant and condition, and mean amplitudes were analyzed with repeated-measures analyses of variance (ANOVAs). As a first step, the variation of effect size over all electrodes was examined, after which a topography-oriented analysis was conducted involving anterior (Fp1/2, F3/4, F7/8, FC1/2, FC5/6, Fz) and posterior distributions (CP5/6, CP1/2, P7/8, P3/4, O1/2, Pz).
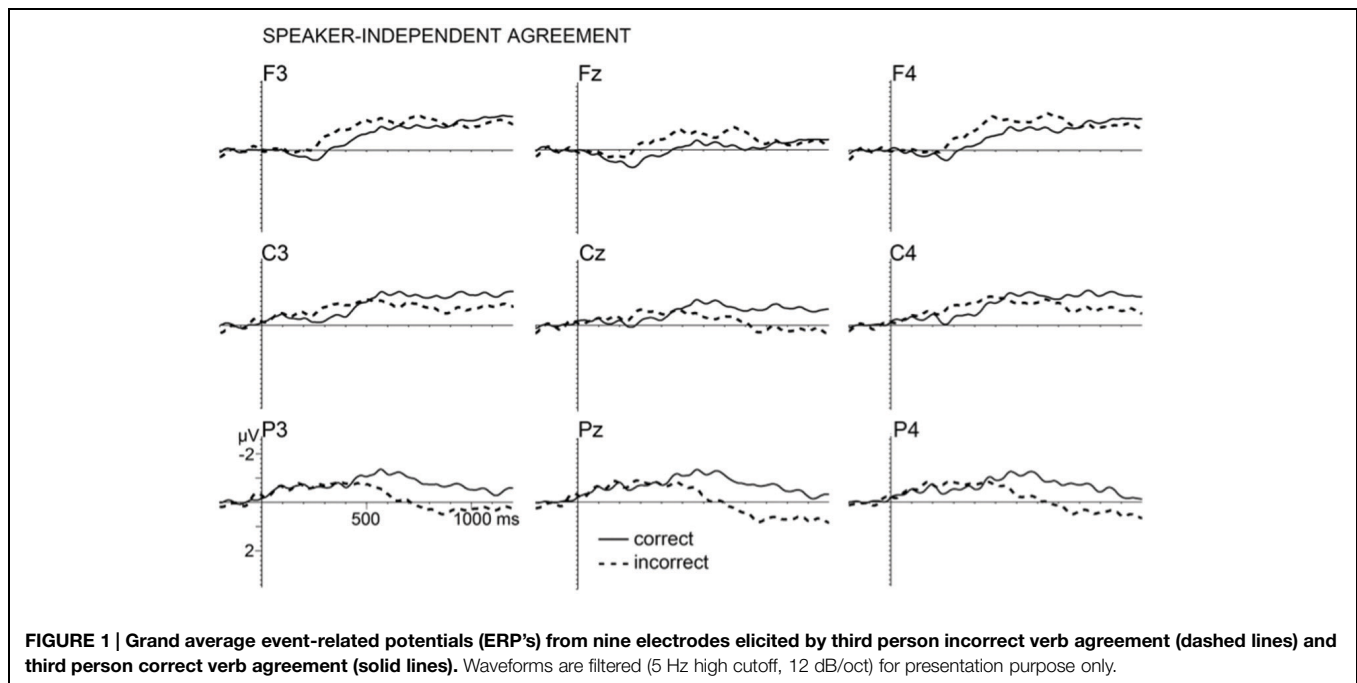
For the statistical analyses, we followed the same analyses steps as in Hanulíková et al. (2012) and the time window was chosen in line with previous research (for an overview of EEG studies in the visual modality, see Molinaro et al., 2011) and on the basis of a visual inspection of the averaged data. For the P600 effect, the time window was 500–1000 ms (for a similar time-window, see e.g., Sassenhagen et al., 2014; for a short review, see Osterhout et al., 2012), for the LAN effect it was 200–500 ms (for

a similar time-window, see e.g., Kutas and Hillyard, 1983; Roehm et al., 2005), and for the N400 effect it was 100–400 ms (for a similar time-window, see e.g., Ye et al., 2006; Shen et al., 2013). The ERP effects within the auditory modality might deviate somewhat from effects observed in the visual modality. Note that the detection of the agreement error is only possible once the critical verb has been heard, recognized, and the gender of the inflectional ending becomes available. Since the verbs varied in length between two to five syllables, we wanted to make sure that violation effects were captured correctly. The critical point within a verb to which the ERPs were time-locked was therefore set to onset of the last syllable that indicated the gender disambiguation (e.g., the onset of the syllable *dla* in the verbs *kradla* 'stole$_{FEM}$,' *dohodla* 'agreed$_{FEM}$' and the syllable *dol* for the verbs *kradol* 'stole$_{MASC}$,' *dohodol* 'agreed$_{MASC}$'). Note that the disambiguation is possible already at the onset of the syllable (i.e., *d*) because phonetic properties of the onset of the critical syllables are affected by the following speech sounds. Similar time-locking procedures to the gender inflection of the critical word or to the ends of verb stems were applied in other auditory ERP studies (e.g., Van Berkum et al., 2005; Shen et al., 2013), resulting in somewhat early onsets of ERP effects.

Results

Speaker-Independent Agreement

As can be seen in **Figure 1**, subject–verb-gender violations in the speaker-independent condition (syntactic agreement in the third person singular) resulted in a larger anterior negativity followed by a larger posterior positivity (P600) compared to correct utterances. The effect size varied over all electrodes as confirmed in a repeated measures ANOVA with the factors *correctness* (violation, correct) and *electrodes* (all 27) showing a significant interaction in the 200–500 time window [$F(1,26) = 3.61$, $p = 0.003$, $\eta_p^2 = 0.104$] and in the 500–1000 time window [$F(1,26) = 7.78$, $p < 0.001$, $\eta_p^2 = 0.201$]. To determine the distribution of the effect, a topography-oriented analysis was conducted by dividing the electrodes into posterior and anterior to the central cross-line and into left and right to the central cross-line. A 2 (*distribution*: posterior, anterior) × 2 (*correctness*) repeated measures ANOVA confirmed a larger effect over the anterior than the posterior area in the 200–500 ms time window [*distribution* × *correctness* interaction: $F(1,31) = 14.22$, $p = 0.001$, $\eta_p^2 = 0.314$], as well as a larger effect over the posterior than the anterior area in the 500–1000 ms time window [*distribution* × *correctness*: $F(1,31) = 27.35$, $p < 0.001$, $\eta_p^2 = 0.469$]. There were no significant interactions between the factor *correctness* and right vs. left *distribution* ($F < 1$), confirming that the effects were not lateralized. Follow-up analyses revealed a significant P600 effect to violations compared to correct sentences across all posterior electrodes [$F(1,31) = 11.11$, $p = 0.002$, $\eta_p^2 = 0.264$] but not across all anterior electrodes ($F < 1$). The anterior negativity was significantly larger for violations compared to correct sentences across all anterior electrodes [$F(1,31) = 5.04$, $p = 0.03$, $\eta_p^2 = 0.140$] but not across all posterior electrodes ($F < 1$).

FIGURE 1 | Grand average event-related potentials (ERP's) from nine electrodes elicited by third person incorrect verb agreement (dashed lines) and third person correct verb agreement (solid lines). Waveforms are filtered (5 Hz high cutoff, 12 dB/oct) for presentation purpose only.

Speaker-Dependent Agreement

Interestingly, subject–verb-gender agreement violations in the speaker-dependent condition (pragmatic agreement in the first person singular) showed a distinct pattern of results (see **Figure 2**). The lack of variation in effect size across all electrodes was confirmed by a non-significant interaction of the factors *correctness* and *electrodes* in the 100–400 ms time window [$F(1,26) = 1.14$, $p = 0.29$, $\eta_p^2 = 0.035$], confirming a broadly distributed negativity. There was, however, a main effect of *correctness* [$F(1,31) = 5.57$, $p = 0.025$, $\eta_p^2 = 0.152$], suggesting that agreement violations resulted in a broadly distributed negativity compared to correct sentences. The topography-oriented analyses showed no significant interactions (all $Fs < 1$) and no other significant differences were found in later time windows.

Discussion

We examined the nature and the time course of the effect of a speaker's biological gender on subject–verb agreement in spoken Slovak. Despite the large literature on ERP effects observed for gender violations, studies on grammatical agreement features that depend on speaker characteristics extracted from spoken language have been missing. The present study fills this gap by contrasting two different types of gender agreement in Slovak, the speaker-dependent/pragmatic gender agreement and the speaker-independent/syntactic gender agreement. As predicted, disagreement between a formally marked subject and a predicate (e.g., 'mother-in-law$_{FEM}$ stole$_{MASC}$ plums') elicited an anterior negativity in the 200–500 ms time window followed by a P600 in the 500–1000 ms time window. The distribution of the anterior negativity was bilateral rather than left lateralised, similar to

some previous studies in the auditory modality (e.g., Hahne and Friederici, 2002; Shen et al., 2013), as well as in the visual modality (e.g., Hahne and Jescheniak, 2001; Hagoort et al., 2003; Yamada and Neville, 2007). In line with previous studies on agreement processes, a possible interpretation of this result is that upon hearing the verb, listeners match features between the predicate and the subject nominal phrase leading to the integration of the syntactic and conceptual representations of the utterance. The anterior negativity could be the result of mismatch detection or failed binding between the verbal morphology and the formally marked subject (e.g., Gunter et al., 1997; Hagoort, 2003). The P600 could then indicate a process of reanalysis, revision or recovery from the mismatch detection (e.g., Osterhout and Holcomb, 1992; Hagoort et al., 1993; Friederici, 1995, 2002; Bornkessel-Schlesewsky and Schlesewsky, 2009). Although the present result does not allow resolving the debate on the functional significance of the observed components, it is in line with prior ERP-studies on grammatical gender agreement and extends the electrophysiological evidence on subject–verb-gender agreement to a new language (Slovak).

In contrast to syntactic agreement, incongruencies between the conceptual gender of a speaker and the predicate in the pragmatic condition (e.g., 'I stole$_{MASC}$ plums' spoken by a female speaker) resulted in a larger centrally distributed N400-like effect in an early 100–400 ms time window relative to the congruent agreement. The distribution of this effect is comparable to previous studies on speaker integration (Van Berkum et al., 2008; Bornkessel-Schlesewsky et al., 2013) and on subject–verb agreement studies involving person features (e.g., Schirmer et al., 2005; Mancini et al., 2011; Shen et al., 2013). One possible interpretation of this result is that speaker characteristics directly impact the computation of the syntactic relations between words in an utterance but lead to a distinct electrophysiological response
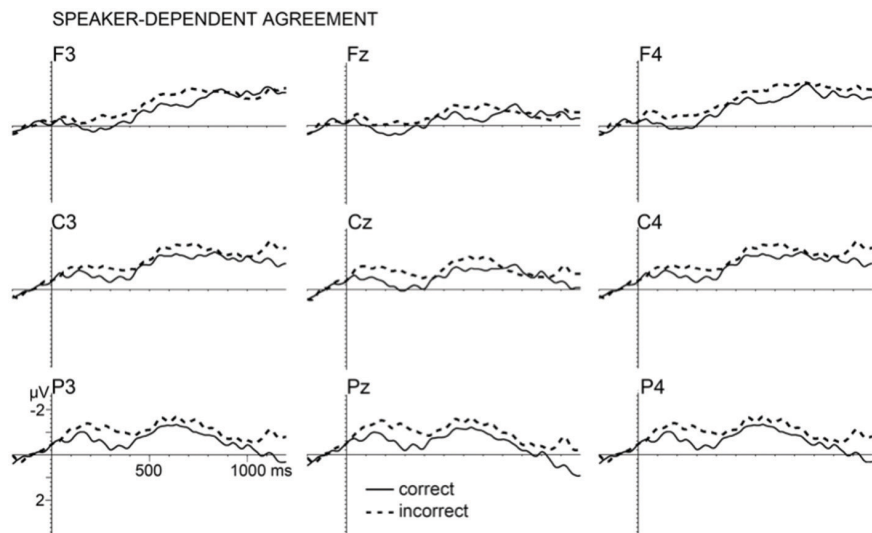
FIGURE 2 | Grand average ERP's from nine electrodes elicited by first person incorrect verb agreement (dashed lines) and first person correct verb agreement (solid lines). Waveforms are filtered (5 Hz high cutoff, 12 dB/oct) for presentation purpose only.

than the computation of formal syntactic features, such as the grammatical gender of the subject in the third person. The absence of an anterior negativity could suggest that pragmatic mismatch between the subject and the predicate in the first person may not be treated as a morphosyntactic violation. Similarly, the presence of an early posterior negativity instead of a P600 effect in the pragmatic agreement would suggest that no pure syntactic re-analysis was triggered. This N400-like effect, as well as the rather early onset of this effect, deserves some more discussion.

Previous research has already shown that listeners take into account speaker identity at early stages of meaning construction (Van Berkum et al., 2008; Bornkessel-Schlesewsky et al., 2013). Speaker inconsistency, such as the biologically improbable utterance *I am pregnant* uttered by a male voice elicited a larger N400 than the same sentence uttered in the more probable context of a female voice (Van Berkum et al., 2008). Similarly, false political statements produced by a well-known politician triggered a larger N400 than the same statements produced by a famous news announcer or a control speaker, such as an unknown professor (Bornkessel-Schlesewsky et al., 2013). In contrast, stereotype-driven beliefs about a speaker in self-referent utterances such as *I like to wear lipstick* produced by a man resulted in a larger P600 relative to the same utterance produced by a woman (Lattner and Friederici, 2003). These studies suggest that violations of stereotypical role nouns (e.g., the use of lipstick referring to a man; Lattner and Friederici, 2003) are likely to elicit a P600 (e.g., Osterhout et al., 1997), while semantic-pragmatic violations (as in Van Berkum et al., 2008) are more likely to elicit an N400 (e.g., Irmen et al., 2010). The present study is a valuable addition to this line of research, showing that grammatical agreement processing involving pragmatic aspects (i.e., speaker-related physical/social information) can be modulated in a similar manner, eliciting an

N400. Although speaker-related effects have been also observed in syntactic processing (e.g., Hanulíková et al., 2012), they did not involve the type of pragmatic agreement as used in the present study. Syntactic gender errors, such as an incorrect use of determiners in Dutch, lead to a P600 when produced by a native speaker but not when produced by a non-native speaker with a foreign accent, suggesting that the late positivity can be modulated by participants' inferences about speakers' linguistic performance. However, the present study is concerned with the pragmatic gender agreement and allows a direct evaluation of the role of speaker identity during processing of sentences that are syntactically correct at the surface level. The results have shown that a speaker's gender modulates syntactic processing: speaker-based agreement violations elicited a larger N400 compared to speaker-based matching agreement. This suggests that listeners integrate conceptual/semantic information about a speaker during syntactic processing comparable to speaker integration during semantic processing. Since the pragmatic agreement examined in the present study relies on semantic-pragmatic congruity between the conceptual gender of the speaker and the predicate verb, it would be plausible to expect that pragmatic agreement involves the evaluation of speaker characteristics and reflects integration difficulties at the conceptual rather than purely syntactic level.

The question remains, however, how exactly listeners integrate speaker information? Following Nevins et al. (2007), the processing of agreement features could start before hearing the verb in a top–down fashion. Such context-driven top–down processing could be the result of expectation formation. Listeners anticipate certain verbal inflections given a speaker's gender, which is immediately available for matching agreement features. Indeed, prior research indicates that listeners anticipate linguistic properties (e.g., Connolly and Phillips, 1994; Van Berkum et al., 2005). Van Berkum et al. (2005) has shown

that the semantic context of an utterance leads to anticipations of the syntactic gender of a noun and elicits a larger broadly distributed negativity around 300–400 ms for an unpredicted noun relative to a predicted one. Similarly, a word deviating from the expected word in the initial phoneme leads to an early negativity around 300 ms (Connolly and Phillips, 1994). It would therefore be plausible to assume that listeners in the present study anticipated a verbal inflection corresponding to the speaker's gender and the early N400 may reflect greater speaker-context dependency when such expectations are violated (see also Ye et al., 2006). This would be also in line with suggestions that the degree of semantic-pragmatic predictability is associated with N400 effects (e.g., Kutas and Hillyard, 1984; Van Berkum, 2008).

Listeners could also create similar predictions in the third person singular. However, since the matching between the formal grammatical gender of a nominal phrase and the verbal inflection entails syntactic integration, a P600 instead of an N400 emerges (as in e.g., Nevins et al., 2007). Consequently, the information used by the linguistic system clearly (and perhaps unsurprisingly) differs in the third and first person. Unlike utterances including the third person subject–verb disagreement that is rendered ungrammatical on surface level, the surface structures of the utterances in the pragmatically violated sentences were all grammatically correct. The sentence *lebo som kradla slivky* ('because I stole_FEM plums') is grammatically incorrect only if spoken by a male speaker. Upon the detection of the incorrect verbal inflection, listeners may trigger a reanalysis of the nominal phrase, but since the personal pronoun is unmarked for gender, no syntactic re-evaluation of the subject takes place. Instead, listeners re-evaluate and revise the speaker information resulting in an N400-like effect, in line with studies on speaker integration during processing of utterance meaning (e.g., Van Berkum et al., 2008; Bornkessel-Schlesewsky et al., 2013), as well as studies on subject–verb agreement involving person feature processing (e.g., Mancini et al., 2011). When the conceptual gender is a salient cue provided by the acoustic signal, the evaluation of the inflectional marker may then be mediated by a pragmatic integration. Since listeners have more time to conceptually and semantically interpret the speaker, the processing of the speaker-dependent verbal inflection could be less syntactically disrupted or the disagreement syntactically less noticeable because the pragmatic feature information decays over time and the sentence *per se* is grammatical (see Deutsch, 1998). The same inflectional marker, however, triggers syntactic integration when the computation is driven by the formally marked grammatical gender of a noun. This interpretation would be in line with proposals according to which the recruitment of non-syntactic information, such as the gender of a speaker, results in N400-like effects, while formally marked syntactic features lead to a structural integration as reflected in a LAN/P600 complex (e.g., Deutsch and Bentin, 2001; Schmitt et al., 2002; Molinaro et al., 2011).

Interestingly, the incongruency effect in the pragmatic agreement arouse very early in time. There are two possible explanations concerning this rather early onset of the negativity. It is plausible to assume that the early onset results from the omnipresent speaker-related information. Semantic-conceptual integration can take place early because gender is present throughout the utterance and listeners might have predictably built expectations about the gender information encoded in the verbal inflectional morpheme (Nevins et al., 2007). Although similar early onsets were observed for semantic integration in the auditory domain (e.g., Holcomb and Neville, 1991; Schirmer et al., 2005; Ye et al., 2006; Bornkessel-Schlesewsky et al., 2013), it should be noted that the time-locking to the inflectional marker of the last syllable of the verb could also contribute to the earlier onset of the effect in the present study. Moreover, the fact that our results do not exactly match the timing of the N400 reported by earlier auditory comprehension work in English (Holcomb and Neville, 1991), Dutch (Hagoort et al., 2003), and German (Friederici et al., 1993), may be due to differences in the investigation of classic semantic violations and pragmatically driven syntactic violations, as well as due to characteristics of the Slovak language. Bornkessel-Schlesewsky et al. (2013) reported an early increased N400 (150–450 ms time-window) to auditory false versus true political statements uttered by a famous politician when compared to the same utterances produced by a famous news announcer or a control speaker (an unknown professor). The result was interpreted in terms of a socially mediated interpretation provided by the speaker identity and suggests that the social status of a speaker influences the neural computation of a linguistic message (Bornkessel-Schlesewsky et al., 2013). In line with their interpretation, the pragmatic N400 in the present study could suggests that listeners combine speaker characteristics and the message (whether semantic or syntactic in nature) very early during linguistic processing in a communicative context.

While the pragmatic agreement computations clearly differ from syntactic agreement, further research is needed to disentangle the possible explanations for the exact nature of the pragmatic agreement compared to the syntactic agreement. It would be interesting to consider further languages with distinct options for social speaker-related grammatical agreement features. Moreover, the future challenge of developing a model of language processing that captures the spread of different types of linguistic information across different languages (e.g., pragmatic vs. syntactic agreement processes) and incorporates both predictive processing and bottom-up feature checking remains. Future studies could also examine individual differences in the computation of pragmatic gender agreement. Specifically, does the gender of the speaker or the gender of the listener and her/his congruency with the speaker changes the sensitivity of detecting pragmatic incongruity between the speaker and the verb participle? And do working memory capacities or empathy (e.g., van den Brink et al., 2012) modulate pragmatic and syntactic agreement computations in different ways?

Conclusion

Taken together, the results of the present study show that the processing of subject–verb agreement is modulated by the gender of a speaker, and that the integration of speaker and

morphosyntax occurs relatively early. Overall, this result extends our knowledge regarding the role of speaker characteristics on the neural correlates of speech processing and is a valuable contribution to cross-linguistic comparisons. Previous research has already shown that listeners integrate a speaker's identity during meaning construction. The present study has further shown that listeners take the speaker into account during syntactic processing in a similar manner. The linguistic brain thus takes into account all information available to achieve an effortless and successful comprehension of spoken language.

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01396

References

Badecker, W., and Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: evidence from gender and case in Slovak. *J. Mem. Lang.* 56, 65–85. doi: 10.1016/j.jml.2006.08.004

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289. doi: 10.1016/j.tics.2007.05.005

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Bates, E., Devescovi, A., Hernandez, A., and Pizzamiglio, L. (1996). Gender priming in Italian. *Percept. Psychophys.* 58, 992–1004. doi: 10.3758/BF03206827

Batterink, L., and Neville, H. (2013). The human brain processes syntax in the absence of conscious awareness. *J. Neurosci.* 33, 8528–8533. doi: 10.1523/JNEUROSCI.0618-13.2013

Boelte, J., and Connine, C. M. (2004). Grammatical gender in spoken word recognition in German. *Percept. Psychophys.* 66, 1018–1032. doi: 10.3758/BF03194992

Bornkessel-Schlesewsky, I., Krauspenhaar, S., and Schlesewsky, M. (2013). Yes, you can? *PLoS ONE* 8:e69173. doi: 10.1371/journal.pone.0069173

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2009). *Processing Syntax and Morphology: A Neurocognitive Perspective.* Oxford: Oxford University Press.

Braun, A., and Cerrato, L. (1999). "Estimating speaker age across languages," in *Proceedings of ICPhS,* San Francisco, CA, 1369–1372.

Brown, C., and Hagoort, P. (1993). The processing nature of the n400: evidence from masked priming. *J. Cogn. Neurosci.* 5, 34–44. doi: 10.1162/jocn.1993.5.1.34

Cerrato, L., Falcone, M., and Paoloni, A. (2000). Subjective age estimation of telephonic voices. *Speech Commun.* 31, 107–112. doi: 10.1016/S0167-6393(99)00071-0

Connolly, J. F., and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal words of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266. doi: 10.1162/jocn.1994.6.3.256

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

Corbett, G. (2006). *Agreement.* Cambridge, MA: Cambridge University Press.

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58. doi: 10.1080/016909698386582

Deutsch, A. (1998). Subject-predicate agreement in Hebrew: interrelations with semantic processes. *Lang. Cogn. Process.* 13, 575–597. doi: 10.1080/016909698386447

Deutsch, A., and Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. *J. Mem. Lang.* 45, 200–224. doi: 10.1006/jmla.2000.2768

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., et al. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain Lang.* 85, 280–296. doi: 10.1016/S0093-934X(03)00055-5

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Friederici, A. D. (1995). As time goes by: the time-course of syntactic activation during language processing. *Brain Cogn.* 28, 186–186.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8

Friederici, A. D., and Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *J. Psycholinguist. Res.* 28, 467–484. doi: 10.1023/A:1023243708702

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Lang. Cogn. Process.* 25, 149–188. doi: 10.1080/01690960902965951

Grosjean, F., Dommergues, J., Cornu, E., Guillelmon, D., and Besson, M. (1994). The gender marking effect in spoken word recognition. *Percept. Psychophys.* 56, 590–598. doi: 10.3758/BF03206954

Gunter, T. C., Friederici, A. D., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568. doi: 10.1162/089892900562336

Gunter, T. C., Stowe, L. A., and Mulder, G. (1997). When syntax meets semantics. *Psychophysiology* 34, 660–676. doi: 10.1111/j.1469-8986.1997.tb02142.x

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage* 20, S18–S29. doi: 10.1016/j.neuroimage.2003.09.013

Hagoort, P., and Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *J. Psycholinguist. Res.* 28, 715–728. doi: 10.1023/A:1023277213129

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1

Hagoort, P., Brown, C. M., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585

Hagoort, P., Wassenaar, M., and Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cogn. Brain Res.* 16, 38–50. doi: 10.1016/S0926-6410(02)00208-2

Hahne, A., and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *J. Cogn. Neurosci.* 11, 194–205.

Hahne, A., and Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Brain Res. Cogn. Brain Res.* 13, 339–356. doi: 10.1016/S0926-6410(01)00127-6

Hahne, A., and Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? *Cogn. Brain Res.* 11, 199–212. doi: 10.1016/S0926-6410(00)00071-9

Hanulíková, A., van Alphen, P. M., van Goch, M. M., and Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 2, 878–887. doi: 10.1162/jocn_a_00103

Hinojosa, J. A., Martin-Loeches, M., Casado, P., Munoz, F., and Rubia, F. J. (2003). Similarities and differences between phrase structure and morphosyntactic violations in Spanish: an event-related potentials study. *Lang. Cogn. Process.* 18, 113–142. doi: 10.1080/01690960143000489

Holcomb, P. J., and Neville, H. J. (1991). Natural speech processing: an analysis using event-related brain potentials. *Psychobiology* 1, 286–300.

Irmen, L., Holt, D. V., and Weisbrod, M. (2010). Effects of role typicality on processing person information in German: evidence from an ERP study. *Brain Res.* 1353, 133–144. doi: 10.1016/j.brainres.2010.07.018

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *J. Psycholinguist. Res.* 31, 165–193. doi: 10.1023/A:1014978917769

Knösche, T. R., Lattner, S., Maess, B., Schauer, M., and Friederici, A. D. (2002). Early parallel processing of auditory word and voice information. *Neuroimage* 17, 1493–1503. doi: 10.1016/nimg.2002.1262

Kutas, M., and Federmeier, K. D. (2007). "Event-related brain potential (ERP) studies of sentence processing," in *The Oxford Handbook of Psycholinguistic,* ed. G. Gaskell (Oxford: Oxford University Press), 385–406.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550. doi: 10.3758/BF03196991

Kutas, M., and Hillyard, S. A. (1984). "Event-related potentials in cognitive science," in *Handbook of Cognitive Neuroscience*, ed. M. S. Gazzaniga (New York: Plenum Press), 387–409.

Kutas, M., and Van Petten, C. (1994). "Psycholinguistics Electrified: event-related potential investigations," in *Handbook of Psycholinguistics*, ed. M. A. Gernsbacher (Waltham, MA: Academic Press), 83–143.

Lattner, S., and Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing - evidence from event-related brain potentials. *Neurosci. Lett.* 339, 191–194. doi: 10.1016/S0304-3940(03)00027-2

Lau, E. F., Weber, K., and Gramfort, A. (2014)). Spatiotemporal signatures of lexical-semantic prediction. *Cereb. Cortex* doi: 10.1093/cercor/bhu219 [Epub ahead of print].

Levinson, S. C. (1983). *Pragmatics*. Cambridge, MA: Cambridge University Press.

MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *J. Verbal Learn. Verbal Behav.* 23, 127–150. doi: 10.1016/S0022-5371(84)90093-8

Mancini, S., Molinaro, N., and Carreiras, M. (2013). Anchoring agreement in comprehension. *Lang. Linguist. Compass* 7, 1–21. doi: 10.1111/lnc3.12008

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject-verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055

Martin, A. E., Nieuwland, M. S., and Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage* 59, 1859–1869. doi: 10.1016/j.neuroimage.2011.08.057

Martin, A. E., Nieuwland, M. S., and Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain Lang.* 135, 42–51. doi: 10.1016/j.bandl.2014.05.001

Migdalski, K. M. (2006). *The Syntax of Compound Tense in Slavic*. Utrecht: LOT.

Molinaro, N., Barber, H. A., Caffarra, S., and Carreiras, M. (2014). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex* 66, 156–159. doi: 10.1016/j.cortex.2014.06.009

Molinaro, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 4, 908–930. doi: 10.1016/j.cortex.2011.02.019

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., and Farnsworth, L. M. (1995). The perceptual representation of voice gender. *J. Acoustical Soc. Am.* 98, 3080–3080. doi: 10.1121/1.413832

Münte, T. F., and Heinze, H. J. (1994). "ERP negativities during syntactic processing of written words," in *Cognitive Electrophysiology*, eds H. J. Heinze, T. F. Münte, and G. R. Magnun (Boston: Birkhäuser), 211-238.

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058

Osterhout, L., Bersick, M., and McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Mem. Cogn.* 25, 273–285. doi: 10.3758/BF03211283

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Osterhout, L., Kim, A., and Kuperberg, G. (2012). "The neurobiology of sentence comprehension," in *The Cambridge Handbook of Psycholinguistics*, eds M. J. Spivey, K. McRae, and M. F. Joanisse (Cambridge: Cambridge University Press), 365–189.

Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033

Roehm, D., Bornkessel, I., Haider, H., and Schlesewsky, M. (2005). When case meets agreement: event-related potential effects for morphology-based conflict resolution in human language comprehension. *Neuroreport* 1, 875–878. doi: 10.1097/00001756-200505310-00019

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain Lang.* 137, 29–39. doi: 10.1016/j.bandl.2014.07.010

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2011). You had me at "Hello": Rapid extraction of dialect information from spoken words. *Neuroimage* 56, 2329–2338. doi: 10.1016/j.neuroimage.2011.04.007

Schirmer, A., Kotz, S. A., and Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: sex differences revisited. *Cogn. Brain Res.* 2, 442–452. doi: 10.1016/j.cogbrainres.2005.02.022

Schmitt, B. M., Lamers, M., and Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cogn. Brain Res.* 14, 333–346. doi: 10.1016/S0926-6410(02)00136-2

Shen, E. Y., Staub, A., and Sanders, L. D. (2013). Event-related brain potential evidence that local nouns affect subject-verb agreement processing. *Lang. Cogn. Process.* 28, 498–524. doi: 10.1080/01690965.2011.650900

Silva-Pereyra, J., and Carreiras, M. (2007). An ERP study of phi-features in Spanish. *Cogn. Brain Res.* 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029

Slovenský Národný Korpus. (2009). *Prim-4.0-Public-All*. Bratislava: Slovenský Národný Korpus.

Steinhauer, K., and Drury, J. E. (2012). On the early left anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. doi: 10.1016/j.bandl.2011.07.001

Tanner, D. (2014). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement. *Cortex* 66, 149–155. doi: 10.1016/j.cortex.2014.04.007

Van Berkum, J. J. A. (2008). Understanding sentences in context: what brain waves can tell us. *Curr. Direct. Psychol. Sci.* 17, 376–380. doi: 10.1111/j.1467-8721.2008.00609.x

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *J. Exp. Psychol.* 31, 443–467.

Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054

van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., et al. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Soc. Cogn. Affect. Neurosci.* 7, 173–182. doi: 10.1093/scan/nsq094

van Dommelen, W. A., and Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identifcation: sex-specifc behaviour. *Lang. Speech* 38, 267–267.

Yamada, Y., and Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Res.* 1130, 167–180. doi: 10.1016/j.brainres.2006.10.052

Ye, Z., Luo, Y., Friederici, A. D., and Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: evidence from event-related potentials. *Brain Res.* 1071, 186–196. doi: 10.1016/j.brainres.2005.11.085

Zawiszewski, A., Santesteban, M., and Laka, I. (2015). Phi-features reloaded: an ERP study on person and number agreement processing. *Appl. Psycholinguist.* (in press). doi: 10.1017/S014271641500017

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Interaction of Morphological and Stereotypical Gender Information in Russian

Alan Garnham * and Yuri Yakovlev

School of Psychology, University of Sussex, Brighton, UK

Previous research, for example in English, French, German, and Spanish, has investigated the interplay between grammatical gender information and stereotype gender information (e.g., that secretaries are usually female, in many cultures), in the interpretation of both singular noun phrases (*the secretary*) and plural nouns phrases, particularly so-called generic masculines—nouns that have masculine grammatical gender but that should be able to refer to both groups of men and mixed groups of men and women. Since the studies have been conducted in cultures with broadly similar stereotypes, the effects generally reflect differences in the grammatical systems of the languages. Russian has a more complex grammatical gender system than the languages previously studied, and, unlike those languages frequently presents examples in which grammatical gender is marked on the predicate (in an inflection on the verb). In this study we collected stereotype norms for 160 role names in Russian, providing a useful resource for further work in this language. We also conducted a reading time study examining the interaction of grammatical and stereotype gender information in the interpretation of both Russian singular noun phrases, and plurals that were (potentially) generic masculines. Our results show that, although both types of gender information are used, when available, the effects of grammatical marking on the predicate are not as strong as those of such marking on subject noun phrases.

Keywords: gender, stereotype, grammatical gender, text comprehension, Russian

INTRODUCTION

The understanding of written text and spoken discourse depends on the integration of information that is explicitly presented in the text with background information, both specific and general, that the comprehender has. The combination of these two types of information should be straightforward according to theories, such as the theory of mental models (Johnson-Laird, 1983), that claim the two types of information have the same format. The combining of the two types of information can produce additional pieces of information, which, if the information is in the form of descriptions of part of real or imaginary worlds, can be regarded as inferences from what is explicit in the text, making use of background information. There has been some debate about the extent to which inference making occurs routinely during comprehension. McKoon and Ratcliff's (1992) minimalist hypothesis, for example, claim that inference making is relatively restricted. Alternative, constructivist accounts (e.g., Graesser et al., 1994) place more emphasis on inference making, attempting, for example, to identify classes of inference that are routinely made, or circumstances in which inferences are made in an "effort after meaning." To some extent

the minimalist vs. constructivist dichotomy is a false one (Garnham, 1992), and it may be more productive, or even necessary, to pursue a common theme in both approaches, which is to specify and investigate factors that determine when inferences are made.

One idea that we have pursued (e.g., Garnham, 2005: 248) is that inferences that depend strongly on the presence of particular lexical items in a text might be particularly easy to make, as the template for the inference is (very likely) retrieved as the lexical item is processed. One domain in which we have investigated this idea is that of social stereotypes associated with occupational and other social roles. We would not claim that our work on stereotypes has entirely supported our ideas. There are, for example, terms (in British English) such as *primary school teacher* that would not typically be thought of as single lexical items, but that behave very similarly, from a gender stereotyping point of view, to lexical items such as *secretary*. The original idea behind our work, which was primarily a study of inferencing rather than a study of stereotypes, was that a word such as *secretary* would be directly linked in memory to the knowledge that, in the society in which our experimental participants moved, a high proportion of secretaries are female. So, in interpreting a statement presenting a particular individual as *the secretary*, this information would probably come to mind, and could result in the individual being represented as either probably a female or even definitely a female (though in a way that could be overridden if later information indicated that the person was male).

Of course, there is a sense in which such inferences should not be made. The core meaning of a word such as *secretary* is the information that defines that role. And it is a role that can be fulfilled by either females or males. From that point of view, simply describing someone as *the secretary* says nothing about their gender. Nevertheless, we found clear evidence that, in reading simple English texts, people do take secretaries to be (probably) female and engineers to be (probably) male (e.g., Carreiras et al., 1996; Garnham et al., 2002; Oakhill et al., 2005). Some of our individual findings do not clearly show that this inference is made as the word *secretary* or *engineer* is read, as the effect shows up when a later coreferential pronoun *he* or *she* occurs. Nevertheless, we believe that our results overall do support the idea of an immediate inference (see particularly the arguments in Reynolds et al., 2006).

As an Indo-European language, English is relatively unusual in that its nouns do not fall into grammatical gender categories and, except for pronouns and certain derivational endings, such as –ess, which are falling out of favor, there is little morphological marking for gender, for example on articles and other determiners, nouns, and adjectives. So, in our English experiments we would typically have a noun phrase, such as *the secretary* or *the engineer* that was not grammatically marked for gender followed by a pronoun *he* or *she* or *him* or *her* that was. At the initial noun phrase, therefore, stereotype information had free reign to determine the representation of the sex and/or gender of the protagonist. In other European languages that we have studied, such as Spanish, French, and German, the situation is different. Determiners are inflected (for example, Spanish: *el/la*; French: *le/la* German: *der/die/das*), and there

may also be indications in the noun ending of likely gender (for example, Spanish: *-o/-a*). Thus, grammatical and stereotype information may be working either together or against each other to provide information about the sex/gender of a person talked about in a text. Indeed, our research (Carreiras et al., 1996; Gygax et al., 2008) shows different patterns of results for each of the languages studied, which can be explained by the interaction of their different patterns of morphological marking of gender and the relatively constant effects of gender stereotyping. A further complication in these languages, which we have studied in the context of French and German, is the generic use of the masculine. A particularly pertinent case, and the one we capitalized on (Gygax et al., 2008), is when a speaker or writer is not sure or does not wish to commit to whether a group of people is all male, all female, or mixed. In such cases, a masculine plural noun phrase can be used (e.g., French: les assistants sociaux; German: Die Sozialarbeiter; English: the social workers). However, at least locally, the use of such an expression will be ambiguous between this generic case and cases in which the composition of the group is known to the entirely male. A difference between German and French, which appeared to influence our results, is that the German masculine plural definite article has the same spelling as the feminine singular *die*.

There is also some evidence from Russian on the interpretation of generic masculine nouns. Doleschal (1993) conducted an experiment in which she investigated how generic masculine nouns denoting persons are interpreted in Russian. The results of the study showed that, when attention shifts from the speaker's to hearer's perspective, neutral-masculine nouns are predominantly perceived and interpreted as male. However, it has been noted that some masculine nouns, such as *bukhgalter*[1] (accountant) and *vrach* (doctor) received higher female scores than male. Doleschal's assumption was that there might be an influence of what she called extra-linguistic factors, knowledge that some occupations are typically female. Moreover, some of the common gender nouns, for example *nedotroga* (touchy person) and *umnitsa* (know-all), also were interpreted as female. Similar results were obtained by Schmid (1998, cited in Doleschal and Schmid, 2001: 266), whose data suggested a strong tendency for interpreting masculine-neutral (generic) nouns as denoting men, though with some exceptions.

In the present study, we further investigate these issues in Russian. In the first part of the study we collect a new and more comprehensive set of stereotype norms for 160 Russian role names. The second part is an experimental study of stereotyping in short Russian texts, using the self-paced reading technique that has been used in previous studies. More specifically, we investigated the interpretation of masculine nouns with either masculine, neutral, or feminine stereotypes, and how the stereotype information interacted with grammatical information, in particular the inflection on the main verb in the sentence containing the stereotyped role name, and the gender of a definite pronoun in the following sentence. We looked at role names in both singular NPs, which we intended to be interpreted as referring to a specific person, and plural NPs, which were

---

[1] All transliterations in this paper follow the BGN/PCGN system.

intended to be interpreted generically. The nouns fell into three different groups, with different gender-related properties, and we were interested in whether nouns from the three groups behaved differently. These groups of nouns are described in detail below. Before describing the studies, we present some information about grammatical gender in Russian, which is more complex than in the other languages we have studied, and can be hypothesized to have different effects on processing.

Grammatical Gender in Russian

Russian is a language in which gender is predictable from semantic and morphological factors. Russian nouns are divided into three gender classes: feminine, neuter, and masculine (though see below for the notion of a fourth, common, gender). All nouns are marked grammatically and agree with adjectives, verbs (past-tense) and pronouns, the form of which depends on the gender of the noun they refer to. Gender agreement with the verb is a major difference between Russian, on the one hand, and English, French, German, and Spanish, on the other. An important question, therefore, will be whether gender marking on the verb acts in the same way, in on-line processing, as the types of gender marking studied in English, French, German, and Spanish (on determiners, nouns, and pronouns).

Russian also contains declension classes for nouns, and there is a strong relationship between gender and declension class. Traditionally nouns are categorized into three major classes:

●Declension I
Nouns in declension I are mainly feminine, though some are masculine. In the nominative singular they end in –*a* or –*ya*:

> *glina*—clay (f), *zemlya*—earth (f), *yunosha*—youngster (m), babushka—grandmother (f).

●Declension II
Nouns in declension II are mainly neuter, though a few are masculine and neutral nouns. In the nominative singular they either end in a consonant or in -*e* or -*o*:

> *vecher*—evening(m), *utro*—morning (n), *zadaniye*—task(n), *bereg*—coast(m).

●Declension III
Nouns in declension III are feminine. In the nominative singular they end in a soft consonant:

> *noch'*—night (f), *rol'*—role (f)

Noun agree in gender with adjectives, participles (in the singular), and verbs (in the past tense), as illustrated in the following examples:

> *nastal teplyy vecher*—a warm(m) evening(m) came(m)
> *nastala teplaya noch'*—a warm(f) day(f) came (f)
> *nastalo teploye utro*—a warm(n) morning(n) came (n)

Each gender category has an animate and inanimate subgender. For inanimate nouns the assignment of grammatical gender is (semantically) arbitrary, as in other gender-marked languages. For example, sun is neutral in Russian (*solntse*), feminine in German (*die Sonne*) and masculine in French (*le solei*) and

Spanish (*el sol*). However, grammatical gender correlates strongly with morphological factors, and in particular declension class (Corbett, 1982, 1991). All nouns denoting human beings are animate and normally belong to either masculine or feminine grammatical gender, depending on their semantic gender. Only a few animate nouns have neutral gender, for example *chudovishche* (monster) and *zhivotnoye* (beast). Nouns that denote people are known as personal nouns. They can be divided into six classes or groups:

a. Paired nouns with independent words for each gender.

> *medsestra—medbrat*
> nurse (f, sg.)—nurse (m, sg.)/(medical sister—medical brother)

> *evropevka—evropevets—evropevki—evropevtsy*
> European (f, sg.)—European (m, sg.)—Europeans (f, pl.)—Europeans (m, pl., generic)

b. Paired nouns with masculine nouns that can denote a female person in contexts when gender is not important.

> *studentka—student—studentki—studenti*
> 'student (f, sg.)—student (m, sg.)—students (f, pl.)—students (m, pl., generic)'

> *uchitel'nitsa - uchitel' - uchitel'nitsy—uchitelya*
> 'teacher(f.sg)—teacher(m.sg)—teacher(f. pl.)—teachers (m. pl. generic)'

c. Masculine nouns that do not have a feminine counterpart.

> *kosmetolog—kosmetologi*
> 'beautician (m., sing)—beauticians (m, pl., generic)'

> *electrik—electriki*
> 'electrician (m, sg.)—electricians (m, pl., generic)'

d. Masculine nouns that have only a so-called "colloquial" feminine counterpart. The use of colloquial feminine nouns goes against the norms of modern literary Russian (and they usually carry a negative connotation, if used). In addition, their use is ambiguous as they were used in the past to refer to a female person who was married, for example, to a doctor or a professor.

> *vrachikha—vrach—vrachi—vrachikhi*
> 'doctor (f. sg. colloq.)—doctor (m. sg.)—doctors (m. pl., generic)—doctors (f. pl. colloq.)'

> *parikmakhersha—parikmakher—parikmakhery— parikmakhershi*
> 'hairdresser (f. sg. colloq.)—hairdresser (m. sg.)—hairdressers (m, pl., generic)—hairdressers (f. pl. colloq.)'

e. Feminine nouns that do not have male counterpart. In contrast to masculine personal nouns that do not have feminine counterpart, they cannot denote a male person. The only exceptions are in metaphorical expressions.

balerina—baleriny
'ballerina (f, sg.)—ballerinas (f, pl.)'

nyanya—nyani
'nanny (f, sg.)—nannies (f, pl.)'

We note however, that it is possible to create masculine counterparts for these nouns (e.g., *baleron/balerun*—male ballet dancer or *nyan'*—male nanny), but their use would be colloquial and almost always humorous. The appearance and popularization of *nyan'* is related to relatively recent release of a comedy film "The Sitter" (2011). The word *"balerun"* was popularized in 90s after several ironical usages in the press.

f. Common gender nouns (mostly in spoken language) denote both males and females. This gender is different from the neutral gender. Common gender nouns all end in *-à/-ya*. So they belong to a declension in which most of the nouns are feminine. However, in terms of modifying adjectives, ordinal numbers, pronouns, and past tense verbs, agreement depends on the semantic gender of the individual in question.

plaksa—plaksy
'weeper (common gender, sg.)—weepers (common gender, pl.)'

kollega—kollegi
'colleague (common gender, sg.)—colleagues (common gender, pl.)'

The gender of a personal noun can be unambiguously determined on the basis of its semantic or morphological agreement with other syntactically dependent words (e.g., adjectives and pronouns). However, semantic agreement is complicated by the existence of classes c, d, and f, described above. For example, masculine nouns that do not have a feminine counterpart may trigger feminine agreement in certain syntactic positions (1):

(1) *kosmetolog skazala*
    'the beautician (m) said (f)'

Corbett (1991) argued that the agreement of these nouns is subject to an agreement hierarchy (attributive—predicate—relative pronoun—personal pronoun), with semantic agreement becoming increasingly common from left to right. So, (2) is relatively acceptable, with sematic agreement in the predicate, but (3) with attributive semantic agreement is not.

(2) *budushchiv filolog skazala*
    'future (m) philologist (m) said (f)'
(3) *ya vstretil tvoyu kosmetologa** (not allowed)
    'I met (m) your (f. accusative) beautician (m. accusative)'

In contrast, common gender nouns have two consistent agreement patterns (feminine and masculine). The choice of gender depends only on the semantic gender of the referent. See example (4) below.

(4) *nash novvy kollega skazal*
    'our (m) new (m) colleague (common gender) said (m)'

nasha novaya kollega skazala
'our (f) new (f) colleague (common gender) said (f)'

In cases where the gender of a person is unknown or irrelevant, Russian uses masculine forms, which are seen as stylistically neutral. For example, in official contexts it is more appropriate to use "masculine-neutral" nouns. In plural forms masculine nouns are regularly used generically.

RATING STUDY

The aim of the rating study was to produce norms for the gender stereotypicality of selected role nouns in Russian. The norms are of interest in themselves, but are also needed to construct items for the main on-line experiment, because appropriate norms for Russian do not exist. Of the six classes of personal nouns described above, only three (b, c, d) were included in rating study: masculine nouns with feminine counterpart ("paired"); masculine nouns with colloquial feminine pair ("colloquial"); masculine nouns without a feminine counterpart ("unpaired")[2]. We looked at whether nouns in the different groups received significantly different ratings.

Ethical approval for this study and for the online study that follows was granted by the University of Sussex Life Sciences & Psychology Cluster-based Research Ethics Committee and all participants provided written consent prior to taking part. All procedures complied with the British Psychological Society's Code of Human Research Ethics.

Methods
Questionnaire and Design
Gender stereotypes for 160 role nouns in Russian were evaluated in an online questionnaire. The selection of role names and the design of the survey were based on previous studies in other European languages (Kennison and Trofe, 2003; Gabriel et al., 2008; Misersky et al., 2014). The role nouns were divided into three groups: gender paired nouns, where the masculine noun can refer both to men and women ($n = 44$); masculine nouns without a feminine counterpart ($n = 55$); masculine nouns which have a feminine counterpart, but only one that is used colloquially ($n = 61$).

Role names were presented in the masculine plural form (serving as generic) on the left side of the screen, slightly separated from an 11-point rating scale, which ranged from 100%

---

[2]We used four electronic dictionaries to allocate nouns to the three noun groups. T.F. Yefremova. The Comprehensive Dictionary of the Contemporary Russian Language. ©2006; http://www.lingoes.net/en/dictionary/dict_down.php?id=79A76AF19099A343BD990EC195719601.
S.A. Kuznetsov. Comprehensive Russian Explanatory Dictionary ©2010; http://www.lingvo-online.ru/en/LingvoDictionaries/Details?dictionary=ExplanatoryBTS%20%28Ru-Ru%29.
Dmitry Ushakov. Explanatory dictionary of Russian. ©≪ ACT, Astrel≫, Moscow, 2000; http://dic.academic.ru/contents.nsf/ushakov/.
S. Ozhegov and N. Shvedova. Explanatory dictionary of Russian. ©Very Ltd., Israel, 1994. http://dic.academic.ru/contents.nsf/ogegova/.
We checked if each masculine noun had a feminine form. If it did not it was added to the unpaired class. If the feminine noun was labeled *razg.* (colloquial), *ustar.* (archaic) or *prost.* (vulgar) the masculine noun was added to colloquial class. The remaining nouns formed the paired class.

women and 0% men on the left to 0% women and 100% men on the right. Previous work (Gabriel et al., 2008) suggests that presenting generic masculine forms of the role names, as opposed to explicitly gender marked versions, can increase the perceived proportion of males. However, we decided not to use specific feminine and masculine personal nouns because of the inclusion of nouns with colloquial feminine counterparts. Some of these feminine forms are archaic, some vulgar, and they are rarely, if ever, used in written form. As in other similar studies (Gabriel et al., 2008), our data showed that some participants interpreted the masculine generic version of some nouns (e.g., florists) as specifically male despite the fact that they were embedded in the series of personal nouns that were generically interpreted. However, the comparability of our results with previous results in other languages (e.g., Misersky et al., 2014) suggests that such responses did not constitute a serious problem.

Another issue is that ratings are influenced by scale direction (Kennison and Trofe, 2003; Gabriel et al., 2008). A scale with 100% male on the right is associated with a numerically small but significant increase in the tendency to rate nouns as referring to males. Given that the effect is small, we decided to use one scale direction only.

After reading the instructions and indicating their consent form, participants were asked to estimate the proportion of females vs. males in each role. The list of the nouns was randomized and presented in the same order to all participants. On the last page of the questionnaire participants were asked to indicate their native language and answer optional demographic questions (age group, gender, education level). The questionnaire was created using Bristol Online Surveys (BOS) and administered via the Web. Its design is demonstrated in Appendix A (Supplementary Material). The list of nouns is given in Appendix B (Supplementary Material), together with summary data from the survey for each noun.

Sample and Procedure

A total of 112 participants took part in the rating study. They were recruited via advertising in Russian social networks, and participation was on a voluntary basis. Data from six participants were excluded from the analysis because Russian was not their mother tongue ($n = 4$) or because they did not understand the instructions ($n = 2$). The final sample, therefore, consisted of 106 participants (16 male, 87 female and 3 who chose not to specify their gender).

Results

Data from the questionnaire was coded so that high values on the scale reflect a higher proportion of men, for example "100% women and 0% men" was recorded as 1, "50% women and 50% men" as 6 and "0% women and 100% men" as 11.

Interparticipant Analyses

For each participant, the mean rating across the role names was calculated ($M = 6.38$, $SD = 0.33$, range 5.26–7.34; scale midpoint = 6). The overall distribution of scores was normal, Kolmogorov–Smirnov's $D_{(106)} = 0.06$, $p = 0.2$. Female participants rated the proportion of men as being slightly, though

not significantly, higher ($M = 6.4$, $SD = 0.3$) than male participants ($M = 6.3$, $SD = 0.48$), [$t_{(101)} = 1.04$, $p = 0.297$]. Overall, the proportion of women and men was rated similarly by participants in four age groups "18–24," $n = 24$ ($M = 6.41$, $SD = 0.31$); "25–34," $n = 43$ ($M = 6.36$, $SD = 0.31$); "35–44," $n = 25$ ($M = 6.43$, $SD = 0.39$); "45–54," $n = 13$ ($M = 6.35$, $SD = 0.37$), [$F_{(3, 101)} = 0.32$, $p = 0.82$] and at two education level groups "high school," $n = 11$ ($M = 6.39$, $SD = 0.27$); "university degree," $n = 94$ ($M = 6.39$, $SD = 0.34$), [$t_{(103)} = 0.06$, $p = 0.95$].

Interitem Analyses

The mean rating and standard deviation were calculated for each noun (see Appendix B in Supplementary Material for overall data, and for mean ratings by female and by male participants). A scatterplot of the mean rating for each noun against its standard deviation is shown in **Figure 1**. Low standard deviations, which reflect consensus in perceived proportions, can be seen in the middle and at the both ends of the scale. These results are in line with previous research in English, German and French (Gabriel et al., 2008, Figures 3–5). As in Gabriel et al.'s study more nouns are found at the male than at the female end of the scale.

Means were also calculated for each noun group. Participants rated unpaired masculine personal nouns as denoting a higher proportion of males than females ($M = 6.78$, $SD = 2.17$), paired masculine nouns as denoting close to equal proportions of females and males ($M = 6.06$, $SD = 1.54$), and masculine personal nouns with colloquial feminine pair as denoting a slightly higher proportion of males than females ($M = 6.25$, $SD = 1.98$). A One-way between-items ANOVA showed that the difference between these means is not significant [$F_{(2, 157)} = 1.92$, $p = 0.15$].

ONLINE EXPERIMENT

The main experiment investigated the role of gender stereotypes associated with role names on the interpretation of grammatically

FIGURE 1 | Scatterplot of mean ratings for each personal noun ($n = 160$) in the rating study against its standard deviation.
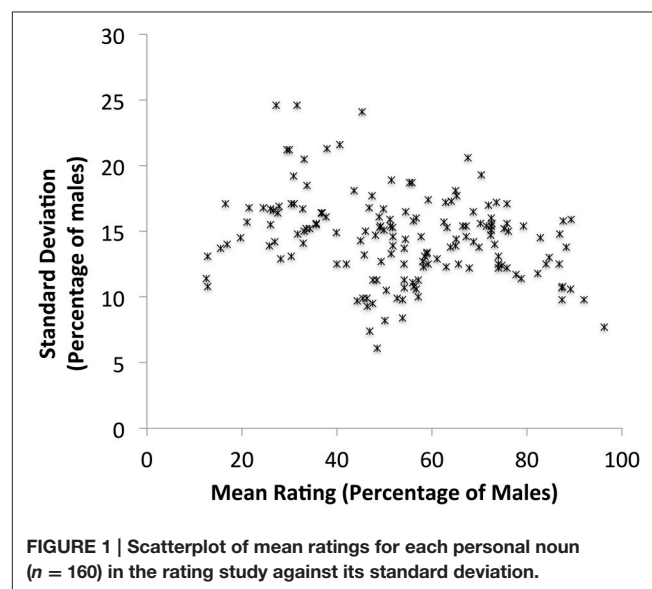
masculine Russian nouns. Three sets of nouns were selected on the basis of rating study: (a) female stereotyped, (b) neutral, and (c) male stereotyped. The experiment investigated the interpretation of masculine nouns both when they are intended as specific and when they are intended as generic, in passages such as those shown in **Table 3**. Nouns that were intended as specific occurred in the singular form. Nouns that were intended as generically occurred in the plural form. We used passages both in the present tense, where the marking on the verb is not gender-specific, and in the past tense, where it is. We expected to see different results in these two cases.

Predictions

Nouns Intended to be Interpreted Specifically

In the experimental items, the subject noun of the first sentence was always masculine singular, though it could be male, neutral or female stereotyped, and it might turn out to refer to a male or

FIGURE 2 | Interaction of stereotype (female, neutral, male) and gender marking (masculine, feminine) in first sentence reading times (ms), for past tense specific sentences only. Error bars represent standard errors calculated by the SPSS MIXED procedure.

a female, depending on the morphological markings in the rest of the passage. The subject pronoun of the second sentence could be masculine or feminine. In addition, in the past tense, the verbs were morphologically marked as masculine or feminine. The markings on both the pronouns and the verbs were determined by the sex of the person being discussed.

The match or mismatch between the grammatical gender of the subject noun of the first sentence and the stereotype was immediately apparent, but any mismatch effect might be mitigated by the use of masculine nouns to refer to females in Russian, both generally, and in the course of this experiment.

The sex of the person referred to, and any clash with other gender information, therefore first becomes apparent at the predicate of the first sentence in the past tense passages and at the subject pronoun in the second sentence in the present tense passages.

Our results in Spanish (Carreiras et al., 1996) suggest that mismatch effects between stereotype and sex of protagonist occur, and can be resolved, at the earliest possible point, predicting a stereotype mismatch effect in the first sentence (only) for past tense passages and in the second sentence (only) for present tense passages. However, as gender marking on verbs has not previously been studied, we cannot be certain it will have the same effect as, for example, the gender marking on the determiners in the Spanish experiment, which is syntactically closer to the noun itself.

Other mismatch effects, in particular between the grammatical gender of the subject noun of the first sentence and the information that determines the sex of the protagonist should also be present, but, for reasons stated above, may not necessarily affect processing. However, we might expect different behavior from the different noun groups. With the paired nouns (i.e., those that have a feminine counterpart), and to a lesser extent with the colloquial nouns (where the feminine counterparts

FIGURE 3 | Three-way interaction of tense (past, present), pronoun (she, he), and noun group (unpaired, paired, colloquial) in second sentence reading times (ms), for specific sentences. Error bars represent standard errors calculated by the SPSS MIXED procedure.
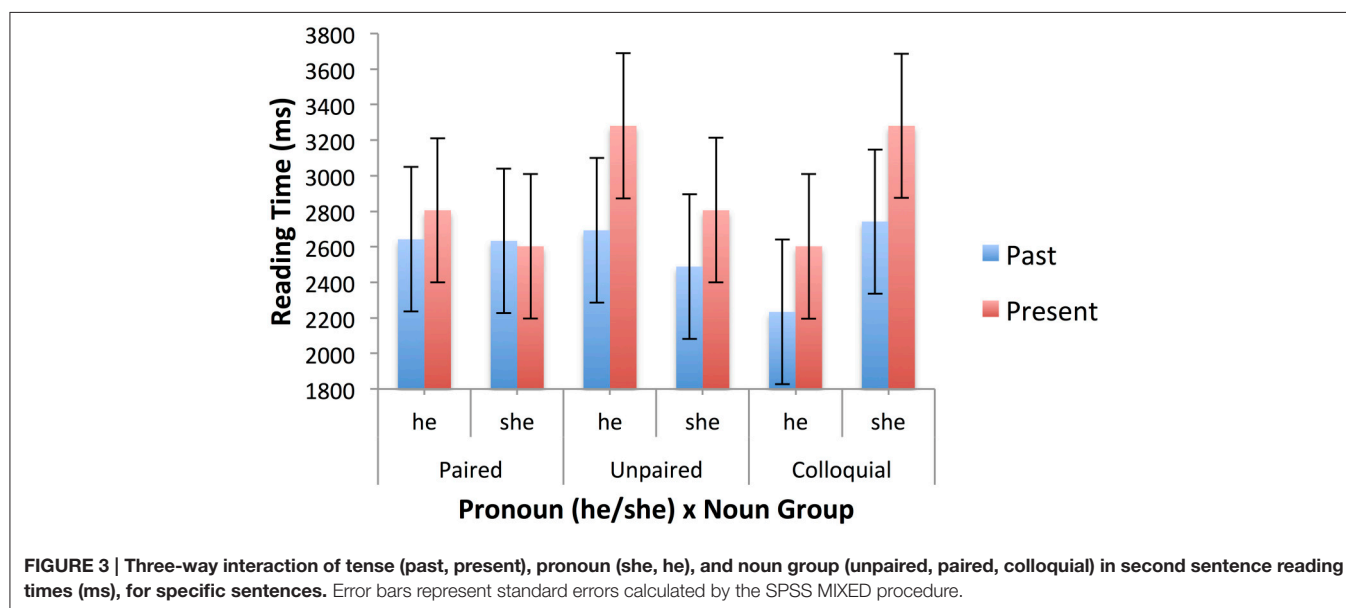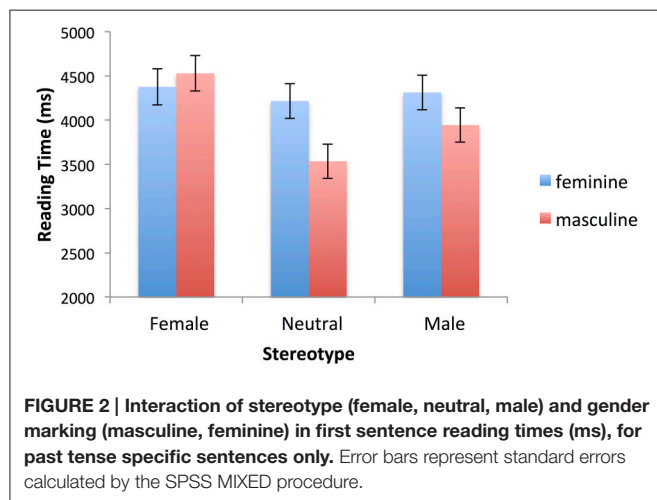
FIGURE 4 | Three-way interaction of tense (past, present), pronoun (she, he), and stereotype (female, neutral, male) in second sentence reading times (ms), for specific sentences. Error bars represent standard errors calculated by the SPSS MIXED procedure.

exist, but have different connotations and/or denotations), the use of the masculine form strongly suggests the person being referred to is male. Thus, a feminine-marked predicate in the first sentence (past tense form) is potentially problematic, and may cause processing difficulties. Furthermore, in the present tense passages, the end of the first sentence is reached with no suggestion that the person being referred to is not male. In these passages, a feminine pronoun in the second sentence is likely to be particularly problematic, as the notion that the protagonist is masculine may have become entrenched. A further complication is that such effects may be modified by the stereotype of the role noun in the first sentence. A female stereotyped noun is likely to make reference to a female person more plausible.

Nouns Intended to be Interpreted Generically

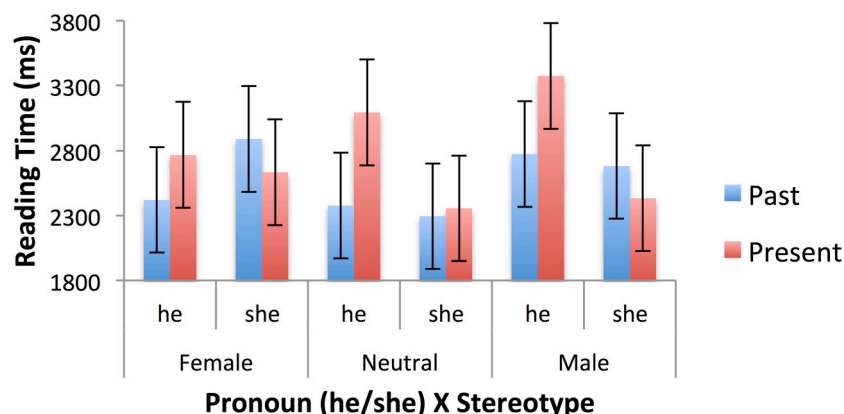Gygax et al. (2008) showed that in French and German morphologically marked generically-intended masculine plural nouns were interpreted as referring to groups of males. Given that generic plurals in Russian are also morphologically marked as masculine, a similar effect can be predicted.

Methods
Participants

Twenty volunteers (8 men and 12 women, mean age = 29.9 years, SD = 8.18) took part in the experiment. Most were students and staff from Sussex University and the University of Brighton. They were people from the former Soviet republics of Belarus, the Russian Federation, Ukraine, Kazakhstan, and Latvia, whose mother tongue was Russian. All subjects were entered into a £25 prize draw for their participation, which lasted for about 35 min.

Materials

Seventy-two nouns (24 male stereotyped, 24 female stereotyped, and 24 neutral) were selected for the texts with specific nouns. A further 36 nouns (12 male stereotyped, 12 female stereotyped, and 12 neutral) were selected for the texts with generic nouns. The mean stereotype ratings and standard deviations for the selected nouns (on the 11-point scale) are shown in **Table 1**.

TABLE 1 | Mean ratings and standard deviations for selected stereotyped nouns.

Stereotype	Specific nouns	Generic nouns
Female	3.8 (0.69)	3.5 (0.69)
Neutral	6.1 (0.29)	6.2 (0.26)
Male	8.3 (0.59)	8.4 (0.43)

TABLE 2 | Mean ratings and standard deviations for paired, unpaired, and colloquial nouns within each group of stereotyped nouns.

Stereotype	Unpaired	Paired	Colloquial
Female	3.5 (0.79)	4.1 (0.53)	3.8 (0.66)
Neutral	6.2 (0.29)	6.0 (0.31)	6.1 (0.28)
Male	8.4 (0.57)	8.1 (0.72)	8.4 (0.43)

The specific nouns were, in addition, taken from three classes (8 paired, 8 unpaired, and 8 colloquial) within each gender-stereotyped group. **Table 2** displays the mean ratings and standard deviations for these groups of nouns.

For each noun a two-line text was constructed. For the specific nouns, the noun, which was always grammatically masculine, though stereotypically female, neutral or male, was the subject of the first sentence. The second sentence began with the (Russian equivalent of the English) pronoun *he* or *she*. Each text could be written in either the present or the past tense. In the past tense the predicate in the first sentence was marked for the real-world gender of the character, thus mismatching the grammatical gender of the noun when it referred to a female person (in the present tense the predicate is gender neutral). Each noun was used in all four types of passages: (1) present tense, referential pronoun *she*; (2) present tense, referential pronoun *he*; (3) past tense, referential pronoun *she*; (4) past tense, referential pronoun *he*. The occurrence of a noun in the four types of passage was counterbalanced between participants.

For generic nouns, the first sentence was used to introduce a group of people using a masculine plural role name (female, male, or neutral stereotyped), and the second talked about some of the men or some of the women in the group. **Table 3** displays sample sentences.

After the two sentences were displayed, participants were asked the following question: "Is it possible to use first and second sentences in that way?" Participants were to answer either Yes or No.

To prevent participants from realizing that only masculine nouns were under investigation, there were 20 filler texts that included specific nouns with feminine grammatical gender, and 10 filler texts with feminine plural nouns, which, unlike masculine plurals, cannot be used generically. It was, therefore, possible to construct texts that required a definite No respond to the evaluation question.

Procedure

All participants were tested individually in a small, quiet room. Their task was to read each passage at a fast but comfortable speed and to decide whether the two sentences fitted together to make a sensible passage. The experiment was built using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA)[3]. There were two untimed breaks during the experiment; participants could continue when they were ready by pressing a continue key, the spacebar. The spacebar was also used to advance the presentation of the sentences. The "C" key was used for No responses to evaluation questions and the "M" key was used for Yes responses. Each participant viewed seven practice passages before the main part of the experiment.

Experimental Design

The experiment investigated two questions in parallel. The first question was about the interpretation of specific nouns followed by pronouns that refer to the male or female person introduced by the noun. We were interested in four factors: (i) Gender stereotype of the noun in sentence 1 (Male, Female, Neutral), (ii) Gender of pronoun in sentence 2 (she vs. he), (iii) class of the noun (paired, unpaired, colloquial), and (iv) Tense of the predicate in sentence 1 (past vs. present; in the past tense the

[3]Psychology Software Tools Inc. [E.-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com.

predicate was gender marked, and the gender marking matched that of the pronoun in the second sentence).

The second question was about the interpretation of generic nouns. There were two factors in this part of the study: (i) Gender stereotype of the generic noun in sentence 1 (Male, Female, Neutral); and (ii) continuation in sentence 2 (*some of the men* vs. *some of the women*). For both types of text reading times of each sentence, responses to the classification question, and response times were recorded.

In each case, there were two or four versions of a passage with a particular role noun (masculine vs. feminine pronoun and past vs. present tense for singulars, "some of the men" vs. "some of the women" for plurals). Therefore, four versions of the experiment were created to counterbalance the allocation of items to conditions. A different random order was selected by E-Prime to present the passages to each participant.

Results and Discussion
Interpretation of Specific Nouns

Examination of the histograms for the reading times suggested that 10 s was a sensible cut off to use for both first sentence and second sentence reading times. The lengths of the sentences showed considerable variability: first sentences mean = 51.1 characters, min = 22, max = 105; second sentences mean = 46.4, min = 22, max = 69. Therefore, for the purposes of analysis, raw reading times were replaced with residual reading times, based on regressions of reading time against length in characters (including spaces) calculated separately for each participant and for each sentence (first and second).

Table 4 shows the mean raw reading times, after truncation, for the first sentences of the passages in the past and present tenses. Mixed effects models, with stereotype (female, neutral, male), pronoun (he, she), tense (past, present), and noun group (unpaired, paired, colloquial) as fixed effects were fitted to the residual reading time data using the SPSS MIXED procedure. A compound symmetry covariance structure was chosen for the repeated effects, and random intercepts (only) were selected for both participants and materials, so no issue of covariance structure for random effects arose[4]. Because we

[4]Ideally, we would have fitted a full model, with random intercepts for all repeated effects, as recommended by Barr et al. (2013). However, neither SPSS MIXED nor lmer from the R package lme4 would converge on such a model a reasonable time (several days).

TABLE 3 | Examples of sample sentences from the reading experiment.

Sentence 1	Sentence 2
SPECIFIC NOUN, PRESENT TENSE PASSAGE	
Kosmetolog govorit po telefonu.	*Ona/On ob"yasnyaet novomu klientu kak ih nayti.*
"The beautician (masc.) is talking on the phone."	"She/He is explaining to a new client how to find them."
SPECIFIC NOUN, PAST TENSE PASSAGE	
Kosmetolog govorila/govoril po telefonu.	*Ona/On ob"yasnyala/ob"yasnal novomu klientu kak ikh nayti.*
"The beautician (masc.) talked (fem./masc.) on the phone."	"She/He explained (fem./masc.) to a new client how to find them."
GENERIC NOUN	
Inzhenery stroili model' 3 chasa.	*Neskol'ko muzhchin/zhenshchin reshili otdokhnut' 15 minut.*
"Engineers (masc., pl., generic) were building the model for 3 hours."	"Some of the men/women decided to take 15 minutes break."

TABLE 4 | Mean reading times (ms) for first sentences for the texts in which the nouns were intended to be interpreted specifically.

Stereotype		Female		Neutral		Male	
Pronoun in second sentence		Fem	Masc	Fem	Masc	Fem	Masc
PAST TENSE							
Noun Group	Unpaired	4914	4692	4258	3321	3884	3443
	Paired	4467	4567	4512	3612	4575	4017
	Colloquial	3748	4328	3877	3674	4479	4373
PRESENT TENSE							
Noun Group	Unpaired	4469	3628	3841	4321	3122	3358
	Paired	4018	3581	3455	3839	3715	4126
	Colloquial	3776	3914	4209	4008	4956	4229

For the past tense passages, the gender of the pronoun in the second sentence is also the gender of the inflection of the verb in the first sentence.

TABLE 5 | Mean reading times (ms) for second sentences for the texts in which the nouns were intended to be interpreted specifically.

Stereotype		Female		Neutral		Male	
Pronoun		She	He	She	He	She	He
PAST TENSE							
Noun Group	Unpaired	2535	2886	2474	2323	2918	2691
	Paired	2722	2828	2391	2060	2966	2578
	Colloquial	2001	2953	2264	2500	2436	2773
PRESENT TENSE							
Noun Group	Unpaired	2430	2595	2959	2666	3028	2547
	Paired	3081	2523	3052	2152	3710	2579
	Colloquial	2786	2780	3269	2248	3387	2172

expected different effects in the present tense and past tense passages, we also fitted mixed models to these two subsets of the data separately. All reported times are based on the actual set of times, after truncation, in the various conditions of the experiment.

For the first sentences there was a main effect of tense, $F_{(1, 1237.181)} = 7.78$, $p = 0.005$, Reading was faster for present tense passages than past tense passages (3920 vs. 4152 ms). For the past tense sentences only there was a main effect of the gender marking on the predicate $F_{(1, 573.102)} = 3.691$, $p = 0.055$., and a marginal interaction of gender marking and stereotype, $F_{(2, 597.831)} = 2.739$, $p = 0.065$. Reading times were higher when the gender marking was feminine (4301 vs. 4003 ms). In addition, times were higher when masculine gender marking mismatched a female stereotype than when it was consistent with a male or neutral stereotype. There were no differential effects of stereotype when the gender marking was feminine. The interaction pattern, including the stereotype match-mismatch effect can be seen in **Figure 2**. In the analysis of the residual reading times, the crucial 3-way interaction of tense, stereotype and gender marking was not significant, $F_{(2,1253.111)} = 1.709$, $p = 0.18$), though it was significant in the analysis of the (trimmed) raw reading times, which we ran to compute the means reported in **Table 4**, $F_{(2, 1314)} = 3.372$, $p = 0.035$.

These results indicate that both grammatical and stereotypical gender information were used in the interpretation of the first sentences. In these sentences a masculine role noun was always used. However, both the sex of the person referred to, and the stereotype of the role could match with or clash with the grammatical gender of the noun (or bear a neutral relation in the case of the neutral stereotypes). When the person was female, the grammatical gender marking on the predicate of the first sentence mismatched the grammatical gender of the noun, when the sentence was in the past tense. In the present tense, the marking on the predicate was uninformative. So, in relation to the use of grammatical information, the main effect of tense on reading times for the first sentences showed that when there was more gender-related information to process (in the past tense), more time was taken to read the sentence, although the sentences did also differ in tense. Furthermore, in the past tense passages where

it was clear in the first sentence whether the semantic gender of the intended referent matched or mismatched the grammatical gender of the subject noun, a clash with the grammatical gender of the subject noun slowed people down.

In relation to the use of stereotype information, in some cases the gender stereotype matched the grammatical gender of the role noun, and in others it mismatched. A standard stereotype mismatch effect would, therefore, have been reflected in a main effect of stereotype. In particular, male stereotype would be match and female mismatch. However, no evidence for such an effect was found. However, there was evidence that stereotype information was used in the processing of the first sentence, but its effects were only seen when it clashed with two pieces of grammatical gender information—a female stereotype plus the masculine noun and masculine gender marking on the verb slowed people down. In the present tense, when the stereotype clashed only with the masculine grammatical gender of the subject noun, no stereotype mismatch effect was found, so it appears that the stereotype effect is driven by the mismatch with grammatical inflections on the predicate.

Table 5 shows the mean raw reading times, after truncation, for the second sentences of the passages in the past and present tenses. Mixed models were fitted using the SPSS MIXED procedure, as for the first sentences, and with the same limitations (see footnote 2). There were main effects of tense, $F_{(1, 1282.039)} = 7.111$, $p = 0.008$, pronoun, $F_{(1, 1282.233)} = 10.734$, $p = 0.002$, and gender stereotype, $F_{(2, 62.080)} = 2.609$, $p = 0.082$. Responses were faster for passages with *he* rather than *she* (2548 vs. 2800 ms), for past tense passages than present tense passages (2572 vs. 2776 ms), and for passages with neutral stereotypes (2530 ms) than female (2677 ms) and male (2815 ms). There were also significant two-way interactions between tense and pronoun, $F_{(1, 1282.089)} = 19.343$, $p < 0.001$, and pronoun and stereotype, $F_{(2, 1295.076)} = 5.605$, $p = 0.004$). Finally, there was a significant three-way interaction of tense, pronoun and noun group, $F_{(2, 1284.519)} = 3.126$, $p = 0.044$. The pronoun effect was restricted to present tense passages 604 vs. −99 ms), and among those it occurred only for paired and colloquial nouns, not unpaired nouns, with the larger effect being for paired nouns. These effects are illustrated in **Figure 3**. The pronoun

by stereotype interaction is clearly of interest to the literature on stereotypes. The pronoun "she" was read more slowly after male stereotyped nouns (3074 ms) than female (2592 ms.) or neutral (2735 ms), whereas "he" was read more quickly after male (2557 ms) and neutral (2325 ms) than female (2761 ms) stereotypes. **Figure 4** shows these effects. In the past tense (blue bars), the standard stereotype match-mismatch effect is seen, including the usual advantage for masculine pronouns following neutral stereotypes, though this effect is numerically small in the current data set. The corresponding effect in the present tense (red bars) is overlaid on the main effect of pronoun, with sentences containing the masculine pronoun being read faster overall in the present tense.

