

---

**Doctoral thesis submitted to  
the Faculty of Behavioural and Cultural Studies  
Heidelberg University  
in partial fulfillment of the requirements of the degree of  
Doctor of Philosophy (Dr. phil.)  
In Psychology**

Title of the publication-based thesis  
*Cognitive Representation of Gender Typicality and its Effects on  
Linguistic Processing*

presented by  
M. Sc. Chiara Reali

year of submission  
2015

Dean: Prof. Dr. Birgit Spinath  
Advisor: Prof. Dr. Lisa von Stockhausen

---

## Contents

---

List of scientific publications for this publication-based dissertation .....	3
<b>1. Introduction .....</b>	<b>4</b>
1.1. Research motivation .....	4
1.2. The anaphor resolution paradigm as a tool to investigate gender typicality .....	6
1.3. Methodology .....	9
1.4. State-of-the-art of international research on the topic .....	11
<b>2. Overview of the present research .....</b>	<b>13</b>
2.1. Paper 1: Isolating typical gender in a grammatical gender language .....	15
2.2. Paper 2: Role descriptions induce gender mismatch effects in eye movements during reading .....	18
2.3. Paper 3: Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution .....	20
2.4. Paper 4: Prominence of Gender Cues in the Assignment of Thematic Roles in German .....	22
2.5. Paper 5: Gender Hierarchies in the Processing of Ambiguous French Anaphors .....	24
<b>3. Conclusion .....</b>	<b>25</b>
<hr/>	
References .....	27
Paper 1 .....	31
Paper 2 .....	62
Paper 3 .....	76
Paper 4 .....	100
Paper 5 .....	148
Erklärung gemäß § 8 Abs. 1 Buchst. b) und c) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften .....	187

## List of scientific publications

---

### **Paper 1**

Reali, C., Esaulova, Y., & von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics*, 36, 977-1006.

### **Paper 2**

Reali, C., Esaulova, Y., Öttl, A., & von Stockhausen, L. (in press). Role descriptions induce gender mismatch effects in eye movements during reading. *Frontiers in Psychology – Cognition*. 6, 1607.

### **Paper 3**

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, 29, 781-803.

### **Paper 4**

Esaulova, Y., Reali, C., & von Stockhausen, L. (2015). Prominence of gender cues in the assignment of thematic roles in German. Manuscript submitted for publication.

### **Paper 5**

Esaulova, Y., Reali, C., & von Stockhausen, L. (2015). Gender Hierarchies in the Processing of Ambiguous French Anaphors. Manuscript submitted for publication.

## 1. Introduction

The goal of the present research is to investigate the cognitive representation of gender in occupational roles with the help of psycholinguistic tools. For this reason, the research can be placed at the intersection between the fields of psycholinguistics, cognitive psychology and social psychology, and approaches the topic of gender typicality on different levels. Chapter 1.1 of the Introduction presents the research topic and its importance from a socio-psychological perspective. Chapter 1.2 describes the psycholinguistic paradigms employed throughout the research and offers theoretical background. Chapter 1.3 presents the research methodology. Chapter 1.4 describes the state-of-the-art in international research on the topic. Chapter 2. presents an overview of the work and explains the general research plan connecting the different studies. Finally, Chapter 3. offers a summary and conclusions for the research.

### 1.1 Research motivation

We define gender typicality in occupational roles as the estimated probability for a role to be ascribed to men or to women. For instance, we often refer to a doctor or an architect of unspecified sex as *he* rather than *she*, thus assigning a typical male gender to these roles, whereas we expect a female referent when talking about a nurse or a model, thus assigning them a typical female gender (Oakhill, Garnham, & Reynolds, 2005). When gender typical beliefs associated to specific roles do not correspond to our actual experience, we feel obliged to provide additional information, even when other cues unequivocally indicate the referent gender, as in this example: *Military rules ban pregnant servicewomen from front-line duties, though last year another female British soldier gave birth two weeks after returning from her six-months deployment to Afghanistan* [BBC News, 24th March, 2013] (Syianova-Chanturia, Warren, Pesciarelli, & Cacciari, 2015).