Looking at the separate analyses of present and past tense passages, the pronoun effect was only significant in the present tense passages, $F_{(1, 599.227)} = 31.677$, $p < 0.001$. The pattern is consistent with the size of the effect in the two types of passage noted above. The pronoun by stereotype interaction was significant in the present tense passages, $F_{(2, 587.973)} = 3.892$, $p = 0.021$, and marginal in the past tense passages, $F_{(2, 626.234)} = 2.53$, $p = 0.081$. A pronoun by noun group interaction, which was marginal in the overall analysis, $F_{(2, 1286.932)} = 2.724$, $p = 0.066$, was significant in the present tense passages, $F_{(2, 612.097)} = 3.408$, $p = 0.034$, and marginal in the past tense passages, $F_{(2, 614.511)} = 2.761$, $p = 0.064$. When the pronoun is "she" there is more difficulty if the noun in the first sentence is paired and so has a feminine counterpart that was not used. This pattern is not seen for the pronoun "he." Finally, for the past tense passages, there was a marginal effect of stereotype, $F_{(2, 76.619)} = 2.828$, $p = 0.065$. Neutral stereotypes led to faster reading of the second sentence in the past tense passages compared with all other conditions defined by stereotype x tense.

The pattern of reading times for the second sentences was complicated, and partly reflected differences in the information available in (and processed from) the first sentences. So, the effect of tense in the second sentence was the reverse of that in the first sentence. Given that more time had been devoted to the first sentences in the past tense passages (presumably because they contained more information), less was devoted to the second sentences in the same passages. Conversely, as the sex of the main character only became apparent in the second sentence of the present tense passages, more time was needed to process this information. As with the first sentences, there was evidence for the use of both grammatical and stereotypical information in the processing of the second sentences.

The initial pronoun in the second sentence carried the information about the actual sex of the character, which had also been indicated by the inflectional ending in the predicate in the first sentence of the past tense passages, but not of the present tense passages. In the present tense passages, therefore, any clash between the grammatical gender of the role name and the sex of the character, or between the stereotype of the role name and the sex of the character, could only become apparent in the second sentence. The effect of pronoun ("she" vs. "he") in the second sentence, which was restricted to the present tense passages, reflects the match or mismatch between the pronoun and the grammatical gender of the role noun, and that fact that

such a match or mismatch had already been observed in the first sentences of the past tense passages.

The pronoun by stereotype interaction in the second sentence is the standard stereotype match-mismatch effect reported elsewhere in the literature. There is some evidence, from the fact that the two-way interaction is only significant in the present tense, that the effect is stronger in the present tense passages, in which the second sentence is the first place in which the effect might be detected. In the past tense passages, the stereotype match-mismatch is present in the first sentence. However, the three-way interaction with tense is not significant and the pattern of the interaction is similar for past and present tense passages. In previous studies (e.g., Carreiras et al., 1996; Duffy and Keir, 2004), if such a clash appears early in a passage, it is dealt with at that point and does not affect later processing. The pattern is not so clear in the present study, which may be because inflections on verbs are less obvious indicator of a person's sex than the gender of a definite article (Carreiras et al., Spanish experiments) or the explicit use of the terms "male" and "female" (Duffy and Keir). One possibility is that, although inflections on past tense verbs are quite obvious gender indicators in spoken Russian, they may be less obvious in the written form as in most cases only a single character is added to a masculine form verb [*skazal—said (masc)*, *skazal*a—said (fem)]. However, such single character inflectional changes are not difficult to notice, at least when attention is drawn to them. More plausibly, it may be that, because of the proximity of the verb to the subject noun in our sentences, grammatical, rather than semantic agreement would be acceptable, so that the inflection on the verb is not a reliable indicator of the protagonist's sex.

Finally, the pronoun by noun group interaction is of some interest as it shows that readers are affected by other ways that a language makes available for expressing the same idea. For example, if a writer knows that a character is female and is going to indicate this fact by inflecting a past tense verb, it seems odd for that writer to use a masculine noun generically when a feminine counterpart is available (as in the case of paired nouns). However, the effect of this anomaly does not appear in first sentence reading times, again suggesting that the inflectional morphology on the predicate plays only a weak role in the representation of gender. Instead, it appears in the second sentence, where a personal pronoun provides more direct evidence of the person's sex.

Interpretation of Generic Nouns
Reading times
As for the specific items, examination of the histograms for the reading times suggested that 10 s was a sensible cut off to use for both first sentence and second sentence reading times. Again, as for the specific items, the lengths of the sentences showed considerable variability: first sentences mean = 68.2 characters, min = 39, max = 109; second sentences mean = 49.9, min = 29, max = 72. For the purposes of analysis, therefore, raw reading times were replaced with residual reading times, based on regressions of reading time against length in characters (including spaces) calculated separately for each participant and for each sentence (first and second).

Tables 6, 7 shows the mean raw reading times, after truncation, for the first and second sentences, respectively. Mixed effects models, with stereotype (female, neutral, male), continuation (women, men), and noun group (unpaired, paired, colloquial) as fixed effects were fitted to the residual reading time data using the SPSS MIXED procedure. A compound symmetry covariance structure was chosen for the repeated effects, and random intercepts (only, for the same reasons as stated in Footnote 2) were selected for both participants and materials, so no issue of covariance structure for random effects arose. All reported times are based on the actual set of times, after truncation, in the various conditions of the experiment.

No significant effects were found for the reading times of the first sentences.

For the second sentence, there was a marginal effect of continuation, $F_{(1, 661)} = 3.721$ $p = 0.054$, with slower reading times for "some of the women" (3511 ms) compared with "some of the men" (3071 ms). There was a also main effect of stereotype, $F_{(2, 661)} = 3.016$, $p = 0.05$, with reading times of 2942, 3190, and 3741 ms for female, neutral, and male stereotypes. However, the effect of stereotype in the reading times for the second sentence should be viewed with caution, even though we analyzed residual reading times, as it is a between-passages effect.

Judgements

Because of the possibility of non-generic interpretations, which would make *some of the women* infelicitous, we analyzed the judgment data for the "generic" passages. **Table 8** shows the mean percentage of positive responses There was a main effect of continuation ("some of the women" vs. "some of the men"), $F_{(1, 658.489)} = 12.178$, $p = 0.001$, with more positive responses for men (79.3%) than for women (67.6%). There was also a two-way interaction between continuation and noun group, $F_{(2, 661.417)} = 7.045$, $p = 0.001$; and a three-way interaction of continuation, noun group, and stereotype, $F_{(4, 661.464)} = 6.519$, $p < 0.001$. Given the main effect of continuation, the two- and three-way

interactions are primarily driven by the fact that "Some of the women" attracted a particularly low number of positive responses when the stereotype was male and the noun was from a paired couple. In this case, the use of the masculine noun, when an equivalent feminine noun is available, together with the male stereotype, reinforces the idea that the people being talked about are men.

The effect of continuation, with "some of the men" continuations being read more quickly than "some of the women" continuations, and being accepted more frequently, parallels the findings of Gygax et al. (2008) for French and German. Those authors interpreted that finding as evidence that the masculine plural noun phrases (in the first sentences) were interpreted as referring to groups of males, rather than as generic references to group of both men and women or of unknown composition.

The interactions of continuation and noun group and continuation, noun group and stereotype in the judgments, despite including between-item comparisons, are relatively unproblematic, as judgment is not directly affected by length. It appears, as with the specific passages, that the use of the masculine member of a pair of nouns, when there is also a male stereotype, makes it particularly difficult to consider that there are women in the group.

GENERAL DISCUSSION

We collected a set of stereotype norms for 160 Russian role names, and used a subset of these role names to construct short passages for an online study of the use of stereotype and other gender information in the interpretation of Russian. The stereotype norms showed a similar distribution to that found in other languages.

In the online study, Russian was of interest because of its morphological gender marking on (past tense) verbs, which typically agrees with the semantic gender of an animate referent, and also because of the existence of different (masculine) noun classes, which may or may not have corresponding feminine forms. When they do, the feminine forms may differ simply (or at least primarily) in referring to females rather than males. Or they may be "colloquial," with their use restricted to the spoken language and often having derogatory connotations.

Our study explored similarities and differences between the processing of gender information in Russian and in other languages (English, French, German, Spanish) that we had previously studied. As in those other languages, we saw

TABLE 6 | Mean reading times (ms) for first sentences for the texts in which the nouns were intended to be interpreted generically.

Stereotype		Female		Neutral		Male	
Continuation		Women	Men	Women	Men	Women	Men
Noun Group	Unpaired	5489	4359	6956	5329	4563	5412
	Paired	4743	4347	3711	3733	5546	6733
	Colloquial	4802	4858	4679	4917	4419	4912

TABLE 7 | Mean reading times (ms) for second sentences for the texts in which the nouns were intended to be interpreted generically.

Stereotype		Female		Neutral		Male	
Continuation		Women	Men	Women	Men	Women	Men
Noun Group	Unpaired	3275	2412	3687	3026	3612	3293
	Paired	2647	2826	3365	3076	4935	3863
	Colloquial	3260	3233	3367	2622	3449	3292

TABLE 8 | Mean percentage of yes responses to questions for the texts in which the nouns were intended to be interpreted generically.

Stereotype		Female		Neutral		Male	
Continuation		Women	Men	Women	Men	Women	Men
Noun Group	Unpaired	65.0	85.0	60.0	90.0	77.5	65.0
	Paired	75.0	70.0	68.0	82.0	30.0	90.0
	Colloquial	70.0	81.3	81.7	81.7	81.4	68.6

immediate deployment of gender information, both grammatical and stereotypical, in Russian. However, some differences did emerge, which could be related to properties of the Russian language. First, although there was evidence that people were sensitive to this information, grammatical gender marking on predicates (verbs) was not immediately and completely used to counteract stereotype information, unlike, for example, grammatical gender marking on Spanish definite articles (Carreiras et al., 1996). In Spanish *la futbolista* (the female footballer) was initially processed more slowly that *el futolista* (the male footballer). However, once the stereotype had been neutralized by the definite article (*la* vs. *el*) it was just as easy to refer to the female footballer with a feminine pronoun as to refer to the male footballer with a masculine pronoun. In our Russian experiment, the detection of a mismatch between a stereotype and the actual sex of the person referred to did not result in the mismatch being completely resolved, so that it did affect later processing.

Second, there was also evidence that the different classes of noun, which do not have direct counterparts in the other languages we have studied, behaved in different ways. In particular, when a masculine noun has a straightforward feminine counterpart, the oddity of using the masculine form when later information in the sentence suggests that the writer knows the person is female causes additional processing difficulty.

Our findings, therefore, show that even in a language like Russian, which has a more complex noun class and gender system than other languages previously studied, gender information is processed quickly and easily to provide detailed representations of the characters described in a text. However, the complexities of Russian do result in effects that were not seen in other languages, effects that can be clearly related to properties of the Russian language.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01720

REFERENCES

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Carreiras, M., Garnham, A., Oakhill, J. V., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol. A* 49, 639–663. doi: 10.1080/713755647

Corbett, G. (1982). Gender in Russian: an account of gender specification and its relation to declension. *Russ. Linguist.* 6, 197–232.

Corbett, G. (1991). *Gender.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166119

Doleschal, U. (1993). *Genus als Grammatische und Textlinguistische Kategorie. Eine kognitiv-funktionalistische Untersuchung des Russischen.* Unpublished PhD thesis, University of Vienna.

Doleschal, U., and Schmid, S. (2001). "Doing gender in Russian," in *Gender Across Languages: The Linguistic Representation of Women and Men,* Vol. 1, eds M. Hellinger and H. Bussmann (Amsterdam: John Benjamins), 253–282. doi: 10.1075/impact.9.16dol

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cognit.* 32, 551–559. doi: 10.3758/BF03195846

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. V. (2008). Au-pairs are rarely male: norms on the gender perception of role names across English, French and German. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206

Garnham, A. (1992). Minimalism versus constructionism: a false dichotomy in theories of inference during reading. *PSYCOLOQUY* 3:63.

Garnham, A. (2005). "Language comprehension," in *The Handbook of Cognition,* eds K. Lamberts and R. Goldstone (London: Sage), 241–254. doi: 10.4135/9781848608177.n10

Garnham, A., Oakhill, J. V., and Reynolds, D. J. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Mem. Cognit.* 30, 439–446. doi: 10.3758/BF03194944

Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101, 371–395. doi: 10.1037/0033-295X.101.3.371

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J. V., and Garnham, A. (2008). Generically intended, but specifically interpreted: when beauticians, musicians and mechanics are all men. *Lang. Cogn. Process.* 23, 464–485. doi: 10.1080/01690960701702035

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press.

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

McKoon, G., and Ratcliff, R. (1992). Inference during reading. *Psychol. Rev.* 99, 440–466. doi: 10.1037/0033-295X.99.3.440

Misersky, J., Gygax, P., Canal, P., Gabriel, U., Garnham, A., Braun, F., et al. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behav. Res. Methods* 46, 841–871. doi: 10.3758/s13428-013-0409-z

Oakhill, J. V., Garnham, A., and Reynolds, D. J. (2005). Immediate activation of stereotypical gender information in reading. *Mem. Cognit.* 33, 972–983. doi: 10.3758/BF03193206

Reynolds, D. J., Garnham, A., and Oakhill, J. V. (2006). Evidence of immediate activation of gender information from a social role name. *Q. J. Exp. Psychol.* 59, 886–903. doi: 10.1080/02724980543000088

Schmid, S. (1998). Zur Bezeichnung weiblicher Personen im Russischen: eine empirische Pilotstudie. *Weiner Slawistischer Almanach* 41, 239–262.

Counter-stereotypical pictures as a strategy for overcoming spontaneous gender stereotypes

Eimear Finnegan[1,2]*, Jane Oakhill[1] and Alan Garnham[1]

[1] School of Psychology, University of Sussex, Brighton, England, [2] School of Psychological Sciences and Health, University of Strathclyde, Glasgow, Scotland

The present research investigated the use of counter-stereotypical pictures as a strategy for overcoming spontaneous gender stereotypes when certain social role nouns and professional terms are read. Across two experiments, participants completed a judgment task in which they were presented with word pairs comprised of a role noun with a stereotypical gender bias (e.g., beautician) and a kinship term with definitional gender (e.g., brother). Their task was to quickly decide whether or not both terms could refer to one person. In each experiment they completed two blocks of such judgment trials separated by a training session in which they were presented with pictures of people working in gender counter-stereotypical (Experiment 1) or gender stereotypical roles (Experiment 2). To ensure participants were focused on the pictures, they were also required to answer four questions on each one relating to the character's leisure activities, earnings, job satisfaction, and personal life. Accuracy of judgments to stereotype incongruent pairings was found to improve significantly across blocks when participants were exposed to counter-stereotype images (9.87%) as opposed to stereotypical images (0.12%), while response times decreased significantly across blocks in both studies. It is concluded that exposure to counter-stereotypical pictures is a valuable strategy for overcoming spontaneous gender stereotype biases in the short term.

Keywords: pictures, spontaneous, gender, stereotype, reduction

Introduction

While English has a number of personal nouns that include maleness or femaleness as part of their lexical definitions (e.g., father, girl, son), or are formally marked for lexical gender through the use of suffixes (e.g., wait<u>ress</u>, land<u>lord</u>, land<u>lady</u>), the majority of human nouns in English are not gender specific. Instead, gender information associated with a human noun is typically indicated through social gender. This term refers to stereotypical assumptions about appropriate male and female social roles and the extent to which those roles are filled by females or males (Hellinger and Bußmann, 2001). Indeed social gender is now more commonly referred to as gender (stereo)typicality and is simply defined as the likelihood of a noun referring to women or men (Irmen and Roßberg, 2004). This gender typicality plays an important role in building cognitive representations of gender and is the reason why people come to expect, for example, surgeons to be male and nurses to be female. It is now well established that such occupational stereotypes

are activated spontaneously and unintentionally when certain gender-biased role nouns are read (Carreiras et al., 1996; Garnham et al., 2002, 2012; Kennison and Trofe, 2003; Duffy and Keir, 2004; Reynolds et al., 2006; Irmen, 2007; Kreiner et al., 2008; Finnegan et al., 2015), thus contributing to the maintenance and propagation of gender stereotypes in English speakers.

Once stereotypes or prejudiced associations are established, they can start to function automatically (Bodenhausen et al., 2009). Automatic processes typically require few attentional resources and are activated spontaneously, often without the perceiver's control or awareness (Bargh, 1994). Moreover, these associations can be activated independently of a person's conscious endorsement of them (Bodenhausen et al., 2009). In contrast, controlled processes operate through conscious intent and involve the attention of the perceiver (Schneider and Shiffrin, 1977). Both types of process are relevant to the current research in which we investigate a strategy for overcoming the spontaneous activation of occupational gender biases so as to ultimately result in lower levels of stereotype application.

Given the subtle pervasiveness of stereotype biases and prejudice, researchers have focused attention on devising means to overcome them. Indeed, Shiffrin and Schneider (1977) argued that with considerable and consistent training automatic responding to a particular stimulus could be "unlearned" and newer responses trained to take their place (Schneider and Shiffrin, 1977). While biased associations have proven difficult to completely overturn, they have proven malleable given appropriate strategies and conditions (see Blair, 2002 for a review). Evidence has been mounting for the positive effect of counter-stereotype promotion in tackling stereotype biases, because such stereotype incongruent information appears to weaken the stereotype itself or access to it (e.g., Kawakami et al., 2000; Blair et al., 2001; Dasgupta and Greenwald, 2001; Lai et al., 2014). Thus, in this article we explore how strengthening counter-stereotype associations may reduce gender stereotyping in relation to occupational role nouns.

While stereotype representations reflect the strongest or most typical group associations, research suggests that these representations may also include information about counter-stereotypes. For instance, it has been shown that people represent subtypes that are inconsistent with a group stereotype, for example that of business woman or female athlete (Devine and Baker, 1991; Green and Ashmore, 1998; Coats and Smith, 1999). Blair et al. (2001) argued that as stereotypes and counter-stereotypes are often polar opposites, it is unlikely that they would be represented completely independently of one another. Indeed increasing the accessibility of one of these constructs could result in a decrease in accessibility of the other (e.g., Dijksterhuis and van Knippenberg, 1996).

Blair et al. (2001) explored whether increasing the accessibility of counter-stereotypes through use of a mental imagery task could result in lower levels of implicit gender stereotype activation. They devised an experiment with four different imagery conditions: stereotypic (participants imagined a weak woman), counter-stereotypic (participants imagined a strong woman), gender neutral (participants imagined a holiday in the Caribbean) and no imagery (participants played with a

simple water game for 5 min; Experiment 2; Blair et al., 2001). Implicit stereotypes were measured both before and after the 5 min mental imagery task using the Implicit Association Test (IAT; Greenwald et al., 1998). Participants in the counter-stereotype condition subsequently produced significantly weaker implicit gender stereotypes than those in the three other mental imagery conditions, thus providing convincing evidence for the moderating effect of counter-stereotype mental imagery on implicit stereotypes. Indeed the same pattern of results was found when this mental imagery strategy was used with two further measures of stereotype bias; the Go/No-go association test (GNAT; Experiment 4) and a false memory paradigm (Experiment 5).

This comprehensive set of experiments by Blair et al. (2001) suggests that implicit associations can be altered by directing participants' attention to subtypes of group members or triggering counter-stereotypical links in the cognitive network (Kunda and Thagard, 1996; Bodenhausen and Macrae, 1998). A more direct approach to increasing counter-stereotype saliency was taken by Kawakami et al. (2000) who devised a non-stereotypic association training method aimed at reducing automatic stereotyping toward racial groups and skin heads. This association training involved presenting participants with counter-stereotypic and stereotypic word pairs relating to the category of interest. The task was to repeatedly affirm (i.e., say 'yes') and negate (i.e., say 'no') the counter-stereotypic and stereotypic pairings respectively. It was found that participants who received extensive training in counter-stereotype affirmation/stereotype negation showed lower levels of automatic stereotyping on a primed Stroop task (Studies 1 and 2) and a person categorization task (Study 3) than those who received little or no training.

Given the extent of research aimed at overcoming stereotypes and prejudice, it is important to compare the efficacy of interventions that have led to significantly lower levels of bias. Estimates of effect magnitude are critical to such comparisons, yet are rarely included in stereotype or prejudice reduction research, where null hypothesis significance testing is dominant. Instead, researchers tend to provide evidence that an intervention results in less implicit prejudice or stereotyping than a control condition (Frick, 1996; Lai et al., 2013). Lai et al. (2014) sought to address this issue by holding a research contest to experimentally compare 17 interventions aimed at overcoming implicit racial preferences. These interventions fell into six categories of which "exposure to counter-stereotypical exemplars" proved the most effective ($d = 0.38$, 95% CI [0.32, 0.44]). The three most successful strategies within this category involved getting participants to imagine a vivid counter-stereotypic scenario, shifting group boundaries through competition (i.e., cooperating with outgroup members and competing against ingroup members in a dodgeball game), and getting participants to practice an IAT with counter-stereotypical exemplars (i.e., positive Black exemplars, and negative White exemplars)[1].

---

[1]Lai et al. (2014) point out that this comparative study allowed for inferences about the interventions to be made in a highly specific experimental context only.

While Lai et al.'s (2014) research was aimed at overcoming implicit racial biases (as opposed to the spontaneous gender biases that we are interested in), their findings are relevant here as stereotype reduction strategies have often proved successful across different domains. For instance, social norm information has been used to successfully reduce biases against groups such as racial minorities (e.g., Stangor et al., 2001), people suffering from obesity (Puhl et al., 2005) and has also lowered levels of spontaneous gender stereotyping (Finnegan et al., 2015).

Broader support for the use of counter-stereotypes in bias reduction is found with Bodenhausen et al. (2009), who advocate the role of diverse environments in undermining biased representations. Evidence suggests that increased interaction with out-group members can substantially weaken biased attitudes, and automatic negative emotional and physiological reactions to these outgroup members (Blascovich et al., 2001; Pettigrew and Tropp, 2006). Similarly, diverse environments can influence automatic stereotypes and attitudes about groups, as it has repeatedly been found that counter-stereotype exemplars of devalued groups can result in more positive attitude and stereotype activation in both lab (e.g., Dasgupta and Greenwald, 2001; Joy-Gaba and Nosek, 2010) and real-world settings (e.g., Dasgupta and Asgari, 2004). For instance, by exposing participants to admired African American individuals and disliked European Americans, implicit preference for Whites compared to Blacks was successfully reduced on an IAT (Dasgupta and Greenwald, 2001).

But how do counter-stereotypes operate to reduce levels of stereotyping? Two potential processes are (1) the bookkeeping process in which stereotypes are hypothesized to change slowly, through encountering numerous counter-stereotype exemplars of a particular category and (2) the conversion process in which stereotypes are thought to change more rapidly, upon encountering fewer, yet more striking counter-stereotype exemplars than postulated in the book-keeping process (Operario and Fiske, 2004). While both of these processes highlight means of achieving stereotype reduction via counter-stereotype information, a third process, subtyping, suggests how such information could work to protect the stereotype. Essentially, subtyping processes may ensure that the original stereotype remains unchanged as new categories are formed to account for counter-stereotype information. However, it is also possible that stereotypes could be weakened and reduced with sufficient category variation and subtyping (Operario and Fiske, 2004).

In the current research, Experiment 1 employs a striking counter-stereotype strategy in which participants are presented with pictures of men and women working in obviously counter-stereotypic roles. We hypothesize that these gender-salient pictures will bring about stereotype reductions through direct and immediate conversion processes. By highlighting category variability we hope to strengthen counter-stereotype representations and remind participants that, for example, a surgeon could be female and a nurse male. This salient counter-stereotypical information should be incorporated into the perceiver's gender and occupation-related cognitive representations so as to update and modify them. The measure of stereotyping that we used in conjunction with this picture training was a judgment task devised by Oakhill et al. (2005).

Oakhill et al. (2005) were interested in whether gender biases are evoked for single words, and the extent to which these biases can be overcome. To explore this question they asked participants to quickly decide whether two terms presented onscreen could refer to one person. These word pairs comprised a role noun that was stereotype biased (e.g., *builder, beautician*) and a kinship term that was definitionally gendered only (e.g., *uncle, aunt*). In order to respond successfully, participants needed to take definitional gender into account (e.g., that an uncle is always male) but to dismiss stereotypical gender (e.g., that most beauticians are female).

Across a series of studies, Oakhill et al. (2005) found that participants consistently rejected stereotype incongruent pairings (e.g., *builder/mother*) to a significantly greater extent than stereotype congruent pairings (e.g., *builder/father*). This was still the case when they were explicitly reminded that nowadays many jobs are not marked for gender and that they should carefully consider whether the first term presented (i.e., the role noun) could be occupied by men, women or both (Experiment 4). Results of this research provide strong evidence that there is an automatic component to responding, as participants struggled to overcome the gender stereotype information associated with the role nouns, even when its activation was detrimental to task performance. Indeed the authors posit that such gender stereotype information is incorporated immediately, and likely automatically, into a perceiver's mental representation.

This research has much in common with that of Finnegan et al. (2015) who sought to overcome such occupational gender biases through the use of social-consensus feedback (again in conjunction with the judgment task of Oakhill et al., 2005). Finnegan et al. (2015) administered three blocks of judgment trials with social consensus feedback provided after each response in Block 2 only. This feedback consisted of a sentence stating the percentage of previous students at the university who had completed the judgment task and agreed with the participant's choice, e.g., '_% of previous students agreed with you.' In reality this feedback was fictitious and constructed so as to strongly and consistently suggest that past participants did not succumb to stereotype biases, i.e., that they accepted stereotype incongruent pairings without a problem. In this way, the social feedback sought to reinforce non-stereotypic responding and highlight any discrepancy between a participant's response and the peer group norm.

For instance, in all cases the Implicit Association Test (IAT) or Multi-Category IAT was used as the main dependent variable, while a number of specific criteria had to be fulfilled in order to win the research contest (as opposed to inclusion in the study) e.g., at least 85% of participants had to finish the intervention in 5 min or less. Such criteria may have made it more likely to find an effect using certain interventions over others. For instance, evaluative conditioning may be more effective over longer intervention formats (Bar-Anan et al., 2010).

Performance on judgments of stereotype incongruent pairings was found to improve significantly following the introduction of social feedback in Block 2. Moreover, this improvement continued in Block 3 when the feedback was no longer given (Experiment 1), thus providing evidence for the use of social consensus information as a useful stereotype reduction strategy.

Other strategies aimed at overcoming gender biases for occupational role nouns have more frequently been examined in sentence comprehension studies. In such cases, a stereotyped term is typically followed by gender congruent or incongruent information in a match/mismatch paradigm, e.g., the *surgeon* went to work early as *he/she* was very busy. Processing difficulty is frequently evident in the incongruent condition relative to the congruent condition as the reader struggles to reconcile the unexpected definitional gender information on the pronoun with the stereotype-biased gender already generated by the occupational role term (e.g., Carreiras et al., 1996; Garnham et al., 2002; Kennison and Trofe, 2003; Duffy and Keir, 2004; Irmen, 2007; Kreiner et al., 2008; Garnham et al., 2012). However, such gender biases have successfully been overcome through establishing the sex of a character *before* a role noun is encountered, e.g., after reminding *himself/herself* about the letter, the *minister* immediately went to the meeting at the office (e.g., Duffy and Keir, 2004; Kreiner et al., 2008; Lassonde and O'Brien, 2013).

We report two studies in which we investigated the influence of counter-stereotype pictures as means of increasing counter-stereotype saliency and reducing levels of gender-based occupational stereotyping on the judgment task of Oakhill et al. (2005).

Overview of Studies

In Experiment 1 participants were presented with two blocks of stereotype judgment trials, with the picture task immediately following the first block. In the picture task participants were presented with 24 pictures of people working in counter-stereotypical roles. The participants' task was to answer a set of four questions for each picture about the character's supposed earnings, leisure activities, job satisfaction and personal life. This was intended to result in deeper processing of the character presented and the counter-stereotypical job this person was depicted as holding.

It was hypothesized that participants would initially respond more slowly and less accurately to trials of stereotype incongruent word pairs (e.g., *nurse/father*) than to stereotype congruent word pairs (e.g., *nurse/mother*) in Block 1. However, following the picture training, it was hypothesized that the processing cost associated with the stereotype incongruent condition in Block 1 would be attenuated and lead to higher accuracy and faster reaction times to the critical trials in Block 2.

Experiment 2 was a control study which differed from Experiment 1 solely in the picture task. Participants were now presented with images of people working in stereotypical

roles (as opposed to counter-stereotypical) to provide a clear basis for explaining the Block 1 to Block 2 changes in performance in Experiment 1. For these studies, 24 pairs of pictures of men and women working in the same occupational roles were first required, for instance, a female make-up artist (stereotypical) and a male make-up artist (counter-stereotypical). These pictures were collected through a web search and from a picture database. A short pilot study was conducted to evaluate (a) the similarity of the male and female versions of the pictures and (b) how realistic the pictures looked.

Experiment 1

Pilot Study

Twenty students (10 male and 10 female) took part in the pilot study that lasted 5 min. Each of the 24 picture pairs was presented as pictures of men and women working in the same roles. The participants' first task was to rate these pairs on "how similar they are (ignoring gender and thinking about features such as the race, age, facial expression of the people, pose and the background)." Ratings were made on a scale ranging from 1 (very similar) to 6 (very dissimilar).

Next, the pictures were re-presented to the participants, who judged how realistic they found the pictures to be – again ignoring gender and thinking about features such as the race, age, facial expression of the people, pose and the background. In this part of the pilot study the 48 pictures were rated individually on a realism measure from 1 (very realistic) to 6 (very unrealistic).

The mean similarity rating across picture pairs was 2.24 (SD = 1.26), thus falling between the points of moderately similar (2) and mildly similar (3). The mean rating of how realistic a picture looked was 1.93 (SD = 1.26), thus falling between the points of very realistic (1) and moderately realistic (2). In two instances males and females were found to have significant differences in their ratings of similarity and in one instance had significant differences in their ratings of realism. Ultimately, however, all pictures were kept for the experimental task, as none were rated as being more dissimilar than similar or more unrealistic than realistic. Furthermore, because no obviously dissimilar or unrealistic pictures were included for rating, participants may have been stricter in their judgments than otherwise expected.

Method
Participants

The participants were 30 monolingual native English speakers (14 male, 16 female) from the student population of the University of Sussex. Participants' ages ranged from 18 to 37 years (*M*: 20.27; SD: 4.12) and they received either £6 or 4 course credits for taking part in the session, which lasted ∼45 min. Ethical approval for both experiments in this paper was obtained from the University of Sussex, School of Psychology Research Ethics Committee, which follows the British Psychological Society guidelines for ethics on human subject testing. All participants signed a consent form prior to participating.

Materials

Gender-Biased Role Nouns

Gender biased role nouns were selected from norms compiled by Gabriel et al. (2008). The chosen items were the 12 most highly male-biased nouns (e.g., bricklayer), the 12 most highly female-biased (e.g., beautician), and the 12 closest to the neutral point on the scale (e.g., pedestrian). As described in Finnegan et al. (2015), the range of the bias ratings for the male terms is narrower than for the female items (11.10% vs. 17.55% respectively), while ratings of the neutral terms have the shortest range of 5.29%. These figures suggest that the neutral terms should prove less problematic for participants than the other role nouns. See Finnegan et al. (2015) for a full list of the stereotyped items, their associated bias ratings and all filler items.

Kinship Terms

As in previous studies, six kinship terms (three male, three female) were selected to be used as one of the terms in the word pairs (Oakhill et al., 2005; Finnegan et al., 2015). These terms were *father, mother, brother, sister, uncle, aunt.* Importantly, these words incorporate a specific gender into their definitions, e.g., the term '*brother*' can only refer to a person of male sex.

Critical Word Pairs

The 12 male-biased, female-biased, and neutral role nouns were each combined once with the six kinship terms to produce a set of stereotype congruent (e.g., *pilot/brother, nurse/sister*), stereotype incongruent (e.g., *pilot/sister, nurse/brother*) and neutral word pairs (e.g., *artist/brother, artist/sister*). There were, therefore, 72 word pairs in each of the three congruency conditions, totaling 216 trials.

Filler Trials

Filler items were 240 word pairs created by pairing the six kinship terms with role nouns that are also gender-specific by definition (e.g., geisha, hero). In this way, filler trials were gender unambiguous pairings to which participants could respond with relative ease and certainty. These items were selected from norming studies conducted by Kennison and Trofe (2003) and Hamilton (2008).

Item Overview

The word pairs used in this study were identical to those of Finnegan et al. (2015) in content although the number of pairings presented differed. While that study had three blocks of trials (and a total of 456 word pairs) the current work used two blocks of trials (and a total of 304 word pairs). Therefore, use of the three original blocks from Finnegan et al. (2015) was counter-balanced so that each of their 456 pairs appeared an equal number of times across participants in the current experiments. This procedure also ensured that each of the six kinship terms appeared with each of the role nouns an equal number of times. In total, 184 items were intended to elicit a *yes* response (including all critical items) while 120 required a *no* response.

Picture Task

Twenty four pictures of a man or a woman working in a counter-stereotypical job environment were selected. Half of the pictures depicted people working in roles that were also mentioned in the judgment task and half depicted 'new' role terms that the participants had not yet been exposed to (six male and six female stereotypical terms in each case)[2].

When displayed on-screen, the pictures were accompanied by two short sentences. These sentences always introduced the character in the picture and their job, e.g., "This is Rebecca. She is a bricklayer" or "This is Christopher. He is a make-up artist." The first names presented were a selection of the most popular baby names from 1994 and 1984 which participants were likely to have been highly familiar with (sourced from Merry, 1995). Upon presentation of a picture and the accompanying sentences, participants were required to answer four questions relating to each characters' probable salary (How much do you think [insert character name] earns each year?), leisure activities (What are [his/her] leisure activities?), job satisfaction (How satisfied do you think [he/she] is with [his/her] job?) and lifestyle (Briefly describe [his/her] personal life). Three different picture lists were created with the pictures presented in a different, but fixed, order in each list. Following this, three response booklets were prepared that matched the presentation order of the pictures. Note that the primary purpose of asking these questions was to focus participants' attention on the pictures presented, and notably the job that each person was doing. However they were also a window to the views that participants hold about people in these different roles. Responses to these questions will be discussed after the results of Experiment 2.

Design and Procedure

In the judgment task, the two nouns were presented individually in the center of a computer screen. A role term was first displayed for 1000 ms, followed immediately by a kinship term (inter-stimulus interval of 0). This kinship term remained onscreen until a response was made. There followed a 500 ms delay before onset of the next trial. As described in Finnegan et al. (2015), the word pairs were divided into three fixed sets of blocks (with two of these chosen for each participant in the current study), with the order of the individual trials randomized separately for each participant. A button box was used to record participants' responses, with one button clearly marked 'Y' for *yes* and another 'N' for *no*. Between the two blocks of trials, participants were asked to complete the picture task.

---

[2]It was hypothesized that the role terms that appeared in the picture booklet would elicit a higher level of accuracy for the stereotype incongruent pairs (i.e., judgments that the two words can refer to the same person) and lower response times in Block 2 of the judgment task than those that did not appear in the picture booklet (as the pictures explicitly depicted a person of counter-stereotypical gender fulfilling the role). However, it was found that accuracy to both sets of terms was identical (at 88%) while RTs were somewhat slower for the role nouns that had previously appeared in the booklet ($M = 731$ ms) compared to those that did not ($M = 686$ ms), although this difference was not significant, $t(22) = 1.34$, $p = 0.193$. Although this latter trend was not anticipated, we suggest that the pictures participants saw may have induced them to think more about these specific occupations as they arose in the judgment task thus leading to increased reaction times.

Participants were tested individually in a quiet laboratory. They were provided with written instructions that informed them to read each pair of words and decide (without excessive deliberation) whether the two terms could apply to the same person. Two examples of (definitional) word pairs were provided – one that required a *yes* response and one that required a *no* response. Participants were further informed that they would be required to make judgments about pictures between the first and second block of trials and told what this task entailed. The instructions and examples were then repeated verbally. Next, a short practice session using a representative sample of fillers and critical word pairs (not subsequently used in the experimental blocks) was given to familiarize the participants with the experimental task. Once familiarized with the procedure, participants were left alone to complete the judgment task.

Results
Data Screening
In the analyses reported below, data for word pairs that contained the neutral term 'adolescent' were excluded because negative responses to such pairings (55% in Block 1, 33% overall) appeared to be based on considerations of age rather than gender. For instance, the pairing *adolescent/father* was much more difficult for participants to accept than *adolescent/brother*, despite both being possible combinations. In total, 1.32% of the data was removed for this reason.

Analysis
In both experiments accuracy of judgments and response times (RTs) were analyzed using two mixed-design analyses of variance (ANOVAs): firstly with participants treated as the random variable and secondly with items treated as the random variable. In the by-participants analysis (F_1), the mixed ANOVA had three repeated factors – stereotype bias of the role name (Stereotype: Male/Female/Neutral), gender of the kinship term (Kinship term gender: Male/Female) and block of trials (Block: Block1/Block2). Participant Sex was included as a between-subjects factor. In the by-items analyses (F_2), Stereotype was included as a between-items factor while Kinship term gender, Block and Participant Sex were included as within-item variables. Where sphericity was not satisfied, Greenhouse–Geisser (when $\varepsilon < 0.75$) or Huynh–Feldt ($\varepsilon > 0.75$) corrected degrees of freedom and *p*-values are presented (as recommended by Girden, 1992). With all paired *t*-tests, within-subject or within-item effect sizes were estimated using Cohen's *dz* while with independent-samples *t*-tests estimates of between-subject or between-item effect sizes were estimated using Cohen's *d*.

Congruency
It is important to note that an interaction of Stereotype by Kinship term gender is an effect of Congruency, i.e., it is the combination of the levels of these two factors that give rise to the three critical conditions of congruent, incongruent and neutral. Therefore, Stereotype by Kinship term gender interactions are referred to as Congruency effects (though

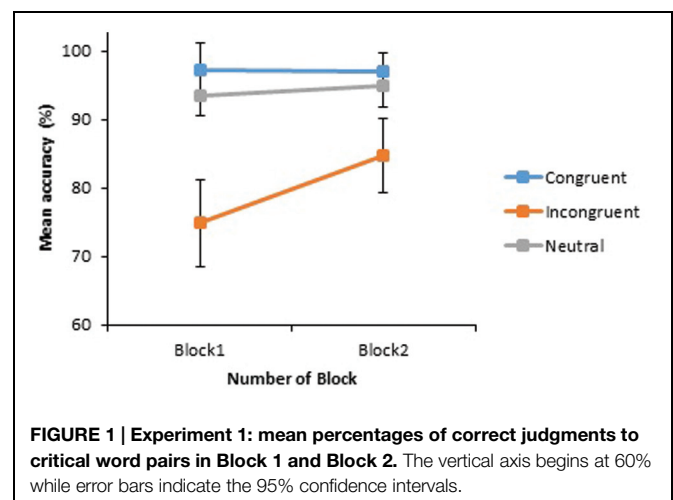primarily in relation to the male and female stereotyped terms).

Accuracy
Analysis revealed a main effect of Stereotype, $F_1(1.67,46.67) = 6.27$, $p = 0.006$, $F_2(2,32) = 9.59$, $p = 0.001$, with higher accuracy to word pairs that contained a neutral role term ($M = 94.3\%$), than those that contained male ($M = 88.2\%$) or female-biased terms ($M = 89.2\%$). A main effect of Block was also found, $F_1(1,28) = 6.90$, $p = 0.014$; $F_2(1,32) = 17.73$, $p < 0.001$, driven by a 3.5% increase in accuracy of critical pairings from Block 1 (88.8%), to Block 2 (92.3%). As anticipated, there was a main effect of Congruency, $F_1(1.18,33.08) = 14.76$, $p < 0.001$; $F_2(2,32) = 67.55$, $p < 0.001$, with significantly lower accuracy to stereotype incongruent word pairs ($M = 79.80\%$), than to congruent ($M = 97.15\%$) and neutral ($M = 94.35\%$) pairings.

Importantly, an interaction of Congruency by Block was also found, $F_1(1.39,38.89) = 8.93$, $p = 0.002$; $F_2(2,32) = 22.00$, $p < 0.001$. This interaction was driven by a substantial 9.87% increase in accuracy for stereotype incongruent pairings across blocks, while accuracy to neutral and stereotype congruent pairings was high from the outset, with little room for improvement (see **Figure 1**).

The increase in accuracy to stereotype incongruent pairings across blocks was significant, $t_1(29) = 3.33$, $p = 0.002$, $dz = 0.61$; $t_2(23) = 5.70$, $p < 0.001$, $dz = 1.16$, revealing the efficacy of the counter-stereotypic picture task as a gender stereotype reduction strategy. However, despite this improvement across blocks, accuracy to stereotype incongruent word pairs remained significantly lower than accuracy to stereotype congruent, $t_1(29) = 3.10$, $p = 0.004$, $dz = 0.57$; $t_2(23) = 9.56$, $p < 0.001$, $dz = 1.95$, and neutrally rated word pairs, $t_1(29) = 2.60$, $p = 0.015$, $dz = 0.47$; $t_2(44) = 6.65$, $p < 0.001$, $d = 2.00$, by the end of the experiment. Thus, this picture training did not completely eradicate the effects of stereotype bias in this judgment task.

Next, an interaction of Participant Sex with Kinship term gender was revealed, $F_1(1,28) = 5.27$, $p = 0.029$; $F_2(1,32) = 5.16$, $p = 0.030$. Female participants displayed marginally higher

FIGURE 1 | Experiment 1: mean percentages of correct judgments to critical word pairs in Block 1 and Block 2. The vertical axis begins at 60% while error bars indicate the 95% confidence intervals.

accuracy in response to female kinship terms (88.5%) as opposed to male kinship terms (86.6%) while male participants displayed the opposite pattern, showing greater accuracy in response to male kinship terms (94.4%) than female kinship terms (92.8%). These mean values show that male participants were also more accurate than females on these kinship terms overall (93.6% vs. 87.6%).

A number of further effects involving Participant Sex emerged in the by-items analysis only[3]. A main effect of Participant Sex was first revealed, $F_2(1,32) = 104.01$, $p < 0.001$, with male participants achieving much higher levels of accuracy than female participants overall (93.6% vs. 87.5%). There was also a highly significant interaction of Participant Sex by Congruency, $F_2(2,32) = 8.08$, $p = 0.001$. While male participants outperformed females in each of the three congruency conditions, this difference was most apparent in response to stereotype incongruent pairings where male participants achieved an average accuracy score of 85.3% while female participants reached only 75.0%. Finally, there was a Participant Sex by Block interaction, $F_2(1,32) = 4.92$, $p = 0.034$, with the accuracy of male participants increasing 2.4% across blocks, compared to 4.8% for female participants (although the females had more scope for improvement from Block 1). That said, the final accuracy of females was still lower than that of the males.

Reasons for this superior male performance remain unknown as (sex aside) there were no obvious differences between the male and female samples. The data suggests that male participants are more accepting of stereotype congruent pairings than past work suggests (Oakhill et al., 2005; Finnegan et al., 2015). This will be returned to in the General Discussion.

Response times

Response times below 150 ms, and above 4,000 ms were excluded from analysis (representing 0.92% of the total) along with times for all errors of judgment (representing a further 10.88%), totaling a loss of 11.8% of the data. These data points were replaced with the Participant by Block mean for each participant. Data points 2.5 standard deviation above or below the Participant by Block mean were replaced with the relevant upper or lower cut off point. Analyses were conducted as with the accuracy data.

A main effect of Stereotype was found in the by-participants analysis, along with a marginally significant effect in the by-items analysis, $F_1(2,56) = 5.50$, $p = 0.007$; $F_2(2,32) = 3.00$, $p = 0.064$, with faster RTs to word pairs that contained a neutral role term ($M = 828$ ms), than those that contained male-biased ($M = 850$ ms) or female-biased terms ($M = 889$ ms). A main effect of Block was also revealed, $F_1(1,28) = 15.50$, $p < 0.001$; $F_2(1,32) = 97.60$, $p < 0.001$, with RTs decreasing 143 ms from

FIGURE 2 | Experiment 1: mean response times (in milliseconds) of correct judgments to critical word pairs across Block 1 and Block 2. The vertical axis begins at 600 ms while error bars indicate the 95% confidence interval.

Block 1 to Block 2 (927 ms vs. 784 ms respectively). Again, there was a main effect of Congruency, $F_1(1.33,37.12) = 12.31$, $p < 0.001$; $F_2(2,32) = 11.62$, $p < 0.001$, with fastest RTs observed in response to stereotype congruent word pairs ($M = 815$ ms), followed by neutral ($M = 829$ ms) and incongruent pairings respectively ($M = 920$ ms).

Importantly, a significant interaction between Block and Congruency also emerged, $F_1(2,56) = 4.87$, $p = 0.011$; $F_2(2,32) = 5.27$, $p = 0.010$. As can be seen in **Figure 2**, RTs decreased across all conditions from Block 1 to Block 2, with the greatest reduction found in response to stereotype incongruent pairings (225 ms). This was found to be a significant improvement across blocks, $t_1(29) = 4.23$, $p < 0.001$, $dz = 0.77$; $t_2(23) = 7.89$, $p < 0.001$, $dz = 1.61$. Furthermore, by the end of the experiment, there was no significant difference between RTs to stereotype incongruent and stereotype congruent, $t_1(29) = 1.59$, $p = 0.122$; $t_2(23) = 1.65$, $p = 0.112$, or neutral pairings, $t_1(29) = 1.36$, $p = 0.183$; $t_2(44) = 1.41$, $p = 0.167$. Overall, the RT data provide further strong support for the use of counter-stereotypical pictures as an effective stereotype-reduction strategy. However, it should be noted that past results suggest that this improvement across blocks is also likely due in part to practice effects (Finnegan et al., 2015).

A main effect of Participant Sex was also observed, $F_1(1,28) = 5.64$, $p = 0.025$; $F_2(1,32) = 93.24$, $p < 0.001$, with male participants typically much slower to respond than female participants (925 ms vs. 786 ms). An interaction of Participant Sex with Kinship term gender also emerged, $F_1(1,28) = 14.23$, $p = 0.001$; $F_2(1,32) = 9.09$, $p = 0.005$. Again, female participants responded faster to female kinship terms over male kinship terms (765 ms vs. 807 ms respectively), while male participants responded faster to male kinship terms over female kinship terms (892 ms vs. 958 ms respectively). These means also reveal that female participants responded faster than male participants in both cases (786 ms vs. 925 ms), a finding which may shed some light on the lower accuracy scores

[3]Given that there were many fewer participants than items in this experiment (30 participants vs. 304 item pairs per participant) it is highly likely that this effect was only significant by-items because the standard errors of the condition means are likely to be much lower in the by-items analysis than in the by-participants analysis, if the variances are roughly equal. For instance, the average standard deviation of responses to critical word pairs was 16.3% in the by-participants data while just 5.3% in the by-items analysis. As a similar imbalance between participant numbers and item numbers runs throughout both studies in this article this pattern (a significant effect by-items but not by-participants) frequently recurs.
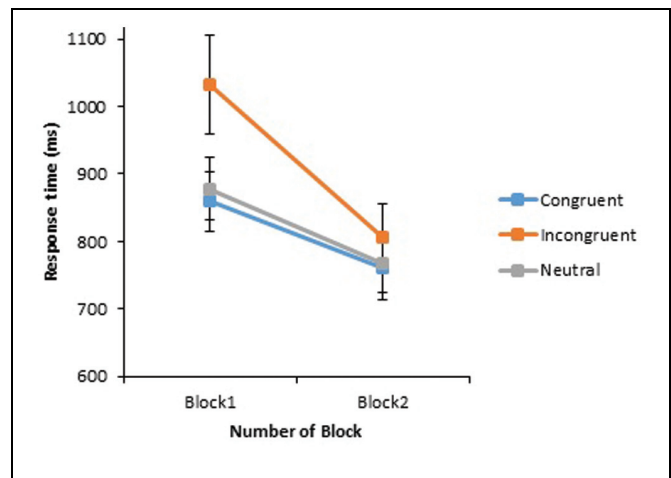
achieved by females in this study; it is possible that accuracy performance may have deteriorated for the sake of faster responding.

Fillers – Accuracy

As responding to the filler trials was not the main focus of this research, tests of significance were not carried out on these data; only a descriptive analysis is presented (with one exception below). Performance on the filler trials was somewhat variable, with an average of 93.0% accuracy on definitionally matching word pairs versus an average of 87.28% on definitionally mismatching word pairs across the experiment. As in Finnegan et al. (2015), accuracy of responses to definitionally mismatching word pairs was lower to trials involving male terms ($M = 80.61\%$) than female terms ($M = 93.94\%$). This pattern was previously thought to result from the generic interpretation of certain terms such as *host* (0.45% accuracy across blocks) or *steward* (0.25% accuracy across blocks) which have female-specific counterparts (i.e., *stewardess* and *hostess*) and which should, therefore, be taken as male specific. However, to investigate whether this effect was driven by male participants (for whom these pairings are more self-relevant than females) we conducted a mixed ANOVA on the definitionally mismatching responses. A significant interaction of Participant sex by Definitional gender was indeed found, $F_1(1,28) = 4.71$, $p = 0.038$; $F_2(1,58) = 5.61$, $p = 0.021$ with males achieving lower accuracy to definitionally male role nouns when presented with a female kinship term (76.0%) than females did to these pairings (84.7%), while both sexes scored more similarly on the definitionally female terms paired with male kinship terms (92.9% vs. 94.9% respectively). While this data suggests that poor performance on the definitionally mismatching pairings was due to the male participants struggling with the male mismatch terms, this interaction was not replicated in the reaction time data described below (ps > 0.3). Nor was it replicated in the accuracy or reaction time data of Experiment 2 (ps > 0.5). For this reason we maintain that it is likely the generic interpretation of certain definitionally male terms that is driving poor performance to definitionally mismatching pairings.

Fillers – Response Times

Average RTs to definitionally matching word pairs were found to be faster than those to definitionally mismatching word pairs (888 ms vs. 950 ms respectively). RTs were also faster in response to female pairings over male in both the definitionally matching (862 ms vs. 914 ms respectively) and mismatching cases (910 ms vs. 989 ms respectively). These findings reflect the accuracy data, with longer processing of male mismatching pairs likely to reflect participants' deliberation over terms that are masculine by definition but often used generically in reference to both sexes.

Discussion

Overall, Experiment 1 provides preliminary evidence for the use of counter-stereotypical pictures as an effective strategy for reducing the immediate activation of gender stereotypes when gender-biased role terms are read. Both accuracy and reaction times to stereotype incongruent word pairs significantly

improved from Block 1 to Block 2 following the counter-stereotypic picture task. While accuracy remained significantly lower to the incongruent pairs than to the stereotype congruent and neutral pairings in Block 2, RTs in Block 2 were similar in all three conditions.

It is hypothesized that exposure to the counter-stereotypical pictures triggered participants' world knowledge that, although there is a strong gender bias associated with certain social roles in society, nowadays both men and women can and do fulfill these roles. The activation of this knowledge is then thought to have helped participants overcome stereotype application in the second block of judgment trials.

Before accepting this picture training as a successful means of stereotype reduction, a control condition against which these results could be compared was required so as to verify that the counter-stereotype manipulation of Experiment 1 was indeed the reason for the improved task performance in Block 2, rather than simply looking at pictures of people carrying out jobs and answering questions about these people. In Experiment 2, therefore gender *stereotypical* pictures replaced the counter-stereotypical pictures in the picture task. If there are stereotype-related effects from processing the pictures, Experiment 2 should see the maintenance of (as opposed to the weakening of) the gender biases associated with many occupational terms in English.

Experiment 2

By providing participants with pictures of people working in gender stereotypical roles, Experiment 2 sought to reinforce participants' world knowledge that women are typically associated with a certain set of roles (e.g., *beautician, secretary*), and men are typically associated with another set (*pilot, mechanic*). The experimental design was exactly as outlined in Experiment 1, but with the counter-stereotypical pictures replaced by stereotypical pictures. The rationale for Experiment 2 was that attending to these gender-stereotypical pictures would lead to deeper adherence to gender biases in the judgment task. Therefore, if there was no improvement in response to stereotype incongruent trials from Block 1 to Block 2, it could be confidently assumed that the reduction in stereotype bias across blocks in Experiment 1 was associated with the presentation of counter-stereotypical pictures.

As in Experiment 1, it was hypothesized that participants would initially respond more slowly and less accurately to trials with stereotype incongruent word pairs (e.g., *nurse/father*) than to stereotype congruent word pairs (*nurse/mother*) in Block 1. However, unlike Experiment 1, it was hypothesized that the processing cost associated with the stereotype incongruent condition in Block 1 would not be attenuated in Block 2 following presentation of the stereotype congruent pictures.

Method
Participants

The participants were 34 monolingual native English speaking students (19 female, 15 male) from the University of Sussex.

Participants' ages ranged from 18 to 32 years (M: 21.23; SD: 4.53). They received either £6 or 4 course credits for taking part in the session which lasted approximately 45 min.

Materials

The same materials and instructions were employed as in Experiment 1, aside from a different set of pictures (and accompanying booklets) for the picture task. The pictures all depicted men and women working in a stereotypical job environment and were accompanied by two sentences introducing the character and stating their job, e.g., *This is Rebecca. She is a make-up artist* or *This is Christopher. He is a bricklayer*. As a reminder, the stereotypic pictures used in this study were previously rated for similarity (to the counter-stereotyped pictures) and realism in the pilot study.

Design and Procedure

The design and procedure were identical to those for Experiment 1, but with participants answering questions about pictures of people working in stereotypical roles as opposed to counter-stereotypical roles.

Results

Data Screening

In this Experiment, the neutral term 'adolescent' was replaced with the term 'swimmer' therefore data for all neutral items were included in the analysis. Accuracy of and RTs for judgments were analyzed as in Experiment 1.

Accuracy

A main effect of Stereotype was found, which was significant by participants and marginally significant by-items, $F_1(1.30, 41.61) = 7.81$, $p = 0.004$; $F_2(2,33) = 3.10$, $p = 0.059$, with greater accuracy for neutral role nouns ($M = 93.1\%$), than male-biased ($M = 90.7\%$) or female-biased terms ($M = 88.8\%$). A main effect of Congruency was also revealed, $F_1(1.03, 33.07) = 12.47$, $p = 0.001$; $F_2(2,33) = 55.04$, $p < 0.001$, with significantly higher accuracy to stereotype congruent ($M = 97.0\%$) and neutral

FIGURE 3 | Experiment 2: mean percentages of correct judgments to critical word pairs in Block 1 and Block 2. The vertical axis begins at 70% while error bars indicate the 95% confidence interval.
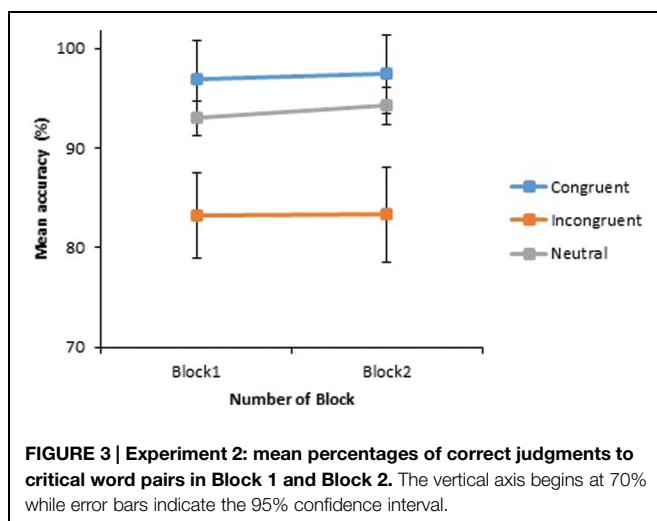
($M = 93.1\%$) word pairs, than to stereotype incongruent pairings ($M = 83.3\%$). However, no significant effect of Block was found, $F_1(1,32) = 0.89$, $p = 0.351$; $F_2(1,33) = 0.67$, $p = 0.417$, with accuracy increasing just 0.5% across the two blocks (Block 1 $M = 90.6\%$ vs. Block 2 $M = 91.1\%$). Importantly, there was also no significant interaction of Congruency by Block, $F_1(2,64) = 1.05$, $p = 0.357$; $F_2(2,33) = 0.74$, $p = 0.490$, with responding across conditions shown in **Figure 3**.
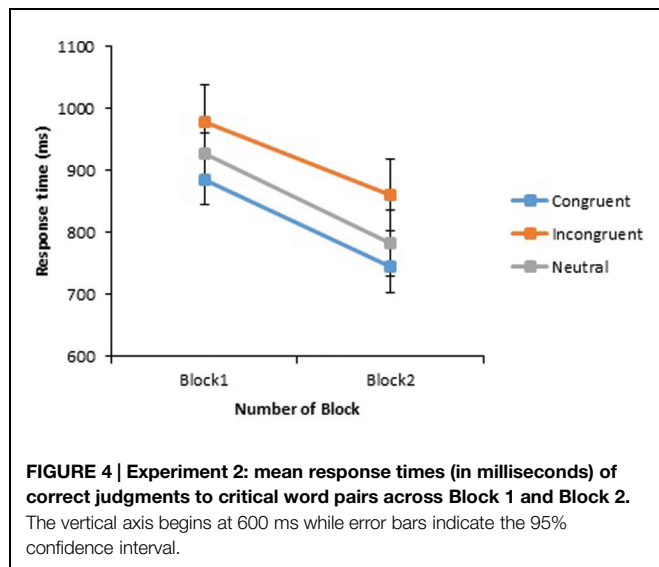
Accuracy for stereotype incongruent pairings failed to significantly increase across the blocks [$(+0.26\%)$, $t_1(33) = 0.10$, $p = 0.918$; $t_2(23) = 0.15$, $p = 0.880$], suggesting that the stereotypical picture training did indeed maintain stereotype biases. However, it is worth nothing that Block 1 accuracy to incongruent pairings in this study was considerably higher than Block 1 accuracy to incongruent pairings in Experiment 1 (83.21% vs. 74.86% respectively), thus leaving less scope for improvement in the current study. This issue is returned to in the General Discussion. Accuracy also remained significantly poorer for stereotype incongruent pairings than for neutral [$t_1(33) = 3.718$, $p = 0.001$, $dz = 0.64$; $t_2(23) = 6.70$, $p < 0.001$, $dz = 1.37$] and stereotype congruent pairings [$t_1(33) = 3.32$, $p = 0.002$, $dz = 0.57$; $t_2(23) = 8.64$, $p < 0.001$, $dz = 1.76$] by the end of the experiment.

Finally, two effects involving Participant Sex were found in the by-items analysis only. There was a main effect of Participant Sex, $F_2(1,33) = 165.93$, $p < 0.001$, with female participants achieving much higher levels of accuracy than male participants overall (94.9% vs. 86.5%). There was also a Participant Sex by Congruency interaction, $F_2(2,33) = 22.14$, $p < 0.001$, with female participants outperforming males in each of the congruency conditions, but particularly in response to incongruent pairings (89.15% vs. 75.85% respectively). In contrast to Experiment 1, it was now females who outperformed males in accuracy performance. The reason(s) for this contrasting performance between both sexes remain(s) unclear as again there were no obvious differences between the two samples.

Response Times

Response times below 150 ms, and above 4,000 ms were excluded from analysis (representing 1.77% of the total data) along with times for all errors of judgment (representing a further 12.85%), totaling a loss of 14.61% of the data. These data points were replaced as in Experiment 1. A significant main effect of Congruency was found, $F_1(2,64) = 15.18$, $p < 0.001$; $F_2(2,33) = 7.22$, $p = 0.003$, with fastest RTs to stereotype congruent word pairs ($M = 817$ ms), followed by neutral ($M = 862$ ms) and incongruent pairings respectively ($M = 920$ ms). A main effect of Block was also observed, $F_1(1,32) = 14.56$, $p = 0.001$; $F_2(1,33) = 130.93$, $p < 0.001$, with average RTs decreasing 128 ms from Block 1 to Block 2. As with the accuracy data, there was no evidence of a Congruency by Block interaction, $F_1(1.82, 58.12) = 0.38$, $p = 0.663$; $F_2(2,33) = 0.01$, $p = 0.988$, as RTs were found to decrease to a similar extent across blocks in all three congruency conditions (see **Figure 4**). While these improvements were each statistically significant ($p < 0.03$), they are taken as evidence for task habituation, as participants got progressively faster at

FIGURE 4 | Experiment 2: mean response times (in milliseconds) of correct judgments to critical word pairs across Block 1 and Block 2. The vertical axis begins at 600 ms while error bars indicate the 95% confidence interval.

responding to all critical word pairs as the task progressed, without any equivalent increase in accuracy performance across critical trials. A significant difference between RTs to stereotype incongruent and congruent pairs remained at the end of the experiment, $t_1(33) = 2.90$, $p = 0.007$, $dz = 0.50$; $t_2(23) = 2.82$, $p = 0.010$, $dz = 0.57$, and also between stereotype incongruent and neutral pairings, $t_1(33) = 2.15$, $p = 0.039$, $dz = 0.37$; $t_2(23) = 2.13$, $p = 0.044$, $dz = 0.43$.

An interaction of Participant Sex by Kinship term gender also emerged, $F_1(1,32) = 8.17$, $p = 0.007$; $F_2(1,33) = 8.32$, $p = 0.007$, with female participants faster when responding to female kinship terms over male kinship terms (796 ms vs. 855 ms respectively). Male participants were faster at responding to male kinship terms than female kinship terms (885 ms vs. 936 ms respectively), although they were slower than females at both.

There was also a main effect of Participant Sex in the by-items analysis only, $F_1(1,32) = 0.71$, $p = 0.406$; $F_2(1,33) = 19.34$, $p < 0.001$, with male participants slower at responding than female participants overall (880 ms vs. 815 ms respectively). Finally, a significant three-way interaction of Block by Congruency by Participant Sex was found in the by-participants analysis, $F_1(1.82,58.12) = 4.21$, $p = 0.023$; $F_2(2,33) = 1.62$, $p = 0.214$. This interaction was driven by performance to stereotype incongruent pairings. Females were faster to respond to these pairings than males in Block 1 by just 7 ms but this improvement jumped to 227 ms in Block 2 as female participants outperformed the males.

Fillers – Accuracy
Performance on the definitionally matching word pairs revealed a high mean accuracy score of 93.6% across blocks, with similar performance on both the male (94.2%) and female pairings (92.9%). However, performance was poorer on the definitionally mismatching word pairs ($M = 83.7\%$). Average accuracy to definitionally female pairings was high at 91.3%, but dropped to 76.2% for the definitionally male pairings. Again, it is hypothesized that this difference in accuracy performance is due

to the generic interpretation of certain male terms that are in fact male-specific by definition.

Fillers – Response Times
The RT data tell a similar story to the accuracy data. Reaction times to both male and female definitionally matching word pairs were similar (926 ms for the male versus 880 ms for the female pairs) with an average RT of 903 ms across blocks. Average RTs in the definitionally mismatching condition were slower, at 982 ms. Female mismatching pairings were responded to faster (943 ms) than male mismatching pairings (1022 ms) in general, again likely because participants considered that certain male terms can be used generically despite their gender specific definitions.

Discussion
This control experiment sought to maintain the stereotypical gender bias associated with certain role terms in English by presenting participants with pictures of men and women working in gender stereotypical roles. The hypothesis that accuracy of judgments on the stereotype incongruent word pairs would not improve across blocks in the judgment task was indeed supported. However, RTs to stereotype incongruent pairings *were* found to speed up across blocks in all congruency conditions. This pattern of results suggests that participants were benefitting from a practice effect and naturally speeding up at the task as it progressed. While RTs in Experiment 2 improved consistently across all conditions, RTs to the stereotype incongruent pairings in Experiment 1 decreased more sharply, with final RTs in line with those of stereotype congruent and neutral pairings.

As accuracy of stereotype incongruent trials did not significantly increase across blocks in Experiment 2, it is concluded that processing and making judgments about stereotype-consistent pictures did not help participants to overcome gender stereotype biases. It, therefore, appears that processing of specifically counter-stereotypical information in Experiment 1 was the reason for the improved performance on counter-stereotypical pairs in Block 2 of the judgment trials. We conclude that increasing exposure to counter-stereotype pictures is a useful means of moderating the effects of immediate gender activation in the judgment task.

Questionnaire Analysis
Participants were asked four questions in relation to each of the pictures presented to them in Experiments 1 and 2. Two of the questions required responses to be made along a Likert scale (earnings and job satisfaction) while two were open-ended (lifestyle and personal life). Responses to the latter two questions varied greatly in the level of detail and content provided by participants and were subsequently rated by two independent raters (one male, one female) along various dimensions described further below. The raters first analyzed the data independently and then met to compare results and try to reach a consensus on conflicting ratings. All inconsistencies were resolved after discussion so all data was kept. This rating procedure allowed for a statistical analysis of the subjective responses provided by participants in questions 3 and 4.

All questionnaire data was analyzed using a 2 (Participant sex: male, female) × 2 (Experiment: stereotype, counter stereotype) × 2 (Character: male, female) independent ANOVA.

Question 1. Earnings: How much do you think [insert character name] earns each year? Response options were on a 6-point scale, ranging from (1) < £10,000 to (6) > £50,000.

There was a significant main effect of Experiment, with those depicted in counter-stereotypical roles ($M = 3.42$) thought to earn more than those in stereotypical roles ($M = 3.12$), $F(1,120) = 6.58$, $p = 0.012$. There was also a significant main effect of Character with male characters deemed to earn more ($M = 3.40$) than female characters ($M = 3.21$), $F(1,120) = 4.52$, $p = 0.036$.

A significant interaction of Experiment by Character [$F(1,120) = 95.58$, $p < 0.001$] also emerged, with males working in stereotypical roles thought to earn more than males working in counter-stereotypical roles ($M = 3.73$ vs. $M = 3.08$ respectively) while females working in stereotypical roles were thought to earn less than females working in counter-stereotypical roles ($M = 2.65$ vs. $M = 3.77$ respectively). This pattern of results may reflect the high status associated with some of the typically male jobs used in this study (e.g., surgeon, judge, architect) compared to the lower status associated with many of the typically female jobs used (e.g., cleaner, hairdresser, au-pair).

Question 2. Job satisfaction: How satisfied do you think [he/she] is with [his/her] job? Response options to the above question were on a 5-point scale ranging from (1) 'Extremely Satisfied' to (5) 'Extremely Dissatisfied.'

No significant differences emerged in relation to job satisfaction, with ratings falling between 2.09 and 2.14 across all comparisons.

Question 3. Leisure: What are [his/her] leisure activities?

As mentioned earlier, responses to questions 3 and 4 were rated along various dimensions by two raters. The dimensions for question 3 were (1) Male typical vs. Female typical (2) Physically oriented or Mentally oriented (3) Social or Solitary activities and (4) High vs. Low cost.

Male typical vs. Female typical. Responses were rated according to the scale: 1 = Male-typical, 2 = Neutral, 3 = Female-typical.

There was a main effect of Character with female characters deemed to engage in more female-typical leisure activities and male characters thought to engage in more male-typical leisure activities ($M = 2.17$ vs. $M = 1.69$ respectively), $F(1,120) = 143.22$, $p < 0.001$. There was also a significant interaction of Experiment by Character [$F(1,120) = 146.025$, $p < 0.001$], with males in stereotypical roles judged as engaging in male-typical activities ($M = 1.43$) while females in stereotypical roles were judged as engaging in female-typical activities ($M = 2.40$). However, males and females in counter stereotypical roles were judged as having similarly rated leisure activities, close to the neutral point of 2 ($M = 1.95$ vs. $M = 1.94$ respectively).

Physically oriented vs. Mentally oriented. Responses were rated according to the scale: 1 = Physical, 2 = Physical and Mental, 3 = Mental.

There was a main effect of Character with female characters judged to engage in more mentally oriented leisure activities than male characters ($M = 2.28$ vs. $M = 2.14$, $F(1,120) = 8.50$, $p = 0.004$. There was also a significant interaction of Experiment by Character [$F(1,120) = 26.72$, $p < 0.001$], with females in stereotypical roles judged as engaging in more mentally oriented leisure activities than males ($M = 2.40$ vs. $M = 2.03$). However, males in counter stereotypical roles were judged as engaging in more mentally-oriented leisure activities than females in these same roles ($M = 2.26$ vs. $M = 2.15$).

Social vs. Solitary. Responses were rated according to the scale: 1 = Social, 2 = Neutral, 3 = Solitary. No significant differences were found to emerge in this category, with typical ratings falling close to 2 (i.e. neutral) across all comparisons. High cost vs. Low cost. Responses were rated according to the scale: 1 = Expensive, 2 = Reasonable, 3 = Cheap.

A main effect of Character was found with female characters judged to engage in somewhat cheaper leisure activities than male characters ($M = 2.33$ vs. $M = 2.23$ respectively), $F(1,120) = 4.24$, $p = 0.042$. There was also a marginal interaction of Participant sex by Character [$F(1,120) = 3.85$, $p = 0.052$] with male participants judging female characters as engaging in cheaper activities than male characters ($M = 2.41$ vs. $M = 2.21$), while female participants gave more similar ratings across female and male characters ($M = 2.26$ vs. $M = 2.25$ respectively).

Question 4. Briefly describe [his/her] personal life.

Responses to Question 4 were first rated along the dimensions (1) Traditional vs. non-traditional and (2) Happy vs. unhappy.

Traditional vs. Non-traditional. Responses were rated according to the scale: 1 = Traditional, 2 = Neutral, 3 = Non-traditional.

There was a main effect of Experiment with those working in stereotypical roles judged to lead more traditional personal lives than those working in counter-stereotypical roles ($M = 1.59$ vs. $M = 1.72$ respectively), $F(1,120) = 9.51$, $p = 0.003$. There was also a significant interaction of Experiment by Participant Sex [$F(1,120) = 5.48$, $p = 0.021$], as female participants deemed that those working in stereotypical roles lived more traditional roles than those working in counter-stereotypical roles ($M = 1.56$ vs. $M = 1.78$ respectively) while male participants judged that those in stereotypical and counter-stereotypical roles lead similarly traditional lives ($M = 1.63$ vs. $M = 1.66$ respectively).

A significant interaction of Experiment by Character was also found [$F(1,120) = 21.37$, $p < 0.001$], with females in counter stereotypical roles thought as leading a more traditional personal life than males working in counter stereotypical roles ($M = 1.63$ vs. $M = 1.81$ respectively). However, females in stereotypical roles were judged as leading a more non-traditional personal life than males working in stereotypical roles ($M = 1.70$ vs. $M = 1.47$ respectively). That said, it is worth noting that all mean values fell between the rating points of Traditional and Neutral as opposed to Non-traditional.

No significant differences were found to emerge with ratings falling between 1.80 and 1.90 across all comparisons, i.e., between the points of Happy (1) and Neutral (2). Happy vs. Unhappy. Responses were rated according to the scale: 1 = Happy, 2 = Neutral, 3 = Unhappy.

Overall, the picture booklets provide interesting supplementary data on the perception of men and women working in stereotypical and counter-stereotypical occupational roles. While integrating the findings into current social psychological literature is beyond the scope of this article, future research could further examine the themes which have emerged in our analysis.

General Discussion

In an effort to build on past research aimed at identifying strategies for overcoming stereotypes and prejudice, the current studies investigated the use of counter-stereotype information as a moderator of gender stereotype use. Experiment 1 involved presenting participants with pictures of people working in gender counter-stereotypical roles, and answering questions about the characters in these pictures. It was hypothesized that the questions would focus participants' attention on the characters presented (specifically their jobs), and that the pictures would be a salient reminder that people can work in gender atypical roles. It was found that accuracy of response to stereotype incongruent pairings did significantly increase after this picture training but importantly did not improve in Experiment 2; a control experiment in which participants were presented with pictures of people working in gender stereotypical roles. RTs decreased across blocks in both Experiments 1 and 2, independently of the type of picture training received. We posit that the decrease in RTs in Experiment 1 was due to the counter-stereotype picture manipulation, however, in Experiment 2, the lack of improvement in the accuracy data suggest this latter decrease was due to practice effects. Indeed this Experiment 2 data reflects unpublished findings from our lab in which RTs improved across blocks in the absence of any experimental manipulation whereas accuracy scores did not change.

However, interpretation of the results is not wholly straightforward as Block 1 accuracy was higher in Experiment 2 (83.21%) than Experiment 1 (74.86%). The reason(s) for such Block 1 differences remain unknown as both experiments were identical up until the picture task (between Block 1 and Block 2 of the judgment trials), and there were no discernible differences between the participant samples. This situation resulted in less scope for improvement across Blocks in Experiment 2 and final accuracy levels were similar across experiments. It is, therefore, not completely clear how the different picture strategies would have affected Block 2 performance if initial performance had been more similar. This same issue arose in Finnegan et al. (2015) and suggests that the judgment paradigm may benefit from further scrutiny when used in between-subject designs. That said, we argue that this Block 1 variability is not crucial to the conclusions we have drawn, as we were primarily interested in participants' response to the counter-stereotype information: specifically, whether this information would lead to a revision of participants stereotyped associations, or whether it would be ignored. While the results should be interpreted with caution because of the differential Block 1 performance, it appears that activating counter-stereotype gender associations did lead to a

revision of participants' stereotyped beliefs and ultimately helped them to control stereotype use in the judgment task.

Effects of Participant Sex were not anticipated in this research based on previous findings (Oakhill et al., 2005; Finnegan et al., 2015). However, in Experiment 1 males showed higher accuracy scores to critical trials, yet female participants were faster to respond. As such it cannot be ascertained whether female participants forsake accuracy so as to complete the test quickly (regardless of inaccurate responding) or whether they were simply weaker at recognizing and overcoming stereotype biases than the male participants. In Experiment 2 female participants were found to have higher levels of accuracy to critical trials than males and were also faster at responding. Therefore, in contrast to Experiment 1, the latter results suggest females are better at quickly recognizing and overcoming occupational stereotypes than males. Reasons for the differential performance of males and females across experiments remain unknown. While it is possible that the counter stereotype training task may have induced male participants to think more about their responding than in the stereotypical condition, it is unclear why female participants would not also respond to this training task.

The use of overt and striking counter-stereotype stimuli as part of the training in Experiment 1 provides evidence for the conversion theory of stereotype change, i.e., that the stereotypes can change rapidly on encountering a few, striking, counter-stereotype exemplars. It is also possible that bookkeeping processes may have played a role. As 24 pictures of people working in counter-stereotypical roles were presented, stereotype change is likely to have been somewhat incremental and become stronger as the participants proceeded through the pictures and questions. However, these findings provide less support for the subtyping theory of stereotype change which stipulates that the original stereotype can be protected through the formation of new categories to account for counter-stereotypical information. Although it cannot be definitively ruled out that participants used subtyping processes to account for the counter-stereotype exemplars, it seems unlikely that such a number of counter-stereotype exemplars could be easily rationalized in this way. On the contrary, the findings suggest that stereotypes can be weakened with sufficient category variation (Operario and Fiske, 2004).

Participants who received the counter-stereotype pictures seem to have been reminded that stereotypes are maladaptive forms of categories in that their content is not always accurate. Indeed, explicit training strategies such as this, in which counter-stereotype saliency is increased, may simply remind participants of specific things they already know, e.g., that a woman can be a surgeon and a man can be a nurse. It is logical to assume that with more frequent exposure, counter-stereotypic associations should become more accessible and the issue of gender 'atypical' roles may become obsolete. If true, this approach shows promise for inducing long-term stereotype change and could, with time, result in perceivers delaying the assignment of gender to a referent when gender-biased occupational terms are encountered (and hold back until more definitive gender information is supplied).

Repeated exposure to cultural images that reinforce automatic stereotypical or prejudiced associations means that these biases can become entrenched and difficult to overcome. Although people can often control and prevent the influences of stereotypes on overt behavior, such correctional efforts can be cognitively demanding and rely on factors such as a perceivers' awareness, motivation and cognitive resources, each of which can be easily undermined. Ideally, stereotype reduction research would aim to combat the initial activation of stereotypes as opposed to controlling the subsequent influence of these biases on behavior (Bodenhausen and Macrae, 1998; Gawronski et al., 2008). Also, although counter-stereotypes are by definition not highly accessible, and are unlikely to be implicitly activated and influence behavior to the extent that stereotypic associations do (Blair et al., 2001), their accessibility and influence can be increased given certain conditions. The use of counter-stereotypical pictures as a stereotype reduction manipulation is an example of a strategy that could be easily applied at a broad, societal level so as to increase exposure to counter-stereotype exemplars, and consequently instigate real change in the cognitive representations of gender-biased terms. For instance, it seems likely that frequent depiction of men and women working in gender atypical roles in educational material would effect change in students' cognitive representations of gender to accommodate this information[4]. Gender-fair pictures could also be used in other contexts where occupational stereotypes may be in use, e.g., in certain job advertisements. Future research could aim to evaluate the efficacy of exposure to counter-stereotypical pictures across a variety of different contexts in both the short- and longer-term.

Macrae and Bodenhausen (2000) suggest that there is an over-reliance on verbal category labels in research investigating the process of category activation. They caution that this over-reliance is problematic as in reality people are complex stimuli that can be classified by perceivers along multiple dimensions. Consequently, it cannot be assumed that the processing of verbal labels equates to the processes involved in person perception. The counter-stereotype picture training of Experiment 1 support this call of Macrae and Bodenhausen (2000) to move beyond the use of verbal stimuli (category labels) and to use more realistic stimuli. As stereotype reduction interventions are often detached from a 'real-life' context, doubt is cast on their usefulness beyond a laboratory setting (Lenton et al., 2009; Paluck and Green, 2009). While future research would undoubtedly benefit from an investigation of the cognitive processes involved in stereotype activation and application upon encountering real people, the use of pictures of people at work is a step in the right direction toward identifying further effective means of stereotype reduction with a training higher in ecological validity than many others.

Although the results of this research provide strong support for the malleability of gender stereotype biases, they also echo previous studies using this judgment paradigm that document the persistency of stereotyping effects. We found that the

processing of stereotype incongruent pairings rarely achieved the same level of effortlessly fast and accurate responding as that of stereotype congruent and neutral pairings. This same level of success (or lack of complete success) at overcoming occupational gender biases was previously found with strategies that included explicitly reminding participants that many jobs are not gender differentiated these days (Oakhill et al., 2005), and providing social consensus feedback that suggested past participants had no problem accepting stereotype incongruent pairings as correct (i.e., they were gender fair in their responding; Finnegan et al., 2015). Thus it appears that gender biases associated with social and occupational role nouns are deeply ingrained and difficult to overcome. Also, the fact that stereotypes are activated even when detrimental to task performance is further evidence that these biases are likely to be automatically activated.

In agreement with the assertion of past authors (e.g., Cohen, 1994; Lai et al., 2014) that it is important to include estimates of effect size in stereotype reduction research so as to assess whether an intervention has practical significance, we compared effect sizes of the current study with the above-mentioned research by Finnegan et al. (2015). We found that the current research led to larger effect sizes for the increase in accuracy to incongruent pairings in the by-participants ($dz = 0.61$ vs. $dz = 0.35$) and by-items ($dz = 1.16$ vs. $dz = 0.87$) analyses respectively. Effect sizes were more similar in the RT data, with larger effects found in the current work in the by-participants data ($dz = 0.77$ vs. $dz = 0.61$) while the opposite pattern was found in the by-items analyses ($dz = 1.61$ vs. $dz = 1.88$). However, as the work of Finnegan et al. (2015) involved three blocks of trials as opposed to two, it is likely to have benefitted more from practice effects. This comparison thus suggests that exposure to pictures of people working in counter-stereotypical occupations can lead to more reduced levels of stereotype application than feedback based on social norm information, at least in the short-term, and on the judgment task of Oakhill et al. (2005). The value of including effect sizes in research on stereotype and prejudice reduction can be seen in such comparisons and we recommend that it become standard procedure in this domain[5].

Overall, the case for reducing gender biases in relation to occupational stereotypes is not a trivial one. On the contrary, such efforts may have important implications for career choice, as exposure to gender stereotypes can influence preference toward jobs and activities from an early age. For instance, Liben et al. (2002) found that children aged 6–11 have quite fixed opinions about whether certain roles can be applied to women and men, typically stating that doctors are men and nurses are women. Moreover, Gottfredson's (1981, 2005) theory on career development asserts that children around 6 years-old begin to lose interest in occupations that are not in line with their gender self-concept. Such research suggests that gender stereotypes lead to inequality by artificially limiting the choices on offer to both

---

[4]Although some studies report that school books have become more gender fair across recent years (e.g., Diekman and Murnen, 2004; Moser and Hannover, 2013), effects of these changes on longer term cognitive representations of gender remain unknown.

[5]However, note that effect sizes like Cohen's dz which control for individual differences can be inflated compared to the effect sizes reported in between-subject designs where individual differences cannot be controlled (Dunlap et al., 1996; Lakens, 2013). Therefore caution should be exercised in comparing these effect sizes with training strategies used in conjunction with between-subject studies.

sexes. As such, it is imperative to devise interventions that challenge people's gendered perceptions and ultimately lead to a reduction in gender stereotyping. Increased exposure to counter-stereotypical exemplars could be a practical step toward achieving this aim.

References

Bar-Anan, Y., De Houwer, J., and Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: a correlational investigation with large samples. *Q. J. Exp. Psychol.* 63, 2313–2335. doi: 10.1080/17470211003802442

Bargh, J. A. (1994). "Four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition," in *Handbook of Social Cognition*, 2nd Edn, Vol. 1, eds R. S. Wyer Jr. and T. K. Srull (Mahwah, NJ: Erlbaum), 1–40.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *J. Pers. Soc. Psychol.* 6, 242–261. doi: 10.1207/S15327957PSPR0603_8

Blair, I. V., Ma, J. E., and Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *J. Pers. Soc. Psychol.* 81, 828–841. doi: 10.1037/0022-3514.81.5.828

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., and Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized individuals. *J. Pers. Soc. Psychol.* 80, 253–267. doi: 10.1037/0022-3514.80.2.253

Bodenhausen, G. V., and Macrae, C. N. (1998). "Stereotype activation and inhibition," in *Advances in Social Cognition*, Vol. 11, ed. R. S. Wyer Jr. (Mahwah, NJ: Erlbaum), 1–52.

Bodenhausen, G. V., Todd, A. R., and Richeson, J. A. (2009). "Controlling prejudice and stereotyping: antecedents, mechanisms, and contexts," in *Handbook of Prejudice, Stereotyping, and Discrimination*, ed. T. D. Nelson (New York, NY: Psychology Press), 111–135.

Carreiras, M., Garnham, A., Oakhill, J. V., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol.* 49A, 639–663. doi: 10.1080/713755647

Coats, S. J., and Smith, E. R. (1999). Perceptions of gender subtypes: sensitivity to recent exemplar activation and in-group/out-group differences. *Pers. Soc. Psychol. Bull.* 25, 515–526. doi: 10.1177/0146167299025004009

Cohen, J. (1994). The earth is round (p<.05). *Am. Psychol.* 49, 997–1003. doi: 10.1037/0003-066X.50.12.1103

Dasgupta, N., and Asgari, S. (2004). Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *J. Exp. Soc. Psychol.* 40, 642–658. doi: 10.1016/j.jesp.2004.02.003

Dasgupta, N., and Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81, 800–814. doi: 10.1037/0022-3514.81.5.800

Devine, P. G., and Baker, S. M. (1991). Measurement of racial stereotype subtyping. *Pers. Soc. Psychol. Bull.* 17, 44–50. doi: 10.1177/0146167291171007

Diekman, A. B., and Murnen, S. K. (2004). Learning to be little women and little men: the inequitable gender equality of nonsexist children's literature. *Sex Roles* 50, 373–385. doi: 10.1023/B:SERS.0000018892.26527

Dijksterhuis, A., and van Knippenberg, A. (1996). The knife that cuts both ways: facilitated and inhibited access to traits as a result of stereotype activation. *J. Exp. Soc. Psychol.* 32, 271–288. doi: 10.1006/jesp.1996.0013

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Mem. Cogn.* 32, 551–559. doi: 10.3758/BF03195846

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., and Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol. Methods* 1, 170–177. doi: 10.1037/1082-989X.1.2.170

Finnegan, E., Garnham, A., and Oakhill, J. (2015). Social consensus feedback as a strategy to overcome spontaneous gender stereotypes. *Discourse Process.* 52, 434–462. doi: 10.1080/0163853X.2015.1026680

Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychol. Methods* 1, 379–390. doi: 10.1037/1082-989X.1.4.379

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au pairs are rarely male: norms on the gender perception of role names

across English, French, and German. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., and Oakhill, J. (2012). Gender representation in different languages and grammatical marking on pronouns: when beauticians, musicians, and mechanics remain men. *Discourse Process.* 49, 481–500. doi: 10.1080/0163853X.2012.688184

Garnham, A., Oakhill, J., and Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Mem. Cogn.* 30, 439–446. doi: 10.3758/BF03194944

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., and Strack, F. (2008). When "just say no" is not enough: affirmation versus negation training and the reduction of automatic stereotype activation. *J. Exp. Soc. Psychol.* 44, 370–377. doi: 10.1016/j.jesp.2006.12.004

Girden, E. R. (1992). *ANOVA: Repeated Measures*. Newbury Park: Sage Publications.

Gottfredson, L. S. (1981). Circumscription and compromise: a developmental theory of occupational aspirations. *J. Counsel. Psychol.* 28, 545–579. doi: 10.1037/0022-0167.28.6.545

Gottfredson, L. S. (2005). "Using Gottfredson's theory of circumscription and compromise in career guidance and counselling," in *Career Development and Counselling: Putting Theory and Research to Work*, eds S. D. Brown and R. W. Lent (New York, NY: Wiley), 71–100.

Green, R. J., and Ashmore, R. D. (1998). Taking and developing pictures in the head: assessing the physical stereotypes of eight gender types. *J. Appl. Soc. Psychol.* 28, 1609–1636. doi: 10.1111/j.1559-1816.1998.tb01692.x

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Hamilton, S. (2008). *Automatic Gender Stereotyping, an ERP Investigation*. Master's thesis, University of Sussex, Brighton.

Hellinger, M., and Bußmann, H. (2001). *Gender Across Languages*, Vol. 1–3. Philadelphia, PA: John Benjamins Company.

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456. doi: 10.1007/s10936-007-9053-z

Irmen, L., and Roßberg, N. (2004). Gender markedness of language: the impact of grammatical and nonlinguistic information on the mental representation of person information. *J. Lang. Soc. Psychol.* 23, 272–307. doi: 10.1177/0261927X04266810

Joy-Gaba, J. A., and Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *J. Soc. Psychol.* 41, 137–146. doi: 10.1027/1864-9335/a000020

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., and Russin, A. (2000). Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *J. Pers. Soc. Psychol.* 78, 871–888. doi: 10.1037/0022-3514.78.5.871

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

Kreiner, H., Sturt, P., and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007.09.003

Kunda, Z., and Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: a parallel-constraint-satisfaction theory. *Psychol. Rev.* 103, 284–308. doi: 10.1037/0033-295X.103.2.284

Lai, C. K., Hoffman, K. M., and Nosek, B. A. (2013). Reducing implicit prejudice. *Soc. Pers. Psychol. Compass* 7, 315–330. doi: 10.1111/spc3.12023

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., et al. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* 143, 1765–1785. doi: 10.1037/a0036260

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

Lassonde, K. A., and O'Brien, E. J. (2013). Occupational stereotypes: activation of male bias in a gender-neutral world. *J. Appl. Soc. Psychol.* 43, 387–396. doi: 10.1111/j.1559-1816.2013.01008.x

Lenton, A. P., Bruder, M., and Sedikides, C. (2009). A meta-analysis on the malleability of automatic gender stereotypes. *Psychol. Women Q.* 33, 183–196. doi: 10.1111/j.1471-6402.2009.01488.x

Liben, L. S., Bigler, R. S., and Krogh, H. R. (2002). Language at work: children's gendered interpretations of occupational titles. *Child Dev.* 73, 810–828. doi: 10.1111/1467-8624.00440

Macrae, C. N., and Bodenhausen, G. V. (2000). Social cognition: thinking categorically about others. *Annu. Rev. Psychol.* 51, 93–120. doi: 10.1146/annurev.psych.51.1.93

Merry, E. (1995). *First Names: The Definitive Guide to Popular Names in England and Wales 1944-1994 and in the Regions 1994.* London: Her Majesty's Stationery Office.

Moser, F., and Hannover, B. (2013). How gender fair are German schoolbooks in the twenty-first century? An analysis of language and illustrations in schoolbooks for mathematics and German. *Eur. J. Psychol. Educ.* 29, 1–21. doi: 10.1007/s10212-013-0204-3

Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cogn.* 33, 972–983. doi: 10.3758/BF03193206

Operario, D., and Fiske, S. T. (2004). "Stereotypes: content, structures, processes, and context," in *Social Cognition*, eds M. B. Brewer and M. Hewstone (Malden: Blackwell Publishing), 120–141.

Paluck, E. L., and Green, D. P. (2009). Prejudice reduction: what works? A review and assessment of research and practice. *Annu. Rev. Psychol.* 60, 339–367. doi: 10.1146/annurev.psych.60.110707.163607

Pettigrew, T., and Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* 90, 751–783. doi: 10.1037/0022-3514.90.5.751

Puhl, R. M., Schwartz, M. B., and Brownell, K. D. (2005). Impact of perceived consensus on stereotypes about obese people: a new approach for reducing bias. *Health Psychol.* 24, 517–525. doi: 10.1037/0278-6133.24.5.517

Reynolds, D. J., Garnham, A., and Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Q. J. Exp. Psychol.* 59, 886–903. doi: 10.1080/02724980543000088

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84, 127–190. doi: 10.1037/0033-295X.84.2.127

Stangor, C., Sechrist, G. B., and Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Pers. Soc. Psychol. Bull.* 27, 486–496. doi: 10.1177/0146167201274009

Cross-linguistic evidence for gender as a prominence feature

Yulia Esaulova* and Lisa von Stockhausen

Department of Psychology, University of Duisburg-Essen, Essen, Germany

This paper discusses recent findings in the online sentence processing research that suggest to consider gender information a prominence feature. Prominence features are hierarchically ordered information types that interact with formal features of arguments (e.g., grammatical functions, thematic roles) and thus determine the readers' ability to efficiently interpret linguistic ambiguities. While previous research addressed a number of prominence features (e.g., animacy, definiteness, person), there is now first empirical evidence indicating that gender information also influences the assignment of thematic roles across languages. Grammatically masculine role nouns are processed faster as agents than patients compared to feminine ones. Stereotypically male role nouns (e.g., electrician) are integrated with an agent role easier than neutral ones (e.g., musician), which in turn are integrated easier than female ones (e.g., beautician). Conceptualizing gender as a prominence feature will not only expand our knowledge about information types relevant for online comprehension but also uncover subtle gender biases present in language. The present work explores the possibility for a theoretical integration of social psychological and psycholinguistic research focusing on gender with research on prominence. Potential advantages an interdisciplinary approach to the study of gender as a prominence feature, open questions and future directions are discussed.

Keywords: prominence, grammatical gender, stereotypical gender, thematic roles, grammatical functions

Introduction

Natural languages often present their users with ambiguities that require an interpretation even in cases when the provided information does not suffice to resolve them. Comprehenders may apply one of the two major strategies to process the ambiguous linguistic input. One strategy would involve computational mechanisms that defer hypotheses about the possible meaning of a sentence until enough information is provided to resolve ambiguities. Another strategy would involve processing the sentence incrementally, on a word-by-word basis, as the linguistic input unfolds. While both strategies have the same goal, the incremental integration seems to offer a more efficient and rapid way to achieve the interpretation of a sentence, most certainly for languages where the (disambiguating) verb occurs in sentence- or clause-final position (Kamide et al., 2003). The model of incremental processing assumes that language users make probabilistic predictions about the syntactic structure and the meaning of a sentence based on a number of constraints (e.g., case, agreement). Prominence is a theoretical notion that is used to identify certain information types as constraints (or prominence features) that are organized hierarchically and interact with formal features of verbal arguments (de Hoop and Lamers, 2006). As a result of this interaction, in the process of incremental interpretation thematic structure and grammatical functions of verbal arguments can be predicted from the position of their prominence features on a scale, where higher

ranked (more prominent) arguments are more likely to be interpreted as agents and subjects rather than patients and objects. Animacy, definiteness, and person are considered such hierarchically organized information types and are referred to as prominence features (Lamers, 2012). Thus, prominence features are discussed in terms of scales, where animate entities are more prominent and outrank inanimate, definite outrank indefinite, and first and second person outrank third. Readers rely on prominence information as a cue to assess the structure and the meaning of a sentence especially in cases where information from case marking and/or word order is ambiguous and cannot be used for interpretation. In this paper, we propose to conceptualize gender information as another prominence feature, i.e., the information type that systematically affects readers' predictions about thematic roles and grammatical functions of arguments. We both suggest a theoretical foundation and evaluate the existing empirical evidence for different types of gender information to function as a prominence feature with the aim to demonstrate that gender influences go beyond the well-known agreement and mismatch effects.

Prominence and Sentence Structure

The role of prominence features for the comprehension of sentences is often discussed in terms of their interaction with grammatical functions or thematic roles. Research has shown that the relative ease or difficulty in the assignment or accessibility of entities performing an action (i.e., grammatical subjects or thematic agents) and those receiving an action (i.e., grammatical objects or thematic patients) depends on the characteristics of prominence features they possess. Generally, entities possessing highly ranked prominence features tend to occupy more syntactically prominent positions, while entities with lower rankings in terms of prominence occupy less prominent syntactic positions. In case of animacy as a most widely studied prominence feature, for instance, animate nouns or noun phrases are rather associated with subject functions and agent roles and inanimate ones with object functions and patient roles (e.g., MacDonald, 1994; Traxler et al., 2002; Bornkessel-Schlesewsky and Schlesewsky, 2009). In other words, prominence hierarchies (e.g., animates over inanimates) and the hierarchies within grammatical functions (e.g., subjects over objects) and thematic roles (e.g., agents over patients) align with or map onto each other. This often results in the so-called "harmonic alignment" (Aissen, 2003) when highly prominent entities (e.g., animates) are matched with highly ranked grammatical functions/thematic roles (subjects/agents) and less prominent entities (e.g., inanimates) are matched with lower ranked objects/patients. Arguments which prominence features are harmonically aligned with their thematic roles/grammatical functions have been shown to be processed faster and even facilitate comprehension difficulties related to syntactic ambiguities (e.g., Traxler et al., 2002, 2005; Gennari and MacDonald, 2008). The theoretical substantiation of the principle of harmonic alignment is offered by the model of Incremental Optimization of Interpretation (de Hoop and Lamers, 2006). This model defines several constraints (e.g.,

agreement, case, etc.) that are used to distinguish subjects from objects (e.g., the verb agrees with the subject and not the object; the subject is in the nominative case, while the object is in the accusative case, etc.). Prominence is defined as one of these constraints and assumes subjects to be associated with higher prominence rankings than objects. In this respect, prominence can be seen as a semantic cue that links grammatical functions to semantic relations between arguments during language comprehension. It is worth noting that prominence has been shown not only to relate formal structure of a sentence to its semantic content, but also to modulate the interpretation of a sentence. One of the examples of such modulation that is widely discussed in literature concerns the interpretation of sentences with subject- and object-extracted relative clauses. It is a well-established finding that subject-extracted relative clauses (e.g., *The reporter that attacked the senator admitted the error*) are easier for comprehension than object-extracted ones (e.g., *The reporter that the senator attacked admitted the error*; e.g., King and Just, 1991). However, Traxler et al. (2002) demonstrated that the difficulty in the processing of relative clauses can be modulated as a function of animate vs. inanimate sentence heads. When sentence heads were inanimate (*The movie that the director watched received the prize*), object-extracted clauses were almost as easy to comprehend as their subject-extracted counterparts (*The director that watched the movie received the prize*). These and similar findings indicate that animacy as a prominence feature is a semantic cue that may reduce or strengthen syntactic complexity effects and thus is a factor significantly influencing the comprehension of a sentence together with syntactic and thematic structures (Traxler et al., 2005).

Gender as a Prominence Feature: Theoretical Motivation

The interaction of animacy with thematic roles and grammatical functions has been confirmed as a cross-linguistic phenomenon in a number of linguistic tasks other than the interpretation of relative clauses (e.g., English—McDonald et al., 1993; German— Van Nice and Dietrich, 2003; Spanish—Prat-Sala, 1997) and established animacy as a prominence feature. Describing animacy as a prominence feature, Yamamoto (1991) regards it as a "supra-linguistic" concept—a fundamental semantic dimension which as such also affects a number of linguistic phenomena (e.g., word order, case marking). This understanding of a prominence feature as a supra-linguistic concept can be applied to a number of information types other than animacy and we would like to argue that gender is one of them. Similarly to animacy, gender is a fundamental semantic dimension expressed on a biological level as a characteristic of individuals and on a social level through social practices (Ridgeway, 2001). As one of the categories essential for social interaction (Fiske, 1998), gender is represented in language in diverse ways: through grammatical gender (i.e., a noun class system where gender may be identified by grammatical markings, such as feminine suffixes, as in German *Musiker*masculine vs. *Musikerin*feminine "musician"), natural gender (i.e., referring to the sex of a referent, as in pronouns), definitional gender (i.e., where it is part of the definition of the word, as in

king vs. *queen*) and stereotypical gender of role nouns (i.e., the likelihood that an activity would be done by either a man or a woman that reflects existing gender stereotypical representations, as in stereotypically male *electrician* vs. stereotypically female *beautician* vs. neutral *musician*; e.g., Gabriel et al., 2008; Kreiner et al., 2008)[1]. The way both biological and social aspects of gender are represented in language makes it plausible to consider gender information—denoted on grammatical (grammatical gender) and conceptual (stereotypical gender) levels in role nouns—a prominence feature influencing the interpretation of thematic roles in a sentence. Whereas there is still no clear understanding which communicative function gender serves, Bates et al. (1996) suggest that it does serve one considering how pervasive and persistent it is in the world's languages despite its linguistic costs. In their experiment, Bates et al. (1996) demonstrated that grammatical gender of an adjective in Italian clearly primed the recognition of the following noun. Some research has found that language users may not always choose the interpretation strategy that would focus on gender information as useful (e.g., Garnham et al., 1992; McDonald and MacWhinney, 1995), however most research has shown that the integration of gender information represented in language is crucial for comprehension and is processed in highly automatized ways (e.g., Irmen, 2007; Cacciari et al., 2011; Esaulova et al., 2014). This research identified various types of gender information (e.g., grammatical gender markings, gender stereotypical representations, definitional gender), as well as the time course of its integration during language processing. To identify influences of different types of gender information person denotations (e.g., *electrician* or *soccer fan*) are often used, since they both entail grammatical gender information (marked morphologically or by the determiner) and are subjects to gender stereotypes (Baudino, 2001). Research paradigms involving such denotations, or role nouns, typically employ reference resolution, which requires the integration of gender information in order to be interpreted (e.g., a masculine or feminine pronoun *he/she* referring to stereotypically masculine role noun *electrician*). This integration may result in mismatch effects that are observed in cases of grammatical disagreement and other gender incongruities (e.g., *electrician—she*) and are reflected in longer processing times indicated by corresponding behavioral measures (e.g., longer fixation times and more regressions in case of eye-tracking measures). Thus, in their reading time study, Kennison and Trofe (2003) presented readers with pairs sentences, where the first one contained a stereotypically male or female role noun and the second one a pronominal reference to this role noun (*he/she*). The results showed significantly longer reading time when the stereotypical gender of the role noun and the pronoun gender mismatched (e.g., *executive*Male . . . *she*; *secretary*Female . . . *he*) compared to when they matched (e.g., *executive*Male . . . *he*; *secretary*Female . . . *she*).

In a similar vein, Duffy and Keir (2004) monitored participants' eye-movements when they read sentences like *a babysitter found*

himself/herself humming while walking up to the door, where stereotypical gender of a role noun either matched or mismatched the reflexive pronoun referring to it (Experiment 1). The results showed the gender mismatch effect when reflexive pronouns were incongruent with the gender stereotype. Interestingly, this effect disappeared in Experiment 2, where a context preceding sentences disambiguated the gender of a character explicitly stating whether it was a man or a woman. Kreiner et al. (2008) further explored instances when the gender mismatch effect can be overridden. After demonstrating that readers slow down when an anaphor (e.g., *herself*) mismatched stereotypical (e.g., *minister*) and definitional (e.g., *king*) gender of the role noun antecedent (Experiment 1), they contrasted the congruity of stereotypical and definitional gender with the reflexive in cataphora sentences, showing that when the reflexive preceded the role noun, the mismatch effect only occurred for definitional but not stereotypical gender nouns (Experiment 2). Kreiner et al. (2008) interpret the results as supporting theoretical perspectives on the nature of different gender types and argue that stereotypical gender is inferred from world knowledge, as suggested by the mental models approach, while definitional gender is defined lexically.

Another theoretical approach was addressed by Sturt (2003), who demonstrated the gender mismatch effect in an eye-tracking study that used paragraphs containing two potential antecedents—one of them a stereotypically male or female role noun—for the reflexive anaphor (e.g., *Jonathan/Jennifer was pretty worried at the City Hospital. He/She remembered that the surgeon had pricked himself/herself with a used syringe needle. There should be an investigation soon.*). Chomsky's binding theory (Chomsky, 1981) predicts that the second character (*the surgeon*) is a grammatical (and the only possible) antecedent, while the character mentioned first is an ungrammatical one. Even though an early effect between the grammatical antecedent and the anaphor supported the binding theory, ungrammatical antecedents also affected processing at a relatively later stage.

Some research reporting gender mismatch-effects detected asymmetries in the processing of gender cues. In an eye-tracking experiment (Experiment 2) in German, Reali, Esaulova and von Stockhausen (2015) analyzed the resolution of pronoun anaphors (*er* "he"/*sie* "she") referring to gender-stereotypical descriptions of an occupation (e.g., stereotypically male *M. F. repariert und stellt Möbel her, arbeitet mit Holz.* "M. F. repairs and produces pieces of furniture, works with wood."). Results revealed an asymmetry in the processing of anaphor gender, as the mismatch effect occurred earlier for masculine and later for feminine pronouns, suggesting that representations of female referents are more flexible and thus are integrated easier into counterstereotypical contexts compared to male referents. In a priming study, Cacciari and Padovani (2007) reported an asymmetry in the same direction on bigender nouns, where the mismatch effect manifested for masculine pronouns following stereotypically female role nouns (e.g., *insegnante— lui* "teacher—he") but not for feminine pronouns after male roles (e.g., *ingegnere—lei* "engineer—she"). Using event-related potentials (ERPs), Siyanova-Chanturia et al. (2012) identified an N400-like effect as an electrophysiological response to

---

[1]Despite the mapping of grammatical gender to natural gender, which is common for such denotations, they should be regarded as neither perfectly correlated with nor completely independent from one another, as grammatically masculine forms may sometimes be used generically and refer to both men and women.

masculine but not feminine pronouns primed by stereotypically incongruent role nouns (e.g., insegnante—lui "teacher—he"). Taking into account the results of the three aforementioned studies, it must be noted that considering specific gender cues (i.e., masculine vs. feminine) can prove beneficial to the understanding of the effects related to gender agreement or congruity.

As some other studies indicate, gender agreement seems to affect language comprehension in ways that go beyond word recognition and anaphor resolution but can also be used as a cue to determine thematic roles, even though gender differs from other aspects of inflectional morphology (e.g., case, person, number, etc.) in that it is an inherent property of nouns. Friederici and Weissenborn (2007) provide an overview of ERP studies demonstrating that subject-verb gender agreement is among other features (number, person) that elicit left anterior negativity effects in the identification of thematic structures. Devescovi et al. (1998) and Kail (1989) also argue that gender agreement seems to play a role in determining "who did what to whom" in a sentence and that the extent to which it is used as a cue may depend on the age of language users and the language itself. This is in line with the competition model (MacWhinney et al., 1984), which evaluates the extent to which readers rely on different cues—word order, (gender) agreement, animacy, etc.,—to interpret the structure of a sentence and predicts that the strength of each of the cues varies across languages. Importantly, the focus of these works is on gender as one of the cues used to determine thematic roles (along with case and word order) and consider gender agreement (e.g., between a noun and a verb) rather than specific gender characteristics of nouns. The question whether particular gender characteristics could make one (role) noun fit a thematic role better than another noun has so far remained open. Thus, while the influences of gender information on language comprehension have been repeatedly demonstrated using various paradigms, research methods and theoretical approaches, considering gender a prominence feature would predict gender to influence the processing of formal relations in a sentence structure on a much more far reaching, rather implicit level. In line with this idea, previous research on biases (discussed in more detail below) has shown that the use of particular linguistic structures (e.g., negations) implies beliefs and expectations corresponding to existing stereotypes (e.g., de Villiers and Flusberg, 1975; Beukeboom et al., 2010). Understanding whether and how gender information is used to predict thematic structures could provide an insight on mechanisms underlying gender biases and stereotyping.

Further indications for why gender information may need to be considered as a prominence feature come from two, at first sight, theoretically distinct areas of research. On the one hand, linguistic theories, such as differential object marking (Aissen, 2003) suggest that the overt case marking of an object reflects its place on a prominence hierarchy, where overtly case-marked objects are more prominent than non-marked ones. In languages with grammatical gender system, case marking often depends on grammatical gender and differs for feminine and masculine entities. In German, for instance, the singular form of the masculine determiner is marked overtly in all four

cases ($der_{\text{Nominative}}$; $den_{\text{Accusative}}$; dem_{Dative}; des_{Genitive}), while the singular form of the feminine determiner only has two forms: one for nominative and accusative cases and one for dative and genitive ($die_{\text{Nominative/Accusative}}$; $der_{\text{Dative/Genitive}}$). According to the differential object marking theory, such differences in case markings indicate a hierarchy where grammatically masculine and feminine entities differ in rankings. The differential marking of grammatical gender information suggests that gender may be considered a prominence feature for which hierarchical organization is typical. On the other hand, research in social cognition relates masculinity and femininity to constructs of status and power, which are described in terms of high and low extremes (high vs. low status/power). Higher rankings on these hierarchies tend to be attributed to masculinity, therefore indicating a gender hierarchy where masculinity outranks femininity (e.g., Spence and Buckner, 2000; Ridgeway, 2001; Koenig et al., 2011). Furthermore, thematic agents reflect to some extent the social psychological agency understood as a modality of human behavior and expressed in the desire to master the environment and experience competence, power, and achievement (Bakan, 1966). This social concept of agency—in turn—relates to gender through sex role characteristics that differ for men and women and become apparent through distinct socialization patterns. The tendency to be socialized to be achievement-oriented, independent, and self-sufficient, for instance, is reported to be typical for men but not women (Cross and Madson, 1997). Similar to the social-role theory, expectation states theory proposes that the gender system is entwined with social hierarchy and leadership through status beliefs (Wagner and Berger, 1997). Status beliefs are commonly held cultural beliefs that associate greater competence and social significance with men than women and are at the core of gender stereotypes (Williams and Best, 1990). Thus, hierarchies within the social gender system that are related to leadership, status, and power can be viewed as representing agency on a social psychological rather than linguistic level of a thematic structure.

The hierarchical organization within the concept of gender suggested by the described linguistic and social psychological phenomena invites an interdisciplinary approach to consider whether gender information as it is represented in languages can be conceptualized as a prominence feature. If this is the case, then grammatical and stereotypical gender of role nouns should affect the processing of their thematic roles in sentences. When the thematic structure of a sentence is ambiguous and allows for more than one interpretation of who produced or received an action, more prominent role nouns (e.g., grammatically masculine and stereotypically male) should be perceived as better agents and less prominent ones (e.g., grammatically feminine and stereotypically female) as better patients. If gender information functions as a prominence feature, differences in the processing of role nouns in specific thematic roles (agent or patient) depending on their gender characteristics should be observed, as some role nouns would be seen as fitting their thematic roles better than others. This hypothesis was addressed in two studies in German and French languages, which identified gender hierarchies that affect the processing of thematic structures. These studies and their experimental results are described below.

Overview of Studies on Gender as a Prominence Feature

The first study (Esaulova, 2015) includes two eye-tracking experiments ($N_1 = 32$; $N_2 = 40$) in German, where locally ambiguous subject- and object-extracted relative clauses were used to examine whether gender information may influence the identification of role nouns as agents and patients in sentences. Readers were presented with sentences like *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet hat/haben, ist aufmerksam* "The flight attendant_Typically female + feminine, who has observed many Tourists_Neutral + feminine/masculine/whom many tourists_Neutral + feminine/masculine have observed, is attentive." These sentences were designed in such a way that agent and patient roles remained ambiguous until readers reached the very final word of the relative clause—the auxiliary verb *hat/haben* "has/have," which then disambiguated agent and patient roles through its number marking. In Experiment 1, all used role nouns were neutral with regard to stereotypical gender. In terms of grammatical gender, main clause role nouns varied and were either grammatically feminine or masculine, while relative clause role nouns were feminine and did not vary. The experimental design thus included grammatical gender of the main clause role noun (RN1) and the relative clause type as factors, which resulted in four conditions: 1. masculine RN1 + SRC; 2. feminine RN1 + SRC; 3. masculine RN1 + ORC; 4. feminine RN1 + ORC. Depending on the type of the relative clause, RN1 served either as a thematic agent or as a patient. According to the hypothesis, grammatically masculine agents were expected to require shorter processing times compared to feminine ones (conditions 1 vs. 2), while masculine patients were expected to require longer processing times compared to feminine ones (conditions 3 vs. 4). The results showed shorter reading times[2] when masculine rather than feminine role nouns served as agents (conditions 1 vs. 2), while no differences were found when the two role nouns served as patients (conditions 3 vs. 4). In Experiment 2, main clause role nouns varied in stereotypical gender and were either stereotypically female (e.g., flight attendant) or neutral (e.g., student), while their grammatical gender was feminine. Relative clause role nouns varied in grammatical gender and were either masculine or feminine, while they were neutral with regards to stereotypical gender. The experimental design included stereotypical gender of the main clause role noun (RN1), grammatical gender of the relative clause role noun (RN2) and the relative clause type as factors. This resulted in either stereotypically female or neutral RN1 in one of the four conditions: 1. masculine RN2 + SRC; 2. feminine RN2 + SRC; 3. masculine RN2 + ORC; 4. feminine RN2 + ORC. In addition to the hypothesis regarding grammatical gender of RN2 (identical to that in Experiment 1), hypothesis concerning stereotypical gender predicted longer reading times for stereotypically female than neutral agents and for neutral than female patients. The results of Experiment 2 again showed that reading times were shorter for grammatically masculine rather

---

[2]Here and below the term reading times is used to refer to fixations and regressions expressed by a number of eye-tracking measures and on various regions of sentences which are not specified in the text (see the original study for these details).

than feminine role nouns when they served as agents (conditions 3 vs. 4) but no significant differences were detected when they served as patients (conditions 1 vs. 2). On the other hand, the effect of stereotypical gender emerged for both agents and patients, with longer reading times for stereotypically female than neutral agents and for neutral compared to female patients. These findings suggest that both grammatical and stereotypical gender is involved in the processing of thematic relations in a sentence, namely the interpretation of agents and patients. Differences in the processing times indicate the relevance of both types of gender cues for the identification of what thematic role a role noun serves in a sentence.

Another study (Esaulova, 2015) also included two eye-tracking experiments ($N_1 = 25$, $N_2 = 33$) that used the French gender-ambiguous indirect object pronoun *lui* "him/her" as a backward anaphor to investigate whether gender information may affect the processing of grammatical functions/thematic roles of role nouns. The pronoun *lui* "him/her" in sentences like *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoreux* "In fact, the dietician_Typically Female + feminine recommended to him/her_gender – ambiguous, so to this_masculine/feminine pharmacist_Neutral + masculine/feminine, a strict plan" indicated an upcoming referent without specifying its gender. According to the design, referent role noun varied in grammatical gender (masculine or feminine) and was neutral with regard to stereotypical gender. The first role noun had a fixed grammatical gender (feminine in Experiment 1 and masculine in Experiment 2) and varied in stereotypical gender (female/neutral in Experiment 1 and male/neutral in Experiment 2). Hypotheses predicted longer processing time for grammatically masculine than feminine objects/patients (the second role noun), as well as neutral than stereotypically male and stereotypically female than neutral subjects/agents (the first role noun). The results of both experiments showed longer reading times for grammatically masculine compared to feminine objects/patients, as expected by the hypothesis regarding the grammatical gender of role nouns. They also supported predictions about the stereotypical gender of role nouns showing longer reading times for stereotypically female than neutral (Experiment 1) and neutral than stereotypically male subjects/agents (Experiment 2). The findings demonstrate a relative difficulty in the processing of masculine compared to feminine referents in both experiments, which indicates that readers do create specific expectations about the gender of the referent role noun relying on its grammatical function of an object in the sentence. Additionally, the findings suggest an easier integration of neutral rather than stereotypically female (Experiment 1) and stereotypically male rather than neutral nouns (Experiment 2) with an subject/agent role in a sentence.

Evidence-Based Interpretation of Gender as a Prominence Feature

The findings of these two studies provide the first evidence that grammatical and stereotypical gender information in role nouns may be conceptualized as a prominence feature. Like other prominence features, gender information appears to map onto

thematic relations and grammatical functions in sentences. The principle of harmonic alignment makes it possible to identify whether an information type is organized hierarchically and how its components are ranked on this scale through the relative ease or difficulty in the processing. When the ranking of the feature on one prominence scale differs with the ranking on another prominence or thematic roles/grammatical functions scale, processing costs are higher compared to when scales are aligned with each other, so that the rankings on one correspond to the rankings on the other and are both either high or low. Esaulova (2015) demonstrated a relative ease in the processing of sentences with relative clauses when masculine rather than feminine grammatical gender of role nouns corresponded to high-ranked thematic agents. Similarly, yet in a different language, the processing of sentences with backward anaphors was easier when low-ranked object referents were grammatically feminine rather than masculine (Esaulova, 2015). Both of these findings suggest the hierarchical organization of grammatical gender information, where masculine gender outranks feminine gender on the prominence hierarchy. Importantly, the results observed in all four experiments above revealed that grammatical gender information is organized hierarchically in the same way (masculine over feminine) in both German and French languages. Since previous research on prominence points at the general characteristic of prominence features as cross-linguistically motivated information types that have the same hierarchical organization across languages and linguistic variations (e.g., as it is for animacy in English—McDonald et al., 1993; German—Van Nice and Dietrich, 2003; and Spanish—Prat-Sala, 1997), the results of both studies can be taken as a cross-linguistic validation of gender information as a prominence feature.

In addition to grammatical gender, stereotypical gender information also appears to map onto thematic structure of sentences revealing its own hierarchical structure. Sentences with relative clauses were processed faster when low-ranked patient roles were assigned to stereotypically female nouns and high-ranked agent roles to neutral ones in Experiment 2 in German (Esaulova, 2015). Similarly, stereotypically male agents were relatively easier to process than neutral ones and neutral agents easier than stereotypically female ones in French (Esaulova, 2015). These processing patterns suggest a hierarchy where stereotypically male role nouns outrank neutral ones and neutral role nouns outrank stereotypically female ones thus providing a complementary prominence scale of gender information.

Gender Prominence: Limitations to be Considered

These findings reveal implicit ways in which grammatical and stereotypical gender affect the interpretation of a thematic structure of a sentence in German and French languages. However, there are several considerations and limitations that need to be taken into account when interpreting the results. Despite differences in the syntactic structure of the experimental materials, grammatical gender appears to be organized in the same hierarchical way and constitute a prominence scale where masculine entities outrank feminine ones. Grammatical gender information affected processing similarly in both German and French languages: feminine entities were perceived as less likely agents/subjects compared to masculine ones. As to the stereotypical gender information, its organization in terms of a prominence scale still remains to be clarified. Due to the properties of the design that allowed the necessary ambiguity in German relative clauses, stereotypically male role nouns were not examined in terms of prominence. Therefore, the position of stereotypically male role nouns is left undefined on the hierarchy in German language, while stereotypically male role nouns outrank neutral ones in French and neutral role nouns seem to outrank stereotypically female role nouns in prominence in both German and French.

Another aspect that needs to be specified regarding gender information as a prominence feature concerns its generalizability. Even though grammatical gender effects appear in sentences with both subject- and object-extracted relative clauses in German, grammatical gender affects the assignment of agent but not patient thematic roles. The design of sentences with backward anaphors in French, on the other hand, allowed for the gender hierarchy regarding the patient but not agent thematic role. Taken together, these effects point at the same hierarchical organization of grammatical gender information in terms of its prominence. However, the direct evidence for grammatical gender to influence the interpretation of agents in French and patients in German sentences is yet to complete the existing results. Moreover, the prominence hierarchy of grammatical gender information observed on role nouns may not be applicable to inanimate or non-human entities. Based on the current evidence, gender hierarchy may be assumed to relate to animacy or even function as its subscale, which brings up further questions about the interaction of gender with other prominence scales.

Gender Prominence as a Bias

Furthermore, the interdisciplinary nature of the research approach with its potential advantages needs to be considered when evaluating the results of the two studies. Consistent with previous research on gender processing, these studies suggest that both grammatical and stereotypical gender information is used during language processing (e.g., Carreiras et al., 1996; Stahlberg et al., 2007; Reali et al., 2015). At the same time, they go beyond previous research in that they show rather implicit ways in which gender information may influence processing even when it is not explicitly required by the rules of grammatical agreement or in order to resolve the reference. The tendency to associate female/feminine entities with less prominent thematic/syntactic roles and neutral/masculine ones with more prominent roles observed during language comprehension in the two studies can be related to gender hierarchies reported in social psychological research (e.g., Koenig et al., 2011) and may provide an insight on mechanisms underlying gender stereotypes. The social cognitive research on biases has shown that the use of some linguistic structures activates stereotypes, or cognitive

expectations and beliefs about a given group of persons. These structures are used preferentially when describing situations that are consistent or inconsistent with a stereotype and thus represent mechanisms that allow stereotypes to be reflected and maintained through language. For instance, Beukeboom et al. (2010) analyzed the use of negations and reported what they call a negation bias—a tendency to use negations when describing behaviors inconsistent with existing stereotypes, such as in saying *not stupid* rather than *smart* when describing a blond girl solving a math problem. Similarly, the use of concrete vs. abstract terms to reflect to which extent behavior was expected or unexpected has been reported as a linguistic intergroup bias (Maass et al., 1989) and an expectancy bias (Wigboldus et al., 2000): adjectives denoting higher levels of abstraction (e.g., *emotional*) tend to be used to encode expected behaviors (e.g., crying women), while unexpected behavior (e.g., crying men) is encoded by action verbs referring to specific events (e.g., *cry*). In the light of this research, the four experiments we described above suggest a bias that reflects and maintains stereotypes about men and women through the thematic structure of a sentence. This bias carries readers' expectations about gender stereotypes and corresponding social hierarchies (e.g., status, power, agency) over the hierarchy in the thematic structure, so that female/feminine nouns are assigned less prominent (i.e., lower-ranked on a hierarchy and rather passive) patient roles while neutral/masculine nouns are assigned more prominent agent roles. The tendency to perceive nouns possessing certain gender characteristics in one or another thematic role cannot be explained by formal linguistic rules, such as gender agreement, and therefore can be regarded as an implicit gender bias. Previous research has recognized that some information types do bias the assignment of thematic roles (e.g., tendency for animate entities to rather function as agents and inanimate entities as patients) and established them as prominence features. Theoretical argumentation and empirical evidence provided above clearly indicate that gender information can also be conceptualized in terms of prominence hierarchies, even though more extensive research is needed to overcome mentioned limitations and establish gender as a prominence feature.

References

Aissen, J. (2003). Differential object marking: iconicity vs. economy. *Nat. Lang. Linguist. Theory* 21, 435–483. doi: 10.1023/A:1024109008573

Bakan, D. (1966). *The Duality of Human Existence: Isolation and Communion in Western Man*. Chicago: Rand McNally.

Bates, E., Devescovi, A., Hernandez, A., and Pizzamiglio, L. (1996). Gender priming in Italian. *Percept. Psychophys.* 58, 992–1004. doi: 10.3758/BF03206827

Baudino, C. (2001). *Politique de la langue et différence sexuelle: La politisation du genre des noms de métier* [Language Policies and Sexual Differences: The Political Awareness of the Gender of Professional Occupations]. Paris: L'Harmattan.

Beukeboom, C. J., Finkenauer, C., and Wigboldus, D. H. J. (2010). The negation bias: when negations signal stereotypic expectancies. *J. Pers. Soc. Psychol.* 99, 978–992. doi: 10.1037/a0020861

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Lang. Linguist. Compass* 3, 19–58. doi: 10.1111/j.1749-818X.2008.00099.x

Conclusion

This paper proposed theoretical reasoning for grammatical and stereotypical gender of role nouns to be considered a prominence feature and discussed to which extent it is supported by recent empirical evidence from studies in German and French (Esaulova, 2015) languages. Conceptualizing gender as a prominence feature appears beneficial in several ways. First, it theoretically integrates findings on gender effects from different categories of language-based gender information. If gender is considered a supra-linguistic, basic semantic category and a prominence feature, then all linguistic expressions of gender, definitional: *king* vs. *queen*, typical: *soldier* vs. *nurse*, grammatical: *un étudiant* vs. *une étudiante*, should underlie the proposed hierarchical structure and be easier to process when aligned with other prominence features and their respective hierarchical ranks than when unaligned. Secondly, the approach offers an interdisciplinary analysis of gender effects that reflect hierarchical structures in as diverse fields as linguistics and social cognition. Thirdly, the approach allows for new predictions of subtle and implicit gender biases that go far beyond the classic mismatch effects. There is first empirical, cross-linguistic evidence for such biases as reported above.

Taken together, the findings suggest that gender information modulates the accessibility of thematic roles and grammatical functions and thus produces effects similar to those that were previously observed for prominence features, such as animacy. In order to validate the notion of gender as a prominence feature future studies should address questions left open such as the place of stereotypically male entities on a gender hierarchy in relation to stereotypically female and neutral ones, the generalizability of gender hierarchies to other linguistic structures and languages, and the interaction of gender information with other prominence features.

Acknowledgment

Cacciari, C., Corradini, P., Padovani, R., and Carreiras, M. (2011). Pronoun resolution in Italian: the role of grammatical gender and context. *J. Cogn. Psychol.* 23, 416–434. doi: 10.1080/20445911.2011.526599

Cacciari, C., and Padovani, R. (2007). Further evidence on gender stereotype priming in language: semantic facilitation and inhibition on Italian role nouns. *Appl. Psycholinguist.* 28, 277–293. doi: 10.1017/S0142716407070142

Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol.* 49A, 639–663. doi: 10.1080/713755647

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Cross, S. E., and Madson, L. (1997). Models of the self: self-construals and gender. *Psychol. Bull.* 122, 5–37. doi: 10.1037/0033-2909.122.1.5

de Hoop, H., and Lamers, M. J. A. (2006). "Incremental distinguishability of subject and object," in *Case, Valency, and Transitivity*, eds L. Kulikov, A. Malchukov, and P. de Swart (Amsterdam: John Benjamins), 269–287.

de Villiers, J. G., and Flusberg, H. B. (1975). Some facts one simply cannot deny. *J. Child Lang.* 2, 279–286. doi: 10.1017/S0305000900001100

Devescovi, A., D'Amico, S., Smith, S., Mimica, I., and Bates, E. (1998). The development of sentence comprehension in Italian and Serbo-Croatian: local versus distributed cues. *Syntax Semantics* 31, 345–377.

Duffy, S. A., and Keir, J. A. (2004). Violating stereotypes: eye-movements and comprehension processes when text conflicts with world knowledge. *Mem. Cogn.* 32, 551–559. doi: 10.3758/BF03195846

Esaulova, Y. (2015). *The Prominence of Gender Information in On-Line Language Processing: Cross-Linguistic Evidence of Implicit Gender Hierarchies.* Ph.D. thesis, University of Heidelberg, Heidelberg, Germany.

Esaulova, Y., Reali, C., and von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Lang. Cogn. Neurosci.* 29, 781–803. doi: 10.1080/01690965.2013.794295

Fiske, S. T. (1998). "Stereotyping, prejudice, and discrimination," in *Handbook of Social Psychology*, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York: McGraw-Hill), 357–411.

Friederici, A. D., and Weissenborn, J. (2007). Mapping sentence form onto meaning: the syntax-semantic interface. *Brain Res.* 1146, 50–58. doi: 10.1016/j.brainres.2006.08.038

Gabriel, U., Gygax, P., Sarassin, O., Garnham, A., and Oakhill, J. (2008). Au pairs are rarely male: norms on the gender perception of role names across English, French, and German. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206

Garnham, A., Oakhill, J., and Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Lang. Cogn. Process.* 7, 231–255. doi: 10.1080/01690969208409386

Gennari, S. P., and MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *J. Mem. Lang.* 58, 161–187. doi: 10.1016/j.jml.2007.07.004

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456. doi: 10.1007/s10936-007-9053-z

Kail, M. (1989). "Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish," in *The Crosslinguistic Study of Sentence Processing*, eds B. MacWhinney and E. Bates (New York: Cambridge University Press), 77–117.

Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

King, J. W., and Just, M. A. (1991). Individual differences in syntactic parsing: the role of working memory. *J. Mem. Lang.* 30, 580–602. doi: 10.1016/0749-596X(91)90027-H

Koenig, A. M., Mitchell, A. A., Eagly, A. H., and Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychol. Bull.* 137, 616–642. doi: 10.1037/a0023557

Kreiner, H., Sturt, P., and Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007.09.003

Lamers, M. J. A. (2012). "Argument linearization in dutch: a multi-factorial approach," in *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension*, eds M. J. A. Lamers and P. de Swart (Dordrecht: Springer), 121–144.

Maass, A., Salvi, D., Arcuri, L., and Semin, G. (1989). Language use in intergroup contexts: the linguistic intergroup bias. *J. Pers. Soc. Psychol.* 57, 981–993. doi: 10.1037/0022-3514.57.6.981

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Lang. Cogn. Process.* 9, 157–201. doi: 10.1080/01690969408402115

MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *J. Verbal Learn. Verbal Behav.* 23, 127–150. doi: 10.1016/S0022-5371(84)90093-8

McDonald, J. L., Bock, J. K., and Kelly, M. H. (1993). Word and world order: semantic, phonological and metrical determinants of serial position. *Cogn. Psychol.* 25, 188–230. doi: 10.1006/cogp.1993.1005

McDonald, J. L., and MacWhinney, B. J. (1995). The time course of anaphor resolution: effects of implicit verb causality and gender. *J. Mem. Lang.* 34, 543–566. doi: 10.1006/jmla.1995.1025

Prat-Sala, M. (1997). *The Production of Different Word Orders: A Psycholinguistic and Developmental Approach.* Ph.D. thesis, University of Edinburgh, Edinburgh.

Reali, C., Esaulova, Y., and von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: evidence from eye movements. *Appl. Psycholinguist.* 36, 977–1006. doi: 10.1017/S0142716414000010

Ridgeway, C. L. (2001). Gender, status and leadership. *J. Soc. Issues* 4, 637–655. doi: 10.1111/0022-4537.00233

Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE* 7:e48712. doi: 10.1371/journal.pone.0048712

Spence, J. T., and Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. *Psychol. Women Q.* 24, 44–62. doi: 10.1111/j.1471-6402.2000.tb01021.x

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language," in *Social Communication*, ed. K. Fiedler (New York: Psychology Press), 163–187.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3

Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836

Traxler, M. J., Williams, R. S., Blozis, S. A., and Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *J. Mem. Lang.* 53, 204–224. doi: 10.1016/j.jml.2005.02.010

Van Nice, K. Y., and Dietrich, R. (2003). Task sensitivity of animacy effects: evidence from German picture descriptions. *Linguistics* 41, 825–849. doi: 10.1515/ling.2003.027

Wagner, D. G., and Berger, J. (1997). Gender and interpersonal task behaviors: status expectation accounts. *Soc. Perspect.* 40, 1–32. doi: 10.2307/1389491

Wigboldus, D. H. J., Semin, G. R., and Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *J. Pers. Soc. Psychol.* 28, 5–18. doi: 10.1037/0022-3514.78.1.5

Williams, J. E., and Best, D. L. (1990). *Sex and Psyche: Gender and Self Viewed Cross-culturally*. Newbury Park, CA: Sage.

Yamamoto, M. (1991). *Animacy and Reference: A Cognitive Approach to Corpus Linguistics*. Amsterdam: John Benjamins.

Can Gender-Fair Language Reduce Gender Stereotyping and Discrimination?

Sabine Sczesny *, Magda Formanowicz and Franziska Moser

Department of Psychology, University of Bern, Bern, Switzerland

Gender-fair language (GFL) aims at reducing gender stereotyping and discrimination. Two principle strategies have been employed to make languages gender-fair and to treat women and men symmetrically: neutralization and feminization. Neutralization is achieved, for example, by replacing male-masculine forms (*policeman*) with gender-unmarked forms (*police officer*), whereas feminization relies on the use of feminine forms to make female referents visible (i.e., *the applicant... he or she* instead of *the applicant... he*). By integrating research on (1) language structures, (2) language policies, and (3) individual language behavior, we provide a critical review of how GFL contributes to the reduction of gender stereotyping and discrimination. Our review provides a basis for future research and for scientifically based policy-making.

Keywords: gender stereotypes, gender-fair language, social discrimination, gender equality, social change

Linguistic gender asymmetries are ubiquitous, as documented in the contributions in Hellinger and Bußmann (2001, 2002, 2003), which analyze 30 languages (e.g., Arabic, Chinese, English, Finnish, Hindi, Turkish, Swahili) from various language families. An almost universal and fundamental asymmetry lies in the use of *masculine generics.* In English, for example, generic *he* can be used when gender is irrelevant (e.g., *the user... he*) and in German, masculine role nouns serve as labels for mixed gender groups (e.g., *einige Lehrer*, masc.pl 'several teachers' for a group of male and female teachers). Thus, masculine forms not only designate men but also mixed-gender groups or referents whose gender is unknown or unspecified (see Stahlberg et al., 2007). Feminine forms, on the other hand, do not function generically but refer to women only (Hellinger and Bußmann, 2001).

That masculine forms are used to represent all human beings is in accord with the traditional gender hierarchy, which grants men more power and higher social status than women (Ridgeway and Correll, 2004). A large-scale content analysis of 800,000 Reuters news messages (published in English between 1996 and 1997) found that the pronoun *he* was more frequent than *she* in the news and also appeared in more positive contexts (Gustafsson Sendén et al., 2014). The interrelation of language and the gender hierarchy has also been documented in a study which analyzed the ratio of male to female pronouns (e.g., *he/she*, *his/hers*) in written texts (full texts of about 1.2 million U.S. books, years 1900–2008; from the Google Books database; Twenge et al., 2012). This ratio was found to reflect the status of women in the United States during the 20th century. When women's status was high (as indicated by educational attainment, labor force participation, etc.), the proportion of female pronouns was higher; when women's status was low, female pronouns were less frequent.

Gender-fair language (GFL)[1] was introduced as a response to this structural asymmetry and as part of a broader attempt to reduce stereotyping and discrimination in language (see Fairclough, 2003; Maass et al., 2013, for the political correctness debate). GFL aims to abolish asymmetries in referring to and addressing women and men, for example, by replacing masculine forms (*policeman*) with gender-unmarked forms (*police officer*), or by using both masculine and feminine forms (i.e., *the applicant... he or she* instead of *the applicant... he*).

In this paper, we review theoretical and empirical work on the role of GFL in sustaining or reducing gender stereotyping and social discrimination, as a follow-up on a comprehensive research program (the *Marie Curie Initial Training Network - Language, Cognition, and Gender, ITN LCG*, http://www.itn-lcg.psy.unibe.ch/content/index_eng.html). In this framework, we survey research on (1) language structures, (2) language policies, and (3) individual language behavior in order to draw conclusions on the effectiveness of GFL and to identify boundary conditions and obstacles for its implementation. Our aim is to critically discuss and integrate research findings to answer the question of whether and under what circumstances GFL contributes to the reduction of gender stereotyping and discrimination. Hopefully, this review will provide a useful basis for future research and for scientifically based policy-making.

LANGUAGE STRUCTURES

Although gender asymmetries exist in most, if not all, languages, they may be more or less conspicuous, depending on the structure of the language. Three types of languages can be distinguished: grammatical gender languages, natural gender languages, and genderless languages (see Stahlberg et al., 2007). **Table 1** gives an overview of this typology, describing the main characteristics of the different types with regard to gender and gender asymmetries as well as preferred strategies of linguistic gender-fairness. German, French, and Czech, for example, are *grammatical gender languages*. In these languages, every noun has a grammatical gender and the gender of personal nouns tends to express the gender of the referent. In *natural gender languages* (English or Swedish)[2] personal nouns tend to be gender-neutral (e.g., *neighbor*) and referential gender is expressed pronominally (e.g., *he/she*). In *genderless languages* such as Finnish or Turkish neither personal nouns nor pronouns signal gender. Here, gender is only expressed through attributes such as 'male/female [teacher]' or in lexical gender words such as 'woman' or 'father.' Consequently, gender and linguistic gender asymmetries are much more visible in grammatical gender languages than in

natural gender languages or genderless languages (Hellinger and Bußmann, 2001).

The way gender is encoded in a language may be associated with societal gender equality (Stahlberg et al., 2007). This assumption was tested empirically for 111 countries with different language systems, controlling for geographic, religious, political, and developmental differences (Prewitt-Freilino et al., 2012). In this research, the *Global Gender Gap Index* of the *World Economic Forum* was used to determine gender equality (GGI; Hausmann et al., 2009). Countries with grammatical gender languages were found to reach lower levels of social gender equality than countries with natural gender languages or genderless languages. This suggests that a higher visibility of gender asymmetries is accompanied by societal gender inequalities. A survey on sexist attitudes yielded additional evidence for this relationship (Wasserman and Weseley, 2009): respondents (native speakers of English as well as bilinguals) exhibited more sexist attitudes when the survey was conducted in a grammatical gender language (Spanish or French) than in a natural gender language (English). These findings document that, from the perspective of gender-fairness or gender equality, grammatical gender languages present a particularly complex and difficult case.

Research has consistently revealed that masculine generics evoke a *male bias in mental representations* and make readers or listeners think more of male than female exemplars of a person category (Stahlberg et al., 2007). Effects of linguistic forms on mental representations were measured with the help of various experimental methodologies, for instance, (1) completing sentences with different pronouns and nouns (e.g., *he, she, he/she, the lawyer, the client*; Jacobson and Insko, 1985), (2) writing stories about fictitious people following an introductory sentence in the masculine or in gender-fair wording (Heise, 2000), (3) naming female or male representatives (e.g., favorite musician) in response to either masculine nouns or combinations of feminine and masculine forms (Stahlberg et al., 2001), (4) estimating the proportion of women and men in certain roles (e.g., participants at a congress of nutritionists versus geophysicists; Braun et al., 1998), (5) measuring reading time as an indicator of fit between sentences about social groups denoted by nouns with different grammatical gender and sentences that contained a reference to the social group that qualified the group members as female, male, or neither one (Irmen and Roßberg, 2004), or (6) measuring reaction times when classifying gender-related (e.g., *she, he*) or neutral pronouns (e.g., *it, me*) as female or male after perceiving gender-related (e.g., *mother, father, nurse, doctor*) or gender-neutral primes (e.g., *parent, student*; Banaji and Hardin, 1996). The masculine bias in language has been observed in English (e.g., Crawford and English, 1984; Hamilton, 1988; Gastil, 1990; Ng, 1990), French (e.g., Chatard et al., 2005; Gabriel et al., 2008), German (e.g., Heise, 2000; Stahlberg et al., 2001; Braun et al., 2005; Irmen, 2007), Italian (e.g., Cacciari and Padovani, 2007), Polish (e.g., Bojarska, 2011), and Spanish (Carreiras et al., 1996). In a study with German and Belgian school children, the grammatical form of job titles was found to influence the children's perceptions of typically male jobs: when occupations were presented in the masculine (e.g., German

---

[1]In the literature, GFL is also referred to with other terms, e.g., *gender-neutral language*: Sarrasin et al. (2012); *gender-inclusive language*: Stout and Dasgupta (2011); *non-sexist language*: Douglas and Sutton (2014).

[2]According to McConnell-Ginet (2013), however, the concept of natural gender language is a myth, and she suggests calling the respective languages "notional" gender languages, since, for example in English, "concepts and ideas about biological sex matter at least as much as sex itself to the choice of English third-person pronouns." (p. 3).

TABLE 1 | Overview of language types regarding expression of gender and gender asymmetries.

Language type	Characteristics	Visibility of gender and gender asymmetries	Preferred strategies for gender-fair language
(1) Genderless (e.g., Finnish, Turkish)	• Neither personal nouns nor pronouns differentiated for gender (e.g., Turkish *öğrenci* 'student,' *o* 'she/he') • Gender expressed only lexically via attributes (e.g., 'male/female [student]') or lexical gender nouns (e.g., 'woman,' 'father')	• Referential gender often not explicit • (Lexical) gender asymmetries exist, but are less frequent than in (2) and (3) Examples: Turkish *adam* 'man' and 'human being' Finnish job titles ending in *-mies* '-man,' *lakimies* 'lawyer,' *lehtimies* 'journalist'	GFL policies generally deemed unnecessary
(2) Natural gender (e.g., English, Swedish)	• Most personal nouns gender-neutral (e.g., *neighbor, student*) • Personal pronouns differentiated for gender (e.g., Swedish *hon/han* 'she/he')	• Referential gender more often explicit than in (1), but less often than in (3) • Lexical and pronominal asymmetries exist, but are less frequent than in (3) Examples: English *chairman*, *the typical student . . . he*	Neutralization
(3) Grammatical gender (e.g., French, German)	• Every noun has grammatical gender • Gender of personal nouns tends to match gender of referent (e.g., German *Student*$_{masc}$/*Studentin*$_{fem}$ 'male/female student') • Personal pronouns differentiated for gender (e.g., German *sie/er* 'she/he') • Pronouns and other grammatically dependent words signal gender of personal noun (e.g., *der*$_{masc}$ *Student*$_{masc}$ 'the (male) student' *eine*$_{fem}$ *kluge*$_{fem}$ *Studentin*$_{fem}$ 'a clever (female) student')	• Referential gender often explicit • All kinds of asymmetries exist and are more frequent than in (1) and (2) Examples: French *homme* 'man' and 'human being' German *der*$_{masc}$ *typische Student*$_{masc}$ . . . *er* 'the typical student (masc) . . . he' German *alle Wähler*$_{masc}$ 'all voters'	Feminization + Neutralization

Ingenieure, masc.pl 'engineers') the mental accessibility of female jobholders was lower than with feminine-masculine word pairs (e.g., *Ingenieurinnen und Ingenieure*, fem.pl and masc.pl '[female and male] engineers'; Vervecken et al., 2013). In another study, adult speakers as well envisaged more men in an occupation when job advertisements included more masculine than feminine forms (Gaucher et al., 2011). In all, both the range of methods as well as the number of languages for which the male bias of masculine generics has been documented attests to the validity of the finding.

In general, different strategies can be used to make language gender-fair and avoid detrimental effects of masculine generics: neutralization, feminization and a combination of the two. Which strategy is the appropriate one depends on the type of language concerned (grammatical gender language, natural gender language, or genderless language, Bußmann and Hellinger, 2003).

In the framework of *neutralization* gender-marked terms are replaced by gender-indefinite nouns (English *policeman* by *police officer*). In grammatical gender languages, gender-differentiated forms are replaced, for instance, by epicenes (i.e., forms with invariant grammatical gender which refer to female as well as male persons; e.g., German *Staatsoberhaupt*, neut. 'head of state' or *Fachkraft*, fem. 'expert' in German). Neutralization has been recommended especially for natural gender languages (e.g., Hellinger and Bußmann, 2003; for English; Norwegian; Danish) and genderless languages (e.g., Engelberg, 2002, for Finnish), as it is fairly easy to avoid gender markings in these languages. Thus, neither generic *he* nor the combination *he/she*, but "singular *they* is the dominant epicene pronoun in modern

written British English. However, despite its use, singular *they* has never been endorsed by institutions of the English language, such as major dictionaries and style guides (although many style guides now reject generic *he...*)" (Paterson, 2014, p. 2). Recently, a gender-neutral third person pronoun was invented in Swedish: *hen*. This neologism first appeared in 2012 in a children's book where it served as an alternative to the gender-marked pronouns 'she' *(hon)* and 'he' *(han;* Gustafsson Sendén et al., 2015).

In contrast, *feminization* is based on the explicit inclusion of women. Thus, masculine generics are replaced by feminine-masculine word pairs (e.g., German *Elektrikerinnen und Elektriker* '[female and male] electricians'; Polish *nauczycielki i nauczyciele* '[female and male] teachers') or abbreviated forms with slashes (e.g., German *Elektriker/in*; Polish *nauczyciel/ka*) or brackets (e.g., *Elektriker[in]*; *nauczyciel[ka]*). Feminization has been recommended for grammatical gender languages such as German, Spanish, Czech, and Italian (Hellinger and Bußmann, 2003; Moser et al., 2011), usually in combination with neutralizing in order to avoid overly complex sentence structures.

However, feminization is not always advantageous for women. The Italian feminine suffix *-essa*, for example, has a slightly derogatory connotation (e.g., Marcato and Thüne, 2002). Accordingly, a woman introduced as *professoressa* 'female professor' was perceived as less persuasive than a man or than a woman referred to with the masculine form *professore* (Mucchi-Faina, 2005). Masculine terms used in reference to a female jobholder were associated with higher status than feminine job titles with *-essa* (Merkel et al., 2012). Another

example is the German (originally French) suffix-*euse* or -*öse*. Feminine terms such as *Masseuse* '(female) masseur' and *Frisöse* '(female) hair dresser' evoke sexual or frivolous associations, so that the neutral suffix -*in* is usually preferred, as in *Ingenieur-in* 'female engineer,' or *Spediteur-in* 'female forwarding agent.' Especially in Slavic languages feminine job titles tend to be associated with lesser status, with rural speech, or with the meaning 'wife of...' rather than 'female job holder' (for Russian: Doleschal and Schmid, 2001; for Serbian: Hentschel, 2003; for Polish: Koniuszaniec and Blaszkowa, 2003). There are also asymmetries in meaning between feminine and masculine forms, as with Polish *sekretarka* 'female secretary,' which designates a personal assistant, whereas the masculine *sekretarz* refers also to a high governmental function. In Polish, the feminine suffix -*ka* not only derives feminine occupational terms (such as *nauczyciel-ka* 'female teacher' from masculine *nauczyciel* 'teacher') but also words for inanimate objects such as *marynar-ka* 'jacket' from masculine *marynarz* 'sailor.' Problems of this kind can limit the possibilities of feminization in some languages. Where feminization faces such structural problems, its use is less widespread and may have negative effects (Italian: Mucchi-Faina, 2005; Polish: Formanowicz et al., 2013, 2015). But where feminine suffixes are productive feminization can became a linguistic norm and can be evaluated positively (German: Vervecken and Hannover, 2012).

The focus of early research on GFL was mostly on the masculine bias associated with masculine generics. But although these findings suggest that linguistic asymmetries may have farther-reaching consequences, this line of research has made no further progress until recently. The latest findings are more comprehensive and indicate how linguistic asymmetries may facilitate *(unintended) forms of social discrimination* (Mucchi-Faina, 2005; Stahlberg et al., 2007). For example, adult women were reluctant to apply to gender-biased job advertisements (e.g., English job titles ending in -*man*) and were more interested in the same job when the advertisement had an unbiased form (Bem and Bem, 1973). Also, the likelihood of naming women as possible candidates for the office of chancellor in Germany was found to depend on the grammatical gender of the word 'chancellor' in the question (Stahlberg and Sczesny, 2001). When the masculine generic (*Kanzler*) was used, fewer respondents suggested female politicians compared to a combination of masculine and feminine form (*Kanzler oder Kanzlerin* '[male or female] chancellor'). Moreover, self-evaluation and evaluations by others were found to be influenced by linguistic forms. Thus, girls assumed women to be less successful in typically male occupations when the jobs were described with masculine rather than gender-fair forms, and they were also less interested in these occupations (see also Chatard et al., 2005; Vervecken et al., 2013). Using feminine-masculine word pairs rather than masculine forms for traditionally male occupations boosted children's self-efficacy (Vervecken and Hannover, 2015). Furthermore, occupations described in pair forms mitigated the difference between ascribed success to female and male jobholders in gendered occupations (Vervecken et al., 2015). Also, women's perceptions of belonging

were found to mediate the effect that women found jobs advertised in the masculine less appealing (Gaucher et al., 2011). Accordingly, women experienced the use of gender-exclusive language during a mock job interview as ostracism (Stout and Dasgupta, 2011). They reported a lower sense of belonging when gender-exclusive language (*he*) was used compared to gender-inclusive (*he or she*) or gender-neutral (*one*) forms. In a study on Austrian German, the wording of job advertisements influenced the evaluation of candidates for leadership positions (Horvath and Sczesny, 2015): men were perceived as fitting a high-status leadership position better than women when a masculine job title was used (*Geschäftsführer*, masc. 'chief executive officer, CEO'). But when the job ad was gender-fair (*Geschäftsführerin/Geschäftsführer*, fem./masc. '[female/male] CEO'), women and men were judged as equally suited. In the context of a lower-status position (project leader) no differences of this kind occurred.

LANGUAGE POLICIES

Many countries have pledged themselves to an equal treatment of women and men (e.g., the member states of the European Union and associated states in the Treaty of Lisbon- European Commission, 2007), and the use of GFL is widely recommended (Schweizerische Bundeskanzlei, 1996, revised in 2009; UNESCO, 1999; National Council of Teachers of English, 2002; European Commission, 2008; American Psychological Association, 2009). But the implementation of GFL has reached different stages in different countries and speech communities.

In the 1970s, guidelines for GFL were introduced in particular professional domains across national and linguistic boundaries, for example, by the American Psychological Association (1975), by the McGraw-Hill Book Company (1974; see also Britton and Lumpkin, 1977; Sunderland, 2011) and the Macmillan Publishing Company (1975). These guidelines demand that authors of (psychological) articles, books, teaching materials, or fiction treat women and men equally, including the language they use (see also Sadker et al., 1991). Publication guidelines of this kind have been effective, because authors need to follow the rules if they want to see their manuscripts published. In texts written by Australian academics (Pauwels, 2003), for example, masculine generic pronouns were infrequent. Similarly, an analysis of American Psychological Association journal articles from the years 1965–2004 revealed a complete absence of generic *he* from 1985 onward, even if the articles still contained other, more subtle gender biases such as androcentric reporting in tables and graphs (Hegarty and Buechel, 2006).

In 1987 representatives of Canada and the Nordic countries argued for an adoption of GFL by the *United Nations Educational, Scientific and Cultural Organization*. This resulted in the creation of guidelines in UNESCO (1999). UNESCO's position in favor of GFL is described in their gender equality guidelines: "This development indicated a growing awareness that language does not merely reflect the way we think: it also shapes our thinking. If words and expressions that imply that women are inferior to

men are constantly used, that assumption of inferiority tends to become part of our mindset; hence the need to adjust our language when our ideas evolve" (UNESCO, 2011, p. 4). The document not only became the most widely recognized international standard for GFL, it also regulates language use in internal documents and publications of UNESCO. Similar guidelines for publications were issued by the European Commission (2008), referring to all working languages of the European Union (EU). Yet, the standards promoted by UNESCO and the EU do not regulate language use in the different countries and are not considered mandatory within their member states.

The availability of GFL policies and the extent of their implementation, that is, their dissemination and execution, also vary considerably between countries (Moser et al., 2011). In Italy, for instance, guidelines for GFL were issued in Sabatini (1987), in the German-speaking area most guidelines appeared in the 1990s (e.g., Hellinger and Bierbach, 1993; Schweizerische Bundeskanzlei, 1996; revised in 2009), and in the Czech Republic guidelines were published only in Valdrová et al. (2010). In other countries such as Poland there are as yet no official guidelines at all. While in some states GFL policies are mentioned only on the website of a ministry (e.g., Czech Republic; Valdrová et al., 2010), use of GFL is mandatory in job ads and public administration in Austria. Since the 1990s the German Duden dictionaries, for example, have included not only the masculine form of personal nouns and job titles but routinely cite the corresponding feminine forms (Kunkel-Razum, 2004). The dictionary lists even feminine forms that are infrequent in texts. An example is the word *Päpstin* 'female pope,' which has been listed in the *Grosses Wörterbuch der deutschen Sprache* (Large dictionary of the German language) from the year 1999 onward, even though obviously there never was a female pope in the history of the Catholic Church (Kunkel-Razum, 2004). Moreover, the Duden editors decided to include a chapter on the "equal treatment of women and men in language" in the ninth volume of the series *Richtiges und gutes Deutsch* (Correct and good German). The chapter describes the linguistic potential which the German language offers for speaking or writing in a gender-fair way.

In the German-speaking countries, language policies have become part of the organizational culture of various institutions such as universities and administrations (e.g., Schweizerische Bundeskanzlei, 1996, revised in 2009; Merkel, 2011; Swiss Federal Institute of Technology Zurich, 2011; Gendup – Zentrum für Gender Studies und Frauenförderung, 2012). Even so, Austria is the only country where the use of GFL in job advertisements is strictly prescribed and companies are fined for failing to address both genders in their job ads (Bundesministerium für Frauen und Öffentlichen Dienst, 2009). This may be the reason why the proportion of job ads worded in GFL differs between Austria and German-speaking Switzerland: only 9% of Austrian job advertisements contain masculine generics, whereas it is 27% in Switzerland (Hodel et al., 2013).