Gender typical representations often correspond to the actual distribution of men and women in the mentioned professions (Garnham, Doehren, & Gygax, 2015). In this perspective, gender typical attributions are an adaptive cognitive strategy which allows the construction of fast and often correct representations of the referents. This anticipatory representation is essential for a quick and smooth discourse comprehension. However, this simplifying strategy entails also important disadvantages. First, eventually wrong gender assignments may lead to comprehension impairments and thus require repair costs to re-establish coherence in the discourse<sup>1</sup>. Second and more important, the gender typicality heuristic can have far-reaching consequences in influencing social expectations and behavior, as shown by socio-psychological studies (for a review see Stahlberg, Braun, Irmen, & Sczesny, 2007). For example, stereotypical beliefs may influence gender expectations in the selection of job candidates, or model a person's professional plan according to gender stereotypes. Empirical evidence shows that (gender) stereotypical beliefs are constantly conveyed, maintained and reinforced through language. For example, in a study with primary school children, Vervecken and Hannover (2012) showed that occupational roles presented in pair form (e.g., *Ingenieurinnen und Ingenieure*, engineers<sub>female/male</sub>), increased the cognitive availability of female jobholders, and strengthened girls' interest in typically male occupations, in comparison to formulations using the generic masculine form (e.g., *Ingenieure*, engineers<sub>masculine</sub>).

A further characteristic of gender stereotypical beliefs resides in the automaticity of their activation. Studies show that cognitive strategies based on gender typicality are often applied automatically in association to specific linguistic cues, rather than as a result of an

---

<sup>1</sup> This classic example by Sanford (1985) illustrates possible comprehension difficulties triggered by gender stereotypes: *A father and son are driving home one day, when they are involved in a serious accident. The father is killed outright, but the son is driven to hospital, where he is about to undergo an emergency operation. However, the surgeon refuses to operate, saying: "I can't operate on him: he's my son." The question is, how can this be?* (Sanford, 1985, p.311).

intentional strategy (see Oakhill et al., 2005; Pyykkönen, Hyönä & van Gompel, 2010). As a consequence, gender stereotypes appear to be difficult to suppress by deliberate decision, counter-stereotypical clarification and explicit instructions.

For these reasons, the present research focuses on the link between language and gender typicality, with special consideration to the subtle and automatic aspects of this cognitive structure. Specifically, the following experiments aim to test the effect of gender typicality on the automatized process of text comprehension during reading. To detect this effect we employ throughout the research the anaphor resolution paradigm, which is explained in detail in the next chapter. The rationale of the paradigm consists essentially in two steps: 1) presentation of gender cues to prime the representation of a personal role, and 2) disambiguation of the actual gender of the referent. If the referent gender matches the reader's expectations, then the comprehension process would proceed smoothly. If the referent gender does not match the reader's expectations, then a disruption in the reading process is expected. Therefore, the anaphor resolution paradigm enables to detect the readers' cognitive expectations about gender, and to test how they impact the process of text comprehension. Specifically, we are interested in analyzing the time course of the activation of different gender cues, in order to understand how the integration of gender information takes place during the reading process. The anaphor resolution paradigm is applied both in the classical version analyzed in psycholinguistic literature, and in a novel experimental version designed to extend the paradigm to a cross-linguistic perspective.

## **1.2 The anaphor resolution paradigm as a tool to investigate gender typicality**

The anaphor resolution paradigm during reading is based on the presentation of a written sentence containing an *antecedent* followed by a later reference to the antecedent (*anaphor*). For example, in the sentence `The executive distributed an urgent memo. She

made it clear that work would continue as normal' (Kennison & Trofe, 2003), 'the executive' represents the antecedent and 'she' the pronominal anaphor. The process of anaphor resolution consists in establishing a correspondence between the anaphoric reference and the initially presented referent. This correspondence is based on the available linguistic cues, which can be represented by grammatical agreement (e.g., number and gender) and by cognitive plausibility (e.g., the possibility that the mentioned executive is a woman). A cognitive mismatch between antecedent and anaphor cues would result in a surprise effect, that is, an impairment in the reading process, commonly expressed through longer fixations on the unexpected text and/or regressions to previous parts of the sentence. This behavior reflects an additional mobilization of cognitive resources to adjust the representations of the referents and update the constructed cognitive model. This process is described and explained by the Mental Model account (Garnham 2001), which represents the theoretical background common to our studies. According to this approach, during language comprehension readers construct a cognitive representation of the situation and characters described in the text, making use of both explicitly stated text cues and implicit information activated on the basis of world knowledge (see Garnham 2001; Pyykkönen et al., 2010).

Linguistic cues relating to gender information are conveyed mostly through two sources: lexical (e.g., mother/father), and conceptual (e.g., nurse/surgeon). Grammatical gender languages, such as German, possess an additional source of gender information, namely grammatical gender (e.g., the suffix *-in* in 'Lehrerin', teacher<sub>female</sub>). In the case of grammatical gender systems, grammatical and stereotypical gender cues are difficult to disentangle, since most role nouns are also marked through morphological suffixes. Thus, in natural gender languages, such as English, professional role nouns are commonly unmarked for gender (e.g., 'the electrician'), whereas in grammatical gender languages a

gender-neutral formulation of a role is rather the exception, and mostly possible only in plural form (e.g., *Die Angestellte*, the employees). These characteristics entail important consequences for the implementation of the anaphor resolution paradigm in different gender systems. In natural gender languages antecedent role nouns are ambiguous with regard to referent gender, therefore they can be used to test the reader's gender expectation on the role. This is achieved by measuring the 'surprise effect' on the anaphor after a gender mismatching antecedent (e.g., 'The engineer... She...'). On the contrary, in grammatical gender languages role noun antecedents may already suggest the referent gender through the antecedent's grammatical markings and thus confound the effect of the role typicality on the anaphor resolution. For instance, the role *Ingenieurin*, engineer, has a male typicality, however the grammatical suffix *-in* suggests the representation of a female engineer; as a consequence, the presentation of a counter-stereotypical anaphor (*sie*, she) may not trigger -or trigger to a minor extent- the surprise effect produced in natural gender languages. For these reasons, the anaphor resolution paradigm classically employed in psycholinguistic literature presents evident limitations in the perspective of a cross-linguistic comparison between grammatical and natural gender languages, due to the different availability of gender cues in the different gender systems.

The present research proposes a novel paradigm which intends to overcome this issue and allow a cross-linguistic comparisons of different gender systems. This goal is reached by replacing role noun antecedents through role descriptions which do not present grammatical cues to the referent gender. For example, the role noun '*Grundschullehrer/Grundschullehrerin*', 'primary school teacher<sub>male/female</sub>' can be replaced by the grammatical-gender free description '*unterrichtet an einer Grundschule*', 'teaches at a primary school'. In our experiments the grammatical subject of the sentences was represented by initials (e.g., '*D.H. unterrichtet...*') in order to conceal the referent gender.



The description-based paradigm allowed us to investigate the effect of gender typicality with a comparable paradigm in both a natural gender language (English) and a grammatical gender language (German). The recording of the exact time-course of the gender stereotype activation and the cognitive processing of the different linguistic cues was made possible by the eye-tracking methodology employed throughout the research.

### 1.3 Methodology

Eye-tracking technique consists in video monitoring eye-movements during the accomplishment of a cognitive task. A high-precision camera tracks and records the movements of the participant's pupil and corneal reflection. This methodology is based on the assumption of a relationship between eye-movements and cognition. Specifically, some aspects of eye behavior, such as the durations of eye fixations on words or on regions of text, are used to infer cognitive processes (Just & Carpenter, 1987; Rayner & Pollatsek, 1989). In relation to our studies, for instance, longer fixations on one item indicate that additional cognitive resources are required to integrate this specific item (e.g., the anaphor *she*) in the current mental model, as compared to the alternative experimental condition (e.g., the anaphor *he*). Eye-tracking offers several advantages in comparison to other methodologies that have been employed to study the anaphor resolution paradigm, for example moving-window self-paced reading<sup>2</sup>. The principal advantage of eye-tracking consists in its high spacial and temporal resolution, which allows a very precise recording of the reading process (see following papers for technical details). This feature is combined with high ecological validity, since readers are allowed to freely process the complete text