School and education are of particular importance for the implementation of GFL. In most countries there are few official GFL guidelines for authors of educational materials (Eurydice, 2009) and regulations concerning schoolbooks exist only in certain countries (e.g., Germany, Ireland, or Iceland). Similarly,

only a few countries require schoolbooks to be officially evaluated or approved. In the UK, for example, educational authorities do not monitor teaching materials and schools choose them autonomously. Today German schoolbooks for mathematics and German mostly use gender-neutral forms, followed by masculine generics and feminine-masculine word pairs, (Moser and Hannover, 2014). The two gender-fair options together (word pairs and neutralizing) outweighed the masculine in the schoolbook sample that was analyzed. Since earlier studies on German schoolbooks (e.g., Lindner and Lukesch, 1994; Preinsberger and Weisskircher, 1997) reported a predominance of masculine generics, this finding indicates an increase of GFL in schoolbooks. In some of the texts, however, feminine-masculine word pairs were mixed with masculine generics (see also Markom and Weinhäupl, 2007). This inconsistency is problematic because in the presence of word pairs masculine forms may be understood as referring to male persons only (e.g., Gabriel et al., 2008).

INDIVIDUAL LANGUAGE BEHAVIOR

Apart from language structures and country-specific aspects, there are a number of factors that make individuals use or reject GFL. One major factor is the novelty of gender-fair forms, which conflicts with speakers' linguistic habits (Blaubergs, 1980). As long as this is the case, people may experience GFL as irritating, and consequentially may refrain from using it. This could explain why negative effects of GFL have been found especially in the initial phases of language reform such as, for instance, in English in the 1990s (McConnell and Fazio, 1996), and in Italian and Polish in the beginning of the 21st century (Mucchi-Faina, 2005; Merkel et al., 2012; Formanowicz et al., 2013).

Moreover, initiatives for GFL were first instigated by activist movements (e.g., Silveira, 1980; Pusch, 1984) and for that reason often met with negative reactions (Blaubergs, 1980; Parks and Roberton, 1998; Formanowicz et al., 2013). It is conceivable that individual reactions toward GFL are not only caused by its novelty, but also depend on attitudes toward gender arrangements (Jost and Kay, 2005; Carney et al., 2008), for conservative political attitudes are associated both with lesser openness for novelty (Carney et al., 2008) and with stronger support for traditional gender arrangements (Jost et al., 2003, 2008; Hoyt, 2012). Thus, speakers of Polish with more conservative attitudes devaluated female job applicants referring to themselves with a feminine job title compared to female and male applicants using a masculine job title (Formanowicz et al., 2013).

Another factor for individual speakers' use of GFL might be speakers' gender: women could be expected to hold more favorable attitudes toward GFL than men and they might be more inclined to use it in their own speech. However, research findings on this point are mixed. While in some studies men rejected GFL more than women did (e.g., Parks and Roberton, 2004; Douglas and Sutton, 2014), other studies found no gender difference in attitudes toward GFL (e.g., Sczesny et al., 2015). Gender differences were mediated by participants' attitudes toward

women, which were, in turn, driven by more comprehensive ideologies that justified the social gender hierarchy (i.e., gender-specific system justification and social dominance orientation; Douglas and Sutton, 2014).

Language use has been viewed as associated with speakers' *sexist attitudes*, so much so that the use of sexist language has been regarded as an example of subtle sexism (Swim et al., 2004). Modern sexism, for instance, is a view that denies that women are still discriminated against and disapproves of policies promoting gender equality (Swim et al., 1995). In fact, participants with modern sexist beliefs were found to use more traditional, gender-unfair language (Swim et al., 2004). Correspondingly, speakers with stronger sexist attitudes toward women used gender-fair pronouns less frequently than speakers with less sexist attitudes (Jacobson and Insko, 1985). Speakers with progressive gender role perceptions, on the other hand, exhibited a tendency to avoid sexist language when writing an essay (McMinn et al., 1991).

This raises the question how sexist or non-sexist ideologies translate into actual language behavior. Spontaneous use of GFL was found to be guided by explicit intentions to use GFL as well as more implicit processes involving use of GFL in the past (Sczesny et al., 2015). GFL use was not predicted directly by sexist beliefs but by intentions and habits. In other words, sexist speakers do not avoid GFL just because they are reluctant to change their linguistic habits, they deliberately employ a form of language that treats males as the norm and makes women less visible. Habits guide speakers' linguistic behavior without their being aware of it (Sczesny et al., 2015), and learning processes play a role for GFL to become a habit. Speakers who grew up with schoolbooks using predominantly masculine generics (e.g., English: Hellinger, 1980; Campbell and Schram, 1995; Lee and Collins, 2008; German: Lindner and Lukesch, 1994; Preinsberger and Weisskircher, 1997) tend not to question this usage. But once speakers have acquired the habit of using GFL they will rely on this language form. Establishing GFL habits via teaching and practicing current linguistic standards (e.g., Duden; Kunkel-Razum, 2004) is a promising approach which should follow the initial phase of GFL implementation and may reduce political controversies. In this sense, a prevalence of GFL in the media could also promote the use of GFL by individual speakers.

So far, few studies have investigated how speakers can be made to use and approve of GFL. After training interventions, speakers of English used slightly more gender-fair pronouns in completing sentences than non-attendants (McMinn and Foster, 1991; McMinn et al., 1991; Prentice, 1994). Their attitudes, however, did not change (Prentice, 1994). German speakers as well used more GFL after being exposed to arguments for GFL than in a control condition (Koeser and Sczesny, 2014), but this did not affect their attitudes toward GFL. Interestingly, merely reading texts in gender-fair wording can also increase speakers' own use of GFL: female speakers of German employed more gender-fair forms after reading a gender-fair text than after other texts, but there was no such effect for men (Koeser et al., 2015). Male speakers increased their use of gender-fair forms only when their attention was drawn to GFL forms. These findings indicate that it is more difficult to change attitudes than to promote speakers' actual use of GFL.

OVERCOMING GENDER STEREOTYPING AND DISCRIMINATION WITH GENDER-FAIR LANGUAGE?

Over the past decades, a large body of research—based on various experimental methodologies, from storytelling to measuring reaction times—has confirmed the influence of linguistic forms on the accessibility of mental representations of women and men (see Stahlberg et al., 2007). Regardless of language structure and of the ease of implementing GFL (Bußmann and Hellinger, 2003), a consistent finding is that speakers do not understand masculine forms as referring to both genders equally but that they interpret them in a male-biased way. This underscores the importance of implementing GFL in everyday language and of using it consistently, so that speakers take up this usage in their own texts and utterances.

How successful have the respective language policies been so far? In *natural gender languages*, neutralization has been fairly easy to adopt and implement (e.g., English, Danish). But even in these language communities people are guided by their knowledge about typical gender distributions in social roles. Thus, English readers tend to associate different occupations or role nouns with men or women, since gender stereotypes are incorporated in their mental representations (Oakhill et al., 2005); and even though there are fewer gender-marked forms in natural gender languages, masculine generics exist and their use can result in social discrimination (Stout and Dasgupta, 2011). In *grammatical gender languages*, feminization as the main strategy of GFL still poses challenges. This is especially true for some languages, e.g., Italian (Merkel et al., 2012) and Slavic languages (Koniuszaniec and Blaszkowa, 2003), where the creation of feminine forms can be problematic, as outlined above. Refusal of GFL can still be observed (Formanowicz and Sczesny, 2014). Such disadvantages are likely to occur while the change is in progress (Formanowicz et al., 2015).

Moreover, our review suggests that—independent of language structure—GFL is more frequent and more accepted when it is backed by official regulations and when the use of biased language is sanctioned in some way (e.g., in official publications or texts; American Psychological Association, 1975, 2009; Bundesministerium für Frauen und Öffentlichen Dienst, 2009; see Hodel et al., 2013). The relationship between policy-making and social change is surely bidirectional. On the one hand, gender equality movements and their demands find their way into legislation. On the other hand, official regulations may stipulate social change by facilitating the internalization of new norms and enforcing their execution. Public discussions over policies also enhance public awareness for GFL (see above the singular pronouns *they* in English and *hen* in Swedish). The contribution of language reforms to gender equality in a society/speech community can best be assessed with investigations that compare countries sharing the same language (e.g., French in Canada and in France) as well as countries with different languages

(e.g., Polish and German, two grammatical languages at different stages of implementing GFL). Although there have been some attempts at this type of research (Formanowicz et al., 2015; Gustafsson Sendén et al., 2015) more research is needed to evaluate the effectiveness of language-related policies and provide an evidence-based rationale for policy-making.

As mentioned above, speakers' use of GFL results from deliberate processes, involving attitudes and intentions, and habitual processes, involving repetition of past behavior (Sczesny et al., 2015). Both types of processes are relevant for the successful implementation of GFL. Despite the various guidelines and legal regulations for GFL that exist on global and national levels, spontaneous use of GFL by individual speakers still seems to be infrequent. For instance, use of GFL in a gap-filling task was quite low among speakers of German from Germany and Switzerland, although GFL policies are fairly advanced in both countries. Most of the participants used more masculine generics than gender-fair forms. As language use is an action performed in a wide range of circumstances, future research should also assess the contiguity between behavior and context. Speakers may employ GFL when writing official texts, for instance, but not when talking or writing to friends. Moreover, attitudes, norms, and intentions concerning GFL in general seem to be only moderately favorable. Even though positive arguments for GFL can help to promote a change in language behavior (Koeser and Sczesny, 2014), future research should attempt to identify factors that are crucial for a deliberate use of GFL. For instance, it might be worthwhile to determine the content and strength of attitudes in different groups of speakers, namely speakers who use GFL regularly compared to speakers who use GFL only occasionally and others who do not use it at all. To gain a more comprehensive understanding of the processes underlying a rejection of GFL, future research could also take a closer look at people's political attitudes (Formanowicz et al., 2013), their preference for status quo, and their acceptance of traditional gender arrangements (Jost et al., 2008).

In any case, attitudes toward GFL may become more favorable the more frequently and longer GFL has been used (in addition to a mere exposure effect, Zajonc, 1968, see also the existence bias: people treat the existence of something as evidence of its goodness; Eidelman et al., 2009). The role of familiarity for an active use of GFL can best addressed with longitudinal studies. In Sweden, for example, speakers' attitudes toward the gender-neutral pronoun *hen* have become more positive over time (Gustafsson Sendén et al., 2015). A meta-analytical approach would constitute another way of capturing the dynamics of GFL implementation, taking into account the time when the studies were conducted but also the availability of policies and the structure of the languages concerned. This approach might help to determine whether a language has left the phase where GFL evokes negative associations as well as the role of other factors (such as language policies).

Interventions aiming to increase the use of GFL could focus on a simple repetition of non-sexist expressions, so that these become established habits (Koeser et al., 2015; Wood and Rünger, 2016). This would be a very subtle and implicit way of promoting use of GFL. The development and evaluation of GFL interventions/trainings has not yet been investigated

systematically. Future research should take both deliberate and habitual processes of GFL use into consideration, for instance, by analyzing whether children—exposed to and trained in GFL at school (with the help of current schoolbooks)—will later use GFL habitually and consequently hold less gender-stereotypic beliefs.

Finally, there are still obstacles that prevent GFL from becoming a linguistic norm/standard and prevent the change toward an equal treatment of women and men. First, the male bias of linguistic asymmetries in mental representations is backed by a *higher prevalence of men in certain social roles* (e.g., heroes, politicians), which facilitates their cognitive accessibility (Stahlberg and Sczesny, 2001). Once women and men occupy all social roles to a similar extent (see social role theory, which poses that gender stereotype content results from observing women and men in certain societal roles; Eagly, 1987; Bosak et al., 2012), this difference in accessibility should decrease and more gender-balanced mental representations should emerge. Ironically, recent research has documented that linguistic asymmetries prevent girls and women from aspiring to male-dominated roles (see Chatard et al., 2005; Gaucher et al., 2011; Stout and Dasgupta, 2011; Vervecken et al., 2013; Vervecken and Hannover, 2015) and thereby perpetuate the higher accessibility of men in these roles.

Second, the use of gender-unfair language, especially of masculine generics, restricts the *visibility of women* and the *cognitive availability of female exemplars* (Stahlberg et al., 2007), which may be disadvantageous for women (e.g., in personnel selection; Stout and Dasgupta, 2011; Horvath and Sczesny, 2015). However, increasing the visibility of women with the help of novel feminine forms may also have negative consequences and may therefore be avoided, for instance, in women's professional self-reference (Merkel et al., 2012; Formanowicz et al., 2013). Thus, the avoidance of GFL by women (e.g., avoidance of feminine job titles in grammatical gender languages), in order to protect themselves from ascriptions of incompetence or lower status, also perpetuates the reduction of gender stereotyping and social discrimination.

Third, arguments against GFL have routinely included the presumed difficulty of understanding GFL texts (Parks and Roberton, 1998). Empirical investigations have refuted this argument and have shown that text quality (Rothmund and Christmann, 2002) and cognitive processing were not damaged (Braun et al., 2007). When GFL texts were compared to (generic) masculine texts, there were no differences in readability and esthetic appeal (Blake and Klimmt, 2010). In all, the empirical evidence does not confirm the alleged disadvantage of GFL. Yet, these findings and the scientific evidence for serious disadvantages of masculine generics (see above) have largely been ignored in political controversies and public discussions about GFL. In all, there is a lack of transfer of scientific knowledge which prevents the understanding of linguistic asymmetries as part of a broader gender imbalance and hinders social change. Education and policy-making therefore need to increase the efforts of circulating new scientific insights about GFL to break the vicious circle of ill-informed controversies and discussions about GFL.

At first glance linguistic gender asymmetries seem to affect mostly women. When masculine forms are used it is women

who are seen as less prototypical category exemplars, it is women who feel less adequate or are less preferred as job candidates, and it is women who profit from GFL. Therefore, the question arises whether GFL benefits men as well. First, the introduction of GFL might represent a particular challenge for men. In a study by Crawford and English (1984) both male and female participants read a text whose title contained either masculine generics (*Psychologist and his work?*) or GFL (*Psychologist and their work?*) and were to recall the text after 2 days. As the results showed, men's recall was better in the masculine and women's recall in the GFL condition. This finding indicates that learning to use GFL involves more than overcoming linguistic novelty. For men, GFL means an unwelcome loss of their privileged position in language. Only in few situations have they something to gain through GFL. If all job advertisements would contain GFL, for instance, men might be more included in traditionally female jobs which used to be referred to in the feminine. Future research should also consider the perspective of men and examine how GFL can turn into a win–win situation for women and men in modern societies.

To conclude, past research has revealed that GFL has the potential to make significant contributions to the reduction of gender stereotyping and discrimination. But as the body of existing evidence is based mainly on experimental paradigms with different kinds of measures, future research should take a closer look on people's actual language use in everyday life (e.g., in conversations, in the classroom, in social media or organizational communication). Moreover, it will be fruitful to further investigate the dynamics of GFL usage and its effects from cross-linguistic and cross-cultural perspectives (see above the *Marie Curie Initial Training Network - Language, Cognition, and Gender, ITN LCG,* which can be regarded as a first step in this direction). Speakers' willingness to use GFL in everyday life is crucial in order to profit from the impact of GFL on the (linguistic and social) treatment of women and men in society. But a deliberate effort is required before the use of GFL can become habitual. Education and policy-making can facilitate these processes. When employed consistently over a longer period, and especially when supported by well-informed controversies and discussions, GFL will contribute even more to the reduction of gender stereotyping and discrimination and may thus function as another barometer for change (like the decrease in gender-stereotypical social perception of leadership, Schein, 2001).

FUNDING

REFERENCES

American Psychological Association (1975). Guidelines for nonsexist use of language. *Am. Psychol.* 30, 682–684. doi: 10.1037/h0076869

American Psychological Association (2009). *Publication Manual of the American Psychological Association: Supplemental Material*, 6th Edn. Washington, DC: American Psychological Association.

Banaji, M. R., and Hardin, C. D. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x

Bem, S. L., and Bem, D. J. (1973). Does sex-biased job advertising "aid and abet" sex discrimination? *J. Appl. Soc. Psychol.* 3, 6–18. doi: 10.1111/j.1559-1816.1973.tb01290.x

Blake, C., and Klimmt, C. (2010). Geschlechtergerechte Formulierungen in Nachrichtentexten [Gender-fair formulations in news texts]. *Publizistik* 55, 289–304. doi: 10.1007/s11616-010-0093-2

Blaubergs, M. S. (1980). An analysis of classic arguments against changing sexist language. *Womens Stud. Int. Q.* 3, 135–147. doi: 10.1016/S0148-0685(80)92071-0

Bojarska, K. (2011). Wpływ androcentrycznych i inkluzywnych płciowo konstrukcji językowych na skojarzenia z płcią [The impact of the androcentric and gender-inclusive language constructions on the gendered asssociations]. *Stud. Psychol.* 49, 53–68. doi: 10.2478/v10167-011-0010-y

Bosak, J., Sczesny, S., and Eagly, A. H. (2012). The impact of social roles on trait judgments – a critical re-examination. *Pers. Soc. Psychol. Bull.* 38, 429–440. doi: 10.1177/0146167211427308

Braun, F., Gottburgsen, A., Sczesny, S., and Stahlberg, D. (1998). Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen [Can geophysicists be women? Generic terms for describing persons in German]. *Z. Ger. Linguist.* 26, 177–195. doi: 10.1515/zfgl.1998.26.3.265

Braun, F., Oelkers, S., Rogalski, K., Bosak, J., and Sczesny, S. (2007). "For reasons of intelligibility." How masculine generics and alternative forms affect the cognitive processing of a text. *Psychol. Rundsch.* 58, 183–189. doi: 10.1026/0033-3042.58.3.183

Braun, F., Sczesny, S., and Stahlberg, D. (2005). Cognitive effects of masculine generics in German: an overview of empirical findings. *Communications* 30, 1–21. doi: 10.1515/comm.2005.30.1.1

Britton, G. E., and Lumpkin, M. C. (1977). For sale: subliminal bias in textbooks. *Read. Teach.* 31, 40–45.

Bundesministerium für Frauen und Öffentlichen Dienst (2009). *Geschlechtergerechte Stellenausschreibungen [Gender-Fair Job Advertisements].* Unabhängiger Bericht der Gleichbehandlungsanwaltschaft iS §3 Abs 5 GBK/GAW-Gesetz, Wien.

Bußmann, H., and Hellinger, M. (2003). "Engendering female visibility in German," in *Gender Across Languages. The Linguistic Representation of Women and Men*, Vol. 3, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins Publishing Company), 141–173.

Cacciari, C., and Padovani, R. (2007). Further evidence on gender stereotype priming in language: semantic facilitation and inhibition on Italian role nouns. *Appl. Psycholinguist.* 28, 277–293. doi: 10.1017/S0142716407070142

Campbell, R., and Schram, P. J. (1995). Feminist research methods. A content analysis of psychology and social science textbooks. *Psychol. Women Q.* 19, 85–106. doi: 10.1111/j.1471-6402.1995.tb00280.x

Carney, D. R., Jost, J. T., Samuel, D., Gosling, S. D., and Potter, J. (2008). The secret lives of liberals and conservatives: personality profiles, interaction styles, and the things they leave behind. *Polit. Psychol.* 29, 807–840. doi: 10.1111/j.1467-9221.2008.00668.x

Carreiras, M., Garnham, A., Oakhill, J. V., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol. A* 49, 639–663. doi: 10.1080/027249896392531

Chatard, A., Guimond, S., and Martinot, D. (2005). Impact de la féminisation lexicale des professions sur l'auto-efficacité des élèves: une remise en

cause de l'universalisme masculin? *Année Psychol.* 105, 249–272. doi: 10.3406/psy.2005.29694

Crawford, M., and English, L. (1984). Generic versus specific inclusion of women in language: effects on recall. *J. Psycholinguist. Res.* 13, 373–381. doi: 10.1007/BF01068152

Doleschal, U., and Schmid, S. (2001). "Doing gender in Russian," in *Gender Across Languages. The Linguistic Representation of Women and Men*, Vol. 1, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins Publishing Company), 253–282.

Douglas, K. M., and Sutton, R. M. (2014). "A giant leap for mankind" but what about women? The role of system-justifying ideologies in predicting attitudes toward sexist language. *J. Lang. Soc. Psychol.* 33, 667–680. doi: 10.1177/0261927X14538638

Eagly, A. H. (1987). *Sex Differences in Social Behavior. A Social-Role Interpretation.* Hillsdale, NJ: Lawrence Erlbaum.

Eidelman, S., Crandall, C. S., and Pattershall, J. (2009). The existence bias. *J. Pers. Soc. Psychol.* 97, 765–775. doi: 10.1037/a0017058

Engelberg, M. (2002). "The communication of gender in Finnish," in *Gender Across Languages. The Linguistic Representation of Women and Men*, Vol. 2, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins Publishing Company), 109–132.

European Commission (2007). *Treaty of Lisbon.* Available at: http://ec.europa.eu/archives/lisbon_treaty/full_text/index_en.htm

European Commission (2008). *Gender-Neutral Language in the European Parliament.* Available at: http://www.europarl.europa.eu/RegData/publications/2009/0001/P6_PUB(2009)0001_EN.pdf

Eurydice (2009). *Gender Differences in Educational Outcomes: Study on the Measures Taken and the Current Situation in Europe.* Available at: http://eacea.ec.europa.eu/education/eurydice/thematicreportsen.php

Fairclough, N. (2003). Political correctness: the politics of culture and language. *Discourse Soc.* 14, 17–28. doi: 10.1177/0957926503014001927

Formanowicz, M., Bedynska, S., Cisłak, A., Braun, F., and Sczesny, S. (2013). Side effects of gender-fair language: how feminine job titles influence the evaluation of female applicants. *Eur. J. Soc. Psychol.* 43, 62–71. doi: 10.1002/ejsp.1924

Formanowicz, M., and Sczesny, S. (2014). Gender-Fair language and professional self-reference: the case of female psychologists in Polish. *J. Mix. Methods Res.* 10, 64–81. doi: 10.1177/1558689814550877

Formanowicz, M. M., Cisłak, A., Horvath, L. K., and Sczesny, S. (2015). Capturing socially motivated linguistic change: how the use of gender-fair language affects support for social initiatives in Austria and Poland. *Front. Psychol.* 6:1617. doi: 10.3389/fpsyg.2015.01617

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au-pairs are rarely male: role names' gender stereotype information across three languages. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206

Gastil, J. (1990). Generic pronouns and sexist language: the oxymoronic character of masculine generics. *Sex Roles* 23, 629–643. doi: 10.1007/BF0028925

Gaucher, D., Friesen, J., and Kay, A. C. (2011). Evidence that gendered wording in job advertisement exists and sustains gender inequality. *J. Pers. Soc. Psychol.* 101, 109–128. doi: 10.1037/a0022530

Gendup – Zentrum für Gender Studies und Frauenförderung (2012). *Leitfaden für einen gerechten Sprachgebrauch [Guideline for Gender-Fair Language].* Available at: http://www.uni-salzburg.at/fileadmin/oracle_file_imports/2103374.PDF

Gustafsson Sendén, M., Bäck, E. A., and Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Front. Psychol.* 6:893. doi: 10.3389/fpsyg.2015.00893

Gustafsson Sendén, M., Lindholm, T., and Sikström, S. (2014). Biases in news media as reflected by personal pronouns in evaluative contexts. *Soc. Psychol.* 45, 103–111. doi: 10.1027/1864-9335/a000165

Hamilton, M. C. (1988). Using masculine generics: does generic he increase male bias in the user's imagery? *Sex Roles* 19, 785–799. doi: 10.1007/bf00288993

Hausmann, R., Tyson, L. D., and Zahidi, S. (2009). *The Global Gender Gap Report 2009.* Geneva: World Economic Forum.

Hegarty, P., and Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965-2004. *Rev. Gen. Psychol.* 10, 377–389. doi: 10.1037/1089-2680.10.4.377

Heise, E. (2000). Sind Frauen mitgemeint? Eine empirische Untersuchung zum Verständnis des generischen Maskulinums und seiner Alternativen [Are women included? An empirical investigation of interpretations of masculine generics and their alternatives]. *Sprache Kogn.* 19, 3–13. doi: 10.1024//0253-4533.19.12.3

Hellinger, M. (1980). For men must work and women must weep: sexism in English language textbooks used in German schools. *Womens Stud. Int. Q.* 3, 267–275. doi: 10.1016/S0148-0685(80)92323-4

Hellinger, M., and Bierbach, C. (1993). *Eine Sprache für beide Geschlechter. Richtlinien für einen nicht-sexistischen Sprachgebrauch. [A Language for Both Genders. Guidelines for a Non-Sexist Language Use].* Available at: http://www.unesco.de/fileadmin/medien/Dokumente/Bibliothek/eine_sprache.pdf

Hellinger, M., and Bußmann, H. (2001, 2002, 2003). *Gender Across Languages: The Linguistic Representation of Women and Men*, Vol. 1, 2, 3. Amsterdam: Benjamins.

Hentschel, E. (2003). "Serbian: the expression of gender in Serbian," in *Gender Across Languages: The Linguistic Representation of Women and Men*, Vol. 3, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins Publishing Company), 287–309.

Hodel, L., Formanowicz, M., Sczesny, S., Valdrova, J., and von Stockhausen, L. (2013). "Gender fair language use in online job Avertisements," in *Proceedings of the 16th Conference of European Association of Work & Organisational Psychology (EAWOP)*, Münster.

Horvath, L. K., and Sczesny, S. (2015). Reducing women's lack of fit with leadership? Effects of the wording of job advertisements. *Eur. J. Work Organ. Psychol.* doi: 10.1080/1359432X.2015.1067611

Hoyt, C. L. (2012). Gender bias in employment contexts: a closer examination of the role incongruity principle. *J. Exp. Soc. Psychol.* 48, 86–96. doi: 10.1016/j.jesp.2011.08.004

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456. doi: 10.1007/s10936-007-9053-z

Irmen, L., and Roßberg, N. (2004). Gender markedness of language. The impact of grammatical and nonlinguistic information on the mental representation of person information. *J. Lang. Soc. Psychol.* 23, 272–307. doi: 10.1177/0261927X04266810

Jacobson, M. B., and Insko, W. R. Jr. (1985). Use of non-sexist pronouns as a function of one's feminist orientation. *Sex Roles* 13, 1–7. doi: 10.1007/BF00287456

Jost, J. T., Glaser, J., Kruglanski, A. W., and Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychol. Bull.* 129, 339–375. doi: 10.1037/0033-2909.129.3.33

Jost, J. T., and Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *J. Pers. Soc. Psychol.* 88, 498–509. doi: 10.1037/0022-3514.88.3.498

Jost, J. T., Nosek, B. A., and Gosling, S. D. (2008). Ideology: its resurgence in social, personality, and political psychology. *Perspect. Psychol. Sci.* 3, 126–136. doi: 10.1111/j.1745-6916.2008.00070.x

Koeser, S., Kuhn, E. A., and Sczesny, S. (2015). Just reading? How gender-fair language triggers readers' use of gender-fair forms. Advance online publication. *J. Lang. Soc. Psychol.* 34, 343–357. doi: 10.1177/0261927X14561119

Koeser, S., and Sczesny, S. (2014). Promoting gender-fair language: the impact of arguments on language use, attitudes, and cognitions. *J. Lang. Soc. Psychol.* 33, 548–560. doi: 10.1177/0261927X14541280

Koniuszaniec, G., and Blaszkowa, H. (2003). "Language and gender in Polish," in *Gender Across Languages. The Linguistic Representation of Women and Men*, Vol. 3, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins Publishing Company), 259–285.

Kunkel-Razum, K. (2004). "Die Frauen und der Duden – der Duden und die Frauen [The women and the Duden – the Duden and the women]," in *Adam, Eva und Die Sprache, Beiträge zur Geschlechterforschung [Adam, Eve and the Language]*, ed. K. M. Eichhoff-Cyrus (Mannheim: Dudenverlag), 308–315.

Lee, J. F. K., and Collins, P. (2008). Gender voices in Hong Kong English textbooks – some past and current practices. *Sex Roles* 59, 127–137. doi: 10.1007/s11199-008-9414-6

Lindner, V., and Lukesch, H. (1994). *Geschlechtsrollenstereotype im Deutschen Schulbuch. [Gender Stereotypes in German Schoolbooks]*. Regensburg: S. Roderer Verlag.

Maass, A., Suitner, C., and Merkel, E. (2013). "Does political correctness make (social) sense?," in *Social Cognition and Communication*, eds J. P. Forgas, O. Vincze, and J. Laszlo (New York, NY: Psychology Press), 331–346.

Macmillan Publishing Company (1975). *Guidelines for Creating Positive Sexual and Racial Images in Educational Materials*. New York, NY: Macmillan.

Marcato, G., and Thüne, E. M. (2002). "Gender and female visibility in Italian," in *Gender Across Languages: The Linguistic Representation of Women and Men*, Vol. 2, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins Publishing Company), 187–217.

Markom, C., and Weinhäupl, H. (2007). *Die Anderen im Schulbuch. Rassismen, Exotismen, Sexismen und Antisemitismus in österreichischen Schulbüchern. [The Others in Schoolbooks. Racism, Exoticism, Sexism, and Anti-Semitism in Austrian Schoolbooks]*. Wien: Braumüller.

McConnell, A. R., and Fazio, R. H. (1996). Women as men and people: effects of gender-marked language. *Pers. Soc. Psychol. Bull.* 22, 1004–1013. doi: 10.1177/01461672962210003

McConnell-Ginet, S. (2013). "Gender and its relation to sex: the myth of 'natural' gender," in *The Expression of Gender*, ed. G. G. Corbett (Berlin: De Gruyter Mouton), 3–38.

McGraw-Hill Book Company (1974). *Guidelines for Equal Treatment of the Sexes*. New York, NY: McGraw-Hill Book Company.

McMinn, M. R., and Foster, J. D. (1991). A computer program to teach nonsexist language. *Teach. Psychol.* 18, 115–117. doi: 10.1207/s15328023top1802_16

McMinn, M. R., Troyer, P. K., Hannum, L. E., and Foster, J. D. (1991). Teaching nonsexist language to college students. *J. Exp. Educ.* 59, 153–161. doi: 10.1080/00220973.1991.10806558

Merkel, E. (2011). *Geschlechtergerechte Sprache in den universitären Gleichstellungskonzepten: Eine vergleichende Analyse [Gender-Fair Language in University Concepts of Gender-Equality: A Comparative Analysis]*. Essen: Netzwerk Frauen– und Geschlechterforschung NRW.

Merkel, E., Maass, A., and Frommelt, L. (2012). Shielding women against status loss. The masculine form and its alternatives in the Italian language. *J. Lang. Soc. Psychol.* 31, 311–320. doi: 10.1177/0261927X12446599

Moser, F., and Hannover, B. (2014). How gender fair are German schoolbooks in the twenty-first century? An analysis of language and illustrations in schoolbooks for mathematics and German. *Eur. J. Psychol. Educ.* 29, 387–407. doi: 10.1007/s10212-013-0204-3

Moser, F., Sato, S., Chiarini, T., Dmitrov-Devold, K., and Kuhn, E. (2011). *Comparative Analysis of Existing Guidelines for Gender-Fair Language within the ITN LCG Network (Work Package B ITN LCG)*. Available at: www.unifr.ch/psycho/itn-lcg/assets/files/ITN_publications/guidelines_final_may2011.pdf

Mucchi-Faina, A. (2005). Visible or influential? Language reforms and gender (in)equality. *Soc. Sci. Inform.* 44, 189–215. doi: 10.1177/0539018405050466

National Council of Teachers of English (2002). *Guidelines for Gender-Fair Use of Language*. Available at: www.ncte.org/positions/statements/genderfairuseoflang

Ng, S. H. (1990). Androcentric coding of man and his in memory by language users. *J. Exp. Soc. Psychol.* 26, 455–464. doi: 10.1016/0022-1031(90)90069-x

Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cognit.* 33, 972–983. doi: 10.3758/bf03193206

Parks, J. B., and Roberton, M. A. (1998). Contemporary arguments against nonsexist language: Blaubergs (1980) revisited. *Sex Roles* 39, 445–461. doi: 10.1023/A:1018827227128

Parks, J. B., and Roberton, M. A. (2004). Attitudes toward women mediate the gender effect on attitudes toward sexist language. *Psychol. Women Q.* 28, 233–239. doi: 10.1111/j.1471-6402.2004.00140.x

Paterson, L. L. (2014). *British Pronoun Use, Prescription, and Processing: Linguistic and Social Influences Affecting 'They' and 'He'*. New York, NY: Palgrave Macmillan.

Pauwels, A. (2003). "Linguistic sexism and feminist linguistic activism," in *The Handbook of Language and Gender*, eds M. Meyerhoff and J. Holmes (Oxford: Blackwell Publishing), 550–570.

Preinsberger, A., and Weisskircher, E. (1997). "Mathematikschulbücher – eine aktuelle Untersuchung [Mathematics books – a recent examination]," in *Schule Weiblich – Schule Männlich. Zum Geschlechterverhältnis im Bildungswesen [Female School – Male School. On Gender Relations in the Educational System]*, eds L. Lassnigg and A. Paseka (Innsbruck: Studien Verlag), 132–143.

Prentice, D. A. (1994). Do language reforms change our way of thinking? *J. Lang. Soc. Psychol.* 13, 3–19. doi: 10.1177/0261927X94131001

Prewitt-Freilino, J. L., Caswell, T. A., and Laakso, E. K. (2012). The gendering of language: a comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles* 66, 268–281. doi: 10.1007/s11199-011-0083-5

Pusch, L. F. (1984). *Das Deutsche als Männersprache. Aufsätze und Glossen zur Feministischen Linguistik [German as a Male language]*. Frankfurt: Suhrkamp.

Ridgeway, C. L., and Correll, S. J. (2004). Unpacking the gender system: a theoretical perspective on gender beliefs and social relations. *Gend. Soc.* 18, 510–531. doi: 10.1177/0891243204265269

Rothmund, J., and Christmann, U. (2002). Auf der Suche nach einem geschlechtergerechten Sprachgebrauch. Führt die Ersetzung des generischen Maskulinums zu einer Beeinträchtigung von Textqualitäten? [In search of gender-fair language. Does replacement of masculine generics impair text quality?]. *Muttersprache* 112, 115–135.

Sabatini, A. (1987). *Il Sessismo Nella Lingua Italiana, Commissione Nazionale per la Realizzazione Della Parità tra Uomo e Donna, Roma*. Available at: http://www.funzionepubblica.gov.it/media/962032/il%20sessismo%20nella%20lingua%20italiana.pdf

Sadker, M., Sadker, D., and Klein, S. (1991). The issue of gender in elementary and secondary eduation. *Rev. Res. Educ.* 17, 269–334. doi: 10.2307/1167334

Sarrasin, O., Gabriel, U., and Gygax, P. (2012). Sexism and attitudes toward gender-neutral language. The case of English, French, and German. *Swiss J. Psychol.* 71, 113–124. doi: 10.1024/1421-0185/a000078

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *J. Soc. Issues* 57, 675–688. doi: 10.1111/0022-4537.00235

Schweizerische Bundeskanzlei (1996). *Geschlechtergerechte Sprache. Leitfaden zum Geschlechtergerechten Formulieren im Deutschen. [Gender-Fair Language. Guideline for Gender-Fair Formulation in German]*. Available at: http://www.bk.admin.ch/dokumentation/sprachen/04915/05313/index.html (revised in 2009).

Sczesny, S., Moser, F., and Wood, W. (2015). Beyond sexist beliefs: how do people decide to use gender-inclusive language? *Pers. Soc. Psychol. Bull.* 41, 943–954. doi: 10.1177/0146167215585727

Silveira, J. (1980). Generic masculine words and thinking. *Womens Stud. Int. Q.* 3, 165–178. doi: 10.1016/s0148-0685(80)92113-2

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language," in *Social Communication. A Volume in the Series Frontiers of Social Psychology*, ed. K. Fiedler (New York, NY: Psychology Press), 163–187.

Stahlberg, D., and Sczesny, S. (2001). Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen [The impact of masculine generics on the cognitive inclusion of women]. *Psychol. Rundsch.* 52, 131–140. doi: 10.1026//0033-3042.52.3.131

Stahlberg, D., Sczesny, S., and Braun, F. (2001). Name your favorite musician: effects of masculine generics and of their alternatives in German. *J. Lang. Soc. Psychol.* 20, 464–469. doi: 10.1177/0261927x01020004004

Stout, J. G., and Dasgupta, N. (2011). When he doesn't mean you: gender-exclusive language as ostracism. *Pers. Soc. Psychol. Bull.* 3, 757–769. doi: 10.1177/0146167211406434

Sunderland, J. (2011). *Language, Gender and Children's Fiction*. London: Continuum.

Swim, J. K., Aikin, K. J., Hall, W. S., and Hunter, B. A. (1995). Sexism and racism: old-fashioned and modern prejudices. *J. Pers. Soc. Psychol.* 68, 199–214. doi: 10.1037/0022-3514.68.2.199

Swim, J. K., Mallet, R., and Stangor, C. (2004). Understanding subtle sexism: detection and use of sexist language. *Sex Roles* 51, 117–128. doi: 10.1023/B:SERS.0000037757.731.06

Swiss Federal Institute of Technology Zurich (2011). *Die Zwölf Sprachregeln [Twelve Language Rules]*. Available at: http://www.equal.ethz.ch/rules

Twenge, J. M., Campbell, W. K., and Gentile, B. (2012). Male and female pronoun use in US books reflects women's status, 1900–2008. *Sex Roles* 67, 488–493. doi: 10.1007/s11199-012-0194-7

UNESCO (1999). *Guidelines for Gender-Neutral Language*. Available at: http://unesdoc.unesco.org/images/0011/001149/114950mo.pdf

UNESCO (2011). *Priority Gender Equality Guidelines*. Available at: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/BSP/GENDER/GE%20Guidelines%20December%202_FINAL.pdf

Valdrová, J., Knotková-Capková, B., and Pacliková, P. (2010). *Kultura Genderove Vyvazeneho Vyjadrovani*. Available at: http://data.idnes.cz/soubory/studium/A100125_BAR_GENDER_PRIRUCKA.PDF

Vervecken, D., Gygax, P. M., Gabriel, U., Guillod, M., and Hannover, B. (2015). Warm-hearted businessmen, competitive housewives? Effects of gender-fair language on adolescents' perceptions of occupations. *Front. Psychol.* 6:1437. doi: 10.3389/fpsyg.2015.01437

Vervecken, D., and Hannover, B. (2012). Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts perception of the speaker. *Eur. J. Soc. Psychol.* 42, 754–762. doi: 10.1002/ejsp.1893

Vervecken, D., and Hannover, B. (2015). Yes I can! Effects of gender fair job descriptions on children's perceptions of job status, job difficulty, and vocational self-efficacy. *Soc. Psychol.* 46, 76–92. doi: 10.1027/1864-9335/a000229

Vervecken, D., Hannover, B., and Wolter, I. (2013). Changing (s)expectations: how gender-fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *J. Vocat. Behav.* 82, 208–220. doi: 10.1016/j.jvb.2013.01.008

Wasserman, B. D., and Weseley, A. J. (2009). Qué? Quoi? Do languages with grammatical gender promote sexist attitudes? *Sex Roles* 61, 634–643. doi: 10.1007/s11199-009-9696-3

Wood, W., and Rünger, D. (2016). Psychology of habit. *Annu. Rev. Psychol.* 67, 11.1–11.26. doi: 10.1146/annurev-psych-122414-033417

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Pers. Soc. Psychol.* 9, 1–27. doi: 10.1037/h0025848

CrossMark

Does Gender-Fair Language Pay Off? The Social Perception of Professions from a Cross-Linguistic Perspective

Lisa K. Horvath[1,2]*, Elisa F. Merkel[1,3], Anne Maass[3] and Sabine Sczesny[1]

[1] Psychology, University of Bern, Bern, Switzerland, [2] TUM School of Management, Technical University of Munich, Munich, Germany, [3] Psychology, University of Padua, Padua, Italy

In many languages, masculine forms (e.g., German *Lehrer*, "teachers, masc.") have traditionally been used to refer to both women and men, although feminine forms are available, too. Feminine-masculine word pairs (e.g., German *Lehrerinnen und Lehrer*, "teachers, fem. and teachers, masc.") are recommended as gender-fair alternatives. A large body of empirical research documents that the use of gender-fair forms instead of masculine forms has a substantial impact on mental representations. Masculine forms activate more male representations even when used in a generic sense, whereas word pairs (e.g., German *Lehrerinnen und Lehrer*, "teachers, fem. and teachers, masc.") lead to a higher cognitive inclusion of women (i.e., visibility of women). Some recent studies, however, have also shown that in a professional context word pairs may be associated with lesser status. The present research is the first to investigate both effects within a single paradigm. A cross-linguistic (Italian and German) study with 391 participants shows that word pairs help to avoid a male bias in the gender-typing of professions and increase women's visibility; at the same time, they decrease the estimated salaries of typically feminine professions (but do not affect perceived social status or competence). This potential payoff has implications for language policies aiming at gender-fairness.

Keywords: social perception, gender-fair language, grammatical gender, gender stereotypes, professional groups, status

INTRODUCTION

Although women have increasingly entered paid employment in the twentieth century and are making their way up the hierarchical ladders (Eagly and Karau, 2002), there are still considerable gender inequalities in the labor market, as documented by many different indices (e.g., the *Gender Gap Index* of the World Economic Forum, the *Gender Inequality Index* of the UNDP, or the *Social Institutions and Gender Index* of the OECD; for an overview, see European Institute for Gender Equality, 2013). Also, different linguistic forms have been found to contribute to gender-(in)equality: Masculine forms used as generics referring to both women and men lead to a male bias in mental representations. In contrast, feminine-masculine word pairs, which are generally considered gender-fair, increase women's visibility (for an overview, see Stahlberg et al., 2007). This seems to suggest that word pairs promote gender equality. However, recent findings also document detrimental effects of gender-fair language in the professional context, especially on status-related measures (e.g., Formanowicz et al., 2013; Vervecken et al., 2015). These contradictory findings concerning effects of gender-fair language (vs. masculine generics) on gender equality were

obtained in different studies under different conditions. The present study is the first to investigate the complex and potentially paradox effects of gender-fair language on the social perception of professional groups within a single paradigm.

The Social Perception of Professions

Occupational gender-stereotyping follows the proportion of women and men holding the respective professions and translates into the classification of professions as typically feminine and typically masculine (Krefting et al., 1978; Glick, 1991; Glick et al., 1995). The social role theory (Eagly, 1987; Eagly et al., 2000; Koenig and Eagly, 2014) provides a theoretical explanation: Social perceivers' views about social (e.g., occupational) groups and the related stereotypes (e.g., occupational stereotypes) follow from perceivers' experiences and observations of the different distributions of women and men in the respective groups. For instance, when men are observed to occupy the majority of leadership roles, perceivers assume that men possess the traits required for successful leadership, such as decisiveness or dominance (*think-manager—think-male*, Schein, 1973, 2001). On the other hand, individuals occupying certain social roles (e.g., homemaker vs. employee) are described with traits that are stereotypical for these roles (Eagly and Steffen, 1984). Experimental research has confirmed these assumptions. In line with social role theory, occupational stereotyping not only goes back to the observation of typical members of the respective occupational groups, but it actually reflects social reality; additionally, occupational stereotypes can change according to fictitious and varied future job holders despite current stereotypes (Koenig and Eagly, 2014).

A highly relevant variable when dealing with gender in the work place is status. A *gender hierarchy* (Ridgeway and Correll, 2001) continues to be widely prevalent, with men and masculinity being ascribed a higher status than women or femininity. This is mirrored by the following facts: Men are ascribed more competence and worthiness (Ridgeway, 2001), men possess more power, men are more likely to be in leadership positions than women (European Commission, 2011), and men have more access to resources than women (Eagly et al., 2000). Moreover, men receive higher salaries for the same work than women (*Global Gender Gap*, Hausmann et al., 2010). When it comes to the social perception of professions, male-dominated professions are accordingly attributed higher prestige (e.g., Glick et al., 1995). Vice versa, people working in male-dominated professions are assumed to have higher salaries than people working in female-dominated professions, which is indeed the case (Cejka and Eagly, 1999; Alksnis et al., 2008). These gender-status beliefs are consistent with the *Stereotype Content Model* (Fiske et al., 2002): High-status groups (e.g., men) are ascribed higher competence than low-status groups (e.g., women) (Cuddy et al., 2007).

Although gender and status are associated in general, the gender-typing of a profession and professional status/prestige are unrelated dimensions in occupational stereotyping (Gottfredson, 1981, 1996). For instance, the distribution of women and men across professions does not predict estimates of occupational prestige (Glick, 1991). Furthermore, these two dimensions were empirically found to form two independent dimensions in

occupational stereotyping (Glick et al., 1995). This is supported by social psychological experiments: Variations in the status of jobs was not reflected in the gender stereotypes ascribed to the job holder (Eagly and Steffen, 1984). This raises the question whether gender-fair language might affect occupational stereotyping concerning both with respect to gender and status.

Language and Gender

Language and cognition are intertwined, with language impacting cognition and vice versa. For instance, color labels and color distinctions that were available in a given language affected native speakers' color perception (Lucy and Shweder, 1979; Winawer et al., 2007) and the order of adjectives and nouns impacts categorization of groups (Percy et al., 2009). With regard to gender, there are differences in how gender is represented in languages (for a detailed overview, see Stahlberg et al., 2007). In *genderless languages* such as Finnish, Turkish, or Chinese gender is mainly expressed through lexical elements of the type "woman," "man," "brother," or "sister." Otherwise, nouns and pronouns lack gender markings. In *natural gender languages* such as English, Danish, and Swedish as well personal nouns are mostly unmarked for gender, but personal pronouns are gendered. In so-called *grammatical gender languages* like French, Italian or German, additionally to gendered pronouns, all nouns have grammatical gender and many other parts of speech (articles, adjectives, or pronouns that depend on the noun) show grammatical agreement; that means, they signal the gender of the noun. Nouns in these languages are either masculine or feminine, in some languages also neuter (e.g., German). For instance, a table in German is masculine (*der Tisch*, the table, masculine), but feminine in French (*la table*, the table, feminine). Interestingly, the grammatical gender of objects affects the way these objects are perceived: people ascribe more typically masculine (vs. feminine) characteristics to objects that are designated with a grammatically masculine (vs. feminine) noun in their native language (Boroditsky et al., 2003). Thus, gender is in general a highly salient feature in these languages. This also applies to the social perception of professions, where grammatical gender is highly relevant. In languages with grammatical gender, masculine and feminine job titles are available to describe professionals (e.g., German *der Lehrer*, the teacher, masculine/male; *die Lehrerin*, the teacher, feminine/female). Masculine forms, however, are also used as generics ("masculine generics") to refer to both women and men, to mixed-gender groups or persons whose gender is unknown or irrelevant in a given context (Braun et al., 2005). This traditional use of masculine generics is not considered gender-fair and alternative forms such as feminine-masculine word pairs (German *die Lehrerinnen und Lehrer*, the teachers, fem. and the teachers, masc.) are recommended as replacements (Stahlberg et al., 2007).

Some authors have argued that the existence of grammatical gender in a language is associated with gender (in)equality on a societal level: Gender inequality or gender gaps tend to be bigger in countries with grammatical gender languages (i.e., where masculine forms are used as generics although feminine forms are available) than in countries with natural gender languages or

genderless languages. This effect even persists when controlling for religion and political system (Prewitt-Freilino et al., 2012).

Reflecting the latter, there is now ample evidence that the conventional use of masculine forms as generics causes a male bias in mental representations. This effect was replicated and confirmed with different methods in investigations from various disciplines such as social psychology, psycholinguistics or cognitive psychology (see Stahlberg et al., 2007, for an overview). Thus, speakers of German, for instance, associate and retrieve predominantly male exemplars when answering a question with a masculine generic (such as "*Wer ist Ihr Lieblingsmusiker?*" "Who is your favorite musician, masc.?"). In contrast, significantly more female exemplars are mentioned when gender-fair forms are used, such as feminine-masculine word pairs (e.g., *Lieblingsmusikerin/-musiker*, favorite musician, fem./musician, masc.) (Stahlberg et al., 2001; see also Braun et al., 2005).

However, psychological studies on the question whether gender-fair language indeed helps to promote gender equality (masculine generics vs. gender-fair forms) have revealed a complex pattern of effects. Some studies show beneficial effects of gender-fair forms (such as an increased visibility of women), while others describe detrimental effects (such as status loss), as will be discussed below. The present research is the first to investigate this mixed outcome that was observed across different studies within a single paradigm using a multidimensional approach. This is done in order to get a broader picture of occupational stereotyping following gender-fair language and not only a focused view on specific outcomes (Glick et al., 1995). We argue that—for the time being—gender-fair word pairs can simultaneously facilitate and hamper gender equality.

Linguistic Forms and the Social Perception of Professions

Interestingly, language may override widespread stereotypes. For example, according to occupational stereotypes the professions of truck driver or physicist are perceived to be typically masculine; social worker or kindergarten teacher are perceived to be typically feminine professions (Kennison and Trofe, 2003; Irmen, 2007). Even these powerful stereotypes are under the influence of linguistic forms (Irmen and Roßberg, 2004; Braun et al., 2005). In German and French—both grammatical gender languages—but not in English—a natural gender language, where no feminine professional nouns are available—masculine generics caused a male bias in mental associations of professions. Participants assumed more men than women to be in a professional group, even for typically feminine professions. However, the male bias was reduced when respondents were presented with masculine and feminine forms of the respective job titles (Gabriel et al., 2008). Furthermore, a set of studies by Braun and colleagues (2005) showed that various gender-fair forms can help to increase women's visibility in general: word pairs (e.g., German *Musikerinnen und Musiker*, musicians, fem. and musicians, masc.), the capital-I form (*MusikerInnen*, musicians with a capital letter marking the feminine ending as generic and including both women and men), or gender-neutral formulations. But the magnitude of the impact depended

on the gender-typicality of professions: When word pairs (vs. masculine forms) were used to refer to a typically masculine profession (geophysicist), more women were assumed to attend a scientific meeting of geophysicists, but this less so for a typically feminine profession (nutritionists). These findings show that linguistic forms have a powerful effect on the social perception of professions and can increase women's visibility.

Linguistic forms also have a tangible impact on behavior in professional contexts. Early research on American English (Bem and Bem, 1973) demonstrated that women and men are more eager to apply for a counter-stereotypical position when the job advertisement refers to both women and men with a gender-neutral form (e.g., *lineworker*) compared to linguistic forms addressing the stereotypical gender. However, use of such gender-specific forms (e.g., *lineman, linewoman*) which were investigated in this study from the 1970s is not permitted any more (UNESCO, 1999). Nevertheless, masculine pronouns (e.g., *he, his, him*) are still used as generics instead of gender-fair forms (e.g., *he/she, her/his, they*). Masculine pronouns—used in reference to an ideal applicant for a vacant position—were found to decrease women's sense of belonging to a professional context, their motivation to pursue the respective position as well as their expected identification with the job compared to gender-neutral forms (*they, the employee*) or word pairs (*he/she, his/her*) (Stout and Dasgupta, 2011). Linguistic forms not only affect potential applicants but also those who make hiring decisions. In a hiring-simulation study on German, decision makers preferred male over female applicants for a high-status leadership position (but not for a middle-management position) when the position was advertised in the masculine (*Geschäftsführer*, CEO, masc.). When word pairs were used (*Geschäftsführerin/Geschäftsführer*, CEO, fem./CEO, masc.), however, women and men were rated as equally suitable for the job (Horvath and Sczesny, 2015).

A number of studies show that children's and adolescents' perceptions of professions and their vocational interests are strongly affected by linguistic forms. For instance, when professions were presented to French adolescents in the masculine, women were perceived to be more successful in typically feminine and men in typically masculine jobs. With word pairs, however, perceptions of success were more balanced: Female and male professionals were perceived as equally likely to succeed in both typically feminine and masculine professions. While linguistic forms did not affect perceived competence, they had an impact on perceived warmth: When professions were presented with masculine forms, holders of typically masculine jobs were perceived as less warm and holders of typically feminine jobs were perceived as warmer compared to the presentation with word pairs. The authors concluded that word pairs shifted perceptions of warmth toward the mid-point and somehow balanced these perceptions, whereas masculine forms tended to evoke gender-stereotypic perceptions of warmth. It should be noted, though, that this was the very first study measuring competence and warmth perceptions of professions (Vervecken et al., 2015). Another study with Belgian and German children showed similar effects for perceptions of success: when professions were presented with word pairs, children estimated female job holders in typically masculine

professions as more successful. Furthermore, girls were more interested in these typically masculine professions (Vervecken et al., 2013). However, beneficial as well as detrimental effects of German gender-fair forms have also been observed in children's perception of professions. While word pairs reduced the perceived difficulty of typically masculine professions, and thus increased vocational self-efficacy, they also reduced the estimated salaries (Vervecken and Hannover, 2015).

In a similar vein, use of feminine titles to introduce female professionals in Italian (e.g., *professoressa*, teacher or professor, fem.) instead of masculine titles (e.g., *professore*, teacher or professor, masc.) made these professionals appear less persuasive (Mucchi-Faina, 2005). It has to be noted, though, that this effect may be caused specifically by the feminine suffix *-essa*, as female professionals described with titles ending in *-essa* (corresponding to the suffix *-ess* in English, e.g., *hostess, authoress*) were perceived as having a lower social status than female professionals described with a title ending in *-a* (e.g., *professora*, teacher or professor, fem.), which is a more modern feminine suffix, or with a masculine form (Merkel et al., 2012). Similar disadvantages of linguistic feminization have been described for Polish: Women applying for a gender-neutral job were perceived as less suitable when referring to themselves with a feminine (vs. masculine) professional title (Formanowicz et al., 2013). However, reactions to linguistic forms may change over time, especially as a function of habituation. Thus, when female proponents of social initiatives were introduced with feminine (vs. masculine) forms in Polish, where gender-fair language is relatively new and uncommon, these initiatives were devalued and were not supported. In German, however, where feminine job titles are common, speakers tended to support the initiatives less when female proponents were introduced in the masculine (Formanowicz et al., 2015). Similarly, in Sweden, negative attitudes toward the newly invented gender-neutral personal pronoun *hen*—additionally to the masculine *han* and the feminine *hon*—have been found to diminish over time (Gustafsson Senden et al., 2015). Thus, a feminization of job titles may be detrimental for women when the implementation of gender-fair language starts, but may become integral part of everyday language once speakers have become accustomed to these (initially unfamiliar) forms.

Taken together, past research on the effects of gender-fair language yields a complex pattern: On the one hand, the visibility of women as a group increases when word pairs are used instead of masculine generics (e.g., Braun et al., 2005; Irmen, 2007; Gabriel et al., 2008). On the other hand, a decrease in status-related measures (e.g., social status, salary) is observed when female professionals are introduced with gender-fair (feminine) job titles compared to masculine forms. However, the different studies are based on a wide range of methods and study designs (between- vs. within-participants), which renders a direct comparison difficult. Some studies, for instance, tested one profession only (e.g., Formanowicz et al., 2013), while others included a larger number of professions (e.g., Vervecken et al., 2015). Also, different participant populations (i.e., children, adolescents, students, adults) have been used. Moreover, the effects of gender-fair language have been studied in different languages (e.g., French, Dutch, German, Italian, Polish,

English), which have their own structural characteristics. Certain effects may therefore be restricted to the respective language, for instance, negative effects of specific feminine job titles in Italian (e.g., Mucchi-Faina, 2005; Merkel et al., 2012) or Polish (Formanowicz et al., 2013). In some cases, opposing reactions to gender-fair language were found in different languages (e.g., support of social initiatives in German, but rejection in Polish; Formanowicz et al., 2015). Also, some studies focused on the individual level (perception of one person, e.g., Formanowicz et al., 2013), others on the group level (perception of professions, Vervecken et al., 2015). Thus, it is unknown whether gender-fair forms decrease adults' perceptions of professional status on a group level, as is true for children (e.g., Vervecken and Hannover, 2015), and on the individual level (e.g., Mucchi-Faina, 2005; Formanowicz et al., 2013). Most importantly, no study so far has tested whether gender-fair language can *simultaneously* lead to a decrease in perceived status and an increase in visibility. Our aim was, therefore, to test these seemingly contradictory effects within a single paradigm with adult participants.

Aim and Hypotheses

The purpose of the present research was to examine whether gender-fair language pays off by increasing women's visibility or whether it also lowers the perceived status of professions. To answer these questions, we used a repeated measures design in a multidimensional approach and tested the effects of linguistic forms on the perception of professional groups. To increase the generalizability of our findings we investigated two grammatical gender languages, namely Italian and German.

Although gender and status are generally associated (Ridgeway, 2001), gender-typicality and social status might constitute independent and orthogonal dimensions when investigating social perceptions of professions (Glick, 1991; Glick et al., 1995; Gottfredson, 1996). On this basis, we assumed that gender- and status-related measures can indeed simultaneously reveal women's visibility and profession's status loss, even when assessed within a single study.

In our study adults evaluated a list of professions with respect to (a) status-related measures (dimensions that tend to suffer when gender-fair language is used) and (b) women's visibility (a dimension that tend to show greater mental inclusion of women when word pairs are used). The question was whether participants exposed to professions designated with word pairs (e.g., German *Mechanikerinnen und Mechaniker*, mechanics, fem. and mechanics, masc.) would form different perceptions of the respective professional group than those exposed to masculine forms (e.g., German *Mechaniker*, mechanics, masc.). The languages under study were Austrian German and Italian, two grammatical gender languages with structural similarities. Most importantly, professional titles are gender-marked in both languages (e.g., German *Fleischerinnen und Fleischer*; Italian *macellaie e macellai*, butchers, fem. and butchers, masc.), so that we expected comparable findings for the two languages. Moreover, to make results comparable with the most relevant and directly related prior studies (e.g., Braun et al., 2005), we adopted methods and dependent variables from these studies wherever possible. Our hypotheses read as follows:

Hypothesis 1: Professional groups are perceived to have a lower social status when designated with word pairs than with masculine forms.
Hypothesis 2: Professional groups are perceived to have lower salaries when designated with word pairs than with masculine forms.
Hypothesis 3: Professional groups designated with word pairs render women more visible than with masculine forms.

In addition, we examined whether the perceived competence and warmth of the professional groups was also affected. But as there was only one published study with French-speaking children (Vervecken et al., 2015), which had produced rather unexpected findings for warmth, we were reluctant to formulate specific hypotheses. Therefore, perceptions of warmth and competence were investigated in an exploratory way.

MATERIALS AND METHODS

Participants

The sample consisted of 391 participants: 195 Austrians (123 women, 72 men; average age 36.03 years, $SD = 10.53$) and 196 Italians (130 women, 66 men; average age 28.55 years, $SD = 7.42$). We recruited participants via snowball sampling and included only individuals over 18 years of age.

Materials
Professions
Pretest and selection of target professions
We selected 27 professions (see Appendix A) based on prior research on professional groups (Kennison and Trofe, 2003; Gabriel et al., 2008). Professional titles were selected only when masculine and feminine plural forms were available in both languages, German (e.g., *Dolmetscherinnen und Dolmetscher*, interpreters, fem. and interpreters, masc.) and Italian (e.g., *traduttrici e traduttori*). The web-based pretest on these professions was run with 100 participants (41 Austrians: 26 women, 15 men; 59 Italians: 36 women, 23 men). The dependent variable was gender-typicality of professions ("*Are the following professions more typical of women or men?*"). As in earlier research (Gabriel et al., 2008), answers for each profession were provided on 7-point bipolar scales with the feminine form (e.g., *Dolmetscherinnen/traduttrici*, interpreters, fem.) as one pole (coded as 1) and the masculine form (e.g., *Dolmetscher/traduttori*, interpreters, masc.) as the other pole (coded as 7). Pole labels were counterbalanced across participants: either the feminine or the masculine label appeared on the left end of the scale. Furthermore, we presented the professions in a random order for each participant. Participants filled out the questionnaire in their native language (German or Italian).

Based on these ratings, professions were categorized as typically feminine (<3.5), gender-neutral (3.5–4.5) or typically masculine (>4.5). In the pretest both Austrian and Italian participants rated seven professions as typically feminine, 13 professions as typically masculine, and three as gender-neutral; judgments of the two national groups were incongruent for four professions (for more details, see Appendix A). For

the main study we selected professions on the basis of the following criteria. First, we aimed at including a broad sample of professions, of different gender-typicality but with matching occupational prestige, in order to avoid a prestige-biased sample of professions (as in Glick et al., 1995). Second, we aimed at selecting a comparable number of typically feminine and masculine professions to avoid a statistical bias in the analyses (Tabachnick and Fidell, 2001). Third, we aimed to avoid making the gender-typicality of professions salient. We therefore decided to present not only strongly stereotyped jobs but also additional, more ambiguous professions (seven slightly masculine, three gender-neutral, and four incongruent ones) as fillers in the main study. In order to reduce the questionnaire to a reasonable length, we split the professions into three lists. We selected six of the seven professions judged as typically feminine professions, balanced for occupational prestige as indicated by average salaries (published by Public Employment Service Austria, Arbeitsmarktservice Österreich, 2015). *Six* of the most typically masculine professions were matched with the feminine professions for occupational prestige. The six typically feminine professions selected were *tailors, hairdressers, dancers, interpreters, nutritionists, pharmacists*, and *psychologists*, the last three being rather high in occupational prestige. The six masculine professions selected were: *truck drivers, electricians, mechanics, computer scientists, physicists*, and *engineers* with the last three being rather high in occupational prestige. These 12 final professions were assigned to three lists, whereby each list contained two typically feminine and two typically masculine professions (matched for occupational prestige). Please find a table of the 12 target professions, distributed over the three experimental lists in Appendix B. The filler professions were randomly distributed across the lists and were not included in the main analyses.

Linguistic Forms
A web-based online questionnaire was used for the main study. Here, all target professions appeared in one of two linguistic versions, namely either in the masculine (e.g., German *Schneider*, Italian *sarti*, tailors, masc.) or in the form of a word pair (e.g., German *Schneiderinnen und Schneider*, Italian *sarte e sarti*, tailors, fem. and tailors, masc.). Each participant was randomly assigned either to the masculine or the word pair condition. If every participant were to rate all professions the questionnaire would have been too long. Therefore, participants were randomly assigned to one of the three lists. In the questionnaire, each profession was followed by a series of items. These items were presented on three separate pages of the online questionnaire. To strengthen the linguistic manipulation, the professions reappeared (in the respective linguistic form) in the heading of each page.

Dependent Variables
We measured the following dependent variables: perceived social status, estimated salary, women's visibility, competence, and warmth, which are described in more detail below. For every dependent variable we aggregated answers for typically feminine

and masculine professions separately. Reliabilities for both types of professions are reported below.

Perceived Social Status

The perceived social status of professions was measured with three items developed by Binggeli et al. (2014): (a) *"How much prestige do* [professional group] *have in our society?"* (b) *"How economically successful have* [professional group] *been?"* (c) *"How is the educational level of* [professional group]*?"* Answers were provided on 7-point bipolar scales (1 = very low; 7 = very high) and item order was randomized. Reliabilities were satisfying for both masculine professions (α = 0.78) and for feminine professions (α = 0.71).

Estimated Salary

The estimated salary was measured by a single item adopted from Becker et al. (2011): *"Please estimate how much* [professional group] *earn compared to the average Italian / Austrian salary."* Participants indicated their responses on an 11-point rating scale ranging from −50% (fifty percent below national average) to +50% (fifty percent above national average), in 10% increments. The midpoint represented the national average salary.

Women's Visibility

Women's visibility was measured with two items which had been used in earlier studies to assess gender typicality: (a) *"How many women and men pursue the profession* [professional group]*?"* (similar to Braun et al., 2005; Gabriel et al., 2008). Answers were provided on an 11-point bipolar scale, ranging from 100% women to 100% men, with 10% increments (90% women, 80% women, 70% women, …); the midpoint was 50% women-50% men; (b) *"For whom is the profession* [professional group] *more typical?"* Answers were provided on a 7-point bipolar scale (ranging from 1 = women to 7 = men, or vice versa). Both items were recoded, so that higher values indicated higher visibility of women. Due to different answering formats we z-standardized the items and merged them. Reliabilities were satisfying for both masculine professions (α = 0.75) and for feminine professions (α = 0.81)

Ascriptions of Competence and Warmth

Ascriptions of competence and warmth were assessed with five items each, adopted from Cuddy et al. (2004) and Cuddy et al. (2009). Participants were asked: *"How would you evaluate* [professional group] *on the following traits? To which degree are they* [competence traits: *able, competent, confident, efficient, skillful*; warmth traits: *warm-hearted, likeable, friendly, altruistic, cordial*]*?"* Answers were provided on 7-point bipolar scales (1 = very little; 7 = very much). The order of the items was randomized. Items for warmth and competence were averaged and reliabilities were satisfying: competence for masculine professions (α = 0.89) and for feminine professions (α = 0.91), warmth for masculine professions (α = 0.91) and for feminine professions (α = 0.93).

An overview of intercorrelations of all dependent variables is provided in **Tables 1** and **2**, for German and Italian, respectively.

TABLE 1 | Intercorrelations of dependent variables perceived social status, estimated salary, women's visibility, ascriptions of competence and warmth for feminine and masculine professions in German.

	1	2	3	4	5
1. Social status	–	0.33***	−0.04	0.64***	0.42***
2. Salary	0.45***	–	−0.14	0.15*	0.00
3. Women's visibility	−0.17*	−0.22**	–	0.12	0.17*
4. Competence	0.60***	0.37***	−0.28***	–	0.68***
5. Warmth	0.40***	0.13	−0.06	0.48***	–

*Intercorrelations for feminine professions are reported above the diagonal, intercorrelations for masculine professions are reported below the diagonal; *p < 0.05, **p < 0.01, ***p < 0.001.*

TABLE 2 | Intercorrelations of dependent variables perceived social status, estimated salary, women's visibility, ascriptions of competence, and warmth for feminine and masculine professions in Italian.

	1	2	3	4	5
1. Social status	–	0.46***	−0.13	0.46***	0.31***
2. Salary	0.40***	–	−0.08	0.22**	0.16*
3. Women's visibility	−0.03	−0.07	–	0.09	0.09
4. Competence	0.37***	0.01	−0.04	–	0.65***
5. Warmth	0.26***	0.00	0.07	0.48***	–

*Intercorrelations for feminine professions are reported above the diagonal, intercorrelations for masculine professions are reported below the diagonal; *p < 0.05, **p < 0.01, ***p < 0.001.*

Procedure

Upon entering the web-based questionnaire, participants were informed that the purpose of the study was to investigate social perceptions of various professional groups. In line with APA guidelines (American Psychological Association, 2010), the main instructions at the beginning of the survey included further information, for instance, on expected duration, procedures, and confidentiality. Participants were then presented with nine professional groups. At the end of the questionnaire, they were debriefed and invited to participate in a lottery for gift vouchers, which had been announced at the beginning[1]. The project was approved by the Ethical Committee of the University of Padova in 2010.

RESULTS

Perceived social status, estimated salary and women's visibility, as well as ascriptions of competence and warmth of typically masculine and feminine professions were analyzed with a 2 (Stereotypicality of Professions: masculine vs. feminine) × 2 (Linguistic Form: masculine forms vs. word pairs) × 2 (Language: German vs. Italian) × 2 (Participant gender: female vs. male) × 3 (List of professions) multivariate analysis

---

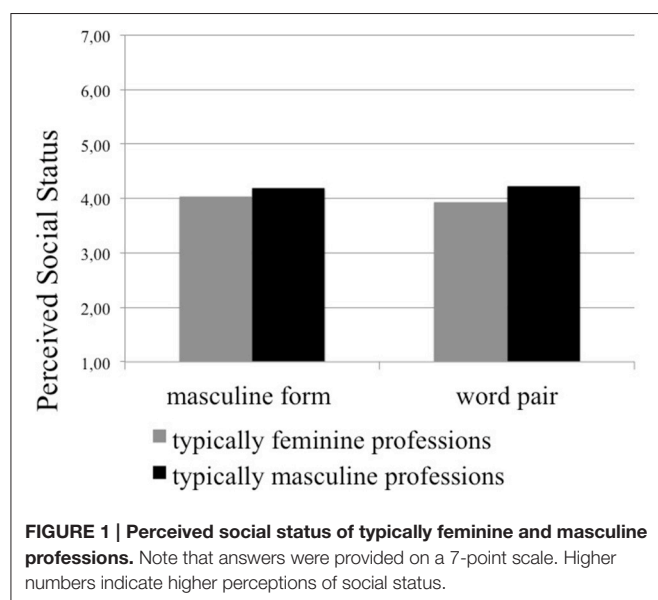[1]After measuring the dependent variables we also assessed participants' attitudes toward gender-fair language (Sczesny et al., 2015) and sexism (with the Ambivalent Sexism Inventory; Glick and Fiske, 1996). Since both attitude scales were correlated with the dependent as well as the independent variables, we could not use them as moderators, as had been intended, and thus do not report them here.

of variance (MANOVA) with repeated measures on the first factor. The MANOVA was followed by ANOVAs with pairwise comparisons (with Bonferroni correction) for each dependent variable. Results with *p*-values of 0.05 or less are considered significant. As we were mainly interested in effects of linguistic forms (masculine forms vs. word pairs) and also to enhance readability, we report only those effects that concern our hypotheses: main effects of stereotypicality of professions and effects involving our core factor, linguistic form. All other effects are reported in Appendix C.

The MANOVA revealed a main effect of linguistic form, $F_{(5, 354)} = 4.57$, $p < 0.001$, $\eta_p^2 = 0.06$, indicating that overall masculine forms and word pairs produced different perceptions of professions. No interaction effect involving linguistic form was significant. For all other multivariate effects not involving linguistic form, please see Appendix C.

Perceived Social Status

The ANOVA for perceived social status revealed a significant interaction between stereotypicality of profession and linguistic form, $F_{(1, 363)} = 4.95$, $p = 0.027$, $\eta_p^2 = 0.01$. Pairwise comparisons showed that typically feminine professions were perceived as having lower status than masculine professions, both when presented with masculine forms ($p = 0.021$, $\eta^2 = 0.02$) and with word pairs ($p \leq 0.001$, $\eta^2 = 0.07$). It is noteworthy that the difference between typically masculine and feminine professions was stronger when word pairs were used. In fact, when word pairs were used, the perceived status of feminine professions declined slightly compared to masculine forms, whereas that of masculine professions increased slightly. These differences are displayed in **Figure 1**. All means and standard deviations are reported in **Table 3**. For all other effects not involving linguistic form, please see Appendix C.

FIGURE 1 | Perceived social status of typically feminine and masculine professions. Note that answers were provided on a 7-point scale. Higher numbers indicate higher perceptions of social status.

Estimated Salary

The main effect for linguistic form, $F_{(1, 359)} = 5.85$, $p = 0.016$, $\eta_p^2 = 0.02$, indicated that professions presented with masculine forms were believed to earn higher salaries than professions presented with word pairs (as predicted in Hypotheses 2). In addition, the interaction between stereotypicality of professions and linguistic form was significant, $F_{(1, 359)} = 4.36$, $p = 0.037$, $\eta_p^2 = 0.01$. Pairwise comparisons showed that feminine professions were estimated to have lower salaries than masculine professions in both linguistic conditions (masculine form: $p < 0.001$, $\eta_p^2 = 0.09$; word pairs: $p < 0.001$, $\eta_p^2 = 0.09$); however, salaries of feminine professions were estimated higher when designated with masculine forms than with word pairs ($p = 0.003$, $\eta^2 = 0.03$). The salary ratings for masculine professions did not differ according to linguistic form ($p = 0.416$, $\eta^2 = 0.002$). All means and standard deviations are reported in **Table 4**. These differences are displayed in **Figure 2**. For all other effects not involving linguistic form, please see Appendix C.

TABLE 3 | Means and standard deviations for perceived social status by stereotypicality of professions, linguistic form, list, and participant gender.

Language	Stereotypicality of professions	Linguistic form	List	Participant gender			
				Female		Male	
				M	*SD*	*M*	*SD*
German	Feminine professions	Masculine	1	3.97	0.74	3.97	0.84
			2	4.44	0.71	3.98	0.45
			3	4.03	0.84	3.60	0.53
		Word pairs	1	3.98	0.73	3.98	0.63
			2	4.06	1.06	3.88	1.02
			3	3.78	0.87	4.03	0.84
	Masculine professions	Masculine	1	4.58	0.69	4.30	0.66
			2	4.57	0.65	4.28	0.58
			3	4.07	0.84	3.67	0.69
		Word pairs	1	4.68	0.62	4.77	0.78
			2	4.43	1.03	4.53	1.28
			3	4.02	0.76	4.17	0.83
Italian	Feminine professions	Masculine	1	3.82	0.62	4.20	0.70
			2	4.10	0.83	4.14	0.76
			3	3.99	0.78	3.98	0.61
		Word pairs	1	4.05	0.54	3.83	0.46
			2	3.79	0.48	3.83	0.83
			3	4.03	0.88	3.83	1.29
	Masculine professions	Masculine	1	3.96	0.71	3.89	0.53
			2	4.25	0.72	4.30	0.62
			3	3.84	0.62	4.18	0.64
		Word pairs	1	4.12	0.61	4.30	0.43
			2	4.07	0.48	4.15	0.64
			3	4.07	0.62	3.85	0.91

Ratings were given on a 7-point scale with higher scores indicating ascriptions of higher social status.

TABLE 4 | Means and standard deviations for estimated salary by stereotypicality of professions, linguistic form, and participant gender.

Language	Stereotypicality of professions	Linguistic form	List	Participant gender			
				Female		Male	
				M	SD	M	SD
German	Feminine Professions	Masculine	1	6.20	0.87	6.10	1.20
			2	6.34	1.11	5.75	0.92
			3	6.47	1.63	6.50	0.41
		Word pairs	1	5.69	0.71	5.86	0.98
			2	5.87	1.09	5.81	1.31
			3	5.85	0.98	5.75	0.88
	Masculine Professions	Masculine	1	7.06	0.77	6.86	0.71
			2	7.34	0.68	7.25	0.82
			3	7.41	0.95	6.93	0.84
		Word pairs	1	7.05	0.87	6.73	1.01
			2	6.97	1.09	7.47	1.22
			3	6.85	0.72	6.88	1.24
Italian	Feminine Professions	Masculine	1	6.53	1.41	6.27	1.27
			2	6.31	0.95	6.29	1.42
			3	6.63	0.68	6.31	1.14
		Word pairs	1	5.97	1.54	6.50	0.94
			2	5.72	0.71	6.15	0.78
			3	6.37	0.87	5.78	1.37
	Masculine Professions	Masculine	1	6.41	1.11	6.36	1.60
			2	6.81	0.94	7.04	0.81
			3	6.96	0.82	6.88	0.99
		Word pairs	1	6.47	1.34	7.20	0.76
			2	6.61	0.80	6.70	0.63
			3	6.98	0.95	6.33	1.25

Ratings were given on a 11-point scale with the midpoint (6) representing the national average salary. Higher scores indicate ascriptions of higher salary.