---

<sup>2</sup> In moving-window self-paced reading studies a text is first displayed as a series of dashes on the screen, with each dash representing a word or a region in the text. Pressing a button causes the first word to appear, replacing the corresponding dashes. Subsequent button presses cause the previous word to be replaced by dashes while the current word is shown, so that only one word is visible at a specific time, thus creating the impression of a moving window on the screen.

on the screen, pressing a button to proceed to the subsequent item. Within an item, it is possible for readers to regress to previous parts of the sentence, skip characters or anticipate the processing of specific words through parafoveal perception. Moreover, eye-tracking offers the opportunity to record a set of different time measures, which in turn provide essential information on the reading behavior and reflect the different stages of text processing: First Fixation Time represents the duration of the first fixation in a given region; First Pass Time reflects 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 Path Time is the time from first entering a region until leaving it to the right, including the time for regressions from this region; Total Fixation 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. Boland, 2004). In general, longer fixation times and a higher probability of regressions indicate comparatively greater difficulty in processing the respective region.

Finally, in comparison to explicit measures of gender stereotypes, such as classical questionnaires, eye-tracking methodology allows to almost completely exclude social desirability effects, since readers are usually unaware of consciously directing their eye movements during reading (see Reichle, Pollatsek, Fisher, & Rayner, 1998).

For these reasons the methodology of eye-tracking during reading was employed in all studies of the present research, in association with reaction-time priming paradigm (paper 1) and explicit measures on gender roles as well as gender typicality ratings (papers 1-5).

#### 1.4 State-of-the-art of international research on the topic

While a more detailed literature review can be found in the papers, we summarize below the most important findings on gender typicality effects during on-line sentence comprehension.

In natural gender languages, studies on sentence processing during reading have focused on the paradigm of anaphor resolution with role noun antecedents in association with personal pronouns (*he/she*, e.g., Kennison & Trofe, 2003; Kennison, Fernandez & Bowers, 2009) or reflexive pronouns (*himself/herself*, see for example Duffy & Keir, 2004; Kreiner, Sturt & Garrod, 2008). Results showed that gender typicality of the anaphor affected the process of pronominal reference resolution and elicited a mismatch effect in the condition of incongruity with the anaphor gender. Research questions still open to debate concern the time course of the activation of the typical gender information (immediate vs. postponed in a later stage after syntactic processing; automatic vs. elaborative) and the storage of gender typicality information (lexical, i.e., associated to the word meaning, vs. conceptual, i.e., related to world-knowledge).

In grammatical gender languages, studies employing the anaphor resolution paradigm with role nouns focused on the complex interaction of grammatical and stereotypical gender cues. Due to the effect of grammatical gender markings on the role noun, a mismatch effect can be displayed in this case already on the antecedent. For example, in a self-paced reading experiment in Spanish by Carreiras, Garnham, Oakhill, & Cain (1996), the grammatical gender of the role noun antecedent could match or mismatch its own stereotypical gender. Moreover, the stereotypical gender of the role noun could either match or mismatch a subsequent anaphor (e.g., *El carpintero/La carpintera tomó las medidas para hacer el armario. Era un encargo bastante urgente. El/Ella tenía que terminarlo en el plazo de una semana.* "The carpenter took measurements to make the

cupboard. It was a quite urgent order. He/She had to finish it in the space of one week.'). Results showed slower reading times on the initial region in the condition of mismatch between grammatical and stereotypical gender (e.g., *La carpintera* 'the carpenter<sub>feminine</sub>'). In the last sentence, which contained the anaphoric reference, no effect of typicality was found when referent gender was already established via morphological features of the role noun and its preceding article. This study shows that when a role noun is encountered, the gender information provided by stereotypicality is immediately compared with, and if necessary overruled by, gender cues provided by the local morphology. However, further studies showed that typical gender cues may not be completely overridden by grammatical gender agreement and thus still have an impact on the pronominal resolution process (e.g., Esaulova, Reali & von Stockhausen, 2014).

Recent research employed further methodologies to explore neural correlates of the gender mismatch effects emerging during on-line anaphor resolution, namely brain event-related potentials (ERPs) during reading. Irmen, Holt, & Weisbrod (2010) investigated the effect produced by typically male and typically female role noun antecedents in masculine plural form on a noun phrase anaphor (e.g., *Viele Informatiker tragen eine Brille, denn diese Männer/ Frauen/ Leute arbeiten viel am Rechner*. Many computer scientists<sub>masculine</sub> wear glasses since these men/ women/ people work often with a computer). Results showed a mismatch effect in the time window between 500 and 700 ms after target word onset, in the condition of mismatch between typical gender of the antecedent and anaphor gender. This effect suggests a difficulty in the integration of a mismatching referent in the existing mental model. The mismatch may be perceived as a syntactic incongruity (classical P-600 effect).

In an ERPs study with single words, Syianova, Pesciarelli, & Cacciari (2012) analyzed the priming effect of masculine and feminine role nouns (e.g., *pensionato/*

*pensionata*, *pensioner*<sub>masculine/feminine</sub>) and of typically male and female bi-gender role nouns (e.g., *insegnante*<sub>masculine and feminine</sub>, teacher) on a following pronoun categorization (*lui/lei*, he/she). Results documented a mismatch effect on the pronoun for trials presenting an incongruity between grammatical gender of the role noun and pronoun gender. Moreover, a mismatch effect emerged when the masculine pronoun was followed by a stereotypically incongruent prime (e.g., *insegnante – lui*, teacher – he). The mismatch was represented in both cases by a N-400-like effect, that is a negative brain potential peaking around 400 ms after target word onset, which may be interpreted as processing of a semantic violation during the comprehension process. Interestingly, the quoted ERPs studies document an asymmetry between masculine and feminine anaphors in the gender mismatch effect, an aspect which has been poorly investigated in literature and which will be further discussed in the present research.

## **2. Overview of the present research**

The present research includes a series of studies which aim to further clarify the effects of gender typicality on the anaphor resolution process during sentence comprehension. The innovative contribution of the first paper consists in the analysis of the gender typicality effect in a grammatical gender language without the confounding interaction of grammatical gender cues. This is achieved through the implementation of a new paradigm based on role descriptions instead of role antecedents. The role descriptions had been empirically developed through a rigorous pre-test phase, which in turn included several experimental steps, leading to the final experimental materials. The second paper further tested the new paradigm in a different gender system, namely in a natural gender language. The replication of the experiment across different gender systems contributed to disambiguate contrasting hypotheses raised in previous literature about the asymmetric

effect of masculine and feminine anaphors, and to clarify the role of specific gender systems in the interpretation of gender cues. The original contribution of the second eye-tracking study of Paper 2 consists in a modulation of the typicality degree of the priming roles. This enabled to test the impact of slightly typical roles, which had been previously neglected in psycholinguist literature, and consequently re-define the cognitive format of the gender typicality effect. The described studies represent the core of the present work. As first author, I was principally responsible for the conception of the research questions and hypotheses, experimental designs, materials development, data collection, analysis, and manuscript writing. Paper 3 added to the experimental design the factor of grammatical gender, employing the classical role anaphor resolution paradigm. The original contribution of Paper 3 consists in the presentation of semantically different types of anaphors (pronominal versus noun phrases), which interestingly proved to affect the gender information processing. The fourth and fifth papers introduce a novel approach to test the effect of gender beliefs on language processing based on the on-line assignment of thematic roles (subject-agent/object-patient). The studies showed that readers tend to assign the role of `agent´ to male rather than female referents. These papers, including four eye-tracking experiments in two languages, have important implications in psycholinguistic theory, proposing the model of gender as a linguistic prominence feature. Moreover, they document through further paradigms the automatic and pervasive impact of gender beliefs on language processing and comprehension. As second author, I contributed to the development of the experimental designs, material selection, interpretation of the results, and to the revisions of different drafts of the manuscripts. The five papers composing the research plan are summarized in the following review.