Women's Visibility

A main effect of linguistic form, $F_{(1, 361)} = 15.10$, $p < 0.001$, $\eta_p^2 = 0.04$, indicated that women's visibility was higher with word pairs than with masculine forms. This is in line with Hypothesis 3. Furthermore, the interaction of linguistic form and list was significant, $F_{(2, 361)} = 3.40$, $p = 0.034$, $\eta_p^2 = 0.02$. Word pairs (vs. masculine forms) increased the visibility of women for List 2 ($p = 0.028$) and List 3 ($p \leq 0.001$), but not for List 1. All means and standard deviations are reported in **Table 5**. For all other effects not involving linguistic form, please see Appendix C.

Ascribed Competence

The ANOVA on competence revealed no significant effects involving linguistic form. All means and standard deviations are reported in **Table 6**. For all other effects not involving linguistic form, please see Appendix C.
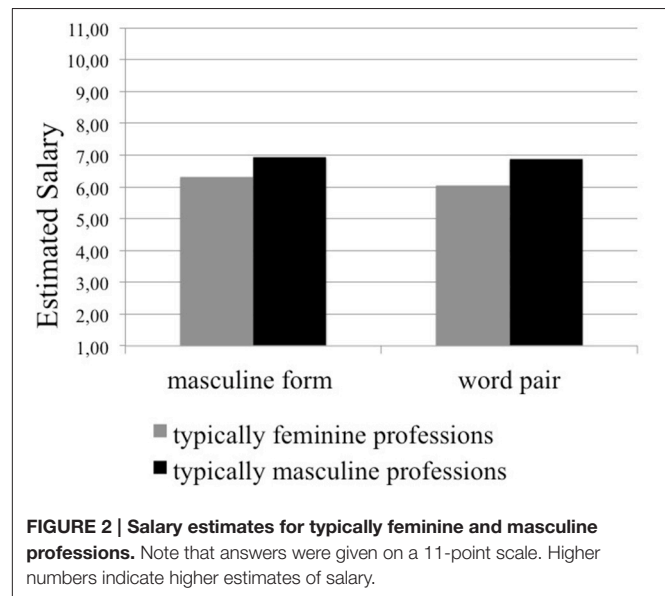

FIGURE 2 | Salary estimates for typically feminine and masculine professions. Note that answers were given on a 11-point scale. Higher numbers indicate higher estimates of salary.

Ascribed Warmth

The ANOVA on ascribed warmth revealed a significant interaction between linguistic form and language, $F_{(1, 363)} = 6.07$, $p = 0.014$, $\eta_p^2 = 0.02$. This was qualified, however, by the three-way interaction of linguistic form, language and participant gender, $F_{(1, 363)} = 6.64$, $p = 0.010$, $\eta_p^2 = 0.02$. Pairwise comparisons within languages revealed the following for Italian: men perceived professions to be warmer when presented with masculine forms in comparison to women ($p = 0.047$, $\eta_p^2 = 0.01$) and in comparison to word pairs ($p = 0.012$, $\eta_p^2 = 0.017$). All means and standard deviations are reported in **Table 7**. For all other effects not involving linguistic form, please see Appendix C.

DISCUSSION

The present research was designed to examine whether gender-fair language increases women's visibility and at the same time lowers status perceptions and salary estimates. We tested these effects in two languages with grammatical gender, Italian and German, within a single paradigm. Results mainly confirmed our hypotheses.

First of all, women's visibility increased for most professions when word pairs were used instead of masculine forms (see Hypothesis 3). This confirms the well-documented male bias in mental representation that is caused by masculine generics (e.g., Braun et al., 2005; Gabriel et al., 2008; Vervecken et al., 2013). With regard to the perceived social status of professions, typically feminine professions were ascribed significantly lower status than masculine professions, independent of linguistic form, which reflects the existing gender hierarchy (Eagly et al., 2000; Ridgeway and Correll, 2001). However, contrary to our expectations (see Hypothesis 1), professions did not lose in status when presented with word pairs compared to masculine forms. Instead, the difference in perceived social status between typically masculine and feminine professions increased when

TABLE 5 | Means and standard deviations for women's visibility by stereotypicality of professions, linguistic form, list, and participant gender.

Language	Stereotypicality of professions	Linguistic form	List	Participant gender			
				Female		Male	
				M	SD	M	SD
German	Feminine professions	Masculine	1	2.13	0.71	1.93	0.67
			2	2.38	0.79	1.98	1.18
			3	1.80	0.79	1.96	0.71
		Word pairs	1	2.53	0.65	2.23	0.68
			2	2.17	0.57	2.22	1.02
			3	2.40	0.66	2.05	0.87
	Masculine professions	Masculine	1	−1.84	0.77	−1.75	0.83
			2	−2.20	0.56	−2.18	0.64
			3	−2.39	0.47	−2.54	0.49
		Word pairs	1	−1.95	0.69	−1.75	0.96
			2	−2.16	0.73	−1.93	1.37
			3	−2.10	0.68	−2.17	0.72
Italian	Feminine professions	Masculine	1	1.99	0.88	2.27	0.79
			2	1.86	0.95	1.65	1.13
			3	1.45	0.77	1.16	0.82
		Word pairs	1	2.37	0.84	1.40	0.78
			2	2.06	0.89	2.13	0.94
			3	1.95	0.76	1.75	0.87
	Masculine professions	Masculine	1	−1.97	0.59	−2.09	0.53
			2	−2.44	0.58	−2.44	0.68
			3	−2.05	0.68	−2.09	0.69
		Word pairs	1	−1.99	0.77	−1.85	0.74
			2	−2.15	0.69	−2.18	1.29
			3	−1.87	0.38	−1.64	0.52

The standardized scale for feminine and masculine professions was calculated by using z-scores of a 11-point and a 7-point scale. Higher values indicate higher visibility of women.

TABLE 6 | Means and standard deviations for ascribed competence by stereotypicality of professions, linguistic form, list, and participant gender.

Language	Stereotypicality of professions	Linguistic form	List	Participant gender			
				Female		Male	
				M	SD	M	SD
German	Feminine professions	Masculine	1	4.70	0.95	4.34	0.86
			2	5.37	0.91	4.87	0.73
			3	5.23	1.06	4.54	0.71
		Word pairs	1	5.03	0.96	4.55	0.55
			2	5.01	1.20	4.92	1.38
			3	5.07	1.04	5.08	0.71
	Masculine professions	Masculine	1	4.89	0.74	4.65	0.53
			2	5.21	0.84	4.83	0.70
			3	5.33	0.63	4.97	0.66
		Word pairs	1	5.34	0.70	5.36	0.85
			2	4.91	1.15	4.85	1.35
			3	4.92	0.95	5.29	0.90
Italian	Feminine professions	Masculine	1	4.57	0.99	4.89	1.11
			2	5.23	0.89	5.42	0.77
			3	4.74	0.87	4.71	0.68
		Word pairs	1	4.75	0.78	4.46	0.54
			2	4.58	0.92	5.03	1.16
			3	4.67	0.75	4.21	1.53
	Masculine professions	Masculine	1	5.02	0.78	5.61	0.85
			2	4.82	1.06	5.17	1.08
			3	4.55	0.75	4.99	0.56
		Word pairs	1	5.34	0.85	4.94	0.67
			2	4.79	0.86	5.10	0.82
			3	4.47	0.86	4.73	1.10

Ratings were given on a 7-point scale with higher scores indicating ascriptions of higher competence.

word pairs were used, as feminine professions slightly lost and masculine professions slightly gained in social status. This finding has to be treated with caution, however, because the differences between masculine forms and word pairs were not significant when typically feminine and typically masculine professions were treated separately. Salary estimates for feminine professions were also generally lower than for masculine professions. For typically feminine professions salary estimates were even lower when word pairs where used rather than masculine forms (see Hypothesis 2). In contrast, masculine professions were not affected by linguistic form. This pattern confirms Hypothesis 2 at least partially. Taken together, we can only partly confirm the detrimental effects of gender-fair word pairs on status-related measures (perceived social status and salary-estimates; e.g., Vervecken et al., 2013; Vervecken and Hannover, 2015).

Our exploration of ascribed competence and warmth showed that—in line with first results of Vervecken et al. (2015)—competence was not affected by linguistic form. This is an important finding in view of the fact that competence is highly relevant in professional contexts. For warmth, we observed an unexpected secondary effect, in that only male Italian participants were affected by linguistic form. They generally ascribed more warmth to professions designated in the masculine and less warmth to professions referred to with word pairs. In view of the means and of an effect on ascribed warmth reported by Vervecken et al. (2015), we would agree with the authors in the guess that word pairs shifted perceptions of warmth toward the midpoint of the scale and thus balanced warmth perceptions. We have no theoretical explanation for this result except for the fact that some studies found men to be more sensitive to linguistic forms than women in certain contexts (Braun et al., 2005). More research would have to be conducted to clarify this issue.

Results from correlational analyses revealed that status-related measures, though correlated with each other,

TABLE 7 | Means and standard deviations for ascribed warmth by stereotypicality of professions, linguistic form, list, and participant gender.

Language	Stereotypicality of professions	Linguistic form	List	Participant gender			
				Female		Male	
				M	SD	M	SD
German	Feminine professions	Masculine	1	4.93	0.98	4.54	0.61
			2	4.53	0.58	4.47	0.31
			3	4.70	1.37	4.69	0.76
		Word pairs	1	5.23	0.90	5.04	0.60
			2	4.34	1.00	4.76	1.44
			3	4.57	1.26	4.87	0.63
	Masculine professions	Masculine	1	3.68	0.79	3.56	0.66
			2	4.12	0.56	3.70	0.67
			3	3.51	0.69	3.73	1.03
		Word pairs	1	3.87	0.88	4.11	0.63
			2	3.83	0.99	3.79	1.36
			3	3.86	1.05	4.16	1.03
Italian	Feminine professions	Masculine	1	4.27	1.14	5.01	1.07
			2	4.27	0.97	4.52	0.69
			3	4.13	0.69	4.38	0.51
		Word pairs	1	4.68	0.83	4.44	1.45
			2	4.08	0.49	4.11	0.54
			3	3.98	0.77	3.79	1.21
	Masculine professions	Masculine	1	3.50	0.99	4.11	1.28
			2	3.80	0.93	3.92	0.75
			3	3.55	0.83	3.44	0.89
		Word pairs	1	3.88	0.76	2.90	1.02
			2	3.64	0.64	3.76	0.67
			3	3.60	1.04	3.28	0.63

Ratings were given on a 7-point scale with higher scores indicating ascriptions of higher warmth.

are largely unrelated with ratings of women's visibility (i.e., gender-typicality). Thus, the two dimensions of status and gender-typicality appear to be independent of each other. This is in line with research by Glick et al. (1995) and Gottfredson (1996), which suggests that with respect to images of occupations status/prestige and gender-typicality are two orthogonal dimensions.

The present findings contribute to social role theory (Eagly, 1987; Eagly et al., 2000) in the following ways: Word pairs increase the inclusion of women in comparison to masculine forms and thus alter the perceived distribution of women and men across professional groups. In this way they affect occupational gender stereotyping. Our results are in accord with findings showing that the social status of professions does not readily translate into gender stereotypes ascribed to professions (Eagly and Steffen, 1984). Furthermore, given that fictitious and experimentally varied distributions of women and men in future professions can change ascribed gender stereotypes despite

currently existing stereotypes (Koenig and Eagly, 2014), the use of word pairs might change occupational gender stereotyping on the long run, too. Further evidence for this idea comes from recent research which shows that linguistic forms in job advertisements for a typically masculine, high-status leadership position changed hiring decisions: Women and men were hired equally when word pairs (vs. masculine forms) were used in the respective job advertisement (Horvath and Sczesny, 2015). Moreover, girls' interest in typically masculine professions was found to be higher when these professions were presented with word pairs instead of masculine forms (Vervecken et al., 2015).

Our findings extend prior research by investigating, for the first time, whether beneficial and detrimental effects of gender-fair language on the social perception of professional groups emerge simultaneously: While previous studies focused mainly on individual professions (e.g., Braun et al., 2005; Gabriel et al., 2008) or individual targets (Merkel et al., 2012; Formanowicz et al., 2013), the current study sheds light on the social perception of a range of typically feminine and masculine professions. More importantly, our study was designed to capture both beneficial and detrimental effects of linguistic forms within a single paradigm. Earlier studies focused either on an increase in women's visibility (e.g., Braun et al., 2005; Gabriel et al., 2008) or on negative side-effects of gender-fair language (Mucchi-Faina, 2005; Merkel et al., 2012; Formanowicz et al., 2013). Our study shows that gender-fair language can simultaneously have positive effects (greater visibility of women) and negative effects (polarization of male-female differences in pay). Note, however, that the present study investigated only descriptive norms (how much status and pay does a given profession currently enjoy?) but not prescriptive norms (how much status and pay should a given profession enjoy?). Thus, it remains to be seen whether word pairs have detrimental effects on prescriptive norms as well.

One limitation of the present research is that we applied a between-participants design and presented participants either with masculine "generic" forms or with word pairs. Current language policies, however, demand the use of a whole range of gender-fair forms, including word pairs as well as other alternatives (for an overview of German gender-fair forms see Braun et al., 2005; Horvath, 2015). Consequently, speakers are likely to encounter many different forms in everyday life. Future research should therefore use a more ecologically valid approach and expose participants to diverse linguistic forms. In particular, future research should include gender-neutral expressions (e.g., German *Lehrkräfte*, teaching staff), which were not investigated here. In contrast to masculine forms and word pairs, gender-neutralizing forms make neither women nor men salient. It remains to be tested whether such neutral forms can increase women's visibility without reducing estimated salaries in comparison to word pairs. The finding that lower salaries are assumed for typically feminine professions designated with word pairs in contrast to masculine forms may simply reflect social reality, given that professions with a high percentage of women tend to be connected with lower salaries, lesser status, and lesser recognition in society. It is therefore conceivable that feminine-masculine word pairs designating typically feminine professions (vs. masculine forms)

automatically activate knowledge about the gender wage gap or associations of men and wealth (Williams et al., 2010), while gender-neutral expressions do not. Such effects could be tested by measuring unconscious associations between professional groups (designated in the masculine, with word pairs or neutralizations) and the gender-wage gap, for example with implicit association tests.

Preliminary evidence has shown that in reference to typically feminine professions, feminine-only forms (e.g., *Kindergärtnerinnen*, kindergarden teachers, fem.) are more frequently used than masculine generics (*Kindergärtner*, kindergarden teachers, masc.) or word pairs (Chiarini, 2013; Hodel et al., 2013). Hence, future research should compare status and salary perceptions for typically feminine professions designated with word pairs compared to feminine forms. In this case, word pairs might actually cause an increase in estimated salaries and status perceptions, because masculine forms are here added to the feminine forms already in use. If this assumption should hold, it would again speak for a consistent use of gender-fair language, which in this case would mean replacing feminine-only forms with word pairs.

Now what are the practical implications of our results? Should word pairs be used to make language gender-fair and to support gender equality? The present findings indicate that, in German and Italian, language reform—and hence use of word pairs—is promising as they are likely to increase women's professional visibility on the one hand. On the other hand, word pairs in comparison to masculine forms may also lower estimated salaries of typically feminine professions. These effects appear to be inevitable for the time being. However, negative consequences of gender-fair language may diminish over time (Formanowicz et al., 2015; Gustafsson Senden et al., 2015). Furthermore, masculine generics are semantically ambiguous and thus problematic: they can refer to men only or to a group of women and men (Stahlberg et al., 2007). Therefore, we would recommend the use of gender-fair forms, such as word pairs or neutralizations in response to the question whether one should use gender-fair forms or masculine generics.

Taken together, our results on the social perception of professions indicate an increase of women's visibility with gender-fair language, but also a decrease in salary estimates of typically feminine professions. Although the latter effect is not negligible, social perceptions of status and competence do not suffer when word pairs are used.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.02018

REFERENCES

Alksnis, C., Desmarais, S., and Curtis, J. (2008). Workforce segregation and the gender wage gap: is "Women's" work valued as highly as "men's"? *J. Appl. Soc. Psychol.* 36, 1416–1441. doi: 10.1111/j.1559-1816.2008.00354.x

American Psychological Association (2010). *Ethical Principles of Psychologists and Code of Conduct.* Available online at: http://www.apa.org/ethics/code/principles.pdf

Arbeitsmarktservice Österreich (2015). *Berufslexikon.* Available online at: http://www.berufslexikon.at

Becker, J. C., Glick, P., Ilic, M., and Bohner, G. (2011). Damned if she does, damned if she doesn't: consequences of accepting versus confronting patronizing help for the female target and male actor. *Eur. J. Soc. Psychol.* 41, 761–773. doi: 10.1002/ejsp.823

Bem, S., and Bem, D. (1973). Does sex-biased job advertising "aid and abet" sex discrimination? *J. Appl. Soc. Psychol.* 3, 6–18. doi: 10.1111/j.1559-1816.1973.tb01290.x

Binggeli, S., Krings, F., and Sczesny, S. (2014). Perceived competition explains regional differences in the stereotype content of immigrant groups. *Soc. Psychol.* 55, 62–70. doi: 10.1027/1864-9335/a000160

Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). "Sex, syntax and semantics," in *Language in Mind: Advances in the Study of Language and Thought*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 61–79.

Braun, F., Sczesny, S., and Stahlberg, D. (2005). Cognitive effects of masculine generics in German: an overview of empirical findings. *Communications* 30, 1–21. doi: 10.1515/comm.2005.30.1.1

Cejka, M. A., and Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Pers. Soc. Psychol. Bull.* 25, 413–423. doi: 10.1177/0146167299025004002

Chiarini, T. (2013). "Occupational terms in German, Italian and Czech: a translation study," in *Presentation at the Conference on Language, Cognition and Gender,* June 13–17 (Bern).

Cuddy, A. J. C., Fiske, S. T., and Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *J. Soc. Issues* 60, 701–718. doi: 10.1111/j.0022-4537.2004.00381.x

Cuddy, A. J. C., Fiske, S. T., and Glick, P. (2007). The BIAS-Map: behaviors from intergroup affect and stereotypes. *J. Pers. Soc. Psychol.* 92, 631–648. doi: 10.1037/0022-3514.92.4.631

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., et al. (2009). Stereotype content model across cultures: towards universal similarities and some differences. *Br. J. Soc. Psychol.* 48, 1–33. doi: 10.1348/014466608X314935

Eagly, A. H. (1987). *Sex Differences in Social Behaviour.* Hillsdale, MI: Lawrence Erlbaum Associates.

Eagly, A. H., and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. doi: 10.1037/0033-295X.109.3.573

Eagly, A. H., and Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *J. Pers. Soc. Psychol.* 46, 735–754. *doi: 10.1037/0022-3514.46.4.735*

Eagly, A. H., Wood, W., and Diekman, A. B. (2000). "Social role theories of sex differences and similarities: a current appraisal," in *The Developmental Social Psychology of Gender*, eds T. Eckes and H. M. Trautner (Mahwah, NJ: Lawrence Erlbaum Associates), 123–174.

European Commission (2011). *Report on Progress on Equality between Women and Men in 2010. The Gender Balance in Business Leadership.* Luxembourg: Publications Office of the European Union. doi: 10.2767/99441

European Institute for Gender Equality (2013). *Gender Equality Index Report.* doi: 10.2839/9948. Retrieved from: http://eige.europa.eu/sites/default/files/documents/Gender-Equality-Index-Report.pdf

Fiske, S. T., Cuddy, A. J. C., Glick, P., and Xu, J. (2002). A Model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902. doi: 10.1037/0022-3514.82.6.878

Formanowicz, M., Bedynska, S., Cislak, A., Braun, F., and Sczesny, S. (2013). Side effects of gender-fair language: how feminine job titles influence the evaluation of female applicants. *Eur. J. Soc. Psychol.* 43, 62–71. doi: 10.1002/ejsp.1924

Formanowicz, M. M., Cislak, A., Horvath, L. K., and Sczesny, S. (2015). Capturing socially motivated linguistic change. How the use of gender-fair language afffects support for social initiatives in Austria and Poland. *Front. Psychol.* 6:1617. doi: 10.3389/fpsyg.2015.01617

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au pairs are rarely male: norms on the gender perception of role names across English, French, and German. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206

Glick, P. (1991). Trait-based and sex-based discrimination in occupational prestige, occupational stalary, and hiring. *Sex Roles* 25, 351–378. doi: 10.1007/BF00289761

Glick, P., and Fiske, S. T. (1996). The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *J. Pers. Soc. Psychol.* 70, 491–512. doi: 10.1037/0022-3514.70.3.491

Glick, P., Wilk, K., and Perreault, M. (1995). Images of occupations: components of gender and status in occupational stereotypes. *Sex Roles* 32, 565–582. doi: 10.1007/BF01544212

Gottfredson, L. S. (1981). Circumscription and compromise: a developmental theory of occupational aspirations. *J. Coun. Psychol. Monogr.* 28, 545–579. doi: 10.1037/0022-0167.28.6.545

Gottfredson, L. S. (1996). "Gottfredson's theory of circumscription and compromise," in *Career Choice and Development*, eds D. Brown and L. Brooks (San Francisco, CA: Jossey-Bass), 179–232.

Gustafsson Sendén, M., Bäck, E. A., and Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Front. Psychol.* 6:893. doi: 10.3389/fpsyg.2015.00893

Hausmann, R., Tyson, L. D., and Zahidi, S. (2010). *The Global Gender Gap Report 2010.* Geneva: World Economic Forum.

Hodel, L., von Stockhausen, I., Formanowicz, M., Sczesny, S., and Valdrová, J. (2013). "How women and men are addressed in job advertisements," in "*Presentation at the EAWOP Small Group Meeting "Gender Equality in Organizations*" (Frankfurt).

Horvath, L. K. (2015). "Gender-fair language in the context of recruiting and evaluating leaders," in *Auswahl und Beurteilung von Frauen und Männern als Führungskräfte in der Wirtschaft – Herausforderungen, Chancen und Lösungen*, eds I. M. Welpe, P. Brosi, L. Ritzenhöfer, and T. Schwarzmüller (Heidelberg: Springer), 263–272.

Horvath, L. K., and Sczesny, S. (2015). Reducing the lack of fit for women with leadership positions? Effects of the wording in job advertisements. *Eur. J. Work Organ. Psychol.* doi: 10.1080/1359432X.2015.1067611. [Epub ahead of print].

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456. doi: 10.1007/s10936-007-9053-z

Irmen, L., and Roßberg, N. (2004). Gender markedness of language: the impact of grammatical and nonlinguistic information on the mental representation of person nouns. *J. Lang. Soc. Psychol.* 23, 272–307. doi: 10.1177/0261927X04266810

Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023599719948

Koenig, A. M., and Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: observations of groups' roles shape stereotypes. *J. Pers. Soc. Psychol.* 107, 371–392. doi: 10.1037/a0037215

Krefting, L. A., Berger, P. K., and Wallace, M. K. Jr. (1978). The contribution of sex distribution, job content, and occupational classification to job sextyping: two studies. *J. Vocat. Behav.* 13, 181–191. doi: 10.1016/0001-8791(78)90043-X

Lucy, J., and Shweder, R. (1979). Whorf and his critics: linguistic and nonlinguistic influences on color memory. *Am. Anthropol.* 81, 581–615. doi: 10.1525/aa.1979.81.3.02a00040

Merkel, E., Maass, A., and Frommelt, L. (2012). Shielding women against status loss. The masculine form and its alternatives in Italian. *J. Lang. Soc. Psychol.* 31, 311–320. doi: 10.1177/0261927X12446599

Mucchi-Faina, A. (2005). Visible or influential? Language reforms and gender (in)equality. *Soc. Sci. Inf.* 44, 189–215. doi: 10.1177/0539018405050466

Percy, E. J., Sherman, S. J., Garcia-Marques, L., Mata, A., and Garcia-Marques, T. (2009). Cognition and native-language grammar: the organizational role of adjective-noun word order in information representation. *Psychon. Bull. Rev.* 16, 1037–1042. doi: 10.3758/PBR.16.6.1037

Prewitt-Freilino, J. L., Caswell, T., and Laakso, E. K. (2012). The gendering of language: a comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles* 66, 268–281. doi: 10.1007/s11199-011-0083-5

Ridgeway, C. L. (2001). Gender, status, and leadership. *J. Soc. Issues* 57, 637–655. doi: 10.1111/0022-4537.00233

Ridgeway, C. L., and Correll, S. J. (2001). Unpacking the gender system: a theoretical perspective on gender beliefs and social relations. *Gender Soc.* 18, 510–531. doi: 10.1177/0891243204265269

Schein, V. E. (1973). The relation between sex role stereotypes and requisite management characteristics. *J. Appl. Psychol.* 97, 95–100. doi: 10.1037/h0037128

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *J. Soc. Issues* 57, 675–688. doi: 10.1111/0022-4537.00235

Sczesny, S., Moser, F., and Wood, W. (2015). Beyond sexist beliefs: how do people decide to use gender-inclusive language? *Pers. Soc. Psychol. Bull.* 41, 943–954. doi: 10.1177/0146167215585727

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language," in *Social Communication*, ed K. Fiedler (New York, NY: Psychology Press), 163–187.

Stahlberg, D., Sczesny, S., and Braun, F. (2001). Name your favorite musician: effects of masculine generics and of their alternatives in German. *J. Lang. Soc. Psychol.* 20, 464–469. doi: 10.1177/0261927X01020004004

Stout, J. G., and Dasgupta, N. (2011). When *he* doesn't mean *you*: gender-exclusive language as ostracism. *Pers. Soc. Psychol. Bull.* 37, 757–769. doi: 10.1177/0146167211406434

Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics, 3rd Edn.* New York, NY: Harper-Collins.

UNESCO (1999). *Guidelines for Gender-Neutral Language.* Available online at: http://unesdoc.unesco.org/images/0011/001149/114950mo.pdf

Vervecken, D., Gygax, P., Gabriel, U., Guillod, M., and Hannover, B. (2015). Warm-hearted businessmen, competitive housewives? Effects of gender-fair language on adolescents' perceptions of occupations. *Front. Psychol.* 6:1437. doi: 10.3389/fpsyg.2015.01437

Vervecken, D., and Hannover, B. (2015). Yes I can! Effects of gender fair job descriptions on children's perceptions of job status, Job difficulty, and vocational self-efficacy. *Soc. Psychol.* 46, 76–92. doi: 10.1027/1864-9335/a000229

Vervecken, D., Hannover, B., and Wolter, I. (2013). Changing (S)expectations: how gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *J. Vocat. Behav.* 82, 208–220. doi: 10.1016/j.jvb.2013.01.008

Williams, M. J., Paluck, E. L., and Spencer-Rodgers, J. (2010). The masculinity of money: automatic stereotypes predict gender differences in estimated salaries. *Psychol. Women Q.* 34, 7–20. doi: 10.1111/j.1471-6402.2009.01537.x

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785. doi: 10.1073/pnas.0701644104

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Warm-hearted businessmen, competitive housewives? Effects of gender-fair language on adolescents' perceptions of occupations

Dries Vervecken[1]*, Pascal M. Gygax[2], Ute Gabriel[3], Matthias Guillod[2] and Bettina Hannover[4]

[1] *Karel de Grote University College, Antwerp, Belgium,* [2] *University of Fribourg, Fribourg, Switzerland,* [3] *Norwegian University of Science and Technology, Trondheim, Norway,* [4] *Free University of Berlin, Berlin, Germany*

Recent studies from countries with grammatical gender languages (e.g., French) found both children and adults to more frequently think of female jobholders and to consider women's success in male dominated occupations more likely when the jobs were described in pair forms (i.e., by explicit reference to male and female jobholders, e.g., *inventeuses et inventeurs*; French feminine and masculine plural forms for *inventors*), rather than masculine only forms (e.g., *inventors*). To gain a better understanding of this phenomenon, we systematically varied the gender connotation of occupations (males overrepresented, females overrepresented, equal share of males and females) and measured additional dependent variables, predicting that gender fair language would reduce the impact of the gender connotation on participants' perceptions. In a sample of 222 adolescents (aged 12–17) from French speaking Switzerland, we found that pair forms attenuated the difference in the ascription of success to male and female jobholders in gendered occupations and attenuated the differential ascription of warmth to prototypical jobholders in male vs. female dominated jobs. However, no effect of language form on the ascription of competence was found. These findings suggest that language policies are an effective tool to impact gendered perceptions, however, they also hint at competence-related gender stereotypes being in decline.

Keywords: gender stereotypes, gender-fair language, adolescence, stereotype content, stereotype change

Introduction

In recent years, the use of so-called gender-fair language has been strongly promoted. This language reform reflects the assumption that language, here gender-fair language, is a tool to influence people's gendered perception of reality. For example, with respect to occupations, studies with adults and primary school children from countries with grammatical gender languages (e.g., French, German, Dutch, Spanish, Italian) suggest that they are perceived in a less gender-typed manner when they are described in gender-fair language, more specifically in pair forms (i.e., by explicit reference to both male and female jobholders, e.g., *inventeuses et inventeurs*; French feminine and masculine plural forms for inventors), rather than masculine plural forms (e.g., inventors; Braun et al., 1998; Heise, 2000, 2003; Stahlberg and Sczesny, 2001; Stahlberg et al., 2001; Rothmund and Scheele, 2004; Vervecken et al., 2013; Vervecken and Hannover, 2015; for

a discussion of this issue for natural gender languages e.g., English; see, e.g., Gabriel et al., 2008; Garnham et al., 2012; Lassonde and O'Brien, 2013).

For example, in an experiment by Stahlberg and Sczesny (2001, Study 1), adult German participants were asked to write down the name of their favorite musicians or athletes. Participants received these instructions with either a masculine only form (*Musiker* [male or generic musician], *Sportler* [male or generic athlete]) or a pair form (Musikerin/Musiker [female/male musician]; Sportlerin/Sportler [female/male athlete]). Results showed that participants who had received the role nouns in a pair form listed more female personalities then participants in the masculine only condition. Similar results have been reported with German and Dutch speaking Belgian primary school children as young as 6 years of age (Vervecken et al., 2013). Vervecken et al. (2013) (Studies 2 and 3) investigated primary school children's perceptions of females' and males' success (i.e., who can succeed?) in traditionally male occupations. The occupational titles were presented to the children either in a masculine only form or a pair form. After being presented with an occupational title, children were asked "*who can succeed in this occupation*" and to indicate their response on a five-point scale (ranging from *1 = only men* to *5 = only women*): children in the pair form condition systematically perceived females' and males' success more equally than children in the masculine only form condition who attributed success predominantly to males.

To gain a better understanding of the above described phenomena, in this study, we wanted to investigate more systematically how language forms interact with the gender connotation of an occupation.

Proportions of Males and Females Working in an Occupation Shape Gendered Perceptions of Prototypical Job Holders

Social role theory predicts that people make inferences about social groups from their typical social roles, for example occupational roles (e.g., Eagly and Koenig, 2014). Applied to gender, professions in which either males or females are clearly overrepresented will be the ones from which perceivers infer gender stereotypes (e.g., attributes which females or males supposedly have): women are traditionally seen as 'communal' (warm), e.g., nurturing or well-intended based on the social roles they are more likely to perform than men (e.g., nurse). By the same token, men are perceived as 'agentic' (competent), e.g., competitive or efficient as a result of the social roles which they more often have than women (e.g., manager; see Diekman and Eagly, 2000).

Empirical support comes from research showing a correspondence between the proportion of males and females working in an occupation and the ascription of gendered attributes to the prototypical job holder (e.g., Cejka and Eagly, 1999; Crawley, 2014; Eagly and Koenig, 2014): to the extent that an occupational group is perceived as dominated by women (e.g., childminder), people tend to believe that feminine qualities are required to be successful within these occupations (e.g., warm-hearted). When an occupational group is perceived as dominated by men (e.g., stock broker), people tend to

believe that masculine qualities are essential for workers to be successful (e.g., competitive). Further support comes from experimental studies manipulating the distribution of males and females in occupations. For example Crawley (2014) varied the percentages of women and men who allegedly worked in different occupations. Participants were more likely to indicate that a university degree was needed if the occupation was, supposedly, primarily occupied by men than if the job was described as dominated by women.

When a social role, like an occupation, is described in a linguistic pair form (e.g., businesswomen and businessmen, housemen and housewives), explicit reference is made to both males and females. Considering the findings described above, we speculated that if it is a profession in which one gender is overrepresented, people should be less inclined to ascribe the characteristics of that occupation to the respective gender group, as the pair form makes them think of both genders when describing the prototypical job holder. For example, descriptions of an occupational group from a male dominated field in a masculine only form, like "businessmen," will most likely trigger associations with stereotypically male, i.e., agentic traits: "businessmen are competent and self-confident people." Describing the same occupational group in a pair form, like "businesswomen and businessmen," may additionally trigger associations with stereotypically female, i.e., communal traits, such as "helpful" and/or "friendly."

Warmth and Competence as Core Dimensions of Gendered Perceptions of Prototypical Job Holders

To investigate whether the ascription of gendered attributes to prototypical job holders is influenced by the linguistic form (pair vs. masculine only) we used the dimensions of *warmth* and *competence*. According to the Stereotype Content Model (SCM, Fiske et al., 2002, 2007; Fiske, 2011), warmth and competence are two universal dimensions that guide people's perception of others. Specifically, these two dimensions are driven by the need to evaluate whether others (a) have beneficial intentions for oneself and for one's group (i.e., warmth dimension) and (b) have the ability to implement their good/bad intentions (i.e., competence dimension). Although women in general often receive more positive evaluations, women occupying more traditional roles (e.g., housewives) are perceived as warm but incompetent and those in non-traditional roles (i.e., businesswomen) as cold yet competent. We suggest that describing an occupational group in its masculine form only, like "successful businessmen," may result in ascriptions of coldness and competence: although they are perceived as having bad intentions (e.g., people who sell something solely for personal gain), they are still perceived as competent (e.g., as very good at making money for themselves). Describing the same occupational group in a pair form ("businesswomen and businessmen") should result in the ascription of comparably more warmth (e.g., people who want to sell useful things) and less competence (e.g., people who, after all, do not earn more money than people in other domains).

Indirect support for the assumption that gender-fair language might affect perceptions of warmth and competence comes

from research comparing the impact of the masculine singular vs. the feminine singular form for job titles on evaluations of those jobs and their suitability for female applicants (Merkel et al., 2012; Formanowicz et al., 2013; Budziszewska et al., 2014). For example Formanowicz et al. (2013) demonstrated with both invented (Studies 1–2) and existing (Study 3) job titles that female applicants described with a feminine job title were evaluated as being less competent than applicants described with a masculine job title. While job titles in the feminine form lead to some devaluation of competence, women are often evaluated more favorably than men on warmth (cf. the "women are wonderful effect," Eagly and Mladinic, 1994). Merkel et al. (2012) illustrated that the "women are wonderful-effect" can be induced by the linguistic form used to describe a professional: female professionals described with a feminine title (e.g., *avvocata* [female lawyer]) were judged as warmer than professionals described by a masculine title (e.g., *avvocato* [male lawyer]). Although these studies illustrate that male vs. female job titles trigger different perceptions of female professionals, no research has explored whether pair forms vs. masculine only forms have a different impact on gendered perceptions of an occupational group in general (i.e., prototypical female and male workers in a given occupation).

Adolescence as Crucial Stage in Vocational Development

We will test our hypotheses with adolescents aged 12–17. Existing research on language effects was primarily done with children, focusing on the emergence of gendered linguistic concepts in the primary school years (e.g., Hyde, 1984; Schau and Scott, 1984; Vervecken et al., 2013; Vervecken and Hannover, 2015), or with adults, focusing on the practical importance of the use of different linguistic forms in everyday life, such as in job advertisements (see Stahlberg et al., 2007 for an overview). In contrast, research with adolescents is almost non-existent (see Chatard et al., 2005 for a noticeable exception). This is an unsatisfactory situation, as the transition from adolescence to adulthood is an important stage in vocational development, in which the gendered perception of occupations can play an essential role (Gottfredson, 2005; Lerner and Steinberg, 2009). Adolescence is a crucial stage in vocational development as youngsters get more realistic about their future career options and start to abandon unrealistic aspirations (Helwig, 2001; Blanchard and Lichtenberg, 2003; Hartung et al., 2005). However, perceptions of what is required to pursue different professions are often biased by gender stereotypes (e.g., Crawley, 2014), which are a result of associating occupations with one of the two genders (White and White, 2006; Eagly and Koenig, 2014).

While not directly investigated in this study, describing potential future professions to adolescents in gender-fair language may help to reduce the restrictions that boys, and more particularly girls, impose on themselves when deciding which occupations to aspire to.

Research Hypotheses

In sum, the present study seeks to investigate the impact of linguistic forms (pair forms compared to masculine only forms) used to describe occupations in which either males are overrepresented, females are overrepresented (male or female gendered occupations), or in which males and females are represented about equally (gender-neutral occupation), on adolescents' perceptions of these occupations. More specifically, we wanted to replicate the finding of previous studies conducted with children or adults, which show that linguistic forms impact the perception of the extent to which women and men can succeed in these occupations. In addition, we wanted to investigate the effect of linguistic forms on the ascription of warmth and competence to people performing these occupations. Against the background of the above described findings, we speculated that when presented with a profession in which one gender is overrepresented, people should be less inclined to ascribe the characteristics of that occupation (warmth, competence) to the respective gender group, as the pair form makes them think of both genders when describing the prototypical job holder.

Our hypotheses were as follows:

The use of pair forms (compared to masculine only forms) to describe occupations:

(1) will attenuate the difference in the ascription of success to males and females in gendered occupations (i.e., the deviation from the midpoint of the answering scale, indicating that males and females alike can succeed in the job, should be smaller);

(2) will attenuate the influence of the distribution of males and females in that occupation on the differential ascription of warmth to prototypical job holders (i.e., the difference in the ascription of warmth to holders of female vs. male jobs should become smaller);

(3) will attenuate the influence of the distribution of males and females in that occupation on the differential ascription of competence (i.e., the difference in the ascription of competence to holders of female vs. male jobs should become smaller).

Materials and Methods

Participants

Two hundred and twenty-two ($N = 222$) French-speaking adolescents from two different schools in Porrentruy (French speaking part of Switzerland) took part in this experiment (mean age = 14; range = 12–17; 114 female, 107 male, one participant did not indicate his/her gender). One female participant was removed from the analyses as she did not follow the instructions. This study was approved by the Ethics Committee of the Department of Psychology (University of Fribourg) and carried out in accordance with their recommendations. All participants have granted informed consent.

Materials and Procedure

Participants (in group sessions) were orally presented with fifteen occupations (i.e., five female stereotyped, five male stereotyped, and five gender-neutral; see **Table 1**). Occupations were presented

Assumed gender distribution within occupations	French	English translations
Male dominated	Camionneurs (et camionneuses)	Male (and female) truck drivers
	Inventeurs (et inventeuses)	Male (and female) inventors
	Maçons (et maçonnes)	Male (and female) bricklayers
	Mécaniciens (et mécaniciennes) sur auto	Male (and female) car mechanics
	Informaticiens (et informaticiennes)	Male (and female) computer scientist
Female dominated	Infirmiers (et infirmières)	Male (and female) nurse
	Babysitters (et babysittrices)	Male (and female) babysitters
	Nettoyeurs (et nettoyeuses)	Male (and female) cleaners
	Esthéticiens (et esthéticiennes)	Male (and female) beauticians Male (and female) preschool teacher
	Educateurs (et éducatrices) de la petite enfance	
Approximately equally distributed	Ecrivains (et écrivaines)	Male (and female) writers
	Chanteurs (et chanteuses)	Male (and female) singers
	Pharmaciens (et pharmaciennes)	Male (and female) pharmacists
	Sportifs (et sportives)	Male (and female) athletes
	Musiciens (et musiciennes)	Male (and female) musicians

one after another in a set random sequence, which was the same for all participants. Participants were instructed to rate each occupation in a booklet on a series of 15 dimensions (for a description of the rating method see below). The experimenter waited until all participants were finished with rating an occupation before going on to the next one.

To manipulate the distributions of males and females in occupations, we selected 15 occupations (see **Table 1**) from a list of 126 role nouns which had been normed with respect to the representation of males and females (in %) in the respective occupational group (Gabriel et al., 2008; Irmen and Schumann, 2011). We used the cut-off value >70% men to define male dominated jobs, >70% women to define female dominated jobs, and both men and women <60% to define gender neutral occupations.

To manipulate linguistic form, for half of the participants ($N = 117$) the occupational titles were orally presented in the masculine only form; the other half ($N = 105$) received the same occupational titles in the pair form. Each job title was orally accompanied by a short description of the jobholder's tasks and activities. These descriptions were identical in both conditions.

To measure ascriptions of gendered attributes to prototypical job holders in the different occupations, we used the six items referring to warmth (e.g., *friendly, well-intended*) and the six items referring to competence (e.g., *efficient, expert*) from the scale of Fiske et al. (2002). Immediately after an occupation had

been presented, participants were asked to indicate on five-point Likert scales (1 = *not at all*, 5 = *extremely*) how competent (e.g., efficient, expert) and warm (e.g., friendly, well-intended) they thought prototypical job holders would be. Cronbach's alpha values were: for competence regarding female dominated occupations ($\alpha = 0.93$), male dominated occupations ($\alpha = 0.87$), and gender-neutral occupations ($\alpha = 0.90$). Cronbach's alpha values were: for warmth regarding female dominated occupations ($\alpha = 0.91$), male dominated occupations ($\alpha = 0.93$) and gender-neutral occupations ($\alpha = 0.93$).

Finally, to measure perceptions of male and female success in the different occupations, we asked participants to indicate on a five-point Likert scale who they thought would succeed in each occupation (1 = *only men, 3 = men and women alike, 5 = only women*, Cronbach's alpha values were: for female dominated occupations: $\alpha = 0.71$; male dominated occupations $\alpha = 0.73$; gender-neutral occupations $\alpha = 0.34$).

For each dependent variable, means were calculated separately for female dominated, male dominated, and gender-neutral occupations[1]. All rating scales were labeled numerically and presented with equidistant markings to ensure that the scales were considered as continuous, hence reliable (Krosnick and Berent, 1993). Data were analyzed using parametric statistics as each subscale consisted of at least five items (Boone and Boone, 2012). In some cases (i.e., ascriptions of success), normality tests (i.e., Kolmogorov–Smirnov) indicated non-normal distributions. In these cases, we ran additional non-parametric statistics. As our sample size was relatively large (all $n > 30$; Hays, 1994), and as there was no difference between the two statistics (unless otherwise stated), we only present the results from the parametric statistics.

Results

Differential Ascription of Success to Male and Female Jobholders in Gendered Occupations

To test our first hypothesis that language forms would impact the ascription of success to men and women, we conducted a 2 (*Form*: Pair form vs. Masculine only) × 2 (*Gender of respondent*: Female vs. Male) × 3 (*Assumed gender distribution within occupations*: Female vs. Male vs. Neutral) factorial mixed ANOVA on the *Perceived success of men and women*, with *Age* as a covariate, *Form* and *Gender of respondent* as between-participant factors and *Assumed gender distribution* as a within-participant factor.

Results showed a main effect of Assumed gender distribution, $F(2,430) = 4.07$, $p < 0.05$, $\eta = 0.02$, suggesting that for female dominated jobs, success was considered more likely for women than for men ($M = 3.56$, $SE = 0.03$). In contrast, men were perceived to more likely succeed in male dominated jobs than

[1]Note that, for each job, we also added a self-efficacy question (i.e., *"Imagine you wanted to become [job], how confident are you that you would pass the qualification test required to do this job?,"* 1 = *not confident at all* to 5 = *extremely confident*), to mimic Chatard et al. (2005) and Vervecken and Hannover (2015). Although we did find, as in previous studies, that girls felt more confident about female stereotyped jobs, and boys more confident about male stereotyped jobs, contrary to what Chatard et al. (2005) had found, our experimental manipulation had no effect on adolescents' self-efficacy beliefs.

women ($M = 2.24$, $SE = 0.03$, $p < 0.001$). As expected, for gender-neutral occupations, i.e., a job in which the genders are represented about equally, participants' mean response reflected the midpoint of the answering scale: women were considered as likely to succeed as men ($M = 2.99$, $SE = 0.01$, pairwise LSD comparison at $p < 0.001$).

In support of hypothesis 1, the analysis revealed an interaction between *Assumed gender distribution* and *Form*: $F(2,430) = 12.73$, $p < 0.001$, $\eta^2 = 0.06$, indicating that compared to the masculine only condition, in the pair form condition participants' mean responses were closer to the midpoint of the answering scale ($3 = $ *women and men can succeed equally*). When occupations were presented in pair form, rather than the masculine only form, the perception that women and men can equally succeed in occupations increased for male dominated occupations [Masculine Form: $M = 2.11$, $SE = 0.04$; Pair Form: $M = 2.38$, $SE = 0.04$, $t(219) = -4.51$; $p < 0.05$] and female dominated occupations [Masculine Form: $M = 3.63$, $SE = 0.04$; Pair Form: $M = 3.48$, $SE = 0.05$, $t(219) = 2.46$; $p < 0.05$[2]]. Hence, the deviation from the midpoint of the answering scale (indicating differential ascription of success to men and women) was more pronounced in the masculine only condition [Male dominated occupations: $M = 3.63$, $SE = 0.03$; Female dominated occupations: $M = 2.11$, $SE = 0.04$, $t(116) = 25.03$; $p < 0.001$; Cohen's $d = 2.32$] than in the pair form condition [Male dominated occupations: $M = 3.48$, $SE = 0.06$; Female dominated occupations: $M = 2.38$, $SE = 0.05$, $t(104) = 11.74$; $p < 0.05$; Cohen's $d = 1.14$]. Interestingly, perceived success of women and men in gender neutral occupations, i.e., jobs in which the genders are represented about equally was also influenced by the linguistic form: [Masculine Form: $M = 2.96$, $SE = 0.02$; Pair Form: $M = 3.04$, $SE = 0.02$, $t(219) = 3.31$; $p < 0.05$]. Neither *Gender of respondent* nor *Age* were significant predictors and none of the other interaction terms with *Form* were statistically significant.

In summary, as expected, adolescents of all ages and regardless of their gender, perceived success in gendered occupations to be more equally shared by women and men when the job had been described to them in a pair form rather than in the masculine form only (see **Figure 1**).

Ascriptions of Warmth and Competence to Jobholders in Male and Female Dominated and in Gender-Neutral Occupations

To test our hypotheses that language form used to describe occupational titles would influence the ascription of warmth and competence to prototypical jobholders, we conducted separate analyses for warmth and competence in line with previous research (e.g., Vervecken and Hannover, 2012; Budziszewska et al., 2014).

Warmth

We performed a 2 (*Form*: Pair form vs. Masculine only) × 2 (*Gender of respondent*: Female vs. Male) × 3 (*Assumed gender*

---

[2]When running a non-parametric statistic, this effect was not significant, $U(220) = 5597$, $Z = -1.05$, $p > 0.05$.

distribution: Female vs. Male vs. Neutral) factorial mixed ANOVA on *warmth*, with *Age* as a covariate, *Form* and *Gender of respondent* as between-participant factors and *Assumed gender distribution* as a within-participant factor.

In support of hypothesis 2, the analysis revealed a statistical interaction between *Assumed gender distribution* and *Form*, $F(2,430) = 3.71$, $p < 0.05$, $\eta^2 = 0.03$. When occupations were presented in pair form, ascriptions of warmth increased for male dominated occupations (Masculine Form: $M = 3.24$, $SE = 0.06$; Pair Form: $M = 3.29$, $SE = 0.06$) but decreased for female occupations (Masculine Form: $M = 3.84$, $SE = 0.04$; Pair Form: $M = 3.78$, $SE = 0.05$). Hence, warmth ascribed to prototypical jobholders differed more strongly between male dominated vs. female dominated jobs when the job had been presented in the masculine only form [difference of 0.59, $SE = 0.04$, $t(116) = 13.54$, $p < 0.01$, Cohen's $d = 1.25$], compared to when the occupation had been described in pair forms [difference of 0.51, $SE = 0.05$, $t(103) = 10.20$, $p < 0.01$, Cohen's $d = 1.00$]. Also, the difference in warmth ascribed to holders of gender-neutral occupations (Masculine Form: $M = 3.40$, $SE = 0.05$; Pair Form: $M = 3.51$, $SE = 0.05$) vs. female dominated occupations decreased when the job had been described in a pair form [difference of 0.28, $SE = 0.04$, $t(103) = 6.51$, $p < 0.01$, Cohen's $d = 0.63$] compared to when it had been presented in the masculine only form [difference of 0.44, $SE = 0.04$, $t(116) = 12.08$, $p < 0.01$, Cohen's $d = 1.12$].

The analysis also revealed a significant statistical interaction between *Assumed gender distribution* and *Gender of the respondent*, $F(2,430) = 5.89$, $p < 0.01$, $\eta^2 = 0.03$, suggesting that girls and boys differed in their attributions of warmth when considering female dominated job [Girls: $M = 3.89$, $SE = 0.04$; Boys: $M = 3.71$, $SE = 0.05$, $t(220) = 2.74$, $p < 0.01$] but not when considering male dominated [Girls: $M = 3.26$, $SE = 0.05$; Boys: $M = 3.27$, $SE = 0.06$, $t(219) < 1$, *ns*], or gender-neutral occupations [Girls: $M = 3.48$, $SE = 0.05$; Boys: $M = 3.41$, $SE = 0.05$, $t(219) < 1$, *ns*]. *Age* was not a significant predictor for adolescents' warmth related attributions toward occupations.

In summary, when occupations were presented in a pair form rather than in the masculine form only, the differential ascription of warmth to prototypical job holders of male dominated, female dominated, and gender-neutral occupations was attenuated, regardless of participants' age and gender. As apparent in **Figure 2**, the effect of linguistic form on warmth-related attributions mirrors the pattern of linguistic form on gendered representations of women's and men's success.

Competence

To test research hypothesis 3, we again conducted a 2 (*Form*: Pair form vs. Masculine only) × 2 (*Gender of respondent*: Female vs. Male) × 3 (*Assumed gender distribution within occupations*: Female vs. Male vs. Neutral) factorial mixed ANOVA on *competence*, with *Age* as a covariate, *Form* and *Gender of respondent* as between-participant factors and *Assumed gender distribution* as a within-participant factor.

There was only a main effect of the *Gender of the respondent*, $F(1,216) = 5.89$, $p < 0.05$, $\eta^2 = 0.03$, showing that girls in general ascribed higher levels of competence ($M = 3.79$, $SE = 0.04$) than
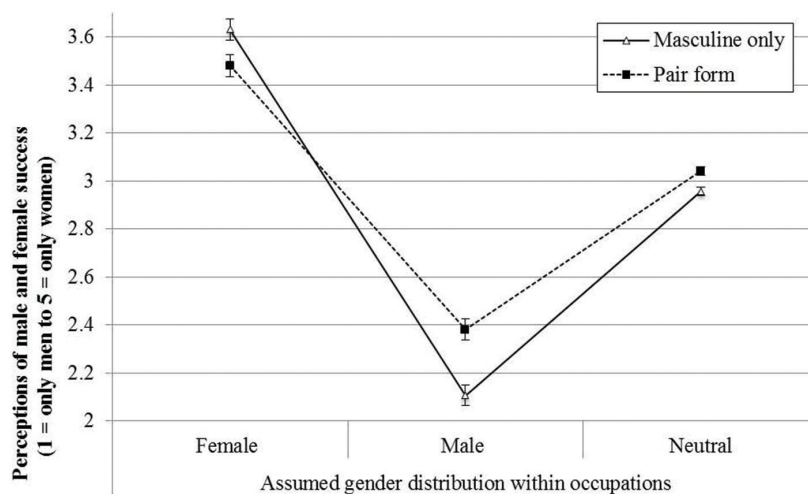
FIGURE 1 | Mean perceptions of success for men and women in occupations with different gender distributions (scale from 1 = only men to 5 = only women).
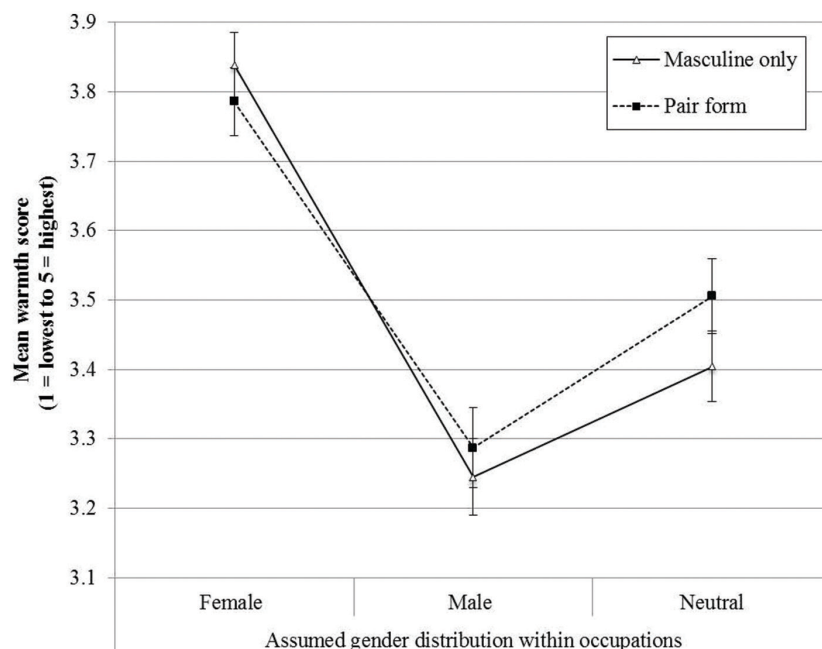

FIGURE 2 | Mean scores on the *warmth* dimension for occupations with different gender distributions.

boys ($M = 3.64$, $SE = 0.06$). There were no other significant main or interaction effects (all $p > 0.20$). Hence, hypothesis 3 was not supported.

Discussion

By combining work on the impact of gender-fair language on mental representations (e.g., Stahlberg et al., 2007) with work on stereotypes (Fiske et al., 2002; Fiske, 2011; Eagly and Koenig, 2014) the present study investigated how different linguistic

forms (i.e., pair form vs. masculine only form) used to present female dominated, male dominated, and gender-neutral jobs impact adolescents' gendered perceptions, in particular their ascriptions of success, warmth and competence to male and female job holders.

Gendered Perceptions of Occupational Success

In our sample of 222 adolescents (aged 12–17) from French speaking Switzerland we found, in line with our first hypothesis,

that regardless of whether males, females or neither gender was overrepresented in an occupation, presentation in the linguistic pair form, rather than the masculine form only, triggered more gender-balanced representations of occupational success. These results broaden the findings of previous studies which typically only investigated male dominated occupations (for a review, see, Stahlberg et al., 2007). To more systematically investigate how language forms interact with the gender connotation of occupations, in our study we not only presented male dominated, but also female dominated and gender-neutral jobs. Interestingly, not only in male stereotyped jobs but also in female and gender-neutral jobs, occupational success was more evenly attributed to males and females when the occupation was described in a linguistic pair form, rather than the masculine generic form only.

These results suggest that subtle linguistic markers, besides other factors potentially influencing gender stereotyping (for a review, see Blakemore et al., 2009), can have an impact on the extent to which adolescents think that women and men can be professionals in the same domains. It seems that masculine only forms vs. pair forms operate like primes, increasing the mental accessibility of either male job holders or, respectively, female job holders. The linguistic markers activate the corresponding mental representations which in turn guide recipients' categorization and interpretation of the information (cf. Bargh, 2014; Stupica and Cassidy, 2014). Our findings suggest that such "natural priming effects" (Bargh, 2014, p. 218) are important in everyday life, as they influence the perception of females' and males' occupational success.

Gendered Ascriptions of Warmth and Competence

Going beyond the scope of previous studies, we not only looked at the effects of linguistic forms on the perception of males' and females' occupational success, but also on the ascription of warmth and competence to prototypical jobholders. We assumed that linguistic pair forms make participants think of both genders, such that the impact of an unequal gender distribution within an occupation on their perceptions of the jobs would be attenuated. As a result, ascriptions of warmth and competence, as two universal dimensions guiding people's perceptions of others (Fiske et al., 2002), should differ less between holders of female dominated vs. male dominated occupations when the jobs are described in a pair form, rather than the masculine form only. Results confirmed our expectation that the difference in the ascription of warmth to holders of female dominated vs. male dominated occupations was smaller in the pair form condition (compared to the masculine only condition). Furthermore, the difference in the ascription of warmth to holders of female dominated vs. gender neutral occupations was also smaller in the pair form condition (compared to the masculine only condition). It seems that when a male dominated occupation (e.g., businessmen) is presented in a pair form, adolescents are inclined to attribute more warmth to the prototypical job holder. However, when a female dominated occupation (e.g., child care taker) is described in pair form, stronger associations with coldness are triggered.

The ascription of warmth being influenced by our experimental manipulation is in line with the results of Merkel et al. (2012), who found that female targets whose job had been described in feminine forms were perceived as warmer than female targets whose job had been described in a masculine only form. Our findings complement the ones reported by Merkel et al. (2012) in that we could show that pair form use influenced attributions of warmth in general (i.e., to female and male workers in a certain occupation), and furthermore, that ascriptions of warmth actually decreased when female dominated occupations had been described in pair forms.

While the stronger ascription of warmth to job holders in male dominated and the weaker ascription of warmth to job holders in female dominated jobs is consistent with our hypotheses, unexpectedly, competence ratings were unaffected by our linguistic manipulation. Interestingly, Merkel et al. (2012) also found that *competence* ratings remained unaffected by the linguistic form in which females' occupations had been presented. In fact, some authors (e.g., Wojciszke et al., 1998; Fiske et al., 2007) have argued that perception of warmth is primary to ascriptions of competence, with others' intentions being more prominent – in an evolutionary perspective – than others' abilities to act on those intentions. From our results, one could then argue that linguistic forms are used to make inferences on moral and social dimensions but not on a person's competency or expertise.

Alternatively, our linguistic manipulation not having an impact on the ascription of competence can hint at competence-related gender stereotypes being in decline. A recent study investigating implicit stereotypes about women in Germany did not replicate the women-incompetence stereotype (Ebert et al., 2014) as described by the SCM (Fiske et al., 2002; for similar findings for Spain see López-Sáez et al., 2008). While such a change in the female stereotype has been demonstrated only in some cultural contexts, poll data from US national surveys point in a similar direction (Newport, 2001; Pew Research Center, 2008): they show that nowadays women are increasingly perceived as more intelligent than men in the general population (see Wood and Eagly, 2012, for a review). In terms of social role theory, our finding that although linguistic form had an impact on perceptions of women and men's success it did not affect competence-related evaluations, could indicate a shift in gender stereotyping: since women are no longer associated with lower competence, differences in the percentage of women in an occupation or variations in the mental accessibility of female job holders – as caused by our linguistic manipulation – can no longer be expected to have an impact on competence perceptions of prototypical job holders.

Practical Relevance of Our Findings

The findings from the current experiment demonstrate that adolescents are sensitive to gender information in occupational titles and use this information to make gendered inferences about the occupations. It seems that the generic use of masculine only forms when describing occupations is likely to lead adolescents to restrictive, gender exclusive associations and perceptions about occupations. This is an especially important finding as the transition from adolescence to adulthood is an important stage in

vocational development in which gender stereotyped perceptions of occupations play an essential role (Gottfredson, 2005; Lerner and Steinberg, 2009).

While changing occupational gender stereotypes has long been recognized as a key for closing the occupational gender gap, few interventions have been investigated to tackle this issue in adolescence. This is especially dissatisfying as adolescents' career aspirations are important predictors for educational and occupational status in adulthood (Sewell and Hauser, 1972; Campbell, 1983; Kao and Thompson, 2003; Feliciano and Rumbaut, 2005; Beal and Crockett, 2010; Lee et al., 2012). For example Beal and Crockett (2010) found in a longitudinal study that adolescents' educational expectations were positively associated with educational attainment in young adulthood. In a similar vein, Sewell and Hauser (1972) demonstrated that post-secondary educational attainment at age 25 was significantly predicted by aspirations students held in adolescence, and educational attainment, in turn, positively predicted earnings at the age of 28.

Any intervention that aims to alter aspects of representational biases may well-contribute to reducing occupational gender segregation (Weisgram et al., 2011; Eagly and Koenig, 2014; Liben and Coyle, 2014). Gender-fair language use by teachers, parents, or the media may thus contribute to an attenuation of adolescents' gender related stereotypes about occupations.

Our findings are also consistent with the view that the current extensive use of the masculine only form (Blaubergs, 1980; Parks and Robertson, 1998; Bußmann and Hellinger, 2003; Mucchi-Faina, 2005; Koeser and Sczesny, 2014; Kuhn and Gabriel, 2014) may well-contribute to shaping, or at least maintaining, gender stereotypes. Consequently, enforcing or encouraging the use of pair forms in grammatical gender languages when referring to mixed gender groups or to groups whose gender composition is unknown or irrelevant seems to be an effective strategy to counter gender stereotypes. Our findings substantiate the effectiveness of recent linguistic reforms as currently promoted by many professional organizations, publishing companies, and governmental organizations (e.g., Duden, 2006; European Commission, 2008; American Psychological Association, 2009): they advocate gender-fair language use and reject the notion that the masculine form can be generic. Unfortunately, these language reforms contrast with the still common use of the masculine only form in various applied settings, for example, in schools, as illustrated by studies on teachers' language practice (e.g., Vervecken et al., 2010) or schoolbooks' contents (e.g., Moser and Hannover, 2014).

It is possible that pair forms might promote wider interest in traditionally constrained disciplines such as the STEM fields (i.e., Science, Technology, Engineering, and Mathematics). In fact, a recent review by Liben and Coyle (2014) suggests that one tangible way to promote interest in STEM fields might be to alter the traditionally masculine image of these occupations to a more feminine one.

Limitations of Our Study and Future Directions

In this article, we argued that the use of gender fair language to describe occupations has an impact on adolescents' perceptions of occupations. Although our evidence is quite compelling, the generalizability of our findings could be discussed.

First, in the present study, we had to restrict the experimental stimulus material to fifteen role nouns. Therefore, we cannot provide by-items analyses, and the generalizability of our findings to other occupations remains to be tested in future studies. However, according to the theory of generalizability (Cronbach et al., 1963), Cronbach's alpha can be viewed as a measure of how well the *sum score* on the selected items captures the expected score in the entire domain, even if that domain is heterogeneous. Hence, the very high Cronbach's alpha values for warmth and competence ratings across the five occupations of each of the three groups of occupations suggest that our findings may be, in fact, generalized to other occupations.

Second, we generally clustered all male, female and gender-neutral occupations together. However, other categorizations and extra subdivisions within occupations are conceivable. For example, one could order occupations using the RIASEC model (cf. Holland, 1997), based on stereotypical personality types (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional), or categorize them according to whether they belong to Science, Technology, Engineering and Mathematics (STEM-fields). Using a broader range of occupations and dividing them into meaningful subcategories could provide a more detailed insight into the effects of gender fair language.

Third, whereas the results of the present cross-sectional experiment illustrate effects of gender fair language shortly after it is presented, it is difficult to make inferences about long-term effects. A full account of the impact of gender fair language on adolescents' development of occupational gender stereotypes and their subsequent educational and vocational development could only be provided by longitudinal study designs. Although there is some evidence suggesting that repetitively combining role nouns with the non-stereotypical gender (e.g., surgeon/mother) indeed may have longer-term impact (e.g., Finnegan et al., 2015), longitudinal research on gender fair language does not exist at this time. Future research may want to explore this. For example, some teachers could be trained in using gender fair language. Simultaneously the development of gender-role beliefs in their students could be monitored over a longer period of time and compared to students whose teachers use traditional language (masculine only forms). Similarly, textbooks using either gender fair or traditional language could be randomly assigned to different school classes. Again, the development of adolescents' gender-role beliefs could be monitored and compared.

Acknowledgment

References

American Psychological Association (2009). *Publication Manual of the American Psychological Association*, 6th Edn. Washington, DC: Author.

Bargh, J. A. (2014). Our unconscious mind. *Sci. Am.* 310, 30–37.

Beal, S. J., and Crockett, L. J. (2010). Adolescents' occupational and educational aspirations and expectations: links to high school activities and adult educational attainment. *Dev. Psychol.* 46, 258–265. doi: 10.1037/a0017416

Blakemore, J. E. O., Berenbaum, S. A., and Liben, L. S. (2009). *Gender Development*. New York, NY: Taylor & Francis.

Blanchard, C. A., and Lichtenberg, J. W. (2003). Compromise in career decision making: a test of Gottfredson's theory. *J. Vocat. Behav.* 62, 250–271. doi: 10.1016/S0001-8791(02)00026-X

Blaubergs, M. (1980). An analysis of classic arguments against changing sexist language. *Womens Stud. Int. Q.* 3, 135–147. doi: 10.1016/S0148-0685(80)9 2071-0

Boone, H. N., and Boone, D. A. (2012). *Analyzing Likert Data. Journal of Extension, 50.* Available at: http://www.joe.org/joe/2012april/tt2.php

Braun, F., Gottburgsen, A., Sczesny, S., and Stahlberg, D. (1998). Können geophysiker frauen sein? Generische personenbezeichnungen im deutschen. [Can geophysicians be women? Generic terms in German]. *Z. Ger. Linguist.* 26, 265–283. doi: 10.1515/zfgl.1998.26.3.265

Budziszewska, M., Hansen, K., and Bilewicz, M. (2014). Men against feminine job titles. The impact of gender – fair language on men's and women's perception of women. *J. Lang. Soc. Psychol.* 33, 681–691. doi: 10.1177/0261927X14544371

Bußmann, H., and Hellinger, M. (2003). "Engendering female visibility," in *German, Gender Across Languages*, eds M. Hellinger and H. Bußmann (Amsterdam: John Benjamins), 141–174.

Campbell, R. T. (1983). Status attainment research: end of the beginning or beginning of the end? *Sociol. Educ.* 56, 47–62. doi: 10.2307/2112302

Cejka, M. A., and Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Pers. Soc. Psychol. Bull.* 25, 413–423. doi: 10.1177/0146167299025004002

Chatard, A., Guimond, S., and Martinot, D. (2005). Impact de la féminisation lexicale des professions sur l'auto-efficacité des élèves: une remise en cause de l'universalisme masculin? [Occupational self-efficacy as a function of grammatical gender in French]. *Année Psychol.* 105, 249–272. doi: 10.3406/psy.2005.29694

Crawley, D. (2014). Gender and occupational prestige: changes in twenty years. *Sage Open* 4, 1–11. doi: 10.1177/2158244013518923

Cronbach, L. J., Nageswari, R., and Gleser, G. C. (1963). Theory of generalizability: a liberation of reliability theory. *Br. J. Stat. Psychol.* 16, 137–163. doi: 10.1111/j.2044-8317.1963.tb00206.x

Diekman, A. B., and Eagly, A. H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Pers. Soc. Psychol. Bull.* 26, 1171–1188. doi: 10.1177/0146167200262001

Duden. (2006). *Richtiges und Gutes Deutsch [Correct and Good German]*, Vol. 9. Mannheim: Dudenverlag.

Eagly, A. H., and Koenig, A. (2014). Evidence for the social role theory of stereotype content: observations of groups' roles shape stereotypes. *J. Pers. Soc. Psychol.* 107, 371–392. doi: 10.1037/a0037215

Eagly, A. H., and Mladinic, A. (1994). "Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence," in *European Review of Social Psychology*, eds W. Stroebe and M. Hewstone (New York, NY: Wiley), 1–35.

Ebert, I. D., Steffens, M. C., and Kroth, A. (2014). Warm, but maybe not so competent? – Contemporary implicit stereotypes of women and men in Germany. *Sex Roles* 70, 359–375. doi: 10.1007/s11199-014-0369-5

European Commission (2008). *Geschlechtergerechter Sprachgebrauch Beim Europäischen Parlament [Gender fair Language Use at the European Commision], Swiss Cabinet.* Available at: http://www.bk.admin.ch/themen/lang/05225/05235/index.html [accessed April 1, 2014].

Feliciano, C., and Rumbaut, R. (2005). Gendered paths: educational and occupational expectations and outcomes among adult children of immigrants. *Ethn. Racial Stud.* 28, 1087–1118. doi: 10.1080/01419870500224406

Finnegan, E., Garnham, A., and Oakhill, J. (2015). Social-consensus feedback as a strategy to overcome spontaneous gender stereotypes. *Discourse Process.* 52, 434–462. doi: 10.1080/0163853X.2015.1026680

Fiske, S. (2011). *Envy Up, Scorn Down: How Status Divides Us*. New York, NY: Russel Sage Foundation.

Fiske, S., Cuddy, A., and Glick, P. (2007). Universal dimensions of social cognition: warmth, then competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Fiske, S., Cuddy, A., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902. doi: 10.1037/0022-3514.82.6.878

Formanowicz, M., Bedyńska, S., Cisłak, A., Braun, F., and Sczesny, S. (2013). Side effects of gender-fair language: how feminine job titles influence the evaluation of female applicants. *Eur. J. Soc. Psychol.* 43, 62–71. doi: 10.1002/ejsp.1924

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au-pairs are rarely male: role names' gender stereotype information across three languages. *Behav. Res. Methods Instrum. Comput.* 40, 206–212. doi: 10.3758/BRM.40.1.206

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., and Oakhill, J. (2012). Gender representation in different languages and grammatical marking on pronouns: when beauticians, musicians, and mechanics remain men. *Discourse Process.* 49, 481–500.

Gottfredson, L. S. (2005). "Using Gottfredson's theory of circumscription and compromise in career guidance and counseling," in *Career Development and Counseling: Putting Theory and Research to Work*, eds S. D. Brown and R. W. Lent (New York, NY: Wiley), 71–100.

Hartung, P. J., Porfeli, E. J., and Vondracek, F. W. (2005). Child vocational development: a review and reconsideration. *J. Vocat. Behav.* 66, 385–419. doi: 10.1016/j.jvb.2004.08.011

Hays, W. L. (1994). *Statistics*, 5th Edn. Fort Worth, TX: Harcourt Brace.

Heise, E. (2000). Sind Frauen mitgemeint? Eine empirische untersuchung zum verständnis des generischen maskulinums und seiner alternativen [Are women included? An empirical study of the generic masculine and its alternatives]. *Z. Spr. Kogn.* 19, 3–13.

Heise, E. (2003). Auch einfühlsame studenten sind männer: das generische maskulinum und die mentale repräsentation von personen [Sensitive students are men too: the generic masculine and the mental representation of persons]. *Verhaltenstherapie Psychosoziale Prax.* 35, 285–291.

Helwig, A. (2001). A test of Gottfredson's theory using a ten-year longitudinal study. *J. Career Dev.* 28, 77–95. doi: 10.1177/089484530102800201

Holland, J. L. (1997). *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*. Odessa, FL: Psychological Assessment Resources.

Hyde, J. S. (1984). Children's understanding of sexist language. *Dev. Psychol.* 20, 697–706. doi: 10.1037/0012-1649.20.4.697

Irmen, L., and Schumann, E. (2011). Processing grammatical gender of role nouns: further evidence from eye-movements. *J. Cogn. Psychol.* 23, 998–1014. doi: 10.1080/20445911.2011.596824

Kao, G., and Thompson, J. T. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annu. Rev. Sociol.* 29, 417–442. doi: 10.1146/annurev.soc.29.010202.100019

Koeser, S., and Sczesny, S. (2014). Promoting gender-fair language: the impact of arguments on language use, attitudes, and cognitions. *J. Lang. Soc. Psychol.* 33, 548–560. doi: 10.1177/0261927X14541280

Krosnick, J. A., and Berent, M. K. (1993). Comparisons of party identification and policy preferences: the impact of survey question format. *Am. J. Pol. Sci.* 37, 941–964. doi: 10.2307/2111580

Kuhn, E., and Gabriel, U. (2014). Actual and potential gender-fair language use: the role of language competence and the motivation to use accurate language. *J. Lang. Soc. Psychol.* 33, 214–225. doi: 10.1177/0261927X1 3504297

Lassonde, K. A., and O'Brien, E. J. (2013). Occupational stereotypes: activation of male bias in a gender-neutral world. *J. Appl. Soc. Psychol.* 43, 387–396. doi: 10.1111/j.1559-1816.2013.01008.x

Lee, J. O., Hill, K. G., and Hawkins, J. D. (2012). The role of educational aspirations and expectations in the discontinuity of intergenerational low-income status. *Soc. Work Res.* 36, 141–151. doi: 10.1093/swr/svs025

Lerner, R. M., and Steinberg, L. (2009). *Individual Bases of Adolescent Development: Volume 1 of Handbook of Adolescent Psychology*, 3rd Edn. Hoboken, NJ: Wiley.

Liben, L. S., and Coyle, E. F. (2014). Developing interventions to address the STEM gender gap: exploring intended and unintended consequences. *Adv. Child Dev. Behav.* 47, 77–115. doi: 10.1016/bs.acdb.2014.06.001

López-Sáez, M., Morales, J. F., and Lisbona, A. (2008). Evolution of gender stereotypes in Spain: traits and roles. *Span. J. Psychol.* 11, 609–617.

Merkel, E., Maass, A., and Frommelt, L. (2012). Shielding women against status loss: the masculine form and its alternatives in the Italian language. *J. Lang. Soc. Psychol.* 31, 311–320. doi: 10.1177/0261927X12446599

Moser, F., and Hannover, B. (2014). How gender-fair are schoolbooks in Germany in the twenty-first century? An analysis of language and illustrations in schoolbooks for mathematics and German. *Eur. J. Psychol. Educ.* 29, 387–407. doi: 10.1007/s10212-013-0204-3

Mucchi-Faina, A. (2005). Visible or influential? Language reforms and gender (in) equality. *Soc. Sci. Inf.* 44, 189–215. doi: 10.1177/0539018405050466

Newport, F. (2001). *Americans see Women as Emotional and Affectionate, Men as more Aggressive: Gender Specific Stereotypes Persist in Recent Gallup Poll.* Available at: http://www.gallup.com/poll/1978/Americans-seewomen-emotional-affectionate-men-moreaggressive.aspx [accessed February 21, 2001]

Parks, J., and Robertson, M. (1998). Contemporary arguments against nonsexist language: Blaubergs (1980) revisited. *Sex Roles* 39, 445–461. doi: 10.1023/A:1018766023667

Pew Research Center (2008). *Men or Women: Who's the Better Leader? A Paradox in Public Attitudes.* Available at: http://www.pewsocialtrends.org/2008/08/25/men-or-women-whos-the-better-leader/

Rothmund, J., and Scheele, B. (2004). Personenbezeichnungsmodelle auf dem prüfstand. lösungsmöglichkeiten für das genus-sexus-problem auf textebene [Putting gender neutral reference terms to the test: constructive solutions to the problem of grammatical vs. referential gender on the textual level]. *Z. Psychol.* 212, 40–54. doi: 10.1026/0044-3409.212.1.40

Schau, C. G., and Scott, K. P. (1984). Review of 21 cause and effect studies. *Psychol. Doc.* 76, 183–193.

Sewell, W. H., and Hauser, R. M. (1972). Causes and consequences of higher education: models of the status attainment process. *Am. J. Agric. Econ.* 54, 851–861. doi: 10.2307/1239228

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language," in *Social Communication. Frontiers of Social Psychology*, ed. K. Fiedler (New York, NY: Psychology Press), 163–187.