## 2.1. Paper 1: Isolating typical gender in a grammatical gender language

The goal of the first study is the investigation of the effects of typical gender in German, a grammatical gender language. In order to disentangle the effect of typical gender from the effect of grammatical gender, classical role nouns were replaced through role descriptions which did not contain any grammatical gender cue to the referent gender. Grammatical gender-free role descriptions were created with the goal to test gender typicality in a comparable paradigm for natural- and grammatical gender languages, and to apply the anaphor resolution paradigm to grammatical gender languages without revealing the referent gender through the antecedent. Therefore, a rigorous construction of the experimental material represented a crucial aspect of the study. The material development presented several challenges. First, the descriptions aimed to describe a role in its meaning and gender typicality, without emphasizing arbitrary details that could convey typical gender cues. For example, describing a *cook* as a person who ‘stays among pots and pans all day long’ conveys a slightly female typicality, whereas a person who ‘works in a posh restaurant and prepares sophisticated dishes’ receives male typicality ratings<sup>3</sup>. Furthermore, materials should be relatively homogeneous in length and structure to prime a comparable amount of information. To meet these criteria, the experimental descriptions were created through a written production pre-test, where participants described role nouns in a standardized format. Further offline pre-tests checked the correspondence between descriptions and role nouns, the gender typicality ratings of the descriptions and the gender neutrality of the sentence containing the anaphor. The complex pre-test procedure ensured an empirical validation of the materials to be employed in the following eye-tracking experiment.

---

<sup>3</sup> Source: unpublished data collected in the pretests.

A further issue concerning the experimental material is the possibility that reading a role description may activate in the reader's mind the corresponding role noun with its grammatical gender markings. If this hypothesis is true, then these grammatical gender markings may have an influence on the anaphor resolution and confound the gender typicality effect. The first experiment of Paper 1 was designed to test the hypothesis of a possible activation of grammatical gender markings through the role descriptions. In a reaction time task, participants had to respond if a given role noun corresponded or not to the previously presented role description. In the crucial condition, male and female role descriptions were followed by semantically unrelated role nouns in masculine or feminine form. If the description activates the corresponding role noun with its grammatical gender markings, then there should be facilitation for target items sharing the grammatical gender primed by the role description, independently from the semantic content of the noun. Results showed no statistical difference for responses to masculine and feminine unrelated role nouns, suggesting that role descriptions did not elicit gender priming through grammatical gender markings. Therefore, they may represent a suitable tool to investigate typical gender in isolation from grammatical gender.

The second experiment of Paper 1 employed the pretested material in an eye-tracking study with the help of the anaphor resolution paradigm. The descriptions were combined with target sentences revealing the referent gender (*er/sie*, he/she). A disruption effect (mismatch effect) was expected in the case of incongruity between description typicality and pronoun gender. After the eye-tracking session, participants performed the Implicit Association Test *Gender-Career* and completed questionnaires on sexism and gender roles. Results showed the expected gender mismatch effect between gender typicality of the role description and referent gender in very early stages of sentence processing. Moreover, the integration of the anaphor in the mismatching conditions



resulted to be more problematic when the referent was male rather than female (masculine vs. feminine personal pronoun). We offer two possible explanations for this asymmetry. A first explanation relies on socio-cognitive theories and postulates that the perception of female professional roles changed in recent years to a bigger extent as compared to the perception of male professional roles. In other words, socio-psychological data may explain the fact that it is more problematic for readers to integrate a male referent in a counter-stereotypical context, than to fit a woman in a gender-untypical professional role. An alternative explanation of the results postulates that the role descriptions did activate the corresponding role noun. This hypothesis had been excluded by the reaction time experiment results. However, this conclusion was limited to semantically unrelated target items, and may not apply to the eye-tracking paradigm employing personal pronoun as target anaphors. If the descriptions do activate the corresponding role noun, then female descriptions would activate role nouns with feminine morphological markings (e.g., the suffix *\_in*), which are constraining for the referent gender. On the contrary, male description would activate role nouns in masculine form, which may be interpreted as generic in German. For this reason, it may be more difficult for the readers to integrate a male referent with the mismatching female description as compared to the alternative mismatching condition. An eventual replication of this finding in a natural gender language would allow to exclude the hypothesis based on linguistic features as a possible explanation for the mismatch effect asymmetry (see Paper 2).

Finally, no correlation emerged between the eye-tracking effect and the results of the offline individual measures on sexism and gender roles, confirming a discrepancy between beliefs on gender roles and the automatic activation of the gender mismatch effect measured through eye-movements.

## **2.2. Paper 2: Role descriptions induce gender mismatch effects in eye movements during reading**

The role description paradigm developed in paper 1 was further tested in Paper 2, where the experimental design was replicated in English, a natural gender language. The first study presents the same experimental design and a translation of the material of the previous eye-tracking experiment in German and aimed to test the effect of the role description gender typicality on the resolution of anaphoric pronouns. The role description paradigm was created to eliminate grammatical gender cues from the German material. Therefore, we expected the results obtained in the ‘artificially’ grammatical gender-free experiment to be comparable with the results of the experiment in the natural gender system language.

The analyses of the present study were based on mixed-effects (multilevel) modeling (see Baayen, Davidson & Bates 2008) instead of  $F_1$  and  $F_2$  ANOVAS (Clark 1973). Unlike general linear models, mixed-effects models are conducted on a trial level and include participants and items as crossed random effects. Moreover, mixed effects models are very robust with respect to missing data and allow better protection against capitalization on chance, or Type I error (Montefinese, Ambrosini, Fairfield & Mammarella, 2014).

The results of the English study replicated the disruption effect for the condition of mismatch between gender typicality of the priming sentence and anaphor gender. However, the effect emerged later (in first pass time) in the English study as compared to the German study (in first fixations time). A possible explanation of this result assumes that readers possessing a grammatical gender system process gender typicality cues differently in comparison to readers possessing a natural gender system. Specifically, the cognitive availability of a grammatical gender system may make typical gender cues more

salient, even if no grammatical gender cue is presented in the stimuli, and thus speed up the processing of the mismatching anaphor. The qualitative comparison between the two studies, however, presents some limitations. First, even though the samples of the two experiments possessed comparable characteristics such as age and education (students at universities of comparable ranking), possible cultural differences in the representation of some professional activities may have played a role in the different processing of the material. To allow a comparison between the two studies, we employed the same gender typicality categorization for the priming descriptions, thus based on ratings collected on the German sample. Follow-up typicality ratings were collected from the English sample immediately after the eye-tracking session. These ratings tended to be more skewed towards neutrality, particularly for female items. However, the previous exposure to gender mismatching material through the eye-tracking session may have influenced the rating task, thus constituting a sort of ‘cognitive training’ in counter-stereotypical gender role thinking. A further issue to be considered for the comparison of the two studies is the slightly different region segmentation of the target sentence in the two experiments, due to the different word order in the two languages. In fact, after the initial adverb of the target sentence, German requires the verb whereas English requires the anaphoric pronoun before the verb.

Despite the differences related to the linguistic features of the two languages, results showed in both experiments an asymmetry in the integration process of male and female referent. Also for the English study, it was easier for readers to integrate a female referent in the counter-stereotypical context, in comparison to fitting a male referent to the female context. This result may support the socio-cognitive explanation rather than the psycholinguistic one, since English does not possess a grammatical gender system that could trigger the difference for male and female roles.