Stahlberg, D., and Sczesny, S. (2001). Effekte des generischen maskulinums und alternativer sprachformen auf den gedanklichen einbezug von frauen [Effects of masculine generics and alternative forms of speech on the cognitive inclusion of women]. *Psychol. Rundsch.* 52, 131–140. doi: 10.1026//0033-3042.52.3.131

Stahlberg, D., Sczesny, S., and Braun, F. (2001). Name your favorite musician: effects of masculine generics and of their alternatives in German. *J. Lang. Soc. Psychol.* 20, 464–469. doi: 10.1177/0261927X01020004004

Stupica, B., and Cassidy, J. (2014). Priming as a way of understanding children's mental representations of the social world. *Dev. Rev.* 34, 77–91. doi: 10.1016/j.dr.2013.12.003

Vervecken, D., and Hannover, B. (2012). Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts perception of the speaker. *Eur. J. Soc. Psychol.* 42, 754–762. doi: 10.1002/ejsp.1893

Vervecken, D., and Hannover, B. (2015). Yes i can! the impact of gender fair descriptions of traditionally male occupations on children's perceptions of job status, job difficulty and vocational self-efficacy beliefs. *Soc. Psychol.* 46, 76–92. doi: 10.1027/1864-9335/a000229

Vervecken, D., Hannover, B., and Wolter, I. (2013). Changing (s)expectations: how gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *J. Vocat. Behav.* 82, 208–220. doi: 10.1016/j.jvb.2013.01.008

Vervecken, D., Moser, F., Sczesny, S., and Hannover, B. (2010). "Entwicklung und validierung eines instruments zur messung der einstellung gegenüber geschlechtergerechter sprache. [Development and validation of an instrument to measure attitudes towards gender fair language]," in *Poster Presented at the 47th General Meeting of the German Psychological Association (DGPs)*, Bremen.

Weisgram, E., Dinella, L. M., and Fulcher, M. (2011). The role of masculinity/femininity, values, and occupational value affordances in shaping young men's and women's occupational choices. *Sex Roles* 65, 243–258. doi: 10.1007/s11199-011-9998-0

White, M. J., and White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles* 55, 259–266. doi: 10.1007/s11199-006-9078-x

Wojciszke, B., Bazinska, R., and Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Pers. Soc. Psychol. Bull.* 24, 1245–1257. doi: 10.1177/01461672982412001

Wood, W., and Eagly, A. H. (2012). "Biosocial construction of sex differences and similarities in behavior," in *Advances in Experimental Social Psychology*, eds J. M. Olson and M. P. Zanna (New York, NY: Elsevier), 55–123. doi: 10.1016/B978-0-12-394281-4.00002-7

The Social Perception of Heroes and Murderers: Effects of Gender-Inclusive Language in Media Reports

Karolina Hansen[1]*, Cindy Littwitz[2†] and Sabine Sczesny[3]

[1] *Faculty of Psychology, University of Warsaw, Warsaw, Poland,* [2] *Friedrich-Schiller-University Jena, Jena, Germany,* [3] *University of Bern, Bern, Switzerland*

The way media depict women and men can reinforce or diminish gender stereotyping. Which part does language play in this context? Are roles perceived as more gender-balanced when feminine role nouns are used in addition to masculine ones? Research on gender-inclusive language shows that the use of feminine-masculine word pairs tends to increase the visibility of women in various social roles. For example, when speakers of German were asked to name their favorite "heroine or hero in a novel," they listed more female characters than when asked to name their favorite "hero in a novel." The research reported in this article examines how the use of gender-inclusive language in news reports affects readers' own usage of such forms as well as their mental representation of women and men in the respective roles. In the main experiment, German participants ($N = 256$) read short reports about heroes or murderers which contained either masculine generics or gender-inclusive forms (feminine-masculine word pairs). Gender-inclusive forms enhanced participants' own usage of gender-inclusive language and this resulted in more gender-balanced mental representations of these roles. Reading about "heroines and heroes" made participants assume a higher percentage of women among persons performing heroic acts than reading about "heroes" only, but there was no such effect for murderers. A post-test suggested that this might be due to a higher accessibility of female exemplars in the category heroes than in the category murderers. Importantly, the influence of gender-inclusive language on the perceived percentage of women in a role was mediated by speakers' own usage of inclusive forms. This suggests that people who encounter gender-inclusive forms and are given an opportunity to use them, use them more themselves and in turn have more gender-balanced mental representations of social roles.

Keywords: heroism, crime, murder, gender, social roles, gender-fair language, cognitive availability, newspaper reports

INTRODUCTION

When we open a newspaper, we often encounter headlines such as "Hometown driver now a local hero" or "A lot of heroes around here." We may also come across a headline that reads "While at war, female soldiers fight to belong" (New York Times, 25 May 2015). What images of men and women do such newspaper articles create? Does the language they contain influence

these images? In our research, we studied the effects of using either only masculine or both masculine and feminine role nouns in newspaper articles. A large body of research documents that women are less visible in the media in general: only 13% of all news stories are about women (Macharia et al., 2010). Furthermore, the media often depict women and men in a stereotyped manner, with 46% of news stories reinforcing gender stereotypes, and only 6% challenging such stereotypes (Macharia et al., 2010). Gender stereotypes that prevail in a society are reflected in the media, but the media also influence how women and men are perceived in the respective society. Also, the way research findings are reported in the popular press may affect readers' beliefs and attitudes and may reinforce stereotyping. For example, a series of studies showed that readers of an article that stressed biological explanations of gender differences endorsed gender stereotypes more strongly than readers of a similar article that focused more on sociocultural explanations for gender differences (Brescoll and LaFrance, 2004).

The image of women is not only influenced by what is said or not said, but also by how it is said. In grammatical gender languages, such as German, French, or Russian, nouns and pronouns have masculine and feminine forms and thus differentiate for gender, for instance, "he" vs. "she" or "hero" vs. "heroine." However, when referring to mixed-gender groups, to persons with unknown gender or persons whose gender is irrelevant, "masculine generics" are used, i.e., grammatically masculine nouns and pronouns (Hellinger and Bussmann, 2003). In contrast, gender-inclusive language makes explicit reference to women and men (word pairs, e.g., "he or she," "firemen and firewomen") or uses gender-neutral forms (e.g., "they," "firefighters," Stahlberg et al., 2007).

Past research has revealed that gender-inclusive language makes women more visible than masculine generics do (Stahlberg et al., 2007). When gender-inclusive forms are used, people assume percentages of women in a profession to be higher than when masculine generics are used (Braun et al., 2005) and more female exemplars of a category are named (Stahlberg et al., 2001; Merkel et al., 2012) or drawn (Bojarska, 2011). Most research of this kind investigated common social roles (e.g., student, professor, physician, Bojarska, 2011) and only some studies tested more prominent roles (e.g., musician, politician, Stahlberg et al., 2001). In one investigation, German participants listed more women when asked to name their favorite "heroine or hero in a novel"[1] (word pair form; German: *Romanheldin oder Romanheld*) than those who were asked to name their favorite "hero in a novel" (*Romanheld*; Stahlberg and Sczesny, 2001). In addition to the impact of language on mental representations of women and men, recent research has shown that after reading a text with gender-inclusive wording, speakers used more gender-inclusive forms themselves in a fill-in-the-blank task (Koeser et al., 2015).

While most research focused on the effects of gender-inclusive language on perceivers (readers or listeners), there is much less

research on the effects of this usage on producing such language in writing or speaking. One study that addressed effects of gender-inclusive language on users showed that participants who were made to employ forms such as *he or she* in a sentence-completion task imagined more female characters as protagonists of the sentences they were completing than participants who were made to use generic *he* (Hamilton, 1988). The author does not explain the mechanisms underlying this effect, but it is conceivable that speakers' own use of gender-inclusive forms enhances the cognitive availability (Tversky and Kahneman, 1973) and causes a more intensive processing of these forms and of their referents.

We were interested in the impact of gender-inclusive language in media texts on mental representations of women and men and in the mechanism underlying this impact. To this end, we investigated reactions to newsworthy, exceptional social roles that are often dealt with in the media: hero and murderer. Both social roles attract much attention and have similarly low percentages of women (ca. 10–20%). In the US, only 9% of the recipients of the Carnegie Hero Medal for saving others are women, and in Germany only about 20% of similar medals are awarded to women. This may be because there are fewer women in professions such as firefighters, soldiers, or police officers—jobs involving dangerous situations where jobholders can act heroically. As for murderers, in 2014 women committed 11% of all homicides in the US (Federal Bureau of Investigation, 2015); in Germany, where the present study was conducted, it was 9% (Statistisches Bundesamt, 2015).

As mentioned above, reading a gender-inclusive text provoked more use of such language in a fill-in-the-blank task (Koeser et al., 2015). To replicate and extend this finding, we examined whether the effect on language use also held for more natural forms of language production, namely for texts written entirely by the participants themselves. We presented participants with "a short science-based press release," and asked them to summarize the text in their own words and to indicate the percentage of women in the described social role. We expected participants summarizing a gender-inclusive text to use more word pairs and other gender-fair forms than those summarizing a text with masculine generics (Hypothesis 1). In line with past research (Stahlberg et al., 2001), we expected that reading a text with gender-inclusive forms would result in higher estimates of the percentage of women in the respective roles than reading a text with masculine generics (Hypothesis 2).

In addition, and more importantly, we examined whether participants' own use of gender-inclusive language would lead to a higher perceived percentage of women in a given social role. In the present study, we aimed at eliciting gender-inclusive forms by having participants read a text containing such forms and by asking them to summarize its content in their own words. Thus, we directly manipulated only the text, but not participants' language use *per se*. We expected that receiving a message in gender-inclusive wording would enhance speakers' inclination to use such forms themselves (Koeser et al., 2015). This use, in turn, was expected to result in higher estimates of the percentage of women in the respective roles than reading a text with masculine generics (Hamilton, 1988). In other words, we

---

[1] We use the expressions "heroines" along with "heroes" as well as "murderesses" along with "murderers" to render the German wording. However, we are aware that in English these expressions are not as gender-fair as in German.

expected speakers' own use of gender-inclusive forms to mediate the relationship between the forms appearing in the text and mental representations of women in the role described in the text (Hypothesis 3).

The present investigation integrates hitherto separate lines of research on gender-inclusive language by studying three effects in the framework of one experiment: (1) the effect of reading gender-inclusive forms on speakers' own language use, (2) the effect of reading gender-inclusive forms on mental representations of women, and (3) the effect of own language use on mental representations of women. By integrating these aspects in one experiment, the present study enables identifying the process underlying effects of gender-inclusive language. In this way, it contributes to a more comprehensive understanding of how gender-inclusive language leads to a reduction of gender stereotyping and discrimination (Sczesny et al., 2016).

MATERIALS AND METHODS

Participants and Design

We recruited participants for an online study via a local newspaper, university mailing lists, and during the *Long Night of Scientists* in Jena, Germany (an event popularizing science among the general public). Participants were 267 native speakers of German. After deleting answers of one participant who took 17 h, the average time for completing the survey was 20 min ($SD = 17$ min). We deleted six participants who completed the survey in less than 3 min ($-1\ SD$) and five who took more than 54 min ($+2\ SD$). The final sample consisted of 256 persons (170 women, 86 men). Participants were of different ages (range: 14–82 years, $M = 32.30$, $SD = 14.32$) and most of them (68%) were living in the German state of Thuringia. About half (55%) of them were university students from different departments.

The experiment had a 2 (social role: heroes vs. murderers) × 2 (linguistic form: masculine generics vs. word pairs) design, with use of gender-inclusive language and estimated percentage of female heroes/murderers as dependent variables. The use of gender-inclusive language in participants' own writing served as a potential mediator of the relationship between linguistic forms used in the text and the estimated percentage of female heroes/murderers.

Procedure and Measures

The study was carried out in accordance with the recommendations of Swiss and German Human Research Acts. It was presented as "a study on the perception of science-based press releases." After providing their informed consent and indicating their age and gender, participants read one (randomly chosen) of four versions of a 250-word media text about the socialization of (1) heroes, (2) heroines and heroes, (3) murderers, or (4) murderesses and murderers. The texts were identical except for the role nouns, which varied in the heading and in three other places in the text. Each text was about a study that attempted to identify factors explaining why some people become heroes/murderers (see Supplementary Materials). After

reading the text, participants were asked to sum it up in their own words in 3–4 sentences.

To rule out that the text quality differed depending on linguistic forms used, we asked participants whether the information in the text was credible (accurate, credible, reliable, reflecting reality, $\alpha = 0.88$) and whether the text was informative (convincing, informative, should be published, relevant, $\alpha = 0.87$; scale: $1 = totally\ disagree$, $7 = fully\ agree$). The respective analyses showed that the texts were perceived as similarly credible and informative in both linguistic versions ($Fs < 2.60$, $ps > 0.10$, $\eta_p^2 < 0.01$). The texts about heroism were perceived as more credible than the texts about murderers, $F(1,251) = 8.23$, $p = 0.004$, $\eta_p^2 = 0.03$, but the texts about murders were perceived as more informative, $F(1,251) = 8.23$, $p = 0.004$, $\eta_p^2 = 0.03$. No interactions were observed, $Fs < 1$.

Then we asked five "memory questions," including the dependent variable. The questions concerned facts and numbers that appeared in the text (How many participants took part in the study? What hobby did most of them have? How well did most of them do at school?), but also "What proportion of heroic deeds/murders are committed by women?"

Finally, participants answered some additional questions, specified demographic data[2], were debriefed, and given the possibility to make comments. Those who left their e-mail addresses had the chance to win three prizes of €20 each.

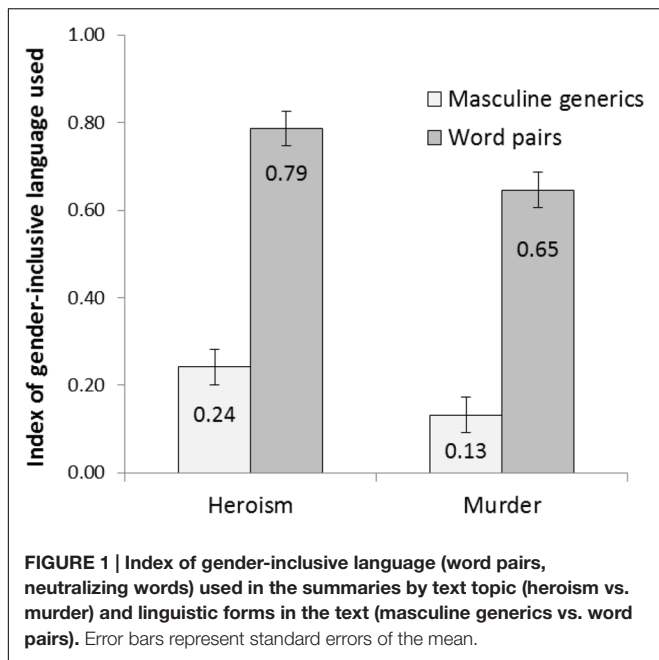## RESULTS

Use of Gender-Inclusive Language

After reading about heroism or murders, participants were asked to summarize the text in their own words. We coded the linguistic forms they used to refer to the roles of hero and murderer. The three linguistic forms most often used to describe the protagonists of the texts were: masculine forms (singular or plural, e.g., German *Helden* "heroes"), word pairs (e.g., German *Heldinnen und Helden* "heroines and heroes"), and neutralization (e.g., German *Menschen* "humans," *Leute* "people," *Personen* "persons"). For each participant we built an index: we computed the proportion of gender-inclusive language by dividing all gender-inclusive forms (word pairs and neutralization) by all forms used for person reference (both gender-inclusive and masculine). The resulting index ranged from 0 to 1; the higher the index, the more gender-inclusive forms were used by the participant.

A 2 (social role: heroes vs. murderers) × 2 (linguistic form: masculine generics vs. word pairs) ANOVA showed that, in accord with Hypothesis 1, participants used more gender-inclusive language after reading texts with word pairs ($M = 0.72$, $SD = 0.38$) than after reading texts with masculine generics ($M = 0.18$, $SD = 0.26$), $F(1,259) = 177.04$, $p < 0.001$, $\eta_p^2 = 0.42$ (see **Figure 1**). Furthermore, participants used more gender-inclusive language when writing about heroism ($M = 0.51$,

---

[2]We also included a few measures that are not the focus of the current article. The description and basic results on these measures are available in the Supplementary Materials.

FIGURE 1 | Index of gender-inclusive language (word pairs, neutralizing words) used in the summaries by text topic (heroism vs. murder) and linguistic forms in the text (masculine generics vs. word pairs). Error bars represent standard errors of the mean.

$SD = 0.42$) than about murders ($M = 0.39$, $SD = 0.42$), $F(1,259) = 9.53$, $p = 0.002$, $\eta_p^2 = 0.04$. The interaction of social role and linguistic form was not significant, $F < 1$.

Given the large age range of the sample (14–82 years), and the possibility of participant gender playing a role, we examined whether age and gender modulated the results. An ANCOVA showed no significant influence of gender ($F < 1$) and a small effect of age, $F(1,246) = 3.30$, $p = 0.07$, $\eta_p^2 = 0.01$: younger participants used slightly more gender-inclusive forms. The remaining results were virtually the same with or without both covariates.

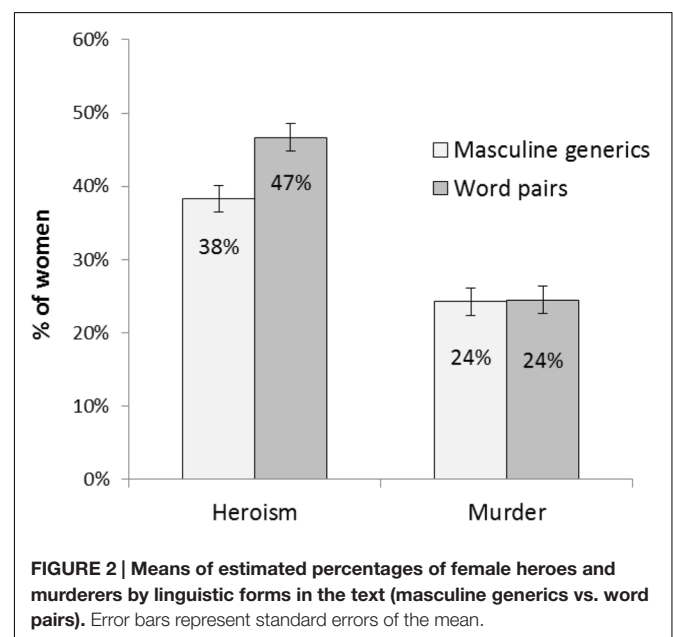## Estimated Percentage of Female Heroes and Murderers

A 2 (social role: heroes vs. murderers) × 2 (linguistic form: masculine generics vs. word pairs) ANOVA revealed that, in accord with Hypothesis 2, participants who read the texts with word pairs estimated a higher percentage of women in both roles ($M = 35.93$; $SD = 16.98$) than participants who read the texts with masculine generics ($M = 31.72$; $SD = 16.49$), $F(1,231) = 5.53$, $p = 0.02$, $\eta_p^2 = 0.02$ (see **Figure 2**). Furthermore, participants estimated a higher percentage of women among heroes ($M = 42.48$; $SD = 14.57$) than among murderers ($M = 24.37$; $SD = 13.78$), $F(1,231) = 98.87$, $p < 0.001$, $\eta_p^2 = 0.30$. An interaction effect, $F(1,231) = 5.00$, $p = 0.03$, $\eta_p^2 = 0.02$, showed that those who read about "heroines and heroes" estimated a higher percentage of female heroes than those who read about "heroes," $p = 0.001$. There was no difference in the estimated percentage of female murderers between the two language conditions, $p = 0.94$. Also, the more gender-inclusive forms participants used in their earlier summaries, the more women they perceived in the respective social role, $r(231) = 0.26$, $p < 0.001$.
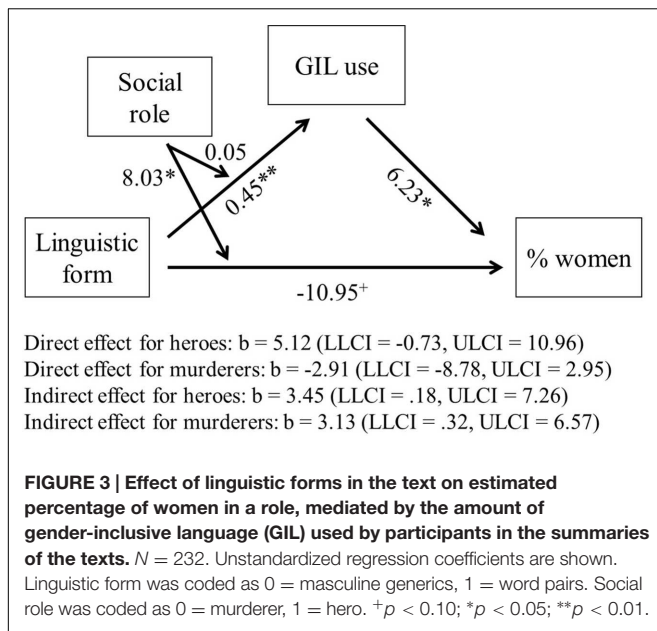
An ANCOVA including participants' gender and age again failed to show a significant influence of gender, $F(1,246) = 2.67$, $p = 0.10$, $\eta_p^2 = 0.01$, but revealed an influence of age, $F(1,246) = 12.88$, $p < 0.001$, $\eta_p^2 = 0.05$: younger participants perceived more women in the social roles described in the texts. However, the age effect was not strong and the results were the same with or without covariates.

The Mediating Role of Gender-Inclusive Language Use

Our hypothesis predicted that gender-inclusive language would make participants use more of such language themselves, think in more gender-inclusive ways, and, in turn, estimate a higher percentage of women in the respective roles (Hypothesis 3). Therefore, we examined the potential mediating effect of own language use in the relationship between linguistic form used in the media text and perceived percentage of women in a role. As the results above showed, gender-inclusive language had a similar effect on speakers' own use of such forms for both social roles (heroes and murderers), but affected the perceived percentage of women only among heroes. In view of this finding, we conducted a moderated mediation model (Hayes, 2013, p. 369), with social role as a moderator.

The analysis showed that after reading the media text containing word pairs participants used more gender-inclusive forms themselves (see **Figure 3**), in line with our previous findings. The more gender-inclusive language they used, the more women they perceived in this role. As shown above, the social role described in the text did not affect participants' own use of gender-inclusive language (no interaction of linguistic form and social role), but affected the perceived percentage of women (higher percentage of women for heroes, but not for murderers). Beyond previous findings and in accord with Hypothesis 3, the

FIGURE 2 | Means of estimated percentages of female heroes and murderers by linguistic forms in the text (masculine generics vs. word pairs). Error bars represent standard errors of the mean.

Direct effect for heroes: b = 5.12 (LLCI = -0.73, ULCI = 10.96)
Direct effect for murderers: b = -2.91 (LLCI = -8.78, ULCI = 2.95)
Indirect effect for heroes: b = 3.45 (LLCI = .18, ULCI = 7.26)
Indirect effect for murderers: b = 3.13 (LLCI = .32, ULCI = 6.57)

FIGURE 3 | Effect of linguistic forms in the text on estimated percentage of women in a role, mediated by the amount of gender-inclusive language (GIL) used by participants in the summaries of the texts. N = 232. Unstandardized regression coefficients are shown. Linguistic form was coded as 0 = masculine generics, 1 = word pairs. Social role was coded as 0 = murderer, 1 = hero. [+]p < 0.10; [*]p < 0.05; [**]p < 0.01.

analysis revealed that when own use of gender-inclusive language was included, the direct effect of linguistic forms in the text disappeared for the category heroes and remained insignificant for murderers. Indirect (i.e., mediational) effects occurred for both roles. Thus, having read a text with word pairs, participants used more such forms themselves, and this, in turn, promoted more gender-balanced representations in both roles.

Post-test

The main effects for estimated percentage of women in the two social roles resembled the results obtained for own language use: participants used more gender-inclusive language and estimated a higher percentage of women after reading the texts with word pairs and after reading the texts about heroism. However, gender-inclusive language increased only the estimated percentage of female heroes, but not the percentage of female murderers. It has been suggested that the number of female exemplars available in a person category may play a role for the changes in gender perceptions that are evoked by gender-inclusive language (Braun et al., 2005). Thus, the observed asymmetry in our results may be due to a different availability of female exemplars among heroes and murderers. To address this possibility, we conducted a post-test.

The post-test assessed how often people thought of women when asked about heroism and murders, without variation in language forms. We asked passers-by (N = 35, 21 women, 14 men, M_{age} = 35.17, SD = 14.82, age range: 19–74) in Jena, Germany, to fill out a short questionnaire in exchange for a chocolate bar. The instruction given in a booklet asked participants to name "people who behaved heroically" and, on the next page, "people who murdered someone" (order counterbalanced). Next, we explicitly asked participants to write down "women who behaved heroically" and "women who killed someone."

The results showed that participants spontaneously named a total of 82 (M = 2.34) male and 12 (M = 0.34) female murderers, as well as 45 (M = 1.29) male and 22 (M = 0.63) female heroes. Thus, they mentioned more men than women for both roles, but the proportion of women was higher for heroes (33%) than for murderers (13%). Furthermore, participants named more female heroes than female murderers, while the opposite was true for male heroes and murderers. When asked explicitly about women, they again named more female heroes (37) than female murderers (23). In all, the post-test revealed that more exemplars of female heroes were available to the participants than of female murderers.

DISCUSSION

The present study shows that the language used in the media to describe social roles affects readers' own language use and that this, in turn, can influence the social perception of groups. In line with and extending earlier research, which relied on fill-in-the-gap sentences (Koeser et al., 2015), participants used more gender-inclusive forms after reading a text with word pairs when summarizing the text in their own words. Moreover, participants who read about "heroines and heroes" estimated a higher percentage of female heroes than those who read about "heroes." Most importantly, we found that gender-inclusive language triggered a more gender-balanced use of language, and this use resulted in a more gender-balanced perception of social roles. In this process, participants' own language use functioned as a mediator between the language of the media texts and the mental representation of social roles. Thus, our results not only replicate and link previous findings (Hamilton, 1988; Stahlberg et al., 2001; Koeser et al., 2015), but also illuminate the mechanism behind the effects of gender-inclusive language. Participants' own use of gender-inclusive language made them think in more gender-inclusive ways.

While readers of a text about "heroines and heroes" perceived more female heroes than readers of a text about "heroes," the estimated percentage of women committing murder did not differ between readers of a text about "murderesses and murderers" and about "murderers." Our post-test showed that people were aware of more female heroes than murderers. It seems that when people know some women in a given role (Becker and Eagly, 2004; Rankin and Eagly, 2008), gender-inclusive language can trigger the female exemplars that are known and can make the mental representation of a role more gender-balanced (Braun et al., 2005). In other words, language can impact speakers' perceptions to some extent, but only within the boundaries of social reality. Other explanations for the observed effect are also conceivable. It is possible, for instance, that norm congruence plays a role here. Speaking of women as murderers might be too negative and less reconcilable with the female stereotype of a caring, nurturing, and selfless mother, than speaking of women as heroes. Perhaps gender-inclusive language makes women more visible in positive roles, but not in negative ones. However, our participants generally overrated the percentage of murderesses, which suggests that

they were not trying to protect women's image. Furthermore, earlier studies have shown that gender-inclusive language has an effect also when talking about disliked personalities (e.g., the least liked politician, Gabriel and Mellenberger, 2004; Gabriel, 2008). Still another explanation of the observed effect could be the different flexibility of the definition of both roles. Perhaps category boundaries are clearer and more straightforward for "murderers" than for "heroes." It might be easier to extend the definition of a hero to include more women than to extend the definition of a murderer (Becker and Eagly, 2004; Franco et al., 2011).

Although the effect of gender-inclusive language on the estimated percentage of women was significant only for heroes but not for murderers, the mediation effect occurred for both roles. The mediation findings suggest that triggering people to use gender-inclusive forms in their writing can make them process information more intensively and enhance cognitive availability of these forms and their referents (Tversky and Kahneman, 1973). In addition, participants may have inferred that because they used gender-inclusive forms in their summaries themselves, there must be a considerable percentage of women in these roles. This self-perception mechanism (Fazio et al., 1977) could be operating in parallel to a more intensive processing due to own use of gender-inclusive language.

Future research should aim at replicating the obtained results and at further contributing to an explanation of the mechanisms underlying them. It could be also studied whether similar effects are present while listening to gender-inclusive language and while using it in one's own speech. Potential mediating variables could be assessed and/or manipulated. The accessibility of exemplars, for instance, could be manipulated by carefully choosing roles that differ in this respect, but not in valence, or by experimentally making some exemplars more accessible. The influence of the roles' valence could be studied by investigating roles with comparable numbers of known female exemplars, but with a different valence. To generalize the effects of valence, it would be advisable to include several positive and negative roles. Future research could also test whether the effect of texts in gender-inclusive language on the own use of such language and on the mental representations "spills over" to other social roles than the ones provided.

Gender-inclusive language can be used to reduce gender-stereotypic images of certain male-typed social roles in a community with a grammatical gender language (Horvath and Sczesny, 2016). Our results show that both perceiving and producing gender-inclusive language can evoke more gender-balanced mental representations of social roles. Such language can be effective not only when read or heard, but it can expand and resonate later when reproduced. The present study also suggests that reality will not be distorted if the media use more gender-inclusive language, but that this type of language may help to present women and men more equally in various social roles.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00369

REFERENCES

Becker, S. W., and Eagly, A. H. (2004). The heroism of women and men. *Am. Psychol.* 59, 163–178. doi: 10.1037/0003-066X.59.3.163

Bojarska, K. (2011). *Wpływ Androcentrycznych i Inkluzywnych Płciowo Konstrukcji Językowych na Skojarzenia z Płcią* [Influence of Androcentric and Gender-Inclusive Lexical Constructions on Gender-Associational Responses]. *Studia Psychol.* 49, 53–68. doi: 10.2478/v10167-011-0010-y

Braun, F., Sczesny, S., and Stahlberg, D. (2005). Cognitive effects of masculine generics in German: an overview of empirical findings. *Communications* 30, 1–21. doi: 10.1515/comm.2005.30.1.1

Brescoll, V., and LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on sex differences. *Psychol. Sci.* 15, 515–520. doi: 10.1111/j.0956-7976.2004.00712.x

Fazio, R. H., Zanna, M. P., and Cooper, J. (1977). Dissonance and self-perception: an integrative view of each theory's proper domain of application. *J. Exp. Soc. Psychol.* 13, 464–479. doi: 10.1016/0022-1031(77)90031-2

Federal Bureau of Investigation (2015). *Murder Offenders by Age, Sex, Race, and Ethnicity 2013*. Available at: http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/offenses-known-to-law-enforcement/expanded-homicide/expandhomicidemain_final [accessed March 17, 2015].

Franco, Z. E., Blau, K., and Zimbardo, P. G. (2011). Heroism: a conceptual analysis and differentiation between heroic action and altruism. *Rev. Gen. Psychol.* 15, 99. doi: 10.1037/a0022672

Gabriel, U. (2008). Language policies and in-group favoritism: the malleability of the interpretation of generically intended masculine forms. *Soc. Psychol.* 39, 103–107. doi: 10.1027/1864-9335.39.2.103

Gabriel, U., and Mellenberger, F. (2004). Exchanging the generic masculine for gender-balanced forms – the impact of context valence. *Swiss J. Psychol.* 63, 273–278. doi: 10.1024/1421-0185.63.4.273

Hamilton, M. C. (1988). Using masculine generics: does generic he increase male bias in the user's imagery? *Sex Roles* 19, 785–799. doi: 10.1007/BF00288993

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach.* New York: Guilford Press.

Hellinger, M., and Bussmann, H. (2003). *Gender Across Languages: The Linguistic Representation of Women and Men.* Amsterdam: John Benjamins Publishing. doi: 10.1075/impact.11

Horvath, L. K., and Sczesny, S. (2016). Reducing women's lack of fit with leadership positions? Effects of the wording of job advertisements. *Eur. J. Work Organ. Psychol.* 25, 316–328. doi: 10.1080/1359432X.2015.1067611

Koeser, S., Kuhn, E. A., and Sczesny, S. (2015). Just reading? How gender-fair language triggers readers' use of gender-fair forms. *J. Lang. Soc. Psychol.* 34, 343–357. doi: 10.1177/0261927X14561119

Macharia, S., O'Connor, D., Ndangam, L., and Gallagher, M. (2010). *Global Media Monitoring Project: 2010 Global Report.* Available at: http://cdn. agilitycms.com/who-makes-the-news/Imported/reports_2010/global/gmmp_global_report_en.pdf (accessed February 23, 2015).

Merkel, E., Maass, A., and Frommelt, L. (2012). Shielding women against status loss: the masculine form and its alternatives in the Italian language. *J. Lang. Soc. Psychol.* 31, 311–320. doi: 10.1177/0261927X12446599

Rankin, L. E., and Eagly, A. H. (2008). Is his heroism hailed and hers hidden? Women, men, and the social construction of heroism. *Psychol. Women Q.* 32, 414–422. doi: 10.1111/j.1471-6402.2008.00455.x

Sczesny, S., Formanowicz, M., and Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Front. Psychol.* 7:25. doi: 10.3389/fpsyg.2016.00025

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language," in *Social Communication*, ed. K. Fiedler (New York, NY: Psychology Press), 163–187.

Stahlberg, D., and Sczesny, S. (2001). *Effekte des Generischen Maskulinums und Alternativer Sprachformen auf den Gedanklichen Einbezug von Frauen* [Effects of the Generic Use of the Masculine Pronoun and Alternative Forms of Speech on the Cognitive Visibility of Women]. *Psychol. Rundsch.* 52, 131–140. doi: 10.1026//0033-3042.52.3.131

Stahlberg, D., Sczesny, S., and Braun, F. (2001). Name your favorite musician: effects of masculine generics and of their alternatives in German. *J. Lang. Soc. Psychol.* 20, 464–469. doi: 10.1177/0261927X01020 004004

Statistisches Bundesamt (2015). *Verurteilte nach Ausgewählten Straftaten, Geschlecht und Altersgruppen* [Convicts by Selected Offenses, Sex, and Age]. Available at: https://www.destatis.de/DE/Publikationen/Thematisch/Rechtspf lege/StrafverfolgungVollzug/StrafverfolgungsstatistikDeutschland (accessed March 14, 2015).

Tversky, A., and Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cogn. Psychol.* 5, 207–232. doi: 10.1016/0010-0285(73)90033-9

Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior

Marie Gustafsson Sendén[1]*, Emma A. Bäck[2] and Anna Lindqvist[3]

[1] *Department of Psychology, Stockholm University, Stockholm, Sweden,* [2] *Department of Psychology, Gothenburg University, Gothenburg, Sweden,* [3] *Department of Psychology, Lund University, Lund, Sweden*

The implementation of gender fair language is often associated with negative reactions and hostile attacks on people who propose a change. This was also the case in Sweden in 2012 when a third gender-neutral pronoun *hen* was proposed as an addition to the already existing Swedish pronouns for *she (hon)* and *he (han)*. The pronoun *hen* can be used both generically, when gender is unknown or irrelevant, and as a transgender pronoun for people who categorize themselves outside the gender dichotomy. In this article we review the process from 2012 to 2015. No other language has so far added a third gender-neutral pronoun, existing parallel with two gendered pronouns, that actually have reached the broader population of language users. This makes the situation in Sweden unique. We present data on attitudes toward *hen* during the past 4 years and analyze how time is associated with the attitudes in the process of introducing *hen* to the Swedish language. In 2012 the majority of the Swedish population was negative to the word, but already in 2014 there was a significant shift to more positive attitudes. Time was one of the strongest predictors for attitudes also when other relevant factors were controlled for. The actual use of the word also increased, although to a lesser extent than the attitudes shifted. We conclude that new words challenging the binary gender system evoke hostile and negative reactions, but also that attitudes can normalize rather quickly. We see this finding very positive and hope it could motivate language amendments and initiatives for gender-fair language, although the first responses may be negative.

Keywords: gender-fair language, gender-neutral pronouns, attitude change, gender, hen

Introduction

Language is seen as an important tool for determining gender, i.e., if something is being perceived as feminine or masculine (Boroditsky et al., 2003; Stahlberg et al., 2007), where gender most often imposes a dichotomy (Ansara and Hegarty, 2014). This implies that language also could be used as a tool for establishing gender-equality and to challenge gender perceptions. In Western culture and languages, actions toward gender-fair languages have primarily focused on making women more salient and reducing the so-called male bias (for a review, see: Stahlberg et al., 2007). For example, in the seventies, the feminist movement questioned the use of a generic masculine pronoun to refer to people in general (Moulton et al., 1978; MacKay, 1980; Phillips, 1981; Murdock and Forsyth, 1985).

The literature describes two types of gender fair language: "balancing/feminization' and 'neutralization.' Feminization implies the use of gender-appropriate forms, and is more often used in languages with grammatical gender (e.g., German, French), for example by adding feminine versions to masculine titles (e.g., Lehrer/Lehrerinnen for masculine and feminine teachers; Stahlberg et al., 2001, 2007). Neutralization is more commonly employed in so called 'natural gender languages' (e.g., English, Swedish, Norwegian), and implies that gender-neutral forms are preferred over gendered forms. Examples are using the word *parents* instead of *mum* and *dad*, and *humankind* instead of *mankind* (at least in official records).

In Swedish, a recent action was to introduce the gender-neutral third person pronoun, *hen*, as a complement to the Swedish words for *she (hon)* and *he (han)* (Ledin and Lyngfelt, 2013; Milles, 2013; Bäck et al., 2015). In current time the word first appeared in 2012, figuring in a children's book. In July 2014, it was announced that *hen* should be included in 2015th edition of The Swedish Academy Glossary (SAOL) constituting the (unofficial) norm of the Swedish language (Benaissa, 2014; Fahl, 2014), after what had been a long, sometimes offensive and heated debate in the media. No other language has so far added a third gender-neutral pronoun that actually has reached the broader population of language users, which makes the situation in Sweden unique. This article presents a review of the process on how *hen* became implemented, including the arguments that were put forward from opponents and proponents, respectively. We present data on attitudes toward *hen* during the recent 4 years and study how time is associated with the attitudes and actual use of the word.

The word *hen* is very similar to, and pronounced as, the Finnish gender-neutral pronoun *hän* with the same meaning, i.e., describing any person no matter their gender – although the language of Sweden's cultural neighbor Finland belongs to the language group without gendered third-person pronouns (Stahlberg et al., 2007; Prewitt-Freilino et al., 2012). Even though the debate about *hen* took off in 2012, the word was first mentioned as early as in the 1960's (Milles, 2013), when linguists proposed that a gender-neutral pronoun would be a more rational choice in comparison to a generic *he* or using double forms (i.e., *he and/or she*). However, these discussions were more of an academic nature limited to small linguistic communities and did not reach a broader public (Milles, 2013). In the beginning of the 21st century people in LGBT-communities (Lesbian-, Gay-, Bi-, Trans-) began to use *hen*, both for people outside the gender dichotomy and as a way of diminishing the salience of gender. A similar movement has been found in the English language, among linguists and among transgender communities, where more than 80 different forms of gender-neutral pronouns have been proposed. Today, one trend in English is to use gender-neutral pronouns such as *zie* and *hir* (Baron, 1986; Ansara and Hegarty, 2014; Love, 2014), although these words have not been very widespread outside the LGBT-communities (Crawford and Fox, 2007).

When the debate took off in 2012, the spark that started it was the publishing of a children's book (Lundquist, 2012) that used *hen* to denote the main character of the book, instead of using a gendered pronoun. The author and the publisher also wrote a debate article in one of the largest newspapers in Sweden together with Karin Milles, a linguist researcher and advisor of gender-fair language planning, arguing for the introduction of a gender-neutral pronoun (Milles et al., 2012; Milles, 2013). Advocates of the word argued that children are too much influenced by gender categories, where non-gendered pronouns allow them to visualize and develop their stories much more freely (Milles et al., 2012). Antagonists argued that children listening to such non-gendered stories would be disoriented not knowing their gender, and that having a (binary) gender (i.e., being a girl or a boy) is something to be proud of (Lagerwall, 2012). At this point in 2012, the use of *hen* was highly controversial, which is illustrated by an incident when a columnist in one of the largest newspapers used *hen*. The reactions led the management of the paper to apply a policy against using *hen* in its news reporting (Cederskog, 2012). In contrast, an entertainment magazine changed all third personal pronouns into *hen* in their second issue in 2012 (Milles, 2013). Later in 2012, the Language Council of Sweden (*Språkrådet*) providing official recommendations about Swedish language, recommended that *hen* should not be used, since it could be irritating and conflict with the content in the text. This illustrates a common argument against gender-fair language reforms – where new forms are commonly described as awkward and potentially steeling attention from the message (Blaubergs, 1980; Parks and Roberton, 1998). For example the publication manual by American Psychological Association (APA) includes guidelines against sexist language stating that '. . .combination forms such as *he/she* or *(s)he* are awkward and distracting and are not ideal' (APA, 2012, p. 74). APA recommends the use of 'neutral' words such as *the person*, or *they*. However, both *they* and *the person* might be associated with gender bias (most often a male bias), which existing literature on gender-fair language has shown is a robust phenomenon (e.g., Hyde, 1984; Stahlberg et al., 2001, 2007; Lenton et al., 2009; Garnham et al., 2012). According to the literature, a gender bias is described as the situation when care is taken to express gender-fairness in the language and people nevertheless seem to create biased perceptions where they associate the gender-neutral expressions with either a masculine or a feminine gender. For example in English, the word *they* could be used as an assumed generic form (Gastil, 1990; Strahan, 2008), but in a study where the generic *he* was replaced by *they*, children still more often associated *they* with a man (Hyde, 1984). Also, supposedly neutral words such as *person*, *mankind,* or even *human* have been associated with a male bias (Stahlberg and Sczesny, 2001; Douglas and Sutton, 2014; Bäck et al., 2015). These results imply that the creation of new words may be needed to override gender and cisgender bias, although it might take some time for language users to get used to them. However, the implementation of newly formed words is not an easy and straightforward enterprise, maybe especially not for a pronoun. From a linguist perspective, it has been argued that pronouns changes more slowly than other words because they belong to the so called 'function words' or 'closed words classes' (Milles, 2013; Paterson, 2014). Function words are used to organize the grammatical structure in a sentence

and their meaning is only derived from how they are used in context (Chung and Pennebaker, 2007; Milles, 2013). Pronouns are organized in a grammatical system, thus adding a new word challenges the whole system (Paterson, 2014), which is not the case when nouns or verbs are added to a dictionary, or when feminine forms of professional roles are added to masculine forms.

Hen can be used in two different ways: either as a third-person pronoun in situations including general descriptions of an individual whose gender is unknown or is considered as irrelevant, or as a third-person pronoun in situations where the described person is not gender-neutral but describing someone identifying themselves outside the gender-dichotomy (Milles, 2013; Bäck et al., 2015). For people with a non-binary gender identity, double forms of pronouns (i.e., *he/she*) and guidelines for gender-fair language are excluding (Ansara and Hegarty, 2014). For example 'APA's binary descriptions of gender reinforce ethnocentric gender ideology that assumes 'woman' and 'man' are the only possible genders' (Ansara and Hegarty, 2014, p. 264).

The different uses of *hen* align with the arguments from its proponents and antagonists. Representatives from LGBT-communities propose a gender-neutral pronoun since it dissolves gender expectations and includes all individuals no matter their gender-identity (Milles, 2013). These arguments have met the strongest reactions where the proponents have been targeted with offensive and hostile attacks. The antagonists have argued that queer people and feminists are trying to change biology, and that gender is one of the most natural categories. A maybe less controversial argument is that the gender-neutral pronoun *hen* is a shorter and more efficient way in comparison to double forms. Accordingly, *hen* could be used when gender is unknown, or as a generic pronoun. These arguments have been put forward by some feminists and linguists (Milles, 2013). Yet, other groups of feminists have been negative toward a gender-neutral pronoun since, they claim, it could be a way of diminishing women. For example, a well-known Swedish feminist and professor in literature has argued that the feminine gender is obscured by the word *hen* (Brattström, 2014). Hence, the use of *hen* and its consequences have not been agreed on, and disputes reside even within the feminist movement.

After 2012 followed a time with progress toward a more official implementation. In 2013, the Swedish Language Council *(språkrådet)* changed their recommendation and proposed that *hen* could be used as a gender-neutral pronoun, although with caution because it may distract attention from the message. The next year, in 2014, it was announced that the word should be included in the 2015th edition of the SAOL that constitutes the (unofficial) norm of the Swedish language (Fahl, 2014). In this year, the language council also formally changed their guidelines for gender fair language in public authorities, and included *hen* as an alternative to other neutral or gender balanced forms. Using *hen* is still not mandatory in official publications; each authority decides themselves whether to use it in public documents or not, and so far very few do (Ledin and Lyngfelt, 2013; Olsson, 2015).

In the Swedish media, the word has become more commonly used (Ledin and Lyngfelt, 2013; Milles, 2013). For example, during the first 6 months of 2012 *hen* was mostly seen in a vivid debate about the word itself, while during the second half of 2012, the word was actually used in texts unrelated to the debate about the word, that is, as a gender-neutral pronoun. In one of the bigger newspapers in Sweden the occurrences of *hen* increased over a year, from 1 in 2010, to 9 in 2011 and to 113 in 2012 (Ledin and Lyngfelt, 2013). This means that though *hen* still is rare, an increase is undisputable. The analyzed paper is among one of the conservative papers, thus it was presumed that occurrences in more progressive papers may be higher, however, a quantification of this hypothesis has not yet been done. In an effort to understand how the media used the word (generic or transgender), Ledin and Lyngfelt (2013) showed that 15% of the occurrences were related to transgender use, whereas 85% corresponded to a practice when gender was unknown, irrelevant or, as a generic form.

Since the pronoun is new, there is still limited research about how the word is perceived and what consequences it might have. A few studies have tested whether *hen* decreases a male- and cisgender bias. In one study (Wojahn, 2013), 150 participants read a story about a cellphone user, referring to the person either as *he, he/she, hen,* or *he or she*. Results showed that *hen* evoked the least male bias and also less cisgender bias. In a previous study, we have shown that a person described as *hen* was more often remembered as a person of unknown gender, whereas a person described by a neutral word is more often remembered as having a masculine gender (Bäck et al., 2015).

Gender-fair language is often implemented over several years. It commonly starts with activist movements who propose a change. Since people have a preference for status quo (Jost et al., 2004; Samuelson and Zeckhauser, 2005; Crandall et al., 2009), and a preservation of traditional gender arrangements (Jost et al., 2008), new linguistic gender word forms may be negatively reacted upon. Proponents of non-sexist language have been attacked, words have been defined as being linguistically wrong or awkward (Blaubergs, 1980; Parks and Roberton, 1998), and feminine occupational titles have been evaluated more negatively than their masculine traditional form (Formanowicz et al., 2013). However, familiarity and exposure breeds liking (Zajonc, 1968), thus the attitudes may change the longer gender-fair language has been used (Eidelman et al., 2009; Moreland and Topolinski, 2010). Whether such attitude change occurs also for gender-neutral pronouns within a country has not been studied before.

In studying the implementation process of gender-fair language reforms and the consequences on population attitudes and use, it is important to consider variables traditionally associated with negative attitudes toward gender fair language. If we are to make a claim that gender fair language reforms will be successful, an important task for the present research is to show that time in use is important to include when studying attitudes and frequency of use, together with other potential explanations. Previous research has identified a number of predictors of

attitudes to gender-fair language and the following section will provide an overview of these.

Sexism in terms of attitudes toward gender equality has been identified as a predictor of negative attitudes toward gender-neutral language use (cf. Sarrasin et al., 2012), together with *political orientation* in terms of right-wing conservatism (Formanowicz et al., 2013; Norton and Herek, 2013). Also in the 'hen-debate,' more left-wing than right-wing politicians used *hen* and promoted that *hen* should be included in the Swedish Dictionary (Milles, 2013). In Sweden, there are feminist movements both on the left and right of the political map, and in the last election more politicians than ever before openly stated that they considered themselves to be feminists (Öhberg and Wängnerud, 2014). Thus, feminist values would be associated with more positive attitudes, no matter of political orientation. Jacobson and Insko (1985) showed that feminist attitudes were associated with a higher use of gender fair language, such as using more double forms of pronouns. Feminist attitudes also mediated the effect between gender and attitudes toward gender fair language. Hence, even though the literature suggests that political right-orientation would predict negative attitudes, this is not entirely straight-forward, and we suggest that interest in gender issues may be a potent predictor as well.

Gender (as coded in a binary system feminine/masculine) as a predictor of attitudes to and use of gender-fair language has been inconclusive so far. Some studies have shown that women are more positive than men to gender-fair language (i.e., Prentice, 1994; Sarrasin et al., 2012) others have shown no differences (i.e., Koeser and Sczesny, 2014). Women tend to use gender-fair language more often than men (Koeser and Sczesny, 2014), and are more easily influenced to adjust to gender-fair language (Koeser et al., 2014). Notably, using a gender-neutral *hen* is not as clearly beneficial for women, as compared to other forms of gender-fair language (e.g., balancing masculine and feminine form, or avoiding masculine generics). Hence, it is not certain how, or even if, gender will affect attitudes to *hen*. Since *hen* challenges the binary gender system that is prevailing in most cultures, it could be argued that some individuals will show stronger resistance than others, depending on how important the gender system is to them. We argue that biological gender is not of greatest importance in this case, but rather the extent to which one identifies as a woman or a man, and how important this identification is. Indeed, arguments in the debate have touched upon such issues; for example, heterosexual people have argued that they are negative toward the word *hen* because it 'restricts their right to express their gender identity,' and that 'romance between men and women will suffer' (Lagerwall, 2012). Very few studies have investigated strength of gender identity as being a woman or a man in relation to gender-fair language. These studies have used forms of BEM Sex Role Inventory (BMSRI; Bem, 1974). The studies showed that a masculine gender identity (no matter of biological gender) was associated with more negative attitudes (Rubin and Greene, 1994), while androgynous gender identity has been associated with more positive attitudes, and higher use of gender-fair language (McMinn et al., 1990; Rubin and

Greene, 1991). Given that Sweden is an egalitarian society, where the distinction between femininity and masculinity is no longer as strongly rooted in traditional feminine and masculine roles, we believe that the strength of *gender identity* is a better measure than gender roles as measured by BMSRI (Bem, 1974).

Finally, we believe that *age* will predict attitudes and the use, because younger people are more susceptible to new ideas and to challenge traditional roles, than older people are (Visser and Krosnick, 1998; Eaton et al., 2009).

The main purpose with the present research is to study how time and other factors are associated with change in attitudes and use of *hen*. In the present research we investigate the effect of time on the outcome variables. However, we do not here study the mechanism (for instance habituation) by which elapsed time can explain such effects, but rather show that other potential explanatory factors are not sufficient to explain the outcome effects alone. To date, time has been proposed as a cause for difference in evaluations (see for example, Sarrasin et al., 2012). However, no studies have followed an implementation over time in one language, with one specific word. It is also known from previous research that time has a positive effect on attitudes such that the longer something has been in effect the better people will like it (Zajonc, 1968; Moreland and Topolinski, 2010). We present data from 2012 to 2015, on the attitudes to *hen* and self-reported use of *hen* from 2013 to 2015. We make the following predictions:

H1. Attitudes towards *hen* will become more positive over time.

H2. Self-reported use of *hen* will increase over time.

H3. Sexism and right-wing political orientation will be associated with negative attitudes, as well a lower use of *hen*

H4. Age will be related to attitudes and use, such that younger people will be more positive, and indicate more use of the word, than older people. Gender is included as a control because some studies have shown that women are more positive to gender-fair language than men.

H5. A strong gender identity (as either a woman or a man) will be associated with more negative attitudes and less use. Interest in gender issues will be associated with more positive attitudes and higher use.

H6. Time will have a significant and independent effect on attitudes and use of *hen*, also when all other variables are controlled for.

Materials and Methods

Participants and Procedure

We have collected data on attitudes and use of *hen* at six points in time since 2012. Participants and the datasets are described in **Table 1**. Dataset 1 and 2 consist of participants that were approached in the waiting hall at the Central station in Stockholm. Dataset 3 and 4 consist of students at Lund University. Dataset 5 consists of participants that were approached in the city of Lund. Participation was rewarded

Year	Dataset	N	Age		Gender		Sample
			M (SD)		Women/men (%)		
2012	1	184	36.6 (18.8)		59/41		Community
2013	2	61	40.3 (17.3)		59/41		Community
2013	3	160	23.6 (6.6)		50/50		Student
2013	4	51	22.7 (3.7)		67/23		Student
2014	5	40	31.0 (12.7)		43/57		Community
2015	6	190	33.5 (9.7)		67/27*		Community
Total		686	31.7 (14.2)		60/39		

In 2015, 4% indicated a gender identity outside the binary system, and 2% did not indicate gender.

with a lottery ticket in all these data collections. All studies from 2012 to 2014 were completed through 'paper-and-pencil' questionnaires. The experimenter distributing the questionnaire was present during the participation, but on a distance to provide confidentiality. Dataset 6 consists of participants recruited through advertisement on different Internet forums. 243 started to fill in the questionnaire, 190 completed it. Participation was not compensated. This study was carried out in accordance with Swedish national ethical standards put forth by the Central Ethical Review board and the Swedish Research Council and with written informed consent from all participants.

Variables

The attitude to *hen* was assessed with one item 'What is your opinion about the gender-neutral pronoun *hen* in the Swedish language?' (Responses were given on a 7-point response scale ranging from '1 = very positive' to '7 = very negative'). A short text introduced to the question and explained that *hen* was a gender-neutral word that can be used as a complement to the Swedish words representing *she* and *he*.

Behavior *(use of hen)* was measured from 2013 and onward with one item 'Do you use *hen* yourself?' (Responses were given on a 7-point response scale ranging from '1 = No, never' to '7 = Yes, always').

From 2013, participants also indicated whether they previously were familiar with the word *hen* from before. Answers ranged on a 7-point scale from '1 = not al all' to '7 = very much'.

Because there were very small variations in the responses from 2013 to 2014, in 2015 we dichotomized this response option into 'yes' and 'no.'

Sexism was measured with five items from the Swedish version of the Modern sexism scale (Ekehammar et al., 2000; e.g., 'Discrimination against women is no longer a problem in Sweden'; 'Humiliating treatments of women in adverts is unusual'; Answers in terms of agreement or disagreement were given on a 7-point scale from '1 = Strongly disagree' to '7 = Strongly agree). Sexism was included in all six datasets. Means and SD over time are included in **Table 2**.

Political orientation was assessed with one item 'On a political scale from left to right, where is your position?' Answers were given on a 7-point scale from '1 = clearly to the left' to '7 = clearly to the right').

Gender identity was included from 2013 and onward. In 2013 and 2014 it was assessed with two items (e.g., 'To be a woman/man is an important part of my identity,' 'To be a woman/man is important to me,' measured on a 7-point scale from '1 = strongly disagree' to '7 = strongly agree'). In 2015, we began measuring gender identity with a validated sub-scale from Luhtanen and Crocker (1992) collective identity scale (e.g., 'My gender identity is an important reflection of who I am,' 'My gender identity is an important part of my self image,' (measured on a 7-graded scale: '1 = strongly disagree' to '7 = strongly agree'). We used the two positively framed items because these were most similar to the items we used in 2013 and 2014. The reason for this shift was to use a more well-established scale.

Interest in gender issues was indicated with one item 'How interested are you in general in gender issues?' The scale ranged from '1 = not at all' to '7 = very much'. This variable was included from 2013 and onward.

Age and *gender* was given by participants in a free-text response in order to avoid cisgenderism (Ansara and Hegarty, 2014). These variables were included in all datasets.

In order to run the analyses, we collapsed all datasets into one. In the regressions, *Time* was included as a continuous variable for the years 2012–2015. This means that dataset 2–4 was collapsed into 1 year, 2013. Because there were different sample types we controlled for that factor.

TABLE 2 | Means and standard deviations for included variables for each year, respectively.

	2012 (N = 184)			2013 (N = 271)			2014 (N = 40)			2015 (N = 190)		
	M	SD	α	M	SD	α	M	SD	α	M	SD	α
Modern sexism	2.31$_a$	0.64	0.75	2.59$_b$	0.86	0.65	2.95$_b$	1.11	0.78	2.11$_a$	1.21	0.83
Political orientation	4.08$_a$	1.85		4.08$_a$	1.73		4.03$_a$	1.53		3.94$_a$	1.77	
Interest gender issues				4.40$_a$	1.73		3.65$_a$	1.89		5.11$_b$	1.73	
Gender identity				4.87$_a$	1.79	0.84	5.46$_a$	1.60	0.96	4.38$_b$	1.66	0.82

Chronbach's alpha is included for scale measures: modern sexism and gender identity.
Means with different subscripts are significantly different from each other.
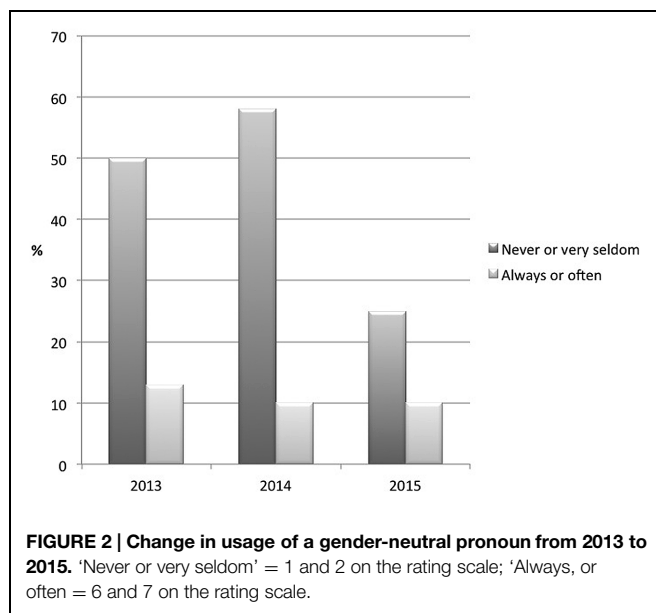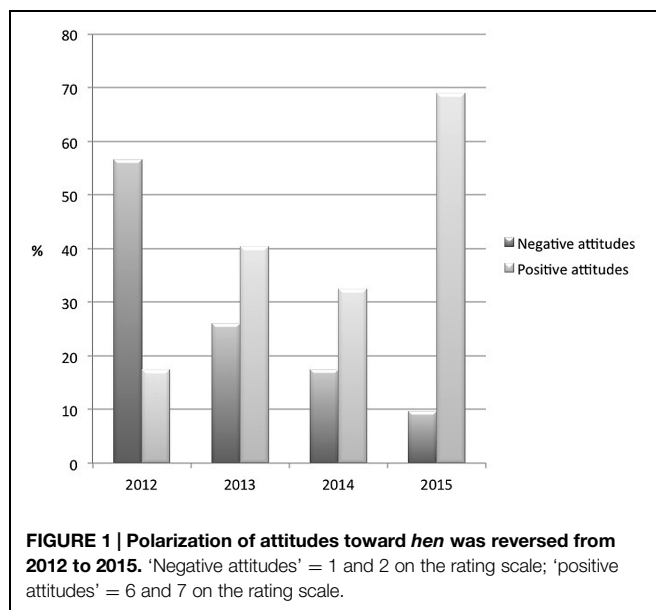
Results

Attitudes to 'hen' and Changes Over Time

Virtually all participants responded that they were familiar with the word *hen*. In 2013 and 2014 more than 95% responded a 6 or 7 on the 7-point scale, while in 2015 99.5% responded 'yes' to the question if they were familiar with the word since before.

The attitudes shifted from negative to positive over time (see **Table 3**). A univariate ANCOVA with year (2012, 2013, 2014, 2015) as the independent variable, sample type (student/community) as covariate, and attitude as the dependent variable, showed that attitudes changed over time, $F(3,679) = 59.22$, $p < 0.001$, = 0.21. *Post hoc* comparisons (Bonferroni adjusted for multiple comparisons), showed that the means did not change significantly from 2012 to 2013, or from 2013 to 2014, but between all other years (p's < 0.004). Furthermore, the attitudes were polarized, such that respondents were either very negative or very positive toward the word *hen*. **Figure 1** shows that the very negative attitudes (i.e., 1 and 2 on the scale) decreased over time (2012 = 56.5%; 2013 = 26.1%; 2014 = 17.5%; 2015 = 9.6%); whereas the very positive attitudes increased (i.e., 6 or 7 on the scale; 2012 = 17.4%; 2013 = 40.4%; 2014 = 32.5%, 2015 = 68.9%).

FIGURE 1 | Polarization of attitudes toward *hen* was reversed from 2012 to 2015. 'Negative attitudes' = 1 and 2 on the rating scale; 'positive attitudes' = 6 and 7 on the rating scale.

Use of 'hen' Over Time

From 2013 and onward respondents also indicated whether or not they used the gender-neutral pronoun *hen* themselves (see **Table 3**). A univariate ANCOVA with year (2013, 2014, 2015) as independent variable, sample type as covariate, and behavior as dependent variable, showed a significant shift in behavior over time, $F(2,498) = 8.56$, $p < 0.001$, $\eta_p^2 = 0.03$. *Post hoc* pairwise comparisons (Bonferroni adjusted for multiple comparisons) showed that the difference was significant between 2013 and 2015, ($p < 0.001$). The responses for behavior were also somewhat polarized but not as much as for the attitudes, and were not reversed over the years (see **Figure 2**). A majority in 2013 (50%) and 2014 (58%) indicated they never or almost never used the word *hen* (as indicated with a 1 or 2 on the rating scale). In 2015, this group had decreased to 25%. However, there was no change in those who very often or always used the word *hen* (as indicated with a 6 or 7 on the rating scale) over time. In 2013, 13% responded they used *hen* often; in 2014 and 2015, 10% indicated they often used *hen*. Thus, both H1 and H2 stating that attitudes will become more positive and the use will increase over time were supported, although the attitudes changed more than the behavior.

FIGURE 2 | Change in usage of a gender-neutral pronoun from 2013 to 2015. 'Never or very seldom' = 1 and 2 on the rating scale; 'Always, or often = 6 and 7 on the rating scale.

Predictors Associated with Attitudes and Use

Hypotheses 3 throughout 6 were related to predictors of attitude and use. To test the influence of time on the attitudes to *hen*, while also controlling for, and investigating effects of, the other

TABLE 3 | Means and SD for 'attitude to *hen*' and 'behavior to use *hen*' over 4 years (2012–2015).

	2012 (*N* = 184)		2013 (*N* = 271)		2014 (*N* = 40)		2015 (*N* = 190)	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Attitude to *hen*	2.88$_a$	2.17	4.38$_{a,b}$	2.19	4.43$_b$	2.02	5.71$_c$	1.89
Behavior use *hen*			2.80$_a$	1.98	2.80$_{a,b}$	1.92	3.30$_b$	1.47

Significance was determined using Bonferroni test for multiple comparisons. Means for attitude and behavior with different subscripts are different from each other.

predictors, a hierarchical regression with all the variables that were measured from 2012 to 2015 was computed. **Table 4** contains the correlations of included variables, collapsed across all years. The regression was performed in three steps (see **Table 5**). Regressions were also computed with dummy coding for time and the results were similar; we chose to present time variable as a continuous variable. The first step included time (2012–2014) and explained 19% of the variance. The longer the word has been known, the more positive were the attitudes. Adding sample type, age, and gender explained further 6% of the variance, such that being a woman, young, and a student was associated with more positive attitudes. Finally, in step 3, modern sexism and political orientation explained an additional 19% of the variance. Those with a right-wing orientation and higher sexism scores were more negative than individuals with left-wing orientation and lower sexism scores. When these factors were included, gender became insignificant, while time was still an important predictor. The total model explained 43% of the variance in the attitude to *hen*. Hence, H3, stating that sexism and political right affiliation would be associated with negative attitudes, H4, stating that younger people would be more positive to *hen*, and H6 stating that time will have an independent and significant effect on attitudes even when controlling for the other predictors, were all supported.

From 2013 and onward three more variables were included in the questionnaires: *behavior (use of hen)*, *gender identity*, and *interest in gender issues*. In order to test whether gender identity and interest in gender issues account for more variance over sexism and political orientation, we calculated two hierarchical multiple regressions for attitude and behavior separately. The correlations, means, and SD are described in **Table 6**.

The regressions were computed in four steps to control for the contribution of variance in each step (see **Table 7**). For the *attitude* to the gender-neutral pronoun *hen*, time explained 9% of the variance in the first step, such that the longer *hen* had been in use, the more positive were the attitudes. The second step, where sample, gender and age were included, explained another 6% of the variance. Again, student samples were more positive than community samples, women were more positive than men, and younger people were more positive than older. The third step included sexism and political orientation, and explained another 21% of the variance, such that modern sexism and being

TABLE 5 | Hierarchical multiple regression analyses predicting attitudes to *hen*.

	Step 1	Step 2	Step 3
	β	β	β
Time	0.435***	0.440***	0.424***
Sample (0 = Community, 1 = Student)		0.097*	0.132***
Gender (0 = Woman, 1 = Man)		−0.115***	−0.021
Age		−0.160***	−0.184***
Modern sexism			−0.270***
Political orientation (high values = right wing)			−0.255***
ΔR²	0.19***	0.06***	0.19***
Total R²			0.43***
N			647

*p < 0.05, **p < 0.01, ***p < 0.001.

right-wing oriented was associated with more negative attitudes. When these variables were included, gender became insignificant. The third step including gender identity and interest in gender issues, explained another 8%. Having a strong gender identity was associated with negative attitudes, whereas being interested in gender issues was associated with a positive attitude. When interest in gender issues and gender identity was introduced, neither gender or sample type were significant predictors, and the beta-weights for modern sexism and political orientation also decreased but remained significant. Although the beta-weight for time decreased in step 4 it remained significant.

In the regression with *behavior* (use of *hen*) as the dependent variable, time itself explained 2% of the variance. When age, gender, and sample type were included in the second step, those variables accounted for another 9% of the variance. Being older and having a masculine gender was associated with less use than being younger and having a feminine gender. The third step included modern sexism and political orientation and explained another 15% of the behavior. Right-wing orientation and sexism was associated with lower use of a gender-neutral pronoun. The fourth step, with gender identity and interest in gender issues explained another 9%, such that a strong gender identity was associated with lower use and being interested in gender issues was associated with a higher use. In the fourth step, time, gender, and sample type were no longer significant

TABLE 4 | Correlations, means, and SDs for variables included in regression.

	1	2	3	4	5	6	7	M	SD
(1) Attitude to *hen*	–							4.35	2.34
(2) Sample (0 = Community, 1 = Student)	0.075	–						1.31	0.46
(3) Age	−0.220**	−0.391*	–					31.67	14.28
(4) Gender (0 = Woman, 1 = Man)	−0.143**	0.057	0.001	–				1.40	0.49
(5) Modern sexism	−0.391**	0.153**	−0.074	0.274**	–			2.41	0.95
(6) Political orientation (high values = right-wing)	−0.374**	0.045	−0.063	0.076	0.448**	–		4.04	1.76
(7) Year	0.431**	−0.189**	−0.024	−0.093*	−0.096*	−0.03	-	2.32	1.14

**Correlation is significant at the 0.01 level (2-tailed).*Correlation is significant at the 0.05 level (2-tailed). N = 672.

TABLE 6 | Correlations, means, SDs for variables included in regressions.

	1	2	3	4	5	6	7	8	9	10	M	SD
(1) Attitude to *hen*	—										4.89	2.16
(2) Behavior, use *hen*	0.701**	—									2.97	1.81
(3) Sample (0 = Community, 1 = Student)	−0.113*	−0.003	—								1.43	0.50
(4) Age	−0.101*	−0.156**	−0.479**	—							29.82	11.66
(5) Gender (0 = Woman, 1 = Man)	−0.170**	−0.209**	0.084	0.028	—						1.39	0.49
(6) Modern sexism	−0.463**	−0.407**	0.149**	−0.069	0.283**	—					2.44	1.04
(7) Political orientation (high values = right wing)	−0.414**	−0.338**	0.064	0.006	0.083	0.469**	—				4.03	1.73
(8) Gender identity	−0.350**	−0.283**	0.028	0.022	0.019	0.213**	0.177**	—			4.74	1.75
(9) Interest gender issues	0.489**	0.460**	−0.087	0.058	−0.297**	−0.477**	−0.391**	−0.188**	—		4.60	1.79
(10) Year	0.292**	0.121**	−0.749**	0.257**	−0.133**	−0.212**	−0.036	−0.128**	0.176**	—	2.81	0.94

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

predictors. Also the beta-weights for political orientation and sexism decreased. The total model explained 32% of the variance. When controlling for all other factors, time contributed to a more positive attitude to a gender-neutral pronoun, although it did not increase the use of the pronoun *hen*. Thus, H6 was partially supported. Again hypothesis H3 and H4 were supported, and as predicted in H5, the strength of gender identity was a stronger predictor than gender itself. In addition, interest in gender issues proved to be a strong and independent predictor of both attitude and use. Even though it did not override the effect of political orientation, it should be taken as an indicator that this is an important aspect to take into consideration in future research.

Our results show that an introduction of a gender-neutral pronoun in the Swedish language was met with high resistance, but that both attitudes and behavior became more positive over time. We found that time predicted the attitude to *hen* also when other factors were controlled for. Other factors that contributed with unique variance to the attitude and the behavior were gender identity (but not gender itself), modern sexism, political orientation, and interest in gender issues.

Discussion

This article has given an overview of the introduction of the new gender-neutral pronoun *hen* in the Swedish language. Data were collected during 4 years, starting in 2012 when the debate about a gender-neutral pronoun began and continued until 2015, 1 year after the word *hen* had been officially included in the Swedish dictionary.

The Impact of Time

The results clearly show how the introduction of *hen* was associated with high resistance (in the media and among lay people), but also that attitudes became positive over time. In 2012, a majority of the study sample was explicitly very negative to the inclusion of a gender-neutral pronoun, whereas only a minority was very positive. However, already in 2013 this polarization was reversed, and in 2015 almost no one was very negative. A similar pattern was found for the use of the gender-neutral pronoun, although this change was smaller.

This is the first study about the introduction of gender-fair language analyzing the attitudes for a specific word over time. Previous research has proposed that variations in attitudes to gender-fair language could be due to how long it has been in use (see for example Sarrasin et al., 2012). This is the first study explicitly testing that hypothesis using data measurements at several time points. Indeed, time was the most important predictor of the attitudes, even after controlling for various other factors. This sends a very important message, because it should motivate language amendments also when there are strong reactions against an implementation.

We found that the attitudes changed faster than the behavior. The debate about *hen* was very wide-spread in the Swedish society, including the broader media landscape, leading to that the familiarity of *hen* very quickly included the large majority.

TABLE 7 | Hierarchical multiple regression predicting attitude to and use (behavior) of a gender-neutral pronoun *hen*.

	Attitude				Behavior			
	Step 1	Step 2	Step 3	Step 4	Step 1	Step 2	Step 3	Step 4
	β	β	β	β	β	β	β	β
Time	0.295***	0.423*	0.383***	0.293***	0.128**	0.204**	0.162**	0.075
Sample (0 = Community, 1 = Student)		0.150**	0.170**	0.100		0.084	0.096	0.029
Gender (0 = Woman, 1 = Man)		−0.130**	−0.028	0.010		−0.203***	−0.111**	−0.064
Age		−0.131**	−0.129**	−0.149***		−0.158**	−0.159***	−0.180***
Modern sexism			−0.282***	−0.178***			−0.270***	−0.162***
Political orientation (high values =right wing)			−0.273***	−0.190***			−0.203***	−0.118**
Interest gender issues				0.258***				0.290***
Gender identity				−0.186***				−0.153***
N				469				470
ΔR^2	0.09***	0.06***	0.21***	0.08***	0.02*	0.09***	0.15***	0.08***
Total R^2				0.44***				0.32***

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Already in 2012, almost 95% of participants were familiar with the word, and in 2015, only 1 out of 190 participants were unfamiliar with *hen*. This may have been of importance for how fast the attitudes changed. For behavior to occur, *hen* must be activated and accessible in a specific moment (Fazio et al., 1989; Fazio and Olson, 2003; Glasman and Albarracín, 2006) as an alternative to, for example, double forms such as *she* or *he*. Because pronouns are often processed automatically (Chung and Pennebaker, 2007) the traditional system with *she* and *he* is probably still cognitively dominant over new forms of pronouns. Accessibility is although likely to increase over time, considering the increasingly widespread use of the word in media (Ledin and Lyngfelt, 2013), and in other arenas. For instance, the word was used in the lyrics of one of the songs to the Swedish contribution to the European Song Contest 2015, indicating its widespread acknowledgment. Social norms also facilitate behavior (Fazio, 1990), and it is plausible that people have been avoiding using *hen* because they still believe that the majority are negative to it. Thus, when people realize that the attitudes have changed, the word may be more common also among lay people and everyday users.

Factors Explaining the Attitudes and Use

The more strongly participants identified themselves with their gender identity, the more negative attitudes they held and the least often they used the word. Women were somewhat more positive toward *hen* and used *hen* more often than men, but gender identity proved to be a much stronger predictor than biological gender. This supports the idea that a gender-neutral pronoun challenges the traditions of a binary gender system. These results also line up with previous research showing that androgynous gender roles were associated with a higher use of gender-fair language than traditional gender roles (Rubin and Greene, 1991). A large body of research indicates that people (especially adults) strongly prefer the system that they currently live in (Jost et al., 2004). People prefer to keep things stable and predictable. Any new word would thus probably elicit some

resistance. However, there is reason to believe that a word explicitly challenging such a basic organizing principle such as the binary gender system elicits even stronger resistance. This resistance may also vary depending on individual factors. As was found in the present research a strong gender identity was negatively associated with attitudes toward *hen*, which can be considered a gender-fair amendment toward neutralization. However a strong gender identity might be positively related to amendments that add feminine alternatives to masculine forms because the binary gender-dichotomy would be even more strongly preserved and perpetuated with such amendments. This is an empirical question.

As in previous research, age, sexism, and political orientation was associated with attitudes to gender-fair language (Parks and Roberton, 2000; Sarrasin et al., 2012; Formanowicz et al., 2013). However, we also found that the influence of those factors decreased when gender identity and interest in gender issues were included. Even though political orientation still proved to be a significant predictor, this may indicate that interest is an important variable that eventually could diminish this effect, considering that there is a growing feminist movement also within the political right in Sweden.

Opinions associated with feminists may evoke higher resistance among people who do not actively endorse such values (Blaubergs, 1980). Thus, when the trendy entertainment magazine in one issue exchanged all third personal pronouns into *hen*, and when newspaper media started to use *hen*, this might have been of more importance than when feminists or linguists debated why *hen* should be used (Cialdini and Goldstein, 2004). Koeser et al. (2014) have also shown that the reading of gender-fair texts increases the use of gender-fair language. Hence, the fact that *hen* occurred more often in ordinary newspapers might have had a positive impact on use and might also imply an increase over the coming years.