The second experiment of Paper 2 originates from the consideration that the selection of highly gender stereotyped professions (e.g., *Schmied/in*, blacksmith) excluded from the experiment many interesting professions, such as *Forscher/in*, researcher, which had received typicality ratings between neutral and gender typical. The research question of experiment 2 tests whether roles rated as 'slightly gender typical' are able to elicit expectations on the referent gender, and thus to trigger a disruption in the process of anaphor resolution. Results showed that even roles that had been explicitly rated as slightly gender typical (that is, applicable almost to the same extent to male and female referents) were able to trigger a surprise effect in case of a mismatching anaphoric pronoun. The effect resulted to be modulated in comparison to the effect of highly typical roles, emerging only in early stages of sentence processing, whereas the effect of highly typical roles involved early as well as late stages of reading processing. Future studies may include different levels of gender typicality in one experiment to allow for a statistical comparison of the elicited mismatch effect. A direct modulation of the mismatch effect according to the typicality degree would indicate that the gender typicality effect is not an all-or-none process, requiring a specific threshold to emerge, but rather a gradual effect which can also be effective at low level of explicit typicality ratings.

### **2.3. Paper 3: Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution.**

The present study adds a further source of gender information to the antecedent in the anaphor resolution paradigm, namely grammatical gender. After examining in the previous papers the isolated effect of typical gender cues, the present paper considers the interaction of both gender information sources with respect to the time course of the effects during sentence processing in German.

The first eye-tracking study presents typically male, female and neutral role nouns followed by a pronominal anaphor, while the second experiment presented a semantically richer noun phrase as anaphor (example: *Oft hatte der Elektriker gute Einfälle, regelmässig plante er (exp. 1) /dieser Mann (exp. 2) neue Projekte*, Often had the electrician good ideas, regularly planned he (exp.1) /this man (exp. 2) new projects). Results showed a pervasive, reliable effect of grammatical gender agreement, with longer fixation time and higher probabilities of regressions for the conditions of mismatch between role noun grammatical gender and referent gender, for both experiments. Interestingly, the grammatical mismatch effect occurred in an earlier stage of sentence processing in the experiment presenting pronominal anaphors. This effect may be due to the processing of the additional semantic information conveyed by the noun phrase anaphor, which may have caused a delay in the mismatch effect.

Moreover, a mismatch effect of typical gender of the role noun antecedent emerged in both experiments. In experiment 1, the gender typicality effect involved the mismatching role noun region, whereas in the second experiment it affected the anaphor region. Importantly, the typical gender effect emerged in later stages of sentence processing as compared to the grammatical gender effect. These results fit the two stage model of reference resolution (Garrod & Sanford, 1995; Garrod & Terras, 2000). This model describes a first stage of resolution (linkage/bonding) which is influenced by lexical information only, and a second stage (verification/resolution) which can also employ semantic information already stored in memory. As proposed by the model, grammatical cues may have been processed first, whereas gender typical cues may have been recruited in a later stage. The two-stage model fits the processing of relatively complex material presenting both grammatical and stereotypical cues, and including not only mismatching conditions but also actual agreement violations (‘grammatical errors’). However, in the

previous experiments (see Paper 1 and 2) typical gender cues were processed immediately, in first pass (English study) or even first fixation time (German), supporting previous literature which documents an immediate and automatic activation of gender stereotypical information (Oakhill et al., 2005; Pyykkönen et al., 2010).

#### **2.4. Paper 4: Prominence of Gender Cues in the Assignment of Thematic Roles in German**

The studies described above analyzed the direct effect of gender typicality cues on the representation of the referent gender. The following two papers adopt instead an indirect approach to test the impact of gender information, namely they test whether different gender cues lead to a subtle preference for a specific syntactic structure ('linguistic bias', see Maass, Salvi, Arcuri, & Semin, 1989; Beukeboom, Finkenauer, & Wigboldus, 2010). The present study focuses on the assignment of thematic roles (agent vs. patient). In the study, the thematic role 'agent' is defined as the person performing the action conveyed by the verb, and is represented by the grammatical subject of the target sentence