Hen in Swedish was adapted from the gender-neutral Finnish word *hän* (Prewitt-Freilino et al., 2012). Maybe there are more words in gender-neutral languages that could be introduced

either in natural or gender-marked languages. Some scholars have pointed to the need to be creative and come up with new words (Wayne, 2004), and borrowing them from other languages could be one strategy. New words might have a potential to override previous problems in applying gender fair language, since they may be less associated with a gender bias, which might be the case with other neutral words (Bäck et al., 2015). It should also be noted that both the Finish word *hän* as well as the Swedish gender-neutral pronoun *hen* very nicely fits into the Swedish system of pronouns, as being literally very close to, as well as alphabetically positioned between *han* ('he') and *hon* ('she'). There is of course a risk that also *hen* could be associated with a male bias in future. Due to our results in this study we believe that such a risk is lower as long as *hen* is used as a generic or a transgender pronoun; however, this is an empirical question. When *hen* is broadly used in society, it is important to replicate studies that investigate how gender is activated when *hen* is used to refer to a person (Wojahn, 2013; Bäck et al., 2015).

Limitations and Future Research

The design of the present study is cross-sectional and not longitudinal, which may imply selection bias in the samples and that other factors such as possible cohort effects may have had an impact on the results. With this in mind, we took care to collect both student and community samples, for which we controlled in the regression analyses. However, these samples were mainly drawn from cities and hence there may still be possible bias in the samples. This implies that generalization from the present study should be done with caution. With respect to cohort effects, our samples were fairly similar, although some minor deviations can be noted. The comparisons of the samples show that sexism was lower in the first year sample and the last year sample, gender interest was higher in the last sample, and gender identity was less strong in the last sample. Political orientation was similar in all samples. Here it can be noted that the last sample was collected using a web survey, which may imply selection bias since those who choose to participate can be expected to be relatively interested in issues of gender and language. With these problems in mind, we computed regressions for 2 years at a time, controlling for sexism, age, gender, and political orientation. Time was a significant factor in all three regressions (2012–2013; 2013–2014; 2014–2015).

This research is fairly explorative and the first of its kind. This entails that the items may not always have been entirely perfectly formulated. For instance, the response scale to the item 'Do you use *hen* yourself?' ranged from 'No, never' to 'Yes, always.' It may not be very clear to the participant what the response option 'Yes, always' entails, and this could be a contributing factor to why the results in general were weaker for the behavioral measure. 'Always' could indicate that one replace all personal pronouns with *hen*, or it could indicate that one always use *hen* when gender is unknown or irrelevant. Another limit is that this measure does not separate between written and spoken language. It is easier to use *hen* in writing than it is to use it in speaking. Future research should take these limitations into account when exploring how *hen* is used.

Language and communication have a large impact on the creation of a common ground and reality, for instance concerning what is considered as normal or desirable (Clark and Brennan, 1991; Hardin and Higgins, 1996). Thus, adding a gender-neutral pronoun to a natural gendered language may influence how individuals with a non-binary gender are perceived. In all our datasets 'gender' was an open-ended question, making it possible to self-categorize as neither woman or man. There were no such responses in 2012–2014, while in 2015, 4% (eight people) indicated a gender identity outside the gender dichotomy. Although it might be a coincidence, it could also be a consequence of the introduction of *hen*. This is something that could be further studied. A related important question that remains is what impact the use of *hen* actually has on representations of gender, and interpersonal attitudes.

We believe it is important to empirically test if common arguments proposed as negative consequences of gender-fair language are true. One such argument that remains to be tested is whether new word forms actually steel attention from the text content. If there is a cognitive load associated with *hen,* reading a text with *hen* should take longer time, and less information should also be recalled from such a text. Finally, it is important to note that we do not argue that time in use operates in isolation from other factors. One important aspect we believe is of great importance is that the Swedish society is becoming increasingly egalitarian and has a strong feminist movement, which includes people of all gender identities, and people with different political opinions. The fact that *hen* has its roots as far back as the 1960's indicates that something else must have sparked the onset of the use in modern day than just time. One factor may be a societal 'readiness' to take this debate. Hence, societies of different levels of such readiness will of course receive a similar implementation differently. However, since there is a strong feminist movement in many societies, as indicated by the UN's 'heforshe' campaign, we believe that the global readiness could be relatively favorable in a near future. Another factor is the word's practical implications. In the Swedish case *hen* was introduced by LGBT communities and within the feminist movement, but clearly it met demands also among lay people as the word became as widespread as soon as it did.

Conclusion

This is the first study analyzing the importance of time in implementing gender fair-language. The introduction of a gender-neutral pronoun in Sweden was firstly met with hostile reactions and negative attitudes, but over the course of only a couple of years, attitudes became largely positive. These results are positive for those working with gender equality and motivate implementations although the initial resistance may be high.

Acknowledgments

References

Ansara, Y. G., and Hegarty, P. (2014). Methodologies of misgendering: recommendations for reducing cisgenderism in psychological research. *Fem. Psychol.* 24, 259–270. doi: 10.1177/0959353514526217

APA. (2012). *Publication Manual of the American Psychological Association,* 6th Edn. Washington, DC: American Psychological Association.

Bäck, E. A., Lindqvist, A., and Gustafsson Senden, M. (2015). Hen can do it: effects of using a gender neutral pronoun in a recruitment situation. *Paper presented at the The 8th Nordic Conferences on Language and Gender*, Stockholm.

Baron, D. (1986). *Grammar and Gender*. New Haven, CT: Yale University Press.

Bem, S. L. (1974). The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* 42, 155–162. doi: 10.1037/H0036215

Benaissa, M. (2014). *Svenska Akademiens Ordlista Inför Hen* [*The Glossary of the Swedish Academy Includes Hen*]. Available at: http://sverigesradio.se/sida/artikel.aspx?programid$=$478&artikel$=$5924958 [accessed July 29, 2014].

Blaubergs, M. S. (1980). An analysis of classic arguments against changing sexist language. *Womens Stud. Int. Q.* 3, 135–147. doi: 10.1016/s0148-0685(80)92071-0

Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). "Sex, syntax, and semantics," in *Language in Mind: Advances in the Study of Language and Thought,* eds D. Getner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 61–79.

Brattström, E. (2014). *Jag Tycker Det Skymmer Kvinnokönet* [*I Think it Obscures the Female Gender*]. Available at: http://www.svd.se/kultur/saol-ingen-censurerande-instans3784668.svd

Cederskog, G. (2012). *Det Lilla Ordet Med Den Stora Laddningen* [*The Small Word With the Great Loading*]. Available at: http://www.dn.se/kultur-noje/det-lilla-ordet-med-den-stora-laddningen/

Chung, C. K., and Pennebaker, J. W. (2007). "The psychological functions of function words," in *Social Communication*, ed. K. Fiedler (New York, NY: Psychology Press), 343–359.

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Clark, H. H., and Brennan, S. E. (1991). *Grounding in Communication*. Washington, DC: American Psychological Association. doi: 10.1037/10096-006

Crandall, C. S., Eidelman, S., Skitka, L. J., and Morgan, G. S. (2009). Status quo framing increases support for torture. *Soc. Influ.* 4, 1–10. doi: 10.1080/15534510802124397

Crawford, M., and Fox, A. (2007). IX. From sex to gender and back again: co-optation of a feminist language reform. *Fem. Psychol.* 17, 481–486. doi: 10.1177/0959353507084333

Douglas, K. M., and Sutton, R. M. (2014). "A giant leap for mankind" but what about women? The role of system-justifying ideologies in predicting attitudes toward sexist language. *J. Lang. Soc. Psychol.* 33, 667–680. doi: 10.1177/0261927x14538638

Eaton, A. A., Visser, P. S., Krosnick, J. A., and Anand, S. (2009). Social power and attitude strength over the life course. *Personal. Soc. Psychol. Bull.* 35, 1646–1660. doi: 10.1177/0146167209349114

Eidelman, S., Crandall, C. S., and Pattershall, J. (2009). The existence bias. *J. Pers. Soc. Psychol.* 97, 765–775. doi: 10.1037/a0017058

Ekehammar, B., Akrami, N., and Araya, T. (2000). Development and validation of Swedish classical and modern sexism scales. *Scand. J. Psychol.* 41, 307–314. doi: 10.1111/1467-9450.00203

Fahl, H. (2014). *Hen Med i Ordlistan* [*Hen in the Dictionary*]. Available at: http://www.dn.se/kultur-noje/spraket/hen-med-i-ordlistan/ [accessed July 29, 2014].

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: the mode model as an integrative framework. *Advan. Exp. Soc. Psychol.* 23, 75–109. doi: 10.1016/S0065-2601(08)60318-4

Fazio, R. H., and Olson, M. A. (2003). "Attitudes: foundations, functions, and consequences," in *The Sage Handbook of Social Psychology*, eds M. A. Hogg and J. Cooper (London: Sage), 139–160.

Fazio, R. H., Powell, M. C., and Williams, C. J. (1989). The role of attitude accessibility in the attitude-to-behavior process. *J. Consum. Res.* 16, 280–288. doi: 10.1086/209214

Formanowicz, M., Bedynska, S., Cisłak, A., Braun, F., and Sczesny, S. (2013). Side effects of gender-fair language: how feminine job titles influence the evaluation of female applicants. *Euro. J. Soc. Psychol.* 43, 62–71. doi: 10.1002/ejsp.1924

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., and Oakhill, J. (2012). Gender representation in different languages and grammatical marking on pronouns: when beauticians, musicians, and mechanics remain men. *Dis. Process.* 49, 481–500. doi: 10.1080/0163853X.2012.688184

Gastil, J. (1990). Generic pronouns and sexist language: the oxymoronic character of masculine generics. *Sex Roles* 23, 629–643. doi: 10.1007/BF00289252

Glasman, L. R., and Albarracín, D. (2006). Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychol. Bull.* 132, 778–822. doi: 10.1037/0033-2909.132.5.778

Hardin, C. D., and Higgins, T. E. (1996). "Shared reality: how social verification makes the subjective objective," in *Handbook of Motivation and Cognition*, eds R. M. Sorrentino and T. E. Higgins (New York, NY: Guilford Press), 28–84.

Hyde, J. S. (1984). Childrens understanding of sexist language. *Dev. Psychol.* 20, 697–706. doi: 10.1037/0012-1649.20.4.697

Jacobson, M., and Insko, W. Jr. (1985). Use of nonsexist pronouns as a function of one's feminist orientation. *Sex Roles* 13, 1–7. doi: 10.1007/BF00287456

Jost, J. T., Banaji, M. B., and Nosek, B. A. (2004). A decade of system justification theory: accumulated evidence of conscious and unconscious bolstering of the status quo. *Polit. Psychol.* 25, 881–919. doi: 10.1037/0022-3514.85.5.823

Jost, J. T., Nosek, B. A., and Gosling, S. D. (2008). Ideology: its resurgence in social, personality, and political psychology. *Perspect. Psychol. Sci.* 3, 126–136. doi: 10.1111/j.1745-6916.2008.00070.x

Koeser, S., Kuhn, E. A., and Sczesny, S. (2014). Just reading? How gender-fair language triggers readers' use of gender-fair forms. *J. Lang. Soc. Psychol.* 34, 343–357. doi: 10.1177/0261927x14561119

Koeser, S., and Sczesny, S. (2014). Promoting gender-fair language: the impact of arguments on language use, attitudes, and cognitions. *J. Lang. Soc. Psychol.* 33, 548–560. doi: 10.1177/0261927x14541280

Lagerwall, K. (2012). *Hen Gör Barnen Förvirrade* [*Hen makes children confused*]. Available at: http://www.dn.se/nyheter/sverige/kritiker-hen-gor-barn-forvirrade/ [accessed February 14, 2012]

Ledin, P., and Lyngfelt, B. (2013). Om bruket av hen i bloggar, tidningstexter och studentuppsatser [The use of hen in blogs, newspaper articles and student work]. *Språk Och Stil* 23, 141–174.

Lenton, A., Sedikides, C., and Bruder, M. (2009). A latent semantic analysis of gender stereotype-consistency and narrowness in American English. *Sex Roles* 60, 269–278. doi: 10.1007/s11199-008-9534-z

Love, A. (2014). A room of one's own: afe placement for transgender youth in foster care. *N Y. Univ. Law Rev.* 89:2265.

Luhtanen, R., and Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Personal. Soc. Psychol. Bull.* 18, 302–318. doi: 10.1177/0146167292183006

Lundquist, J. (2012). *Kiwi Och Monsterhunden* [*Kiwi and the Monster Dog*]. Stockholm: Olika Förlag.

MacKay, D. G. (1980). Psychology, prescriptive grammar, and the pronoun problem. *Am. Psychol.* 35, 444–449. doi: 10.1037/0003-066x.35.5.444

McMinn, M. R., Lindsay, S. F., Hannum, L. E., and Troyer, P. K. (1990). Does sexist language reflect personal characteristics? *Sex Roles* 23, 389–396. doi: 10.1007/BF00289227

Milles, K. (2013). En öppning i en sluten ordklass? Det nya användandet av pronomenet hen [An opening in a closed word class? The new use of the pronoun hen]. *Språk Och Stil* 23, 107–140.

Milles, K., Salmson, K., and Tomicic, M. (2012). *Det Behövs Ett Nytt ord i Det Svenska Språket* [*A New Word is Needed in the Swedish Language*]. Available at: http://www.svd.se/opinion/brannpunkt/det-behovs-ett-nytt-ord-i-svenska-spraket6784859.svd

Moreland, R. L., and Topolinski, S. (2010). The mere exposure phenomenon: a lingering melody by Robert Zajonc. *Emot. Rev.* 2, 329–339. doi: 10.1177/1754073910375479

Moulton, J., Robinson, G. M., and Elias, C. (1978). Psychology in action - sex bias in language use - neutral pronouns that arent. *Am. Psychol.* 33, 1032–1036. doi: 10.1037/0003-066X.33.11.1032

Murdock, N. L., and Forsyth, D. R. (1985). Is gender-biased language sexist? A perceptual approach. *Psychol. Women Q.* 9, 39–49. doi: 10.1111/j.1471-6402.1985.tb00859.x

Norton, A. T., and Herek, G. M. (2013). Heterosexual's attitudes toward transgender people: findings from a national probability sample of U.S. adults. *Sex Roles* 68, 738–753. doi: 10.1007/s11199-011-0110-6

Öhberg, P., and Wängnerud, L. (2014). Testing the impact of political generations: the class of 94 and pro-feminist ideas in the Swedish Riksdag. *Scand. Polit. Stud.* 37, 61–81. doi: 10.1111/1467-9477.12014

Olsson, I. (2015). "Hen Och Andra Sätt Att Skriva Könsneutralt [Hen and Other Ways of Writing Gender-Neutral]," in *Klarspråak* (*A newsletter from the Swedish Language Council*).

Parks, J. B., and Roberton, M. A. (1998). Contemporary arguments against nonsexist language: blaubergs (1980) revisited. *Sex Roles* 39, 445–461. doi: 10.1023/A:1018827227128

Parks, J. B., and Roberton, M. A. (2000). Development and validation of an instrument to measure attitudes toward sexist/nonsexist language. *Sex Roles* 42, 415–438. doi: 10.1023/A:1007002422225

Paterson, L. (2014). *British Pronoun use, Prescription, and Processing* (Basingstoke: Palgrave Macmillan).

Phillips, J. L. (1981). More on the pronoun problem. *Am. Psychol.* 36, 694–694. doi: 10.1037/0003-066x.36.6.694

Prentice, D. A. (1994). Do language reforms change our way of thinking? *J. Lang. Soc. Psychol.* 13, 3–19. doi: 10.1177/0261927x94131001

Prewitt-Freilino, J. L., Caswell, T. A., and Laakso, E. K. (2012). The gendering of language: a comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles* 66, 268–281. doi: 10.1007/s11199-011-0083-5

Rubin, D. L., and Greene, K. L. (1991). Effects of biological and psychological gender, age cohort, and interviewer gender on attitudes toward gender-Inclusive/exclusive language. *Sex Roles* 24, 391–412. doi: 10.1007/BF002 89330

Rubin, D. L., and Greene, K. (1994). Adopting gender-inclusive language reforms. *J. Lang. Soc. Psychol.* 13, 91–114. doi: 10.1177/0261927X94 132001

Samuelson, W., and Zeckhauser, R. (2005). *Status Auo Bias in Decision Making.* Northampton, MA: Edward Elgar Publishing.

Sarrasin, O., Gabriel, U., and Gygax, P. (2012). Sexism and attitudes toward gender-neutral language: the case of English, French, and German. *Swiss J. Psychol.* 71, 113–124. doi: 10.1024/1421-0185/a000078

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language," in *Social Communication*, ed. K. Fiedler (New York, NY: Psychology Press), 163–187.

Stahlberg, D., and Sczesny, S. (2001). Effects of the generic use of the masculine pronoun and alternative forms of speech on the cognitive visibility of women. *Psychol. Rundsch.* 52, 131–140. doi: 10.1026//0033-3042.52.3.131

Stahlberg, D., Sczesny, S., and Braun, F. (2001). Name your favorite musician - effects of masculine generics and of their alternatives in German. *J. Lang. Soc. Psychol.* 20, 464–469. doi: 10.1177/0261927x01020004004

Strahan, T. E. (2008). 'They' in Australian English: non-gender-specific or specifically non-gendered? *Aus. J. Ling.* 28, 17–29. doi: 10.1080/07268600701877473

Visser, P. S., and Krosnick, J. A. (1998). Development of attitude strength over the life cycle: surge and decline. *J. Pers. Soc. Psychol.* 75, 1389–1410. doi: 10.1007/BF00289330

Wayne, L. D. (2004). Neutral pronouns: a moderst proposal whose time has come. *Can. Women Stud.* 24, 85–92.

Wojahn, D. (2013). De personliga pronomenens makt: en studie av hur pronomen styr våra föreställningar om personer [The power of the personal pronouns. A study on how pronouns direct person perception]. *Svenskans Beskrivning* 32, 356–367.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *J. Personal. Soc. Psychol.* 9, 1–27. doi: 10.1037/h0025848

Capturing socially motivated linguistic change: how the use of gender-fair language affects support for social initiatives in Austria and Poland

Magdalena M. Formanowicz[1]*, Aleksandra Cisłak[2], Lisa K. Horvath[3] and Sabine Sczesny[1]

[1] Department of Psychology, University of Bern, Bern, Switzerland, [2] Faculty of Psychology, Department of Social Psychology, University of Social Sciences and Humanities, Warsaw, Poland, [3] TUM School of Management, Technische Universität München, Munich, Germany

Gender-fair language consists of the symmetric linguistic treatment of women and men instead of using masculine forms as generics. In this study, we examine how the use of gender-fair language affects readers' support for social initiatives in Poland and Austria. While gender-fair language is relatively novel in Poland, it is well established in Austria. This difference may lead to different perceptions of gender-fair usage in these speech communities. Two studies conducted in Poland investigate whether the evaluation of social initiatives (Study 1: quotas for women on election lists; Study 2: support for women students or students from countries troubled by war) is affected by how female proponents (lawyers, psychologists, sociologists, and academics) are referred to, with masculine forms (traditional) or with feminine forms (modern, gender-fair). Study 3 replicates Study 2 in Austria. Our results indicate that in Poland, gender-fair language has negative connotations and therefore, detrimental effects particularly when used in gender-related contexts. Conversely, in Austria, where gender-fair language has been implemented and used for some time, there are no such negative effects. This pattern of results may inform the discussion about formal policies regulating the use of gender-fair language.

Keywords: grammatical gender, gender-fair language, political correctness, social change, feminism

INTRODUCTION

The line it is drawn
The curse it is cast
The slow one now
Will later be fast
As the present now
Will later be past
The order is
Rapidly fadin'
……..
For the times they are a-changin'.

> *Bob Dylan, 1963*

These lyrics by Dylan capture a rarely examined phenomenon in social psychology, that is, social reality changes over time and may do so even within a fairly short period. Here, we will look at socially motivated changes in language such as language reforms introduced to instigate and promote changes in social reality. To our knowledge, such language policies' effectiveness has never been examined. Such an examination would require a longitudinal approach with measurements being taken over several points of time. The disadvantage of such an approach is that the time within which changes are to happen is unspecified, which constitutes a serious challenge to the budgetary and time framework of any research. We tried to overcome this disadvantage by using cross-sectional research that compares different speech communities at different stages of implementing a specific language reform.

In our research, we focused on gender-fair language (or "non-sexist language," UNESCO, 1999 or "gender-exclusive language," Stout and Dasgupta, 2011), which is a type of language use that aims to represent women and men symmetrically and equally. In languages where gender-fair language was or is still a matter of debate (English in the 1990s: McConnell and Fazio, 1996; Polish and Italian in the first two decades of the twenty-first century: Mucchi-Faina, 2005; Merkel et al., 2012; Formanowicz et al., 2013), the use of feminine job titles for individual women was found to reduce women's professional credibility and evaluation, suggesting that gender-fair language and other regulations for political correctness may be counter productive. However, this conclusion may be premature as we still do not know the popular reaction to politically correct language after it has been implemented for a longer time. Positive effects of gender-fair language have been reported only for Germany, where this usage has been in practice for a longer time (Vervecken and Hannover, 2012). Hence, over time, politically correct language can be reasonably assumed to become a linguistic standard and thus may then trigger positive evaluation among its users.

We tested this assumption by comparing two speech communities where grammatical gender languages are spoken (Polish in Poland and German in Austria), which substantially differ with respect to gender-fair usage. While pertinent language reforms have been implemented and acknowledged in Austrian German, gender-fair language is rarely accepted and is often rejected in Polish. Using the same research paradigm to examine these two countries and languages representing different stages of linguistic reform, allowed the indirect study of the longitudinal effects of socially motivated language reform.

Gender-fair Language in Poland and Austria

In languages with grammatical gender (such as German and Polish), most human nouns and pronouns are differentiated as feminine or masculine. Therefore, the principle strategy employed to make a language gender fair is to have feminine forms of human nouns used more frequently and systematically to make female referents visible. This means masculine generics, that is, grammatically masculine forms meant to represent both genders (e.g., German *Leser,* Polish *czytelnicy* "readers, masc.") are replaced by feminine–masculine word pairs (e.g., German *Leserinnen und Leser;* Polish *czytelnicy i czytelniczki* "readers, fem. and readers, masc."). Additionally, feminine role names or job titles are introduced to designate female job holders explicitly (e.g., German *Psychologin* or Polish *psycholożka* "female psychologist").

However, across the two countries and languages, differences persist in the adoption of gender-fair language. Two main reasons account for these differences. The first concerns the time of implementation. The debate about gender-fair language was a hot topic in Germany in the 1970s (Trömel-Plötz, 1978). Since then, official regulations have been adopted in German-speaking countries. The implementation of gender-fair language has progressed so far that there is even a special Microsoft add-in for gender-fair German[1]. In Austria, almost all universities and government institutions have their own guidelines for gender-fair language (e.g., University of Salzburg: Gendup, 2012; Technical University of Vienna: Arbeitskreis für Gleichbehandlungsfragen der TU Wien, 2010; e.g., Ministry of Education, Science and Culture: Bundesministerium für Bildung, Wissenschaft und Kultur, 2001; Ministry of Science and Research: Bundesministerium für Wissenschaft und Forschung, 2011). Presently, job advertisements must be phrased in a gender-fair way, e.g., with word pairs (e.g., German *Psychologin oder Psychologe* "psychologist, fem. or psychologist, masc.") to signal that applications from both genders are welcome (Europäisches Parlament, 2009). Furthermore, according to legal regulations for equal treatment in Austria (Gleichbehandlungsgesetz, 2004; Bundesministerium für Frauen und Öffentlichen Dienst, 2009), organizations are fined if they do not advertise jobs in a gender-fair way. However, in Poland, official regulations or guidelines for gender-fair language are absent and its use is rare. According to numerous researchers, the implementation of gender-fair language has reached different stages in Austria and Poland. In an analysis of job advertisements from four European countries, only 9% of Austrian job advertisements (but as many as 83% of Polish advertisements) were found to contain masculine generics (Hodel et al., 2013).

The second reason why countries/languages differ in their implementation of gender-fair language is the relative ease with which feminine forms can be introduced. While creating feminine human nouns is fairly easy in German (mostly by adding the feminine suffix *-in* to the masculine form, e.g., *Lehrerin* "teacher, fem."), this is much more complicated in Slavic languages (Koniuszaniec and Błaszkowska, 2003). In Polish, feminine forms of some role nouns can easily be derived with the suffix *-ka* (e.g., *psycholoż-ka* "psychologist, fem."); however, other feminine job titles with this suffix coincide with diminutive forms (e.g., Polish *fizyczka* "physicist, fem." or "little physics"). Moreover, some feminine forms of job titles denote not only a feminine job holder but also an object (e.g., Polish *drukarka* "printer, fem." = "female printer" and "printing machine" from *drukarz* "printer, masc."). Other job titles show a semantic asymmetry: Polish *professorka* "professor, fem." usually refers to a high school teacher, whereas the masculine form *professor* designates a prestigious academic position. Certain feminine

---

[1]http://gendering.codeplex.com/

forms also traditionally mean "wife of" rather than "female job holder" (e.g., *krawcowa* "tailor, fem." or "wife of a tailor").

Considering these differences, we hypothesized that reactions to gender-fair language would differ in Poland and Austria. In line with earlier findings, we assumed that reactions to gender-fair language would be more negative than reactions to traditional masculine forms in Poland, where gender-fair usage is still novel. However, in Austria, where gender-fair language is well known and fairly established, we expected gender-fair forms to trigger highly positive reactions than the traditional use of the masculine. We conducted three studies (Studies 1 and 2 in Poland and Study 3 in Austria) with a similar design to examine how the use of gender-fair language or masculine forms affected respondents' support for social initiatives (Studies 1–3) addressing gender-related (Studies 1–3), or non-gender-related topics (Studies 2 and 3).

STUDY 1

Methods
Participants
Study 1 was conducted in Poland via Internet. The website hosting the study was accessed by 331 individuals, 122 of whom left the first page without completing it. Thus, the final sample consisted of 209 individuals (120 women, 89 men, M age = 33.73, $SD = 10.33$ years). Of the total participants, 63% had higher education degrees, 36% secondary education, and 1% primary education.

All of the described research was conducted according to the recommendations for online research of Eynon et al. (2008). Participants were anonymous, expressed their consent to participate in the study, and were provided with the opportunity to obtain additional information on the study. The first study was a pilot study, and at the time, no institutional approval was needed in Poland for pilot studies. As the study yielded interesting results, we decided to include them in the manuscript and applied for ethical approval for subsequent studies. The study protocol was reviewed and authorized by the University of Social Sciences and Humanities Institutional Review Board (Decision record: 30/2013).

Measures and Procedure
The study was conducted shortly before the elections of regional authorities in Poland and immediately before the deadline for the parties to submit lists of candidates to the Election Committee (in October 2010). The elections were preceded by a nationwide debate about introducing quotas for women for the election lists. Leading women in Polish society demanded a legal act according to which 50% of positions on the list would be reserved for women. This was supported with over 150,000 signatures from Polish citizens. At the time of the study, no quota system had been legally introduced; however, the topic was very popular. In fact, a legal act reserving a quota of 35% of all positions on the election lists for women and/or men was adopted shortly, thereafter by the Polish Parliament on January 5, 2011. On the website, the study was announced as a 3-min survey concerning democracy. The introduction read as follows:

"The regional elections are forthcoming, and shortly the deadline for submitting the list of candidates to the Election Committee will be reached. The legal act for a quota of 50% women candidates on the election list is under inspection by the Parliament but has not yet been decided upon. Nevertheless, women leaders (among them lawyers, psychologists, sociologists, and academics) are proposing to assign 50% of the positions on the election list to women as a societal grassroots initiative. According to this initiative, including women in the election lists would signal genuine support for gender equality in a modern Poland."

The introduction contained the following manipulation. Half of the participants received the description of women proponents in the masculine form (Polish *adwokatów, psychologów, socjologów i nauczycieli akademickich* "lawyers, masc., psychologists, masc., sociologists, masc., and university professors, masc."), the other half in the feminine form (*adwokatek, psycholożek, socjolożek i nauczycielek akademickich*). The original version of this manipulation (as well as of Study 2 and 3) is presented in the Supplementary Material available online. As women and men sometimes react differently to linguistic forms (e.g., Braun et al., 2005), participant gender was included as another factor.

After reading the introduction, participants answered two questions: "What are your feelings about the introduction of a gender quota in Poland?" and "What are your feelings about the social initiative presented?" They were asked to use a slider to answer these questions. The slider was preset to the mid-point position and the answers were recorded at 1-point intervals ranging from 0 (very negative) to 100 (very positive). Both items were averaged and formed a reliable scale (Cronbach's α = 0.87). This scale served as a dependent measure indicating the evaluation of the gender equality initiative[2]. To assess participants' actual support for the quota, they were also asked whether they had signed the support sheet for the quota act during the previous months. The matrix of correlation coefficients of the main variables of interest for all three Studies is available in **Table 1**. Finally, the participants who provided demographical data were asked for comments and were provided with debriefing information about the study.

Results and Discussion
To test our assumptions, we conducted a regression analysis with evaluation of the social initiative as a dependent variable. In the first step, we used linguistic form (coded 0 for masculine and 1 for feminine) and participant gender (0 for male and 1 for female) as predictors, and support for the quota by signing the support sheet (0 for no and 1 for yes) as a covariate variable in the analysis[3]. The reason to use political attitudes as a covariate

---

[2]To assess behavioral intentions, we also asked participants whether they would support the presented social initiative with their signature, with possible answers being no (coded as 1), I don't know (coded as 2), and yes (coded as 3). When this variable was used as a dependent variable, the pattern of results matched those presented in the main analysis. The interaction term was significant at the 0.09 level.
[3]To examine the independence of the covariate and other predictors, a logistic regression was used to test whether the log odds of support for the parity act depended on experimental condition and participant gender as well as their

in our analysis was that political views can have an impact on the main dependent variable used in our studies, that is support for social equality initiatives. This assumption stems from the fact that liberals do support social equality much more than the conservatives (Jost et al., 2003). In the second step, we added an interaction term (linguistic form and participant gender), since the effects of gender-fair language may be affected by this factor (e.g., Braun et al., 2005). The results indicated that the effects of linguistic form were moderated by participant gender. An examination of the conditional effects of the linguistic form using the Hayes (2012) macro revealed that the effect occurred only among the male participants: $b = -13.08$, $SE = 5.77$; $p = 0.02$; it did not occur among the women participants: $b = 4.70$, $SE = 4.99$; $p = 0.35$. In other words, while women's evaluations of the gender equality initiative were independent of the linguistic form employed, men's evaluations were less favorable when the proponents were referred to in the feminine than in the masculine. The means and SD for all the three studies are presented in **Table 2** and the results of the regression analysis are presented in **Table 3**.

Study 1 showed that the gender-related social initiative was evaluated less favorably by men when framed in a feminine than in a masculine form. However, no such difference was

product term. The overall regression was not significant, $\chi^2_{(3)} = 1.26$, $p = 0.74$, Nagelkerke $R^2 = 0.01$. This indicates that support for the parity act was similar across genders and experimental conditions and the use of covariate was justified. Information on the support for parity act was provided by 205 participants.

observed for women. Earlier studies on gender-fair language already observed that men are less supportive of gender-fair language (Jacobson and Insko, 1985; Matheson and Kristiansen, 1987; Parks and Roberton, 2002, 2004), and our results are consistent with these findings. Moreover, it must be emphasized that Study 1 was performed at a time when a heated debate on quotas was ongoing in Poland. Several issues regarding gender equality were raised at the time, and gender was a salient concept. This may have increased the intergroup divides between men and women as well as men's opposition to gender-fair language, which is often mediated by attitudes toward women in general (Parks and Roberton, 2004). However, a serious limitation of Study 1 is that the social initiative presented was about gender equality. This topic may have reinforced the effect of feminine forms in the description. Language reform in the direction of gender-fairness was indeed a political act and originated from the feminist movement (Pauwels, 2003). Thus, novel feminine forms used in a gender context may be perceived as signaling feminism. This could be problematic, since even individuals who support gender equality often avoid calling themselves feminists, as reflected in utterances such as "I'm not a feminist but …" (Buschman and Lenart, 1996; Williams and Wittig, 1997; Burn et al., 2000).

In general, if gender-fair language is perceived as questioning traditional gender arrangements, negative effects should occur mostly in connection with gender issues. However, if gender-fair language is rejected solely because of its novelty, then the effect observed in Study 1 should be independent of the goal of an

TABLE 1 | Matrix of correlation coefficients across all three studies.

	Study 1			Study 2				Study 3			
	PG	PV	E	PG	PV	GI	E	PG	PV	GI	E
Linguistic form	−0.03	0.04	−0.04	−0.04	0.06	0.04	−0.04	−0.02	0.01	0.01	0.14*
Participant Gender (PG)		0.06	0.26***		−0.26***	0.07	0.17***		−0.26***	−0.03	0.17*
Political Views (PV)			0.36***			0.00	−0.28***			−0.07	−0.25***
Goal of the Initiative (GI)							−0.17***				−0.10

*p < 0.05, ***p < 0.001.
E—in the correlation matrix refers to the Evaluation of the Initiative that is the main Dependent Variable used across the three studies.

TABLE 2 | Means and standard deviations of evaluation of initiatives presented with masculine or feminine forms for gender and non-gender related initiatives according to participant gender across all three studies.

		Gender initiative				Non-gender initiative			
		Women		Men		Women		Men	
		M	SD	M	SD	M	SD	M	SD
Study 1	Feminine forms	64.33	29.15	40.42	27.70				
	Masculine forms	59.21	26.44	52.19	32.71				
Study 2	Feminine forms	3.73	1.37	2.62	1.25	4.35	1.10	3.87	1.54
	Masculine forms	4.01	1.51	3.13	1.55	4.12	1.07	3.85	1.32
Study 3	Feminine forms	4.62	1.50	4.21	1.68	5.01	0.96	4.43	1.31
	Masculine forms	4.26	1.53	3.67	1.82	4.52	1.40	3.97	1.67

All means were adjusted for the covariate used in the analysis namely political views.

TABLE 3 | Study 1. Regression model for the evaluation of the initiative.

	ΔR^2	B	SE B
MODEL 1			
Intercept		41.76***	3.68
Linguistic Form (LF)		−2.91	3.82
Participant Gender (PG)		14.55***	3.86
Support for the parity act		22.79***	4.24
MODEL 2	0.02*		
Intercept		47.15***	4.31
Linguistic Form (LF)		−13.08*	5.77
Participant Gender (PG)		5.20	5.53
Support for the parity act		22.99***	4.19
LF × PG		17.78*	7.62

*$p < 0.05$, ***$p < 0.001$.
Model 1: Adjusted $R^2 = 0.17$; $F_{(3, 201)} = 15.45$; $p < 0.001$.
Model 2: Adjusted $R^2 = 0.19$; $F_{(4, 200)} = 13.20$; $p < 0.001$.

initiative. In Study 1 the support for a social initiative might have been influenced by both, the linguistic form and the readiness to accept gender quotas. Study 2 was designed to address this possible confound.

STUDY 2

Study 2 aimed to replicate the effect of linguistic form found in Study 1. In addition, it examined the question of whether the goal of the initiative, a gender-related vs. other issue, moderated the effect. To avoid associations with in-group interests and to stay clear of ongoing debates about quotas, the gender-related issue in Study 2 involved women professionals helping young female students. The non-gender-related goal was helping students from countries affected by war. The initiative supported affirmative action either for women or for students from countries affected by war. The initiative was presented as follows:

Methods
Participants
Study 2 was again conducted online in Poland and was advertised in the academic forums of two universities in Warsaw. The website of the study was accessed by 744 persons. However, many individuals left the page without completing it; thus, the final sample consisted of 577 students (474 women, 103 men; mean age = 25.50, SD = 6.40 years).

Measures and Procedure
The study was presented as a part of a research project investigating "possibilities for the development of the system of higher education in Poland." The announcement described the study as a 5-min survey concerning the development of the Polish system of higher education. Participants were to evaluate a grassroots campaign that concerned the system of higher education. To support the cover story, the initiative was described in the layout of a popular opinion magazine in Poland. The initiative supported affirmative action either for women or for students from countries affected by war. The initiative was presented as follows:

"Female leaders, including many lawyers, psychologists, and academics, have proposed the introduction of scholarships and additional positions in the areas favored by the Ministry of Higher Education[4] for women/students from countries at war. According to psychologist Magda Leska, initiator of the campaign, this would promote the development of economic life, science, and factual gender equality [gender goal] vs. equality [non-gender goal] in access to higher education and the labor market in the world."

Similar to Study 1, the female proponents of the fictitious initiatives were referred to either with the masculine or the feminine form of their professional title; correspondingly, reference was made to either *psycholog Magda Leska-inicjator akcji* (masculine forms) or to *psycholożka Magda Leska-inicjatorka akcji* (feminine forms).

After reading the introduction, participants were asked to evaluate the proposal by answering seven questions. Participants indicated whether the initiative (1) was generally popular, (2) was governed by genuine concern for other people, (3) was good for the system of higher education; and had the potential of increasing (4) the prestige of higher education in Poland, (5) the quality of schooling, (6) the competitiveness of Polish institutions of higher education, and (7) should be implemented at all Polish institutions of higher education. Answers to these questions could vary from 1 (definitely not) to 7 (definitely yes)[5]. The answers were averaged to form a scale evaluation for the initiative (α = 0.94), which served as dependent measure. In contrast to Study 1, we also measured participants' political attitudes (one item with answers from 1 (very liberal) to 7 (very conservative). Moreover, we asked their opinions on factors influencing women's positions in the job market. For this purpose, we provided seven items from the Neosexism Scale (Tougas et al., 1995), which included such items as "Women will make more progress by being patient and not pushing too hard for change." After recoding several items, we combined the items into a reliable scale (α = 0.74) that captured participants' political attitudes, including gender-related features. Finally, participants were asked for their comments and were provided with debriefing information about the study.

Results and Discussion
Similar to Study 1, we conducted a regression analysis. In the first step, we used linguistic form (coded 0 for masculine and 1 for feminine), goal of the initiative (0 for non-gender and 1 for gender), and participant gender (0 for male and 1 for female) as predictors. As in Study 1, we included participants' political views (mean-centered) as a covariate in the analysis. In the second step, we added three two-way interaction terms derived by multiplying the initial predictors; and in the third step, we added one three-way interaction term. The analysis of the full model with all two-way interactions and the three-way interaction of participant gender, linguistic form, and goal of the initiative revealed that

---

[4]Most often, majors in technical and natural sciences receive financial support from the Ministry of Higher Education to ensure there are enough suitably qualified experts in strategically important economic domains.
[5]To assess behavioral intentions, we asked participants whether they would support the presented social initiative with their signature, possible answers being no (coded as 0) and yes (coded as 1).

TABLE 4 | Study 2. Regression model for the evaluation of the initiative.

	ΔR^2	B	SE B
MODEL 1			
Intercept		3.82***	0.15
Linguistic Form (LF)		−0.03	0.11
Participant Gender (PG)		0.43**	0.14
Goal of the Initiative (GI)		−0.46***	0.11
Political views		−0.36***	0.06
MODEL 2	0.02**		
Intercept		4.03***	0.22
Linguistic Form (LF)		0.07	0.27
Participant Gender (PG)		0.03	0.24
Goal of the Initiative (GI)		−0.81**	0.28
Political views		−0.36***	0.06
LF × PG		0.16	0.28
LF × GI		−0.47*	0.21
PG × GI		0.71*	0.28

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
Model 1: Adjusted $R^2 = 0.11$; $F_{(4, 570)} = 19.28$; $p < 0.001$.
Model 2: Adjusted $R^2 = 0.13$; $F_{(7, 567)} = 12.93$; $p < 0.001$.

the three-way interaction did not improve the model, $\Delta R^2 = 0.000$. Results of the regression analysis are presented in **Table 4**. Because we predicted only the two-way interaction of linguistic form and goal of the initiative and that the interaction of participant gender and linguistic form observed in Study 1 would be replicated, we used Model 2 in the Hayes (2012) SPSS macro, testing the interactions of the focal predictor (linguistic form) with the two remaining factors[6]. The hypothesized interaction of linguistic form and goal of the initiative was significant and indicated that the conditional effect of linguistic forms was close to significant for the gender initiative: $b = −0.27$, $SE = 0.15$; $p = 0.07$. The gender initiative was evaluated less favorably when presented with feminine forms than when masculine forms were used. However, evaluation of the non-gender initiative was not affected by linguistic form: $b = 0.21$, $SE = 0.15$; $p = 0.16$.

For the gender initiative, Study 2 replicated the results of Study 1 and showed that both male and female participants evaluated the initiatives less favorably when it was framed in the feminine than in the masculine form. Having documented the negative effects of gender-fair language in a country where this linguistic usage is novel, we then examined the effects in a country where use of gender-fair language is already well established.

STUDY 3

Study 3 was conducted in Austria to clarify whether the long-term practice of gender-fair language is reflected in positive reactions to this usage in the evaluation of social initiatives.

[6]The remaining two-way interaction of goal of the initiative and participant gender was also significant. The non-gender initiative was evaluated similarly by both genders $b = 0.10$, $SE = 0.19$; $p = 0.62$, whereas the gender initiative was evaluated more favorably by female than male participants $b = 0.83$, $SE = 0.21$; $p < 0.001$.

Methods
Participants

Study 3 was conducted online and was advertised in Austrian forums and via email. We offered participants the opportunity to take part in a lottery for five 10-Euro vouchers. The website of the study was accessed by 309 individuals; the final sample of those who completed the study comprised 210 students (113 women, 96 men, and one individual who did not provide information on gender). To ensure that the participants had sufficient linguistic competence to notice the subtle linguistic manipulation, we excluded four individuals whose native language was not German from further analysis. Thus, the final sample consisted of 206 students (110 women, 95 men, M age $= 31.92$, $SD = 9.48$ years).

Measures and Procedure

The announcement described the study as a 5-min survey on the development of the system of higher education in Austria. In daily life, it is quite common to use academic titles when introducing people. In Austria, the feminization of academic titles is available (Universitätsgesetz, 2002), although masculine forms were used for women in former times. Therefore, the woman in the social initiative was either introduced as "Dr. Martina Winkler (Psychologe)" (masculine forms) or as "Dr. Martina Winkler (Psychologin)" (feminine forms) "Dr. Martina Winkler (psychologist)."

The initiative was evaluated using five questions, a shortened version of the scale used in Study 2[7]. Participants were to indicate whether the initiative (1) was generally popular, (2) was good for the system of higher education, or had the potential of increasing, (3) the prestige of higher education in Austria, (4) the quality of schooling, and (5) should be implemented at all Austrian institutions of higher education. Answers to these questions could vary from 1 (definitely not) to 7 (definitely yes). The answers were averaged to form a scale evaluation of the initiative ($\alpha = 0.90$), which served as dependent measure. Participants also answered the yes/no question "Would you support this initiative if it were to be implemented at your university?" Similar to Study 2, we assessed participants' political attitudes [one item with answers from 1 (very liberal) to 7 (very conservative)]. Moreover, we asked their opinions on factors influencing the situation of women. Thus, we provided seven items from the gender-specific system justification scale (Jost and Kay, 2005, adapted for German by Ullrich and Cohrs, 2007)[8]. After recoding several items, we combined them into a reliable scale ($\alpha = 0.78$; one item was removed due to very low inter-item correlations), which captured participants' political attitudes, including their attitudes

[7]In comparison to Study 2, two questions were omitted. One question (assessing the competitiveness of Polish institutions of higher education) was omitted because we have encountered considerable problems with translation and back-translation of this item into German. Second question (whether the initiative was governed by genuine care for other people) was omitted because we have included two additional similar questions examining for exploratory purposes judgments of fairness of the initiative (see also Footnote 9).

[8]In Study 2 we have used a scale that was previously used in Poland (Formanowicz et al., 2013). However, for the German language we referred to the related scale for which the translation was available for us.

on gender-related issues. Finally, participants were asked for demographic data and for their comments and were provided with information about the study[9].

Results and Discussion

We conducted a regression analysis. In the first step, we used linguistic form (coded 0 for masculine and 1 for feminine), goal of the initiative (0 for non-gender and 1 for gender), and participant gender (0 for male and 1 for female) as predictors. Similar to Studies 1 and 2, we included participants' political views (mean-centered) as covariates in the analysis. In the second step, we added three two-way interaction terms derived by multiplying the initial predictors, and in the third step, we added one three-way interaction term. The analysis of the full model with all two-way interactions and the three-way interaction of participant gender, linguistic form, and goal of the initiative revealed that neither the second ($\Delta R^2 = 0.001$) nor the third iterations ($\Delta R^2 = 0.000$) improved the model. Results of the regression analysis are presented in **Table 5**. The results in the first iteration showed that an initiative presented in gender-fair language was evaluated more positively than an initiative presented in the masculine.

Study 3 was conducted in Austria, a country where, in contrast to Poland, gender-fair language is well established in everyday life. The results demonstrate that in German, gender-fair language has lost its association with feminism because there were practically no differences in favorability due to the linguistic forms used. There was a gender difference in the evaluation of the initiatives, as men rated the initiatives less favorably than women. However, more important was the finding that initiatives received better evaluations when feminine forms were used for the female proponents than masculine forms regardless of participant gender. This indicates that the use of masculine forms in referring to women appears odd when speakers are accustomed to gender-fair language, even if masculine generics were formerly common in the respective country.

TABLE 5 | Study 3. Regression model for the evaluation of the initiative.

	B	SE B
MODEL 1		
Intercept	4.12***	0.22
Linguistic Form (LF)	0.47*	0.20
Participant Gender (PG)	0.34	0.21
Goal of the Initiative (GI)	−0.35*	0.20
Political views	−0.33***	0.10

*$p < 0.05$, ***$p < 0.001$.*
Model 1: Adjusted $R^2 = 0.10$; $F_{(4, 200)} = 6.42$; $p < 0.001$.

[9]For exploratory purposes, we also examined judgments of fairness ("Is this initiative fair?" and "Does it contribute to the public good?"; $\alpha = 0.78$), of the annoyance the initiative might cause ("Could anybody be annoyed by this initiative?" and "Is this initiative irritating?"; $\alpha = 0.78$), and associations of feminism. In addition, we applied a scale measuring attitudes toward affirmative action programs (Bell et al., 2000; $\alpha = 0.84$). Responses to the latter were in the format of a semantic differential and could vary, for instance, from 1 (negative) to 5 (positive). Finally, participants were asked how strongly they supported feminism.

GENERAL DISCUSSION

The present research used an indirect approach to examine socially motivated linguistic change and more specifically, changes in the use of gender-related forms. For this purpose, we compared the effects of gender-fair language in Poland and Austria. Although grammatical gender languages are spoken in both countries, they differ considerably in the use of gender-fair language. While this usage is well established in Austria, it is relatively novel in Poland. Across the first two studies, our results show that in a country where gender-fair language is not common (Poland), social initiatives are evaluated less favorably when gender-fair (i.e., feminine) vs. traditional masculine forms are used. In Study 1, Polish men (but not women) evaluated the initiative for gender quotas on election lists less favorably. Study 2 replicated this effect for both male and female participants. In addition, it showed that the effect depended on the goal of the initiative. Feminine job titles led to less favorable evaluations of the initiative when its goal was gender equality (support of female students), but not when the initiative was aimed at achieving other forms of equality (supporting students of countries affected by war).

Consistent with other studies on German (Vervecken and Hannover, 2012), Study 3 on Austrian German showed that designating women with gender-fair (feminine) forms led to higher support for all types of initiative than when the female proponents were labeled with masculine forms. In line with Vervecken and Hannover (2012), we assumed that the use of gender-fair language in German is currently associated with higher education or competence and has lost its novelty as well as its associations with feminism. Violating a linguistic norm, as well as the gender-fairness norm, may be considered a sign of incompetence (e.g., Giles and Coupland, 1991) and is thus stigmatized. Moreover, the positive effect of gender-fair language, especially of feminine forms referring to a group of women only, on evaluations of the initiative suggests that this usage has become so familiar to speakers of (Austrian) German that failing to use it decreased participants' support for the initiative. Although we do not have direct evidence, participants made several comments in that direction. For instance, some of those who read the text with the masculine forms commented that "the wording was wrong, because masculine forms were used although this was about women!" or "It is very irritating for me that you used masculine forms for women!"

Our studies are the first to investigate different stages in the implementation of gender-fair language by applying the same research design in two different countries. Although we did not directly study the effects of language policies or familiarity with gender-fair language, our results helped elucidate the changes in reactions to gender-fair language and approximate the process that occurs over time as a language changes. When gender-fair language is new, it may face general resistance as it is unfamiliar to speakers and may be perceived as hampering the fluency of everyday speech. Past research suggests that objections to gender-fair language are predominantly due to its novelty (Blaubergs, 1980; Parks and Roberton, 1998). In addition, this usage can be associated with feminism since feminists have fiercely advocated

its use in public discourse (see Blaubergs, 1980; Parks and Roberton, 1998). Proponents of gender-fair language were also judged as overly sensitive and preoccupied with non-essential matters (Parks and Roberton, 1998). This was corroborated by the observation that proponents of gender-fair language have been subject to "hostility and ridicule." Participants in the Parks and Roberton (1998) study, for example, believed that …" the only people who really take offense to any such things are the feminist activists who do nothing but protest all day long" (p. 453). Arguments of this kind were uttered not only by students (Parks and Roberton, 1998) but also in the scientific community (see Maass et al., 2014). Thus, we assume that in the period following the introduction of gender-fair language, its co-occurrence with a gender equality issue may be perceived as strongly indicating a feminist position. Opposition to gender-fair language may be particularly strong in a gender equality context, where both the topic of discussion and language use may suggest a feminist stance. However, when either gender-fair language or a gender-equality issue is presented separately, the association with feminism may be sufficiently unobtrusive to not affect evaluations of the social cause. In an advanced stage, when the use of gender-fair forms has become standard, gender-fair wording is likely to be evaluated as positively as traditional language—or even more positively, once the habit of referring to a woman in the masculine becomes outdated (cf. "policeman Anne Schmidt").

Our findings may offer an explanatory framework for the results of earlier studies, which report both positive and negative speaker perceptions of gender-fair wording (e.g., McConnell and Fazio, 1996; Vervecken and Hannover, 2012). These seemingly contradictory results reflect different stages of adaptation to gender-fair language in the respective societies investigated. In the 1980s, when gender-fair language was new everywhere, the negative effects of gender-fair forms probably occurred regardless of the topic of under discussion as a spill-over effect. However, in our studies, we did not find such negative effects on the non-gender context in Poland, which may suggest that this country is already on its way to adopting, or at least accepting, gender-fair language. Using gender-fair language outside the feminist context may help to make it "normal." Reformed language may then contribute to gender equality. The negative effect of gender-fair language for gender-related initiatives described for Poland can be considered temporary and can be assumed to persist until gender-fair language has become more common and less associated with the feminist context. This development appears to have occurred in Austrian German. Nevertheless, future studies should try to capture the change in attitudes toward gender-fair language more directly. Additionally, future studies should tackle other samples and languages in order to assess the generalizability of the obtained effects.

The most important conclusion to draw from our studies is that language policies aiming at political correctness should not be evaluated rashly. As Bob Dylan said, the times they are a-changin'. Accordingly, negative attitudes toward reformed language may become more positive. What once was new may then become the norm. This conclusion may be helpful for activists and policymakers when advocating changes that at first appear to have detrimental side-effects.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01617

REFERENCES

Arbeitskreis für Gleichbehandlungsfragen der TU Wien. (2010). *Sprache ist nicht neutral [Language isn't neutral]*. Available online at: http://www.tuwien.ac.at/fileadmin/t/akg/Folder_geschlechtergerechte_Sprache_HP.pdf

Bell, M. P., Harrison, D. A., and McLaughlin, M. E. (2000). Forming, changing, and acting on attitude toward affirmative action programs in employment: a theory-driven approach. *J. Appl. Psychol.* 85, 784. doi: 10.1037/0021-9010.85.5.784

Blaubergs, M. S. (1980). An analysis of classic arguments against changing sexist language. *Wom. Stud. Intern. Q.* 3, 135–147. doi: 10.1016/S0148-0685(80)92071-0

Braun, F., Sczesny, S., and Stahlberg, D. (2005). Cognitive effects of masculine generics in German: an overview of empirical findings. *Communications* 30, 1–21. doi: 10.1515/comm.2005.30.1.1

Bundesministerium für Bildung, Wissenschaft und Kultur. (2001). *Macht und Sprache [Power and language]*. Available online at: http://www.eduhi.at/dl/MachtSprache.pdf

Bundesministerium für Frauen und Öffentlichen Dienst. (2009). *Geschlechtergerechte Stellenausschreibungen [Gender-fair job advertisements]. Unabhängiger Bericht der Gleichbehandlungsanwaltschaft iS §3 Abs 5 GBK/GAW-Gesetz*. Wien. Available online at: http://www.bundesregierung.at/DocView.axd?CobId=34384

Bundesministerium für Wissenschaft und Forschung. (2011). *Geschlechtergerechte Sprache [Gender-fair language]*. Available online at: http://www.bmwf.gv.at/uploads/tx_contentbox/Leitfaden_ueber_geschlechtergerechte_Sprache.pdf

Burn, S. M., Aboud, R., and Moyles, C. (2000). The relationship between gender social identity and support for feminism. *Sex Rol.* 42, 1081–1089. doi: 10.1023/A:1007044802798

Buschman, J. K., and Lenart, S. (1996). "I am not a feminist, but…": college women, feminism, and negative experiences. *Polit. Psychol.* 17, 59–75. doi: 10.2307/3791943

Europäisches Parlament. (2009). *Geschlechtergerechter Sprachgebrauch im Europäischen Parlament [Use of gender-fair language in the European Parliament]*. Available online at: http://www.europarl.europa.eu/RegData/publications/2009/0001/P6_PUB(2009)0001_DE.pdf

Eynon, R., Fry, J., and Schroeder, R. (2008). "The ethics of Internet research," in *The SAGE Handbook of Online Research Methods*, eds G. Blank, N. G. Fielding, and R. M. Lee (London: Sage), 23–41.

Formanowicz, M., Bedynska, S., Cisłak, A., Braun, F., and Sczesny, S. (2013). Side effects of gender-fair language: how feminine job titles influence the evaluation of female applicants. *Eur. J. Soc. Psychol.* 43, 62–71. doi: 10.1002/ejsp.1924

Gendup. (2012). *Leitfaden für Einen Gerechten Sprachgebrauch [Guideline for Gender-fair Language]*. Available online at: http://www.uni-salzburg.at/fileadmin/oracle_file_imports/2103374.PDF

Giles, H., and Coupland, N. (1991). *Language: Contexts and Consequences*. Keynes: Open University Press.

Gleichbehandlungsgesetz. (2004). Available online at: http://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2004_I_66/BGBLA_2004_I_66.pdf

Hayes, A. F. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling*. Available online at: http://www.afhayes.com/public/process2012.pdf

Hodel, L., Formanowicz, M., Sczesny, S., Valdrova, J., and von Stockhausen, L. (2013). "Gender-fair language use in online job advertisements," in *16th Congress of the European Association of Work and Organizational Psychology*, Münster.

Jacobson, M. B., and Insko, W. R. (1985). Use of nonsexist pronouns as a function of one's feminist orientation. *Sex Rol.* 13, 1–7. doi: 10.1007/BF00287456

Jost, J. T., Glaser, J., Kruglanski, A. W., and Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychol. Bull.* 129, 339–375. doi: 10.1037/0033-2909.129.3.339

Jost, J. T., and Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. *J. Pers. Soc. Psychol.* 88, 498–509. doi: 10.1037/0022-3514.88.3.498

Koniuszaniec, G., and Błaszkowska, H. (2003). "Language and gender in Polish," in *Gender Across Languages. The Linguistic Representation of Men and Women*, Vol. 3, eds M. Hellinger and H. Bußmann (Amsterdam; Philadelphia, PA: John Benjamins Publishing Company), 260–285.

Maass, A., Suitner, C., and Merkel, E. (2014). "Does political correctness make (social) sense?" in *Social Cognition and Communicationm*, eds J. Lazlo, J. Forgas, and O. Vincze (Sydney: Symposium of Social Psychology), 331–346.

Matheson, K., and Kristiansen, C. M. (1987). The effect of sexist attitudes and social structure on the use of sex-biased pronouns. *J. Soc. Psychol.* 127, 395–398. doi: 10.1080/00224545.1987.9713721

McConnell, A. R., and Fazio, R. H. (1996). Women as men and people: effects of gender-marked language. *Pers. Soc. Psychol. Bull.* 22, 1004–1013. doi: 10.1177/01461672962210003

Merkel, E., Maass, A., and Frommelt, L. (2012). Shielding women against status loss: the masculine form and its alternatives in Italian language. *J. Lang. Soc. Psychol.* 31, 311–320. doi: 10.1177/0261927X12446599

Mucchi-Faina, A. (2005). Visible or influential? Language reforms and gender (in)equality. *Soc. Sci. Inform.* 44, 189–215. doi: 10.1177/0539018405050466

Parks, J. B., and Roberton, M. A. (1998). Contemporary arguments against nonsexist language: blaubergs (1980) revisited. *Sex Rol.* 39, 445–461. doi: 10.1023/A:1018827227128

Parks, J. B., and Roberton, M. A. (2002). The gender gap in student attitudes toward sexist/nonsexist language: implications for sport management education. *J. Sport Managem.* 16, 190–208.

Parks, J. B., and Roberton, M. A. (2004). Attitudes toward women mediate the gender effect on attitudes toward sexist language. *Psychol. Wom.Q.* 28, 233–239. doi: 10.1111/j.1471-6402.2004.00140.x

Pauwels, A. (2003). "Linguistic sexism and feminist linguistic activism," in *The Handbook of Language and Gender*, eds M. Meyerhoff and J. Holmes (Oxford: Blackwell), 550–570.

Stout, J. G., and Dasgupta, N. (2011). When *he* doesn't mean *you*: gender-exclusive language as ostracism. *Pers. Soc. Psychol. Bull.* 37, 757–769. doi: 10.1177/0146167211406434

Tougas, F., Brown, R., Beaton, A. M., and Joly, S. (1995). Neosexism: plus ça change, plus c'est pareil. *Pers. Soc. Psychol. Bull.* 21, 842–849. doi: 10.1177/0146167295218007

Trömel-Plötz, S. (1978). Linguistik und Frauensprache [Linguistics and women's language]. *Linguist. Berich.* 57, 49–69.

Ullrich, J., and Cohrs, J. C. (2007). Terrorism salience increases system justification: experimental evidence. *Soc. Just. Res.* 20, 117–139. doi: 10.1007/s11211-007-0035-y

UNESCO. (1999). *Guidelines for Gender-neutral Language*. Paris: UNESCO. Available online at: http://unesdoc.unesco.org/images/0011/001149/114950mo.pdf

Universitätsgesetz. (2002). *Eintragung akademischer Grade in Urkunden, Eintragungsrichtlinien 2012 [Recording Academic Degrees in Documents, Recording Guidelines]*. Available online at: http://www.bmwf.gv.at/fileadmin/user_upload/wissenschaft/naric/eintragungsrichtlinien.pdf

Vervecken, D., and Hannover, B. (2012). Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts perception of the speaker. *Eur. J. Soc. Psychol.* 42, 754–762. doi: 10.1002/ejsp.1893

Williams, R., and Wittig, M. A. (1997). "I'm not a feminist, but…": factors contributing to the discrepancy between pro-feminist orientation and feminist social identity. *Sex Rol.* 37, 885–904. doi: 10.1007/BF02936345

The Responders' Gender Stereotypes Modulate the Strategic Decision-Making of Proposers Playing the Ultimatum Game

Eve F. Fabre[1,2]*, Mickael Causse[2], Francesca Pesciarelli[1] and Cristina Cacciari[1]

[1] Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy,
[2] Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse, France

Despite the wealth of studies investigating factors affecting decisions, not much is known about the impact of stereotypical beliefs on strategic economic decision-making. In the present study, we used the ultimatum game paradigm to investigate how participants playing as proposer modulate their strategic economic behavior, according to their game counterparts' stereotypical identity (i.e., responders). The latter were introduced to the participants using occupational role nouns stereotypically marked with gender paired with feminine or masculine proper names (e.g., linguist-Anna; economist-David; economist-Cristina; linguist-Leonardo). When playing with male-stereotyped responders, proposers quickly applied the equity rule, behaving fairly, while they adopted a strategic behavior with responders characterized by female stereotypes. They were also longer to make their offers to female than to male responders but both kinds of responders received comparable offers, suggesting a greater cognitive effort to treat females as equally as males. The present study explicitly demonstrates that gender stereotypical information affect strategic economic decision-making and highlights a possible evolution of gender discrimination into a more insidious discrimination toward individuals with female characteristics.

Keywords: ultimatum game, gender stereotypes, proposer, strategic decision-making

INTRODUCTION

The Proposer Playing the Ultimatum Game: Between Social Preference and Strategy

One topic of major interest in economic decision-making studies is the strategic behavior adopted by individuals faced with economic decisions (Camerer, 2003). The interest in how social and emotional information affects economic decision-making has steadily grown over past decades (for overviews see Frith and Singer, 2008; Rilling and Sanfey, 2011). The ultimatum game has provided, for now many years, a fruitful paradigm for assessing the social aspects of economic decision-making (Güth et al., 1982). In the standard version of this two-player game, a proposer offers to split a fixed amount of money (e.g., 10€) with a responder. Both receive their shares only if the responder accepts the offer. Game theory predicts that in order to maximize their outcome, proposers should behave in a rational and self-interested way, offering the smallest share possible to the responder (e.g., 1€ out of 10€; Neumann and Morgenstern, 1947). But psychological research

on judgment and decision-making has produced a wealth of evidence that, in practice, this theory does not provide a satisfactory description of human behavior (e.g., Güth et al., 1982; Camerer, 2003). Indeed, on average, proposers offer about 40% of the total amount of money (e.g., 4€ out of 10€) to the responders (for a meta-analysis see Oosterbeek et al., 2004). Such a behavior has been under scrutiny for now decades, and various theories have been proposed to explain the proposers' fair behavior.

Social norms are defined as "the customary rules that govern behavior in groups and societies." Honesty, loyalty, reciprocity, or promise keeping, to name a few, do guarantee smooth interactions between individuals of a given social group or society (Bicchieri, 2006). This common value system plays a crucial role in individual choices because, by shaping individual needs and preferences, these norms serve as criteria for selecting among alternatives. People choose what they prefer, and what they prefer conforms most of the time to social expectations (Bicchieri, 2006). According to *social preference* models (Fehr and Schmidt, 1999; Falk and Fischbacher, 2006), proposers, who embrace social norms, make almost fair offers because they have altruistic concerns toward the responders and care about the distribution of payoffs among players. On the other hand, according to other authors, the apparent fairness preference of the proposers may in turn reflect strategic concerns (for an overview see Nelissen et al., 2011). Indeed, the proposers are aware that responders are likely to have social expectations and thus may reject unfair offers (Güth et al., 1982; Camerer, 1999, 2003; Sanfey, 2009). Consequently, because they want to maximize their gain, the proposers have to accurately determine the smallest amount of money the responders may accept (i.e., the minimum acceptable; Blount, 1995). Then, proposers may feign altruism, offering almost fair shares to the responders (Kagel et al., 1996; Nelissen et al., 2011), aiming at lowering the occurrence of emotionally painful rejections (Güroğlu et al., 2009). The results of various studies attest that it is more likely that proposers' fair behavior reflects both altruistic and strategic concerns (Blount, 1995), with a certain inter-variability among individuals in the altruistic/strategic balance (Morishima et al., 2013).

While rarely reported in the ultimatum game studies, the time needed by the participants to make a decision (i.e., response times) may be a good indicator of both the complexity of the decision and the cognitive processes involved in the decision-making. Recent studies (e.g., Polezzi et al., 2008; Fabre et al., 2015) found that responders playing the ultimatum game were faster answering easily classifiable offers, both fair (i.e., 5€ out of 10€) and unfair (i.e., 1€ out of 10€), than hardly classifiable mid-value offers (i.e., 3€ of 10€; Sanfey et al., 2003). The decision on both unfair and fair offers appears to rely on fast heuristic-based judgments (i.e., refusing unfair offers and accepting fair offers), while deciding on mid-value offers may be more complex (Sanfey et al., 2003) and therefore requires a more time consuming and cognitively costly deliberative reasoning (Civai, 2013; Fabre et al., 2015). It is plausible that the proposers' decision-making works in a similar way. Indeed, when proposers have

social preferences, their decision-making may be relatively fast, since they would follow social norms and apply the equity rule (i.e., offering a fair share). In contrast, when proposers engage in a strategic decision-making, they have to accurately evaluate all the information available to maximize their gain, which may be cognitively costly and time consuming. To this extent, we assume that studying the proposers' response times may provide critical information concerning the decision process and the balance between altruistic and strategic concerns.

The Impact of Social Information on the Proposers' Strategic Concerns

Several studies investigated how proposers modulate their behavior depending on the responders' social characteristics (e.g., Solnick and Schweitzer, 1999). These studies allowed us to evaluate the proposers' internal representations of the world. In other words, how much is a specific responder "worth" to a proposer? Eckel and Ball (1996) investigated the effect of social status information on the proposers' behavior. In their study, participants were attributed a star or not depending on their performance to a trivia quiz before playing the ultimatum game as either proposer or responder. The priming task (i.e., trivia quiz) enabled to allocate artificially a high (i.e., star) or a low status (i.e., no star) to the participants. Both high and low status proposers offered higher shares to high status responders than to low social status responders, confirming the impact of social status on economic decision-making (for an overview see Heffetz and Frank, 2008). According to *status characteristic* theory (Wagner and Berger, 1993), a status characteristic (i.e., gender, age, race, physical attractiveness, intelligence or occupation) affects people's expectations of reward, and high status individuals expect to receive higher reward than low status individuals. To this extent, proposers may adapt their behavior, offering higher shares to high status responders compared with the low status responders in order to limit the risk of suffering a rejection (Güroğlu et al., 2009). Proposers were also found to be influenced by the responders' attractiveness offering higher shares to more attractive responders (Solnick and Schweitzer, 1999; Zaatari et al., 2009). Finally, some studies investigated the impact of the responders' gender on the proposers' decision. Overall, these studies demonstrated that proposers offer more to male responders than to female responders (e.g., Solnick and Schweitzer, 1999; Eckel and Grossman, 2001; Solnick, 2001). Saad and Gill (2001) found that male proposers were on average more generous with female than male responders, while the responders' gender did not affect the behavior of female proposers, who offered equally fair shares to responders of both sexes. A greater variability in the behavior of male proposers was also found with altruistic male proposers or aggressive male proposers compared to female proposers, who showed less variability in their behavior (Castillo and Cross, 2008).

Because both gender and attractiveness are considered as status characteristics, the observation of the increased shares proposed to attractive responders (Solnick and Schweitzer, 1999; Zaatari et al., 2009) and to male responders in most gender

studies (e.g., Solnick and Schweitzer, 1999; Eckel and Grossman, 2001; Solnick, 2001) may be interpreted in terms of social status differences.

Why Investigate the Impact of Gender Stereotypes on Economic Decision-Making?

Conflict theories postulate that because men have greater social status and power (Reskin, 1988), they allocate occupations that open access to resources (e.g., money, stocks, contacts, information) predominantly to men, thus favoring themselves over women, which creates an occupational segregation (Pratto et al., 1997). For this reason, stereotypically male occupations (e.g., engineer, electrician) are associated with higher social status and power compared with those stereotypically female (e.g., teacher, beautician; Eagly, 1987; Ridgeway, 2001). Nowadays, more and more women can enter the professions with a highly marked male stereotype (e.g., lawyer, banker, doctor), and more men access to occupations with a highly marked female stereotype (e.g., nurse, "mid-wife," teacher; Eagly, 1987; Eagly and Karau, 2002; Phelan et al., 2008). Given the frequent interactions of men and women, it is critical to understand how gender stereotypical beliefs – i.e., a form of social knowledge linked to actions, attitudes, rules and other forms of knowledge attributed to individuals based on their biological gender (Greenwald and Farnham, 2000; Wheeler and Petty, 2001; Quadflieg and Macrae, 2011) – modulate economic decision-making.

A recent study investigated how proposers' gender stereotyped descriptions (i.e., occupations marked with either a male or a female stereotype) influenced the responders' decision-making (Fabre et al., 2015). When playing with female-stereotyped proposers (e.g., linguist), responders were longer to make their decision, reflecting a more deliberative reasoning (Sanfey and Chang, 2008) associated with an increase in acceptation rates. In contrast, participants were found to answer more quickly and to reject more frequently male-stereotyped proposers' offers (e.g., economist) than those of female-stereotyped proposers. That study demonstrated that gender stereotype information of the proposer modulates the economic decision-making in the ultimatum game and the cognitive processes underpinning the decision-making. Therefore, we may reasonably expect gender stereotypical information of responders to modulate the behavior of proposers playing the ultimatum game. To our knowledge, this impact of the responders' stereotypical identity on the proposers' decision-making has never been investigated.

The Present Study

In the present behavioral study, we adapted the study of Fabre et al. (2015) and focused our analyses on the effect of the gender stereotypical beliefs on the proposers' economic strategic behavior. Participants played a repeated one-shot ultimatum game as proposers against 120 simulated different responders. The latter were introduced to the participants by occupational nouns stereotypically marked with gender paired with either feminine or masculine proper names (e.g., *linguist-Anna; economist-David; economist-Cristina; linguist-Leonardo*;

Fabre et al., 2015). We assumed that reading occupational role nouns stereotypically marked with gender leads to automatic and hard-to-suppress activation of gender stereotypical beliefs (Banaji and Hardin, 1996; Irmen and Roßberg, 2004; Oakhill et al., 2005).

We hypothesized that (1) participants would assign a higher minimum acceptable to both male responders and male-stereotyped responders (i.e., described with an occupation stereotypically marked with male gender) – who may be associated with a higher social status – than to respectively female responders and female-stereotyped responders (i.e., described with an occupation stereotypically marked with female gender), who may in turn be associated with a lower social status (Eckel and Ball, 1996; Rudman and Kilianski, 2000; Ridgeway, 2001). We also hypothesized that (2) proposers would be faster in making their offers to both male and male-stereotyped responders following social norms (i.e., equity rule), while they would take more time to decide when interacting with respectively female and female-stereotyped responders following strategic concerns. According to the *Status Incongruity Hypothesis* (Rudman et al., 2012), socially *atypical* male and female individuals (i.e., not conformant to gender rules; Eagly and Karau, 2002), are judged more negatively than socially *typical* ones, all other things being equal, and may sometimes undergo penalties (i.e., *backlash effect*; Rudman and Glick, 1999; Eagly and Karau, 2002; Rudman and Fairchild, 2004; Phelan et al., 2008; Rudman et al., 2012). Hence, we finally hypothesized that (3) proposers would make higher offers to responders who conform to gender rules (e.g., *linguist-Cristina*, economist-*Leonardo*), than to responders who violate gender rules (e.g., *linguist-David*, economist-*Anna).*

MATERIALS AND METHODS

Participants

Thirty-four students of Modena University (17 females; age range 19–26 years $M = 21.5$, $SD = 2.26$) were recruited to play a repeated one-shot ultimatum game as proposer. They participated for 5% of the total amount of money they won and were proposed at the end of the experiment to swap this money for course credits. All were Italian native speakers with normal or corrected-to-normal vision. None of them reported a history of prior neurological disorder. Participants were informed of their rights and gave written informed consent for participation in the study. This study was carried out fulfilling ethical requirements in accordance with the standard procedures of the University of Modena and Reggio Emilia.

Materials

The same two groups of 30 occupational role nouns each (one male, one female) with comparable stereotypicality, wealth and valence, lexical frequency and length used in the study of Fabre et al. (2015) were used in the present study (see Supplementary Material). In order to select the experimental materials, a written questionnaire listing 258 occupational role nouns, ending in –e,–ista or a consonant to avoid cues to the gender of the referent in the word form, was presented to 112 students not

further involved in the experiment (56 females; age range 19–27 years; M_{age} = 23.6, SD = 2.92). Eighty of these students rated to what extent each role noun was stereotypically associated with male or female individuals on a 7-point Likert scale (i.e., stereotype strength; from 1 = only men to 7 = only women), 16 of them to what extent each role noun was associated with a positive or negative value (i.e., valence: from 1 = very negative to 7 = very positive) and 16 the wealth of a person described with each role noun (i.e., wealth: from 1 = very rich to 7 = very poor). The labels of the scale poles were reversed for half of the participants. The final rating assigned to each word was calculated by combining the ratings obtained with both directions of each rating scale. The 60 role nouns selected as experimental materials received comparably high ratings of stereotypicality (the experimental material and the associated ratings are available in Supplementary Material). In order to compare the stereotype strength of the two role noun groups, the ratings of the role nouns ranging from 4 to 7 (i.e., feminine stereotypes) were translated and ranged from 1 to 4 (i.e., X' = 8-X, with X: initial rating and X': translated rating). Stereotype strength (Female Stereotypes: M = 2.81, SD = 1.21; Male Stereotypes: M = 2.77, SD = 1.19), valence (Female Stereotypes: M = 4.42, SD = 0.69; Male Stereotypes: M = 4.36, SD = 0.71), wealth (Female Stereotypes: M = 3.81, SD = 0.94; Male Stereotypes M = 4.10, SD = 0.88), lexical frequency (Female Stereotypes: M = 5.66, SD = 0.87; Male Stereotypes M = 6.09, SD = 1.05) and length (i.e., number of characters; Female Stereotypes: M = 9.77, SD = 2.2; Male Stereotypes M = 8.83, SD = 1.82) of male and female occupational role nouns were comparable (ps > 0.05). The mean stereotypicality rating of feminine role nouns reported in the Supplementary Material Table S1 is the translated rating (i.e., X').

Experimental materials also included 120 Italian familiar proper names (60 feminine) without any unisex names. With the final two groups of 30 occupational role nouns along with the 120 proper names, we created four experimental conditions, two stereotype-matching conditions: female stereotypical occupational role nouns followed by feminine proper names (e.g., *linguista-Anna*) and male stereotypical occupational role nouns followed by masculine proper names (e.g., *economista-Davide*); and two stereotype-mismatching conditions: female stereotypical occupational role nouns followed by masculine proper names (e.g., *linguista-Leonardo*) and male stereotypical role nouns followed by feminine proper names (e.g., *economista-Cristina*). Participants interacted once with 30 different responders of each kind (i.e., 120 different responders in total).

Procedure

Participants were seated comfortably in a darkened sound-attenuated room. They played a one-shot ultimatum game as proposers. An introduction explaining the rules of the ultimatum game was given to each of them. Stimuli were presented in light white upper case letters (Courier font, size 13) against a black background on a high-resolution computer that was positioned at eye level about 70 cm in front of each participant. A fixation cross appeared in the middle of a computer screen and remained until participants pressed a button to start a trial. Each occupational

role noun was displayed for 700 ms followed by a blank screen for 300 ms. Then a proper name appeared and remained on the screen until the participants pressed the key on the keyboard corresponding to the numerical value of the offer they wanted to make (i.e., from 1€ to 9€ out of 10€). Each response was followed by a 1000 ms blank screen. No feedback on the responder's answer was provided to the participants in order to avoid a modulation of their behavior along the experiment. Participants were asked to respond as fast as possible.

Before conducting the game, participants were told that they were playing against 120 real different responders of whom they would know their occupations and proper names. Each participant was presented with 30 trials in each of the four experimental conditions for a total of 120 trials. As a matter of fact, the responder was simulated by the computer. However, in order to make the participants believe they were playing against real responders, they were told that responders had been contacted prior to the game and that they had indicated the offers they were willing to accept if proposed by a student (i.e., a shifted in time ultimatum game). Moreover, participants were indicated that we were thanking the different partners involved in this experiment (e.g., the firefighters of Modena, the Oenology School of UNIMORE University, etc...). Participants were informed that at the end of the game, responders would receive the sum corresponding to a percentage of the accepted offers.

Data Analysis
Mean Offers
Mean offers were submitted to a 2 × 2 × 2 (Responders' Occupation [male-stereotyped, female-stereotyped] × Responders' Gender [male, female] × Participants' Gender [male, female]) analysis of variance (ANOVA). Participants' Gender was a between-subject factor, and the two remaining factors were within-subject factors.

Response Times
Log transformed mean response times were submitted to a 2 × 2 × 2 (Responders' Occupation [male-stereotyped, female-stereotyped] × Responders' Gender [male, female] × Participants' Gender [male, female]) ANOVA. Participants' Gender was a between-subject factor, and the two remaining factors were within-subject factors.

Questionnaires
The *Interpersonal Reactivity Index* (IRI; Albiero et al., 2006) was designed to measure empathy and is composed of four subscales: (1) the Perspective Taking scale (pt) measuring the tendency to spontaneously adopt the psychological point of view of others; (2) the Empathic Concern scale (ec) assessing "other-oriented" feelings of sympathy and concern for unfortunate others; (3) the Personal Distress scale (pd) measuring "self-oriented" feelings of personal anxiety and unease in tense interpersonal setting; and (4) the Fantasy scale (f) that taps respondents' tendencies to transpose themselves imaginatively into the feelings and actions of fictitious characters in books, movies, and plays. The *Bem Sex Role Inventory* (BSRI; De Leo and Villa, 1986) assesses the participants' degree of masculinity/femininity and to what

extent they embrace traditional sex roles. We examined whether the scores at the IRI and the BSRI questionnaires predicted the differences in mean offer and response time observed for the different responders. To that aim, we conducted Pearson correlation analyses between the scores of the IRI, the BSRI, and the resultants of the differences in mean offer and response times participants for (1) responders' occupations marked with male vs. female stereotypes (i.e., Stereotype [M – F]); and (2) male vs. female responders (i.e., Gender [M – F]).

Post hoc Rating Study of Social Status

In order to further our argumentation, a final rating study was realized aiming at measuring the social status associated with the occupational stereotypes used in the present study. Fifty one participants (24 females; $Mage = 31.40$, $SD = 6.11$) were asked to rate the social status associated with the individuals practicing each of the 60 occupations used in our experiment (i.e., from 1 = very low social status to 7 = very high social status). This study was realized online via Google Forms. In order to evaluate the impact of the social status associated with the occupational stereotypes on the proposers' decision-making process, we run two one-tailed partial correlations: one between the occupations stereotypicality (from 1 = very masculine to 7 = very feminine) and the mean offer and one between the occupations stereotipicality and the response times, each time controlling for social status. These results were compared to the results same correlations analysis not controlled for social status.

RESULTS

Mean Offer

On average, participants proposed 3.72€ ($SD = 0.17$) to the responders (see **Table 1**). The ANOVA on mean offers showed a main effect of stereotype [$F(1,32) = 52.53$, $p < 0.001$, $\eta_p^2 = 0.62$, see **Figure 1**]. Participants proposed higher offers to responders presented with male than with female stereotypical occupational role nouns ($M = 3.94€$, $SD = 0.11$; $M = 3.49€$, $SD = 0.13$; respectively). Participants' gender [$F(1,32) = 0.30$, $p = 0.59$, $\eta_p^2 = 0.01$], responders' gender [$F(1,32) = 0.05$, $p = 0.82$, $\eta_p^2 = 0.00$] main effects and Participants' Gender × Responders' Gender [$F(1,32) = 1.60$, $p = 0.22$, $\eta_p^2 = 0.05$], Stereotype × Participants' Gender [$F(1,32) = 0.04$, $p = 0.84$, $\eta_p^2 = 0.00$], Stereotype × Responders' Gender [$F(1,32) = 0.06$, $p = 0.82$, $\eta_p^2 = 0.00$], Stereotype × Responders' Gender × Participants' Gender [$F(1,32) = 1.76$, $p = 0.19$, $\eta_p^2 = 0.05$] interactions were not significant.

Response Times

The ANOVA on log-transformed response times showed a significant stereotype main effect [$F(1,32) = 43.87$, $p < 0.001$, $\eta_p^2 = 0.58$, see **Figure 2A**] with participants making their offers faster to male-stereotyped responders than to female-stereotyped responders (respectively, $M = 847$ ms, $SD = 477$; $M = 1142$ ms, $SD = 706$, see **Table 1**). Participants were also faster when making their offers to male responders
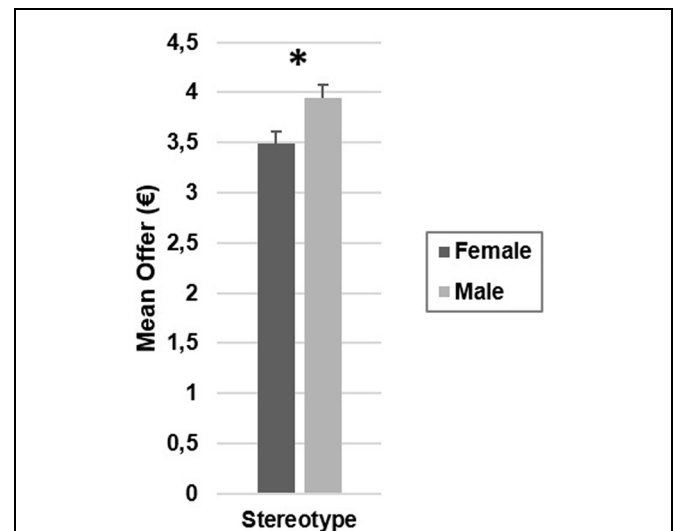

FIGURE 1 | Mean offers as a function of responders' gender stereotypes. Error bars represent standard errors. *$p < 0.001$.
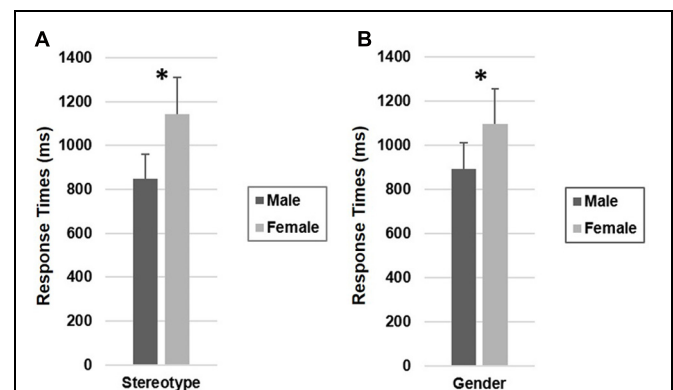

FIGURE 2 | (A,B) Response times as a function of **(A)** responders' gender stereotypes and of **(B)** responders' genders. Error bars represent standard errors. *$p < 0.001$.

than to female responders (respectively, $M = 893$ ms, $SD = 520$; $M = 1096$ ms, $SD = 693$), as shown by a significant responder's gender main effect [$F(1,32) = 54.95$, $p < 0.001$, $\eta_p^2 = 0.63$, see **Figure 2B**]. The participants' gender [$F(1,32) = 0.00$, $p = 0.96$, $\eta_p^2 = 0.00$] main effect and Participants' Gender × Responders' Gender [$F(1,32) = 0.04$, $p = 0.85$, $\eta_p^2 = 0.00$], Stereotype × Participants' Gender [$F(1,32) = 0.02$, $p = 0.86$, $\eta_p^2 = 0.00$], Stereotype × Responders' Gender [$F(1,32) = 2.66$, $p = 0.11$, $\eta_p^2 = 0.08$], Stereotype × Responders' Gender × Participants' Gender [$F(1,32) = 0.25$, $p = 0.62$, $\eta_p^2 = 0.01$] interactions were not significant.

Questionnaires

The correlations between the scores obtained by each participant in the BSRI and in the IRI and the resultants of the various differences in mean offer and in response time revealed only one

significant result (see Supplementary Material). The difference in response times between responders characterized by a male vs. female stereotype [Stereotype (M – F)] was positively correlated to the score of the perceptive taking scale: the higher the score, the greater the Stereotype (M – F) difference [$r = 0.393$, $p < 0.05$, see **Table 2**].

Post hoc Rating Study of Social Status

A dependent *t*-test analysis was conducted on the social status ratings revealing that the thirty occupations stereotypically male were on average associated with a higher social status ($M = 4.19$, $SD = 0.60$) than are the 30 occupations stereotypically female [$M = 3.60$, $SD = 0.61$, $t(50) = 10.41$, $p < 0.001$]. The occupations stereotypicality and the mean offer were significantly correlated ($r = -0.396$, $p < 0.001$), however, the significance dropped when controlling for social status ($r = -0.225$, $p < 0.05$). The occupational stereotypicality and the response times were also found to be significantly correlated ($r = 0.507$, $p < 0.001$) and lightly less when controlling for social status ($r = 0.492$, $p < 0.001$, see Supplementary Material).

DISCUSSION

In the present study we investigated whether both the responders' stereotypical identity and gender modulated the behavior of proposers playing a repeated one-shot ultimatum game. We predicted to observe a modulation of both mean offers and response times depending on the social description of the responders (i.e., gender and occupational stereotype marked with gender).

On average, participants proposed 37.2% of the total amount of money to the responders, which is slightly less than the average offer, i.e., 40% of the share, reported in the meta-analysis of Oosterbeek et al. (2004). This difference is explained by the fact that while participants offered male-stereotyped responders a share similar to the one usually proposed in the ultimatum game (i.e., about 4€ out of 10€; Oosterbeek et al., 2004), they offered female-stereotyped responders on average 45 cents less. The results also revealed that the mean offer was correlated with the degree of stereotypicality of the responders' occupations. Indeed, the more masculine the

TABLE 1 | Offers and response times means and standard deviations for each experimental condition.

	Female stereotype				Male stereotype			
	Female gender		Male gender		Female gender		Male gender	
	M	SD	M	SD	M	SD	M	SD
Mean offer (€)								
Female Participants	3.40	0.72	3.47	0.69	3.87	0.86	3.87	0.86
Male participants	3.58	0.59	3.53	0.55	4.01	0.66	4.01	0.59
Response times (ms)								
Female participants	1359	977	1081	770	947	657	792	464
Male participants	1193	494	936	273	884	339	764	343

Italic values are de standard deviations associated with the mean offers and the mean response times.

TABLE 2 | Correlations between the questionnaires' scores and the differences in mean offers and response times.

	Questionnaires					Mean offer		Response Times	
	Interpersonal Reactivity Index				Bem sex	Stereotype (M – F)	Gender (M – F)	Stereotype (M – F)	Gender (M – F)
	PT	EC	DP	F	Role Inventory				
Questionnaires									
IRI (pt)	_____								
IRI (ec)	0.318	_____							
IRI (dp)	**−0.456****	0.103	_____						
IRI (f)	0.175	**0.517****	105	_____					
BSRI	−0.140	−0.226	−0.314	−0.429*	_____				
Mean offer									
Stereotype (M – F)	0.090	−0.092	−0.090	0.118	0.056	_____			
Gender (M – F)	−0.033	0.061	0.192	−0.139	−0.135	−0.184	_____		
Response times									
Stereotype (M – F)	0.393*	0.198	−0.090	0.000	−0.042	−0.174	−0.020	_____	
Gender (M – F)	−0.205	0.033	0.227	−0.096	0.058	−0.194	−0.008	**0.340***	_____

Bold values are statistically significant.

occupation was, the higher was the offer; and the more feminine the occupation was, the lower was the offer. These results support the idea that female-stereotyped responders are associated with a lower minimum acceptable offer than are male-stereotyped responders. This difference in behavior appears to be partly due to the fact that male-stereotyped occupations are on average associated with a higher social status compared to male-stereotyped occupations according to both our *post hoc* rating study and literature (Eagly, 1987; Ridgeway, 2001). Participants were also longer to make their offers to female-stereotyped responders than to male-stereotyped responders. Again, the correlation analysis supported the results of the analysis of variance and showed that response times were correlated with the stereotypicality of the responders' occupations. The more masculine the occupation was, the shorter were the response times; and the more feminine the occupation was, the longer were the response times. Taken together, these results suggest that participants behaved more or less strategically depending on the responders' stereotypical identity. When facing male-stereotyped responders (e.g., economist), proposers followed social norms, applying the equity rule, offering quickly fair shares. In contrast, when facing female-stereotyped responders, proposers adopted a more strategic and cognitively costly deliberative reasoning, trying to accurately determine their minimum acceptable offer in order to maximize their gain. These behaviors were more extreme when the gender stereotype strength was higher: the more feminine the occupational stereotype was, the more strategic was the decision; and the more masculine the occupational stereotype was, the more altruistic was the decision.

Participants were also found to take more time to make an offer to female responders than to male responders, to this extent it may have been more complex for participants to interact with female responders than with male responders. However, in contradiction with our predictions, participants made comparable offers to both female and male responders. We assume that offering female responders shares equivalent to those offered to male responders, may have had a cognitive cost for participants. The increase in response times may reflect a strategic behavior inhibition when interacting with female responders.

Finally, our predictions concerning the observation of a backlash effect directed at the responders who violate gender rules (e.g., economist-Anna; linguist-David) were not fulfilled, since the results revealed no economical penalization toward these specific responders. The studies reporting backlash effects used diverse experimental protocols testing either the fit of hiring (e.g., Rudman and Glick, 2001; Heilman et al., 2004; Heilman and Okimoto, 2007; Moss-Racusin et al., 2010; Rudman et al., 2012), the selection of a partner game (e.g., Rudman, 1998), the salary recommendation (e.g., Heilman et al., 2004) or the opportunity to sabotage a line manager (e.g., Rudman et al., 2012), to name a few (for a review on backlash effect see Rudman and Phelan, 2008). In these studies, participants were given the possibility to commit backlash but were not taking any risk in doing so. Indeed, the backlashed individuals were not able to punish the participants in return for their behavior.

We assume that no backlash effect was observed in the present experiment because participants may have feared to be punished by the responders, who may have rejected their offer for being backlashed in the first place. A second possibility might be that participants were simply not willing to backlash mismatching responders. The present study does not enable to status on the absence of backlash effect. We plan to address this question in further studies.

CONCLUSION

The present study continues the long list of works investigating the impact of social information on economic decision-making. As far as we know, this study is the first to demonstrate that both men and women modulate their strategic behavior according the gender-marked stereotype of their counterparts during economic interactions. Proposers were found to apply quickly the equity rule when interacting with male-stereotyped responders, while they behaved more strategically at a greater cognitive cost with female-stereotyped responders proposing them lower shares. Proposers were longer to make their offers to female than to male responders but both kinds of responders received comparable offers, suggesting a greater cognitive effort to treat females as equally as males. Taken together, these results suggest that in real life, individuals practicing a profession stereotypically female may suffer discrimination during economic interactions, while female individuals may not, at least when these individuals are given the possibility to punish in return their counterpart. The present experiment highlights an evolution of society in that gender discrimination, which is nowadays strongly decried, may be converting into a more insidious discrimination toward individuals with female characteristics. More work is now needed to confirm this tendency.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00012

REFERENCES

Albiero, P., Ingoglia, S., and Lo Coco, A. (2006). A contribution to the Italian validation of the Interpersonal Reactivity Index. *Test. Psicometria Metodol.* 13, 107–125. doi: 10.1016/j.adolescence.2008.01.001

Banaji, M., and Hardin, C. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York, NY: Cambridge University Press.

Blount, S. (1995). When social outcomes aren't fair: the effect of causal attributions on preferences. *Organ. Behav. Hum. Decis. Process.* 63, 131–144. doi: 10.1006/obhd.1995.1068

Camerer, C. (1999). Behavioral economics: reunifying psychology and economics. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10575–10577. doi: 10.1073/pnas.96.19. 10575

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press. doi: 10.1257/jep.9.2.209

Castillo, M., and Cross, P. (2008). Of mice and men: within gender variation in strategic behavior. *Games Econ. Behav.* 64, 421–432. doi: 10.1016/j.geb.2008.01.009

Civai, C. (2013). Rejecting unfairness: emotion-driven reaction or cognitive heuristic? *Front. Hum. Neurosci.* 7:126. doi: 10.3389/fnhum.2013.00126

De Leo, D., and Villa, A. (1986). *Il Problema del Rilevamento delle Tipologie Sessuali e il Bem Sex Role Inventory*. Firenze: Organizzazioni Speciali.

Eagly, A. H. (1987). *Sex Differences in Social Behavior: A Social-Role Interpretation*. HillSDale, NJ: Erlbaum.

Eagly, A. H., and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. doi: 10.1037/0033-295X.109.3.573

Eckel, C. C., and Ball, S. B. (1996). Buying status: experimental evidence on status in negotiation. *Psychol. Market.* 13, 381–405. doi: 10.1002/(SICI)1520-6793(199607)13:4<379::AID-MAR4>3.0.CO;2-7

Eckel, C. C., and Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Econ. Inq.* 39, 171–188. doi: 10.1111/j.1465-7295.2001.tb00059.x

Fabre, E. F., Causse, M., Pesciarelli, F., and Cacciari, C. (2015). Sex and the money: how gender Stereotypes modulate economic decision-making. An ERP study. *Neuropsychologia* 75, 221–232. doi: 10.1016/j.neuropsychologia.2015.06.013

Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games Econ. Behav.* 54, 293–315. doi: 10.1016/j.geb.2005.03.001

Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003355399556151

Frith, C. D., and Singer, T. (2008). The role of social cognition in decision making. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3875–3886. doi: 10.1098/rstb.2008.0156

Greenwald, A. G., and Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *J. Pers. Soc. Psychol.* 79, 1022. doi: 10.1037/0022-3514.79.6.1022

Güroğlu, B., Haselager, G. J. T., van Lieshout, C. F. M., and Scholte, R. H. J. (2009). Antagonists in mutual antipathies. *J. Res. Adolesc.* 19, 35–46. doi: 10.1111/j.1532-7795.2009.00580.x

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7

Heffetz, O., and Frank, R. H. (2008). "Preferences for status: evidence and economic implications," in *Handbook of Social Economics*, Vol. 1A, eds J. Benhabib, A. Bisin, and M. Jackson (Amsterdam: Elsevier), 69–91.

Heilman, M. E., and Okimoto, T. G. (2007). Why are women penalized for success at male tasks? The implied communality deficit. *J. Appl. Psychol.* 92, 81–92. doi: 10.1037/0021-9010.92.1.81

Heilman, M. E., Wallen, A. S., Fuchs, D., and Tamkins, M. M. (2004). Penalties for success: reactions to women who succeed at male gender-typed tasks. *J. Appl. Psychol.* 89, 416–427. doi: 10.1037/0021-9010.89.3.416

Irmen, L., and Roßberg, N. (2004). Gender markedness of language: the impact of grammatical and nonlinguistic information on the mental representation of person information. *J. Lang. Soc. Psychol.* 23, 272–307. doi: 10.1177/0261927X04266810

Kagel, J. H., Kim, C., and Moser, D. (1996). Fairness in ultimatum games with asymmetric information and asymmetric payoffs. *Games Econ. Behav.* 13, 100–111. doi: 10.1006/game.1996.0026

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., and Fehr, E. (2013). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75, 73–79. doi: 10.1016/j.neuron.2012.05.021

Moss-Racusin, C. A., Phelan, J. E., and Rudman, L. A. (2010). When men break the gender rules: status incongruity and backlash against modest men. *Psychol. Men Masc.* 11, 140–151. doi: 10.1037/a0018093

Nelissen, R. M., Leliveld, M. C., van Dijk, E., and Zeelenberg, M. (2011). Fear and guilt in Proposers: using emotions to explain offers in ultimatum bargaining. *Eur. J. Soc. Psychol.* 41, 78–85. doi: 10.1002/ejsp.735

Neumann, L. J., and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cogn.* 33, 972–983. doi: 10.3758/BF03193206

Oosterbeek, H., Sloof, R., and van de Kuilen, G. (2004). Differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* 7, 171–188. doi: 10.2139/ssrn.286428

Phelan, J. C., Link, B. G., and Dovidio, J. F. (2008). Stigma and prejudice: one animal or two? *Soc. Sci. Med.* 67, 358–367. doi: 10.1016/j.socscimed.2008. 03.022

Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., et al. (2008). Mentalizing in economic decision-making. *Behav. Brain Res.* 190, 218–223. doi: 10.1016/j.bbr.2008.03.003

Pratto, F., Stallworth, L. M., and Sidanius, J. (1997). The gender gap: differences in political dominance and social dominance orientation. *Br. J. Soc. Psychol.* 36, 49–68. doi: 10.1111/j.2044-8309.1997.tb01118.x

Quadflieg, S., and Macrae, C. N. (2011). Stereotypes and stereotyping: what's the brain got to do with it? *Eur. Rev. Soc. Psychol.* 22, 215–273. doi: 10.1080/10463283.2011.627998

Reskin, B. F. (1988). Bringing the men back in: sex differentiation and the devaluation of women's work. *Gend. Soc.* 2, 58–81. doi: 10.1177/08912438 8002001005

Ridgeway, C. L. (2001). Gender, status, and leadership. *J. Soc. Issues* 57, 627–655. doi: 10.1111/0022-4537.00233

Rilling, J. K., and Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62, 23–48. doi: 10.1146/annurev.psych.121208.131647

Rudman, L. A. (1998). Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *J. Pers. Soc. Psychol.* 74, 629–645. doi: 10.1037/0022-3514.74.3.629

Rudman, L. A., and Fairchild, K. (2004). Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *J. Pers. Soc. Psychol.* 87, 157–176. doi: 10.1037/0022-3514.87.2.157

Rudman, L. A., and Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *J. Pers. Soc. Psychol.* 77, 1004–1010. doi: 10.1037/0022-3514.77.5.1004

Rudman, L. A., and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* 57, 743–762. doi: 10.1111/0022-4537.00239

Rudman, L. A., and Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Pers. Soc. Psychol. Bull.* 26, 1315–1328. doi: 10.1037/a0027356

Rudman, L. A., and Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Res. Organ. Behav.* 28, 61–79. doi: 10.1016/j.riob.2008.04.003

Rudman, L. A., Phelan, J. E., Nauts, S., and Moss-Racusin, C. A. (2012). Status incongruity and backlash toward female leaders: defending the gender hierarchy motivates prejudice against female leaders. *J. Exp. Soc. Psychol.* 48, 165–179. doi: 10.1016/j.jesp.2011.10.008

Saad, G., and Gill, T. (2001). Sex differences in the ultimatum game. *J. Bioecon.* 3, 171–193. doi: 10.1023/A:1020583425623

Sanfey, A. G. (2009). Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses. *Mind Soc.* 8, 93–107. doi: 10.1007/s11299-009-0053-6

Sanfey, A. G., and Chang, L. J. (2008). Multiple systems in decision making. *Ann. N. Y. Acad. Sci.* 1128, 53–62. doi: 10.1196/annals.1399.007

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Solnick, S. (2001). Gender differences in the ultimatum game. *Econ. Inq.* 39, 189–200. doi: 10.1111/j.1465-7295.2001.tb00060.x

Solnick, S. J., and Schweitzer, M. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organ. Behav. Hum. Decis. Process.* 79, 199–215. doi: 10.1006/obhd.1999.2843

Wagner, D. G., and Berger, J. (1993). "Status characteristics theory: the growth of a program," in *Theoretical Research Programs: Studies in the Growth of Theory*, eds J. Berger and M. Zelditch Jr. (Stanford, CA: Stanford University Press), 23–63.

Wheeler, S. C., and Petty, R. E. (2001). The effects of stereotype activation on behavior: a review of possible mechanisms. *Psychol. Bull.* 127, 797–826. doi: 10.1037/0033-2909.127.6.797

Zaatari, D., Palestis, B. G., and Trivers, R. (2009). Fluctuating asymmetry of responders affects offers in the ultimatum game oppositely according to

attractiveness or need as perceived by proposers. *Ethology* 115, 627–632. doi: 10.1111/j.1439-0310.2009.01648

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Reading is for girls!? The negative impact of preschool teachers' traditional gender role attitudes on boys' reading related motivation and skills

Ilka Wolter[1†], Edith Braun[2] and Bettina Hannover[1*]

[1] Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany, [2] International Centre for Higher Education Research at Universität Kassel, Kassel, Germany

According to gender stereotypes, reading is for girls. In this study, we investigated the role of preschool teachers in transmitting such gendered expectations. We suggest that boys are less motivated to read in preschool, and less competent in reading 1 year later in primary school, if their preschool teacher holds a traditional gender role attitude than if the teacher has egalitarian beliefs. In 135 independent dyads of a female preschool teacher ($N = 135$) and one boy ($n = 65$) or one girl ($n = 70$) we measured teacher's gender role attitude, child's reading related motivation as well as precursors of reading skills in preschool, and child's reading skills at the end of first grade in primary school. As expected, the more traditional preschool teachers' gender role attitude was, the weaker was boys' motivation to (learn to) read while girls' motivation was unrelated to teachers' gender role attitude. In either gender, motivation in preschool predicted reading skills at the end of first grade.

Keywords: preschool teachers, gender role attitude, boys' underachievement, precursors of reading skills, reading related motivation, reading skills, gender stereotypes

Introduction

Reading skills are essential for individuals to gain an understanding across subject domains in school and hence are an important predictor of their future socioeconomic success (e.g., Duncan et al., 2007; Ritchie and Timothy, 2013). Sadly, boys have consistently been found to be less competent readers than girls, across different countries and languages (e.g., in PIRLS 2011, reported in Mullis et al., 2012).

The goals of this research are as follows: We wanted to test the assumption that preschool teachers have an influence on gender typing in children's development of reading related skills. While the causes for boys' underachievement in reading are manifold (see Lynn and Mikk, 2009; Martin and Ruble, 2010, for reviews), in this research we want to focus on one potentially relevant factor: the gender role attitude of the preschool teacher. We predicted that preschool teachers who share traditional views with respect to gender roles shape children's attitudes and behaviors in a gender typed manner: Boys whose preschool teacher has a traditional gender role attitude should be less motivated to read than boys with a preschool teacher

holding more egalitarian beliefs, with this lowered motivation in turn having a negative impact on boys' reading achievement in primary school 1 year later.

While reading related skills have extensively been studied in samples of children in (mostly later) primary school grades and secondary school (see Morgan and Fuchs, 2007, for a review), research on emergent readers attending preschool is scarce. Precursors of reading competence, such as phonological awareness or phonological recoding in lexical access, start to develop long before school entry, i.e., during the preschool years (see Townsend and Konold, 2010; Hulme and Snowling, 2013, for reviews), with some studies reporting gender differences to already appear in these early precursor competences (e.g., Lundberg et al., 2012; Wolter et al., 2014), continuing into girls outperforming boys in their reading achievements in first and second grade of primary school (e.g., McCoach et al., 2006; Niklas and Schneider, 2012).

Becoming a skilful reader not only requires precursor competences but also a sufficiently strong motivation to (learn to) read (see Gambrell and Gillis, 2007, for a review). Boys typically describe their motivation to (learn to) read as less strong than girls do (e.g., for first, second, and third graders: Guay et al., 2010; for fourth, fifth, and sixth graders: McGeown et al., 2012). This is particularly worrisome as reading related motivation has been found to covary with reading skills in primary school students (see Morgan and Fuchs, 2007, for a review): The motivation to engage in reading related activities predicts the amount of reading (e.g., Cox and Guthrie, 2001) which in turn predicts growth in reading competence during the primary school years (e.g., Guthrie et al., 2007; Taboada et al., 2009). To summarize, boys' lower reading attainments are related to a weaker motivation to read.

Against this background, in addition to investigating the potential influence of preschool teachers' gender role attitudes, we wanted to test (a) whether the relation between reading related motivation and reading skills can already be found in preschool aged children, (b) whether reading related motivation captured as early as in preschool predicts the development of reading skills during children's first year in primary school, and (c) whether girls and boys already differ in their reading related motivation and precursors of reading skills before school entry.

Reading Related Gender Stereotypes

Boys' lower attainments and motivation in reading seem to reflect gender stereotypes according to which reading is for girls. Developmental research shows that children start to acquire gender stereotypes as early as 2–3 years old, with that knowledge accumulating until about school entry (e.g., Trautner et al., 2005; Banse et al., 2010). Dwyer (1974) found that already in second grade, children endorse reading related gender stereotypes: throughout the school years until grade 12, girls and boys were inclined to describe reading as a feminine activity. Pottorff et al. (1996) found children from grades 2, 4, 6, and 8 to associate reading books with mothers rather than fathers and to consider girls as better able to read books than boys. Similarly, Millard (1997) interviewed seventh graders and found them to associate their mothers rather than their fathers with reading activities

at home and with being taught to read. Martinot et al. (2012) found boys and girls from fifth grade to believe that people conceive of girls as superior to boys in reading. Several studies found secondary school students to conceive of subjects related to reading as "for girls" (e.g., Hannover and Kessels, 2002; Colley and Comber, 2003; Plante et al., 2009; Steffens and Jelenec, 2011).

Transmission of Gender Stereotypes by Preschool Teachers

An extant body of research attests to stereotypes operating in a self-fulfilling manner (e.g., Hamilton et al., 1990): Expectations accompanying a stereotype, e.g., that reading is for girls, trigger a chain of events that finally result in its confirmation, in our case that boys are less motivated to learn to read and therefore, finally, become less skillful readers. The many ways in which children acquire gender stereotypes have not yet been fully identified. One important mechanism of transmission are significant others, such as parents, preschool or school teachers (see Schoon and Eccles, 2014, for a review). They exert their influence via social modeling, via the expression of their own beliefs about the genders, or via direct socialization practices.

First, significant others serve as role models via their own gendered behaviors. For instance, a female teacher engaging in both female (e.g., cooking and baking) and male stereotyped activities (e.g., playing soccer) can be assumed to foster less gender typed attitudes and behaviors in the children of her preschool group than a female teacher offering female stereotyped activities only (cf. Wolter et al., 2014). Indirect evidence for this assumption comes from research showing that children from mothers being employed in the workforce hold less traditional gender role attitudes and that girls with employed mothers are more inclined to strive for an occupation than girls whose mothers stay at home (e.g., Willetts-Bloom and Nock, 1994; Jackson and Tein, 1998).

Second, significant others influence children's assumptions about gender roles by expressing their own gendered expectations. Imagine, for instance, a teacher saying in front of her preschool group of children that the girls probably want to stay inside while the boys would probably prefer to play outside. Indirect evidence for such a transmission of gender stereotypes comes from research on parents' or teachers' gendered ability related expectations (e.g., Tiedemann, 2000; Rouland et al., 2013; see Jacobs et al., 2005, for a review). For instance, Upadyaya and Eccles (2014) found primary school teachers to rate boys as higher in math ability than girls, and girls as putting more effort into reading than boys, with these perceptions in turn predicting children's ability self-concepts in math and reading. Similarly, Retelsdorf et al. (2015) found teachers' reading related gender stereotypes (do boys or girls read better, read more, and have more fun reading?) to favor girls and, when measured at the beginning of grade 5, to predict boys' more negative reading self-concept at the end of grade 6 (controlling for previous self-concept and reading attainments). Other research has shown that gender related attitudes and stereotypes are transmitted across generations within families (e.g., Carlson, 2011; Endendijk, 2013; Farré and Vella, 2013; Hess et al., 2014). For instance, investigating 244 German families, Hess et al. (2014) found that

the more traditional fathers' gender role attitudes were, the more traditional were the attitudes of their sons and daughters 5 years later.

Third, significant others also directly shape gendered behaviors by either reinforcing, ignoring or punishing children's affiliation toward or engagement in different activities, or children's expression of certain self-beliefs (e.g., their ability self-concepts). For instance, a preschool teacher may be more inclined to comfort a girl crying than a boy crying by giving the child a cuddle. Evidence for such means of transmission comes from Lytton and Romney's (1991) classic meta-analysis: they found systematic differences in parents' socialization practices toward their sons vs. daughters in that they tended to encourage gender-typed activities, i.e., different activities depending on their child's gender. To give one example: Simpkins et al. (2005) found mothers of sons to report more encouragement and provision of materials for computer, math, and science related activities than mothers of daughters.

Preschool Teachers' Gender Role Attitudes as Indicator of Gendered Socialization Practices

As explained above, we assume that preschool teachers influence children's gender development via social modeling, expression of expectations about the genders, and via direct socialization practices. To capture the extent to which teachers are inclined to—directly or indirectly—support gender typed attitudes and behaviors in their preschool group of children via these mechanisms we measured their gender role attitudes.

Individuals differ in their gender role attitudes (also: gender role orientation, gender role beliefs, gender role ideology), i.e., in the extent to which they have internalized societal norms about the traditional division of labor between the genders—with men being the breadwinners and women taking care of household and children—and societal expectations according to which behaviors differ in their appropriateness or desirability depending on the actor's biological sex. While individuals with a *traditional gender role attitude* differentiate traits, attitudes, and behaviors as either more typical/desirable for males or for females and, accordingly, endorse the gendered distribution of labor, individuals with an *egalitarian gender role attitude* reject differences between the genders as described by gender stereotypes and disapprove of individuals being assigned to different tasks and resources based on their biological gender. Many different instruments measuring gender role attitudes have been developed (see Frieze and McHugh, 1998, for a review on scales in English; for scales in German: Krampen, 1983; Athenstaedt, 2000).

We expected that teachers' gender role attitudes would systematically relate to the socialization practices they deploy in their preschool group of children. Indirect support for this assumption comes from a study by Cahill and Adams (1997) who found preschool teachers' attitudes toward adult gender roles (e.g., work roles, parental responsibilities) to strongly correlate with their attitudes toward children's gender roles, and child rearing practices (e.g., encouragement of gendered behaviors). Teachers who expressed non-traditional adult gender roles also endorsed non-traditional gender role childrearing practices.

Preschool teachers' gender role attitudes should translate into their treating children in a more or less gender stereotyping manner, such that teachers with a traditional gender role attitude foster gender differences, here in reading motivation, more than teachers with egalitarian beliefs. Indirect evidence for this assumption comes from a study by Kingsbury (2012). Mothers of preschool-aged girls and boys were asked to report their gender role attitudes and to describe how they would react if their child displayed extremely aggressive or shy behaviors. As expected, particularly mothers with traditional gender role attitudes were inclined to respond with negative emotions in response to children's gender incongruent behaviors (i.e., shyness in boys, aggressiveness in girls). Further support for our assumption that preschool teachers with traditional gender role attitudes treat children in a more gender stereotyping manner than teachers with egalitarian beliefs comes from the study by Hess et al. (2014) which, however, investigated a sample of older children. The authors did not only find fathers' gender role attitudes (measured when the child was about 14 years old) to directly impact the child's gender role beliefs 5 years later, but also indirectly via parenting practices (as reported by the child) which were the more gender typed the more traditional fathers' gender role attitudes had been.

Hypotheses and Study Overview

The goals of our study were two-fold:

(A) First, we wanted to identify a potential impact of preschool teachers' gender role attitudes on children's skill development in reading. To that end we measured teachers' gender role attitudes and linked it with data on children's reading related motivation.

Our hypothesis was as follows: The preschool teacher's gender role attitude has a differential effect depending on the child's gender in that: (1) the more traditional the teacher's gender role attitude is, the less motivated boys are to learn to read while (2) for girls, the more traditional the teacher's gender role attitude is, the more motivated girls are to learn to read, or their motivation is unrelated to their teacher's gender role attitude.

(B) Second, we wanted to investigate the interrelatedness between motivation to read and reading skills in a sample of preschoolers. Available reading motivation instruments typically target older groups of children and assume that children can read (see Sperling et al., 2013, for a review). The few studies investigating preschoolers have typically used parents' (e.g., Yeo et al., 2014) or teachers' (e.g., Lepola, 2004; see Morgan and Fuchs, 2007, for a review) judgements on the child's reading motivation. As teacher judgements on children's reading motivation may be biased by their gender role attitudes, in our study we wanted to use children's self-reports. We captured the child's reading related motivation in individual standardized interviews, based on a set of seven questions tailored for our age group of preschoolers. In addition, we measured precursors of reading skills using standardized tests. Reading related motivation and precursors of reading skills were measured

during the child's last year in preschool. Children were followed up after transition to primary school and tested for their reading skills at the end of first grade.

Our hypotheses were as follows: Reading related motivation in preschool is correlated with precursors of reading skills in preschool. Reading related motivation in preschool predicts reading skills at the end of first grade of primary school (direct effect). Precursors of reading skills impact later reading skills via reading related motivation in preschool (indirect effect).

Methods

Sample

Our sample consisted of 135 dyads of a female preschool teacher ($N = 135$) and one boy ($n = 65$) or one girl ($n = 70$) from their group of children. To obtain independent child-teacher dyads, only one dyad was drawn from randomly selected preschool groups in Berlin (for a more detailed description on the sampling strategy see Wolter et al., 2014). All parents were asked for their written consent prior to the first data collection. We only included children whose parents and teachers had given their consent to take part in the study. Teachers and parents were informed that participation was voluntary, and that they could opt-out at any time[1].

Every child and every teacher was investigated in an individual session. Children were first tested during their last 2 months of preschool (t1), and followed up at the end of first grade (t2), on average 13 months after the first measurement, Min = 10 months, Max = 14 months. At t1, children were aged $M = 71.4$ months, $SD = 3.3$, range = 65–78 months. From t1 to t2, our sample was reduced by 28 children (in most cases because they could no longer be reached after transition to primary school). Therefore, our sample size at t2 was $N = 106$ children. The preschool teachers were investigated only once, at t1. They were aged $M = 43.51$ years, $SD = 8.15$ years, range = 23–58 years. At t1, children had been with this particular teacher in their preschool group between 2 and 65 months, $M = 2.4$ years, $SD = 19$ months. On average the children spent about 8 h per day in preschool, $M = 7.27$, $SD = 0.94$, range 5–10 h.

The mean level of socioeconomic background of the children's families, as measured at t1 and operationalized by the HISEI (Highest International Socio-Economic Index of Occupational Status; Ganzeboom et al., 1992), was $M = 57.73$, $SD = 14.75$, and thus somewhat higher than in representative samples of German families (e.g., the German sample of PISA 2009, Klieme et al., 2010, p. 235: $M = 48.9$, $SD = 15.6$).

Data Analysis

Although the sample basically consisted of independent child-teacher dyads, from 16 preschools more than one dyad was drawn—always from a different group led by a different teacher. As we did not analyse variables from different levels

simultaneously, however, there was no need for multilevel-modeling (e.g., Hox, 2002). As we measured different variables in teachers (gender role attitudes) and children (motivation and skills), the data are also independent within dyads, i.e., there is no need to correct for non-independence (e.g., Kenny, 1996). According to MacKinnon (2008), when testing the significance of indirect paths, common methods of significance testing (e.g., joint significance test; Baron and Kenny, 1986) are somewhat inaccurate due to the non-normality of products of two paths. We therefore evaluated the confidence intervals of indirect paths and applied a bootstrapping method (using 1000 bootstrap samples) to correct for potential sample-bias.

To test our hypotheses we conducted a path model using the statistical software *MPlus* 5.1 (Muthén and Muthén, 1998–2010). All predictor variables were entered as grandmean-centered variables into the regression analyses. Missing values were treated as missing at random by using full information maximum likelihood estimators.

Research Instruments
Preschool Teachers' Gender Role Attitudes
To measure gender role attitudes, preschool teachers were asked to fill in the scale by Athenstaedt (2000) at t1. The unidimensional scale consists of 29 items (five-point answering scales: 1 = strongly disagree, 5 = strongly agree) describing either traditional (e.g., "Both boys and girls should undertake household chores"; "For a good first impression a neat appearance is more important for a woman than for a man") or egalitarian views toward gender roles (e.g., "Women are as qualified as men for a leadership position in an engineering company"; "When it comes to politics, men should listen to women to a greater extent"; "A higher number of male preschool teachers would be pleasing").

Out of the 29 items, the 12 items describing an egalitarian view were recoded. Afterwards, we calculated a mean score for every teacher, with higher scores indicating more traditional gender role attitudes. The scale reached Cronbach's $\alpha = 0.703$. On average, teachers responded below the scale's mean, $M = 1.94$, $SD = 0.34$, range = 1.14–2.86, indicating rather egalitarian gender role attitudes.

Reading Related Motivation in Preschool
At the end of preschool (t1) we measured reading related motivation via seven items regarding children's liking for different reading related tasks ("How much do you like to learn new rhymes or poems or songs?"; see Appendix; cf., Bachmann and Burock, 2008). Children had to respond on scales consisting of three "smiley" faces: one with a big smile, (3 = I like it very much), one with a slight smile 2 = I like it , and one with a neutral expression (1 = I don't like it much). The scale reached an internal consistency of Cronbach's $\alpha = 0.70$, with an overall mean of $M = 2.16$, $SD = 0.48$.

Precursors of Reading Skills in Preschool
Precursors of reading skills were measured at the end of preschool (t1) using the "Bielefelder Test zur Früherkennung von Lese-Rechtschreibschwäche" (Jansen et al., 2002; *BISC*;

---

[1]This study was carried out in accordance with the recommendations of the ethical guidelines of the DGPs (German Psychological Society), as approved of by the ethics committee of the DGPs on December 13, 2004.

Bielefelder Screening for Early Detection of Dyslexia). The BISC taps into four different phonological information processing skills (phonological awareness, phonological recoding in lexical access, phonetic recoding in working memory, visual spatial attention regulation); using eight different tasks (e.g., repeating pseudo-words, finding rhyming words, naming the colors of uncoloured or incongruent objects, segmenting syllables, matching phonemes and words). Children achieved a mean sore of $M = 67.90$, $SD = 7.47$, out of a possible 82 points. Internal consistencies for the subscales were ranging between Cronbach's α = 0.63 and 0.85.

Reading Skills in Primary School

To investigate children's reading skills at the end of first grade in primary school (t2), we used the "Wuerzburger Leise Leseprobe," a one scale speed-test of silent reading applicable for first to fourth graders (Küspert and Schneider, 1998). The test consists of 140 tasks children have to work on. In our sample, reliability of the scale was very good (Kuder-Richardson formula for dichotomous items: 0.96). The children mastered between 13 and 106 tasks, $M = 42.34$, $SD = 18.09$. Test scores of our sample are comparable to norm values for this age group (norm sample from item development: $N = 646$, $M = 42.82$, $SD = 17.15$; $t_{(779)} = 0.0817$, n.s., $d = 0.01$).

Results

Gender Differences in Reading Related Motivation and (Precursors of) Reading Skills

In a first step, we tested whether our measures of reading related motivation and precursors of reading skills were invariant across the genders by comparing the measurement models via a confirmatory factor analysis approach. For reading related motivation, the model fit, $\chi^2 = 44.979$, $df = 39$, $p = 0.236$; CFI = 0.927, TLI = 0.921, RMSEA = 0.048, 90%CI 0.000–0.102; SRMR = 0.107, and an insignificant χ^2-difference to the prior metric model, $\Delta\chi^2 = 7.709$, $df = 4$; $p = 0.102$, confirmed partial scalar invariance, such that the prerequisite condition for latent mean comparisons was met. The only restriction from total scalar invariance was the free estimate for two items' intercepts (Items 1 and 7). Similar findings were obtained for precursors of reading skills: the model fit, $\chi^2 = 36.023$, $df = 39$, $p = 0.606$; CFI = 1.00, TLI = 1.05, RMSEA = 0.000, 90%CI 0.000–0.076; SRMR = 0.132, and an insignificant χ^2—difference to the prior metric model, $\Delta\chi^2 = 9.103$, $df = 6$, $p = 0.168$, confirmed partial scalar invariance. The only restriction from total scalar invariance was the free estimate of two subscales' factor loadings (i.e., "Laut-zu-Wort" and "Silben-Segmentieren").

These findings suggest that the measurement models for boys and girls are comparable, such that differences in latent mean values can be interpreted as actual gender differences. Girls, $M = 2.25$, $SD = 0.42$, indicated higher reading related motivation than boys, $M = 2.06$, $SD = 0.52$, at the end of preschool, $b_{(diff)} = 0.536$, $SE = 0.266$, $z = 2.015$, $p = 0.044$. Also as expected, girls obtained significantly higher scores, $M = 69.67$, $SD = 5.16$, than

boys, $M = 65.99$, $SD = 8.94$, in precursors of reading skills at the end of preschool, $b_{(diff)} = 0.597$, $SE = 0.296$, $z = 2.018$, $p = 0.044$.

Further, we tested gender differences in reading skills in primary school. As reading skills were measured by a speed test with a single value (sum score), as indicator of participants' actual skills, measurement invariance is assured. As expected, boys, $M = 38.32$, $SD = 15.63$, were outperformed by girls, $M = 46.16$, $SD = 18.74$, in their reading skills at the end of first grade in primary school, $t_{(133)} = 2.63$, $p < 0.01$, $d = 0.46$.

The Impact of Preschool Teachers' Gender Role Attitudes on Children's Reading Related Motivation

To test our hypothesis referring to the impact of motivation on later reading achievement, we regressed children's reading skills at the end of first grade on reading related motivation in preschool. To further test our hypothesis that preschool teachers' gender role attitudes would impact the development of children's reading related motivation and skills, we regressed children's reading related motivation in preschool on their precursors of reading skills and their preschool teacher's gender role attitude. As we expected preschool teachers' attitudes to be differentially related to reading motivation in boys vs. girls, we also included the term for the statistical interaction of teachers' gender role attitudes and children's gender into the analysis. In addition, as there was considerable variation in the time span that the children in our sample spent with their preschool teacher, we included it as a control variable. As previous research found children's reading attainments to depend on their families' socioeconomic background (e.g., Mullis et al., 2012), we included HISEI as a control variable for children's precursors of reading skills in preschool. Furthermore, we also included the time children spent in preschool per day as an additional control variable for precursors of reading skills.

Zero-order correlations (i.e., Pearson's correlation coefficients) among the predictor variables were as follows: As expected, children's reading related motivation was correlated with precursors of reading skills in preschool, $r = 0.28$, $p < 0.01$. Socioeconomic family background neither correlated with reading related motivation, $r = -0.08$, n.s., nor with precursors of reading skills, $r = 0.13$, n.s.

The model fit of the path model was $\chi^2 = 18.037$, $df = 12$, $p = 0.115$; CFI = 0.902, TLI = 0.828, RMSEA = 0.061, 90%CI 0.000–0.115; SRMR = 0.057. Results (see **Table 1**) revealed, as expected, that reading related motivation in preschool predicted reading skills 1 year later, $b = 8.38$, $SE = 3.49$, $\beta = 0.21$, $p < 0.05$ $d = 0.80$. Moreover, the expected two-way interaction was observed, indicating that teachers' gender role attitudes had a differential effect on reading related motivation in boys vs. girls, $b = 0.62$, $SE = 0.21$, $\beta = 0.34$, $p < 0.01$, $d = 0.98$.

Further, our findings showed that the time the child had already been with the respective teacher was irrelevant for the prediction of reading related motivation, $b < 0.01$, $SE < 0.01$, $\beta = 0.05$, n.s. Results also showed that children's socioeconomic background was related to precursors of reading skills in

TABLE 1 | Path model for (a) reading skills as predicted by reading related motivation, and (b) reading related motivation as predicted by children's precursor of reading skills, preschool teacher's gender role attitudes, and interaction of child's gender and preschool teacher's gender role attitudes, with time spent with preschool teacher controlled for, and (c) children's precursor of reading skills with socioeconomic background (HISEI) and time spent in preschool per day controlled for.

|	b		SE	β	d
READING SKILLS (t2) ON					
Intercept	43.49	***	1.67		
Reading related motivation	**8.38**	*	3.49	0.21	0.80
R^2	0.045				
READING RELATED MOTIVATION (t1) ON					
Child's gender	**0.14**	*	0.07	0.16	0.67
Time spent with teacher (months)	0.00		0.00	0.05	0.20
Precursors of reading skills	**0.02**	***	0.01	0.29	1.17
Teacher's gender role attitude	−0.11		0.11	−0.09	−0.34
Interaction child's gender * teacher's gender role attitude	**0.62**	**	0.21	0.24	0.98
R^2	0.198				
PRECURSORS OF READING SKILLS (t1) ON					
Child's gender	**2.33**	*	1.03	0.19	0.75
Socioeconomic background (HISEI)	**0.07**	+	0.04	0.17	0.60
Time spent in preschool per day (hours)	0.66		0.64	0.13	0.34
R^2	0.072				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$; predictor variables grandmean-centered; significant coefficients are in boldface.

preschool, however, only marginally significantly so, $b = 0.07$, $SE = 0.04$, $β = 0.17$, $p = 0.057$, $d = 0.63$. Precursors of reading skills in preschool and reading related motivation were significantly correlated, $b = 0.02$, $SE = 0.01$, $β = 0.29$, $p < 0.001$, $d = 1.17$.

Post-hoc Probing (cf. Aiken et al., 1991). The results from *post-hoc* probing (i.e., simple-slopes analyses) showed that consistent with our hypothesis, the more traditional preschool teacher's gender role attitudes were, the less boys—but not girls—were motivated to learn to read in preschool (boys: $b = -0.42$, $SE = 0.16$, $β = -0.33$, $p < 0.01$, $d = -0.89$; girls: $b = 0.20$, $SE = 0.14$, $β = 0.16$, n.s.).

To further investigate this differential effect, we conducted additional simple-difference tests to compare the reading related motivation of children whose preschool teachers had either quite traditional (1 SD above the scale's mean) or quite egalitarian attitudes toward gender roles (1 SD below the scale's mean) (see **Figure 1**). As expected, simple difference-tests revealed gender differences in reading related motivation for those children who were with a rather traditional preschool teacher, $b = 0.35$, $SE = 0.10$, $β = 0.40$, $p < 0.001$, $d = 1.11$, but not for children whose teacher was rather egalitarian, $b = -0.07$, $SE = 0.09$, $β = -0.08$, n.s. In fact, in groups with an egalitarian teacher, boys were as motivated to read as girls in preschool.

Indirect Effects. To test the significance of the indirect paths we evaluated the confidence intervals of the expected indirect effect. The indirect effect of teacher's gender role attitudes on children's reading skills, mediated by children's reading related motivation, was significant for boys, $b = -3.50$, $SE = 2.07$, $β = -0.07$, $p = 0.091$, $d = -0.56$, but not for girls, $b = 1.69$, $SE = 1.36$, $β = 0.03$, n.s. Furthermore, though only marginally significant, children's precursor skills in preschool were linked to their later reading

FIGURE 1 | Reading related motivation of girls and boys in preschool as predicted by their preschool teacher's gender role attitude.
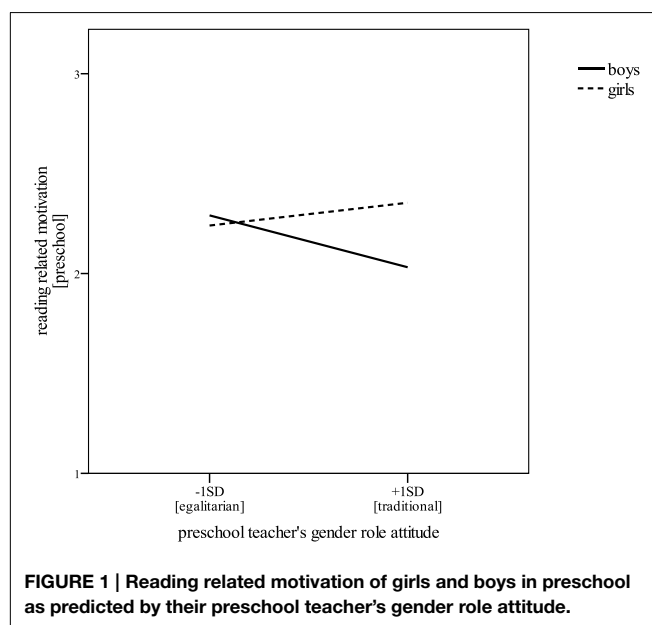
skills in primary school, with this effect again being mediated by children's reading related motivation, $b = 0.17$, $SE = 0.09$, $β = 0.06$, $p = 0.070$, $d = 0.60$.

To sum up, the results showed a differential impact of preschool teachers' gender role attitudes on boys' vs. girls' reading related skill development. The more strongly the preschool teacher endorsed traditional gender roles, the less were boys interested (to learn) to read in preschool and the lower were boys' reading skills at the end of primary school. In contrast,

teacher's gender role attitudes neither had a substantial effect on girls' reading related motivation in preschool nor on their later reading skills in primary school.

Discussion

In this research we wanted to find out whether preschool teachers play a role in children's gendered skill development. In a sample of 135 independent dyads of a female preschool teacher and one boy or one girl from their group of children we found that in fact, gender differences in reading related motivation were, to some extent, explained by teachers' gender role attitudes. Interestingly, our preschool teachers indicated rather egalitarian views on the gender roles, replicating what an earlier study found in teachers for preschool aged children in Turkey (Erden, 2004) and consistent with studies showing that teachers and teacher students hold more egalitarian views when they are female than when they are male (Lasonen and And, 1991; Togrol and Onur, 2000). The more traditional the gender role attitudes of the preschool teachers participating in our study had been, the less were boys motivated to (learn to) read while in preschool and the poorer they performed on a reading skill test one a year later in primary school. While boys and girls did not differ in the strength of their reading related motivation when with a preschool teacher with an egalitarian gender role attitude, boys were significantly less motivated to read when their teacher endorsed traditional gender beliefs. In contrast, girls' reading related motivation was the same, irrespective of how their preschool teachers thought about gender roles. This finding is in line with previous research: With reading being perceived as a female activity (e.g., Dwyer, 1974; Pottorff et al., 1996; Millard, 1997; Hannover and Kessels, 2002; Colley and Comber, 2003; Plante et al., 2009; Steffens and Jelenec, 2011; Martinot et al., 2012; McGeown et al., 2012), traditionally oriented preschool teachers should be less likely to encourage boys to engage in reading, as it is a "gender incongruent" activity. In contrast, both teachers with traditional and egalitarian gender role attitudes should consider reading an appropriate activity for girls, and thus not affect girls' reading motivation differently.

As research on the interrelatedness of reading related skills and motivation in preschoolers is scarce, we also wanted to investigate (a) whether the interrelatedness between reading motivation and reading related (precursor) skills which previous studies have reported for older groups of children can already be found before school start, (b) whether reading motivation in preschool forecasts reading skills in first grade of primary school, and (c) whether gender differences in reading related motivation and precursors of reading skills can already be found in preschool aged children.

In our sample, reading related motivation and precursors of reading skills were moderately correlated. Precursor skills in preschool impacted children's reading related motivation which in turn predicted their reading skills in first grade. While this indirect path had a medium effect size ($d = 0.60$), due to our rather small sample, it turned out to be only marginally significant. Even so other studies typically investigated precursors and reading skills in older children and sometimes measured

both variables at one point in time, the covariation we found between precursors and reading skills 1 year later was comparable in strength to previous findings (e.g., Niklas and Schneider, 2012; Oakhill and Cain, 2012; see Swanson et al., 2003, for a review).

We had included parents' highest occupational status (i.e., HISEI) into our analyses, to measure the socioeconomic background of children's families. Consistent with previous research (e.g., Mullis et al., 2012), socioeconomic background and precursors of reading skills in preschool were positively correlated, and even though only marginally significantly so, with a strong effect size ($d = 0.60$).

These findings go beyond previous research in several respects. Reading related motivation has typically been measured and related to reading skills in older groups of children. One possible cause is that particularly large scale assessments require to capture reading motivation via questionnaires which presuppose that the test person can read. In our study, we measured reading related motivation in single session interviews. To assess reading related skills before school entry, we captured precursors of reading skills. In this way, we were able to show that even before the onset of schooling, motivation to read is more pronounced in girls than in boys, is related to precursors of reading skills, and predicts children's future skill development in reading during their first year in school. These findings complement the ones published by Harwardt-Heinecke et al. (2014) who found precursors of reading competence measured in preschool to predict literacy related motivation at the end of first grade. Our findings are also consistent with the ones from a study by Lepola (2004) who found that particularly boys with poor precursors of reading competence in preschool showed a negative motivational trajectory toward literacy related activities and poor reading skills when followed up until the end of first grade. The only study we are aware of which measured reading related motivation via children's self-report and captured precursors of reading skills in children comparable in age to our sample was conducted by Sperling et al. (2013). These authors' reading motivation measure turned out, however, to be unrelated to children's reading related skills. Accordingly, to our knowledge our study is first to show that reading related motivation as described by preschool aged children coincides with strong reading related precursor skills which and predicts reading skills at the end of their first year in school.

Our results also broaden previous work in that we found gender differences in precursors of reading skills in preschool aged children. While some of the studies investigating such precursors in children of this age group found gender differences favoring girls (Camarata and Woodcock, 2006; Below et al., 2010; Lundberg et al., 2012; Wolter et al., 2014), other studies found the genders not to differ in their reading related precursor skills (Fröhlich et al., 2010; Niklas and Schneider, 2012). Future studies need to further investigate under which circumstances gender differences do or do not emerge in precursor skills measured during the preschool years.

Limitations of Our Study
HISEI-scores for the families having participated in our study were somewhat higher than previous studies had found in

representative samples (e.g., Klieme et al., 2010). While the preschools participating in our studies had randomly been chosen from a list including all 1100 Berlin preschools and while we randomly selected a different one whenever a preschool refused to participate, we cannot rule out that preschools from more affluent districts of the city were more likely to participate. This may have restricted the range of HISEI-scores of the families whose children were included into our study. While our sample may thus not be representative for the Berlin population, we have no reason to suspect that our findings—the interrelatedness between reading related motivation and reading skills, or the impact of preschool teachers' gender role attitudes on children's reading related motivation—would not have turned out, had our sample included more children from families with comparably lower socioeconomic backgrounds. On the contrary, given the slightly limited variance in our HISEI-data, it can be assumed that our findings would have been even stronger, had our sample included more children from families with low HISEI-scores.

In our study we measured preschool teachers' gender role attitudes via self-report, suggesting that a teacher with traditional beliefs more likely deploys gender typing socialization practices in her group of preschool children than a teacher with an egalitarian view on gender roles. Future studies certainly need to substantiate this claim by including variables more directly reflecting teachers' everyday practices, preferably observational data. Also, future studies should include measures of children's reading related gender stereotypes. Such data would allow for a more direct test of the assumption that—via the above described socialization practices—teachers' traditional gender role attitudes impact boys' and girls' reading related motivation and skill development by nourishing children's gender stereotype that "reading is for girls."

Conclusions

The findings of our study suggest that preschool teachers' gender role attitudes can have long-term consequences for boys' reading related skill development: Boys who had been with a preschool teacher endorsing traditional gender attitudes were less motivated to learn to read, with motivation in turn predicting their future skill development. In this way, preschool teachers' gender role attitudes operate in a self-fulfilling manner.

In order to improve boys' reading attainments in school, preschool teachers (and school teachers, whom we did not investigate in our study though) should be sensitized that they can contribute to a gender fair learning environment by carefully monitoring their own views on the gender roles, to make sure they do not reinforce gender typed attitudes and behaviors. There is first evidence that gender role attitudes may even be changed systematically by education classes or training (Erden, 2009; Lucier-Greer et al., 2012).

References

Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Athenstaedt, U. (2000). Normative geschlechtsrollenorientierung: entwicklung und validierung eines fragebogens [Normative gender role attitudes: the development and validation of a new questionnaire]. *Z. Differ. Diagn. Psychol.* 21, 91–104. doi: 10.1024//0170-1789.21.1.91

Bachmann, K., and Burock, K. (2008). *Der Zusammenhang Zwischen der Lernmotivation und den Sprachlichen und Mathematischen Kompetenzen von Jungen und Mädchen im Vorschulalter [The Relationship of Learning Motivation and Verbal and Mathematical Competence of Boys and Girls of Preschool Age]*. Unpublished Diploma thesis, Freie Universität, Berlin.

Banse, R., Gawronski, B., Rebetez, C., Gutt, H., and Morton, J. (2010). The development of spontaneous gender stereotyping in childhood: relations to stereotype knowledge and stereotype flexibility. *Dev. Sci.* 13, 298–306. doi: 10.1111/j.1467-7687.2009.00880.x

Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Below, J., Skinner, C., Fearrington, J., and Sorrell, C. (2010). Gender differences in early literacy: analysis of kindergarten through fifth-grade dynamic indicators of basic early literacy skills probes. *School Psychol. Rev.* 39, 240–257.

Cahill, B., and Adams, E. (1997). An exploratory study of early childhood teachers' attitudes toward gender roles. *Sex Roles* 36, 517–529. doi: 10.1007/BF02766688

Camarata, S., and Woodcock, R. (2006). Sex differences in processing speed: developmental effects in males and females. *Intelligence* 34, 231–252. doi: 10.1016/j.intell.2005.12.001

Carlson, D. (2011). Family structure and the intergenerational transmission of gender ideology. *J. Fam. Issues* 6, 709–734. doi: 10.1177/0192513X10396662

Colley, A., and Comber, C. (2003). School subject preferences: age and gender differences revisited. *Educ. Stud.* 29, 59–67. doi: 10.1080/03055690303269

Cox, K. E., and Guthrie, J. T. (2001). Motivational and cognitive contributions to students' amount of reading. *Contemp. Educ. Psychol.* 26, 116–131. doi: 10.1006/ceps.1999.1044

Duncan, G., Dowsett, C., Claessens, A., Magnuson, K., Huston, A., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Dwyer, C. (1974). Influence of children's sex role standards on reading and arithmetic achievement. *J. Educ. Psychol.* 66, 811–816. doi: 10.1037/h0021522

Endendijk, J. (2013). Gender stereotypes in the family context: mothers, fathers, and siblings. *Sex Roles* 9, 577–590. doi: 10.1007/s11199-013-0265-4

Erden, F. T. (2004). Early childhood teachers' attitudes toward gender roles and toward discipline. *Hacettepe Univ. J. Educ.* 27, 83–90.

Erden, F. T. (2009). A course on gender equity in education: does it affect gender role attitudes of preservice teachers? *Teach. Teach. Educ.* 25, 409–414. doi: 10.1016/j.tate.2008.11.001

Farré, L., and Vella, F. (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica* 80, 219–247. doi: 10.1111/ecca.12008

Frieze, I., and McHugh, M. C. (1998). Measuring feminism and gender role attitudes. *Psychol. Women Q.* 22, 349–352. doi: 10.1111/j.1471-6402.1998.tb00159.x

Fröhlich, L. P., Metz, D., and Petermann, F. (2010). *Förderung der phonologischen Bewusstheit und sprachlicher Kompetenzen: das Lobo-Kindergartenprogramm.* Göttingen: Hogrefe Verlag.

Gambrell, L. B., and Gillis, V. (2007). "Assessing children's motivation for reading and writing," in *Classroom Literacy Assessment: Making Sense of What Students Know and Do,* eds J. Paratore and R. McCormack (New York, NY: Guilford Press), 50–61.

Ganzeboom, H. B. G., De Graaf, P. M., and Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Soc. Sci. Res.* 21, 1–56. doi: 10.1016/0049-089X(92)90017-B

Guay, F., Chanal, J., Ratelle, C. F., Marsh, H. W., Larose, S., and Boivin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *Br. J. Educ. Psychol.* 80, 711–735. doi: 10.1348/000709910X499084

Guthrie, J. T., Hoa, A. W., Wigfield, A., Tonks, S. M., Humenick, N. M., and Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemp. Educ. Psychol.* 32, 282–313. doi: 10.1016/j.cedpsych.2006.05.004

Hamilton, D., Sherman, S., and Ruvolo, C. (1990). Stereotype-based expectancies: effects on information processing and social behavior. *J. Soc. Issues* 46, 35–60. doi: 10.1111/j.1540-4560.1990.tb01922.x

Hannover, B., and Kessels, U. (2002). Monoedukativer anfangsunterricht in physik: auswirkungen auf motivation, selbstkonzept und kurswahlverhalten von gesamtschülerinnen und gesamtschülern [Physics in single-sex or mixed teaching groups. Impact on motivation, self-concept of abilities, and course enrollment of 8th grade comprehensive school students]. *Z. Entwicklungspsychol. Pädagog. Psychol.* 34, 201–215. doi: 10.1026//0049-8637.34.4.201

Harwardt-Heinecke, E., Milatz, A., and Ahnert, L. (2014). Die herausbildung erster leistungsprofile nach schuleintritt. Zusammenhänge zu motivation, vorläuferkompetenzen und beziehungsqualitäten [Academic achievements post school entry as related to motivation, precursor competencies, and relationship-quality]. *Psychol. Erzieh. Unterr.* 61, 267–280. doi: 10.2378/peu2014.art21d

Hess, M., Ittel, A., and Sisler, A. (2014). Gender-specific macro- and micro-level processes in the transmission of gender role orientation in adolescence: the role of fathers. *Eur. J. Dev. Psychol.* 11, 211–226. doi: 10.1080/17405629.2013.879055

Hox, J. (2002). *Multilevel Analysis. Techniques and Applications.* Mahwah, NJ: Lawrence Erlbaum.

Hulme, C., and Snowling, M. J. (2013). Learning to read: what we know and what we need to understand better. *Child Dev. Perspect.* 7, 1–5. doi: 10.1111/cdep.12005

Jackson, D., and Tein, J. (1998). Adolescents' conceptualization of adult roles: relationships with age, gender, work goals, and maternal employment. *Sex Roles* 38, 987–1008. doi: 10.1023/A:1018826626335

Jacobs, J. E., Davis-Kean, P., Bleeker, M., Eccles, J. S., and Malanchuk, O. (2005). "'I can, but I don't want to': the impact of parents, interests, and activities on gender differences in math," in *Gender Differences in Mathematics: An Integrative Psychological Approach,* eds A. M. Gallagher and J. C. Kaufman (New York, NY: Cambridge University Press), 246–263.

Jansen, H., Mannhaupt, G., Marx, H., and Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten [BISC; Bielefelder Screening for Early Detection of Dyslexia], 2nd Edn.* Göttingen: Hogrefe.

Kenny, D. A. (1996). Models of non-independence in dyadic research. *J. Soc. Pers. Relat.* 13, 279–294. doi: 10.1177/0265407596132007

Kingsbury, M. (2012). Mothers' gender-role attitudes and their responses to young children's hypothetical display of shy and aggressive behaviors. *Sex Roles* 66, 506–517. doi: 10.1007/s11199-012-0120-z

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., et al. (2010). *PISA 2009. Bilanz Nach einem Jahrzehnt. [PISA 2009. A Decade in Retrospect].* Münster: Waxmann.

Krampen, G. (1983). Eine kurzform der skala zur messung normativer geschlechtsrollen-orientierungen [A short version of the sex role orientation scale]. *Z. Soz.* 12, 152–156.

Küspert, P., and Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP) [Wuerzburger Silent Reading-test].* Göttingen: Hogrefe.

Lasonen, J., and And, O. (1991). *Finnish Comprehensive Vocational Institute Teachers' Gender-Role Attitudes [e-book].* Ipswich, MA: ERIC. (Accessed February 13, 2015).

Lepola, J. (2004). The role of gender and reading competence in the development of motivational orientations from kindergarten to grade 1. *Early Educ. Dev.* 15, 215–240. doi: 10.1207/s15566935eed1502_5

Lucier-Greer, M., Ketring, S. A., Adler-Baeder, F., and Smith, T. (2012). Malleability of gender role attitudes and gendered messages in couple and relationship education. *Fam. Consum. Sci. Res. J.* 41, 4–17. doi: 10.1111/j.1552-3934.2012.02125.x

Lundberg, I., Larsman, P., and Strid, A. (2012). Development of phonological awareness during the preschool year: the influence of gender and socio-economic status. *Read. Writ.* 25, 305–320. doi: 10.1007/s11145-010-9269-4

Lynn, R., and Mikk, J. (2009). Sex differences in reading achievement. *TRAMES* 13, 3–13. doi: 10.3176/tr.2009.1.01

Lytton, H., and Romney, D. M. (1991). Parents' differential socialization of boys and girls: a metaanalysis. *Psychol. Bull.* 109, 267–296. doi: 10.1037/0033-2909.109.2.267

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis.* Mahwah, NJ: Erlbaum.

Martin, C. L., and Ruble, D. N. (2010). Patterns of gender development. *Annu. Rev. Psychol.* 61, 353–381. doi: 10.1146/annurev.psych.093008.100511

Martinot, D., Bagès, C., and Désert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: when girls improve their reputation in math. *Sex Roles* 66, 210–219. doi: 10.1007/s11199-011-0032-3

McCoach, D. B., O'Connell, A. A., Reis, S. M., and Levitt, H. A. (2006). Growing readers: a hierarchical linear model of children's reading growth during the first two years of school. *J. Educ. Psychol.* 98, 14–28. doi: 10.1037/0022-0663.98.1.14

McGeown, S., Goodwin, H., Henderson, N., and Wright, P. (2012). Gender differences in reading motivation: does sex or gender identity provide a better account? *J. Res. Read.* 35, 328–336. doi: 10.1111/j.1467-9817.2010.01481.x

Millard, E. (1997). Differently literate: gender identity and the construction of the developing reader. *Gend. Educ.* 9, 31–48. doi: 10.1080/09540259721439

Morgan, P. L., and Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Except. Child.* 73, 165. doi: 10.1177/001440290707300203

Mullis, I., Martin, M., Foy, P., and Drucker, K. (2012). *PIRLS 2011 International Results in Reading.* Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

Muthén, L. K., and Muthén, B. O. (1998–2010). *Mplus User's Guide, 6th Edn.* Los Angeles, CA: Muthén and Muthén.

Niklas, F., and Schneider, W. (2012). Die anfänge geschlechtsspezifischer leistungsunterschiede in mathematischen und schriftsprachlichen kompetenzen [The beginning of gender-based performance differences in mathematics and linguistic competencies]. *Z. Entwicklungspsychol. Pädagog. Psychol.* 44, 123–138. doi: 10.1026/0049-8637/a000064

Oakhill, J. V., and Cain, K. (2012). The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Sci. Stud. Read.* 16, 91–121. doi: 10.1080/10888438.2010.529219

Plante, I., Théorét, M., and Favreau, O. E. (2009). Student gender stereotypes: contrasting the perceived maleness and femaleness of mathematics and language. *Educ. Psychol.* 29, 385–405. doi: 10.1080/01443410902971500

Pottorff, D. D., Phelps-Zientarski, D., and Skovera, M. E. (1996). Gender perceptions of elementary and middle school students about literacy at school and home. *J. Res. Dev. Educ.* 29, 203–211.

Retelsdorf, J., Schwartz, K., and Asbrock, F. (2015). "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. *J. Educ. Psychol.* 107, 186–194. doi: 10.1037/a0037107

Ritchie, S., and Timothy, B. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* 20, 1–8. doi: 10.1177/0956797612466268

Rouland, K., Rowley, S., and Kurtz-Costes, B. (2013). Self-views of African-American youth are related to the gender stereotypes and ability attributions of their parents. *Self Identity* 12, 382–399. doi: 10.1080/15298868.2012.682360

Schoon, I., and Eccles, J. (2014). *(Eds). Gender Differences in Aspirations and Attainment. A Life Course Perspective.* Cambridge: Cambridge University Press.

Simpkins, S. D., Davis-Kean, P. E., and Eccles, J. S. (2005). Parents' socializing behavior and children's participation in math, science, and computer out-of-school activities. *Appl. Dev. Sci.* 9, 14–30. doi: 10.1207/s1532480xads0901_3

Sperling, R. A., Sherwood, T. I., and Hood, A. M. (2013). Relating motivation to read and emergent reading skills: a measurement validation study. *Read. Psychol.* 34, 461–485. doi: 10.1080/02702711.2012.658143

Steffens, M. C., and Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles* 64, 324–335. doi: 10.1007/s11199-010-9924-x

Swanson, H. L., Trainin, G., Necoechea, D. M., and Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: a meta-analysis of the correlation evidence. *Rev. Educ. Res.* 73, 407–440. doi: 10.3102/00346543073004407

Taboada, A., Tonks, S. M., Wigfield, A., and Guthrie, J. T. (2009). Effects of motivational and cognitive variables on reading comprehension. *Read. Writ.* 22, 85–106. doi: 10.1007/s11145-008-9133-y

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *J. Educ. Psychol.* 92, 144–151. doi: 10.1037/0022-0663.92.1.144

Togrol, A. Y., and Onur, A. (2000). Education students' attitudes toward gender roles. *Mediterr. J. Educ. Stud.* 5, 103–113.

Townsend, M., and Konold, T. R. (2010). Measuring early literacy skills: a latent variable investigation of the phonological awareness literacy screening for preschool. *J. Psychoeduc. Assess.* 28, 115–128. doi: 10.1177/0734282909336277

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., and Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: developmental or differential? *Infant Child Dev.* 14, 365–381. doi: 10.1002/icd.399

Upadyaya, K., and Eccles, J. (2014). "Gender differences in teachers' perceptions and children's ability self-concepts," in *Gender Differences in Aspirations and Attainment. A Life Course Perspective,* eds I. Schoon and J. Eccles (Cambridge: Cambridge University Press), 79–100.

Willetts-Bloom, M., and Nock, S. (1994). The influence of maternal employment on gender role attitudes of men and women. *Sex Roles* 30, 371–389. doi: 10.1007/BF01420599

Wolter, I., Glüer, M., and Hannover, B. (2014). Gender-typicality of activity offerings and child–teacher relationship closeness in German "Kindergarten." Influences on the development of spelling competence as an indicator of early basic literacy in boys and girls. *Learn. Individ. Dif.* 31, 59–65. doi: 10.1016/j.lindif.2013.12.008

Yeo, L. S., Ong, W. W., and Ng, C. M. (2014). The home literacy environment and preschool children's reading skills and interest. *Early Educ. Dev.* 25, 791–814. doi: 10.1080/10409289.2014.862147

Appendix

Items of the scale "reading related motivation" measured in preschool (t1):

1. How much do you like to learn new rhymes or poems or songs?
2. There are both simple, short poems and songs and some that are longer and more difficult. How much do you like to learn difficult poems and songs too?
3. How much do you like to cite poems or to sing songs without someone helping you?
4. How much to you like someone reading stories to you?
5. How much do you like to tell stories? For example about things that you have experienced, or about something that you have seen on TV?
6. How much do you like to write or draw alphabetic characters?
7. How much do you like to write alphabetic characters without someone helping you?

Advantages of publishing in Frontiers

OPEN ACCESS

Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW

Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION

Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS

No limit to article distribution and re-use

TRANSPARENT

Editors and reviewers acknowledged by name on published articles

SUPPORT

By our Swiss-based editorial team

IMPACT METRICS

Advanced metrics track your article's impact

GLOBAL SPREAD

5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK

Our network increases readership for your article

Find us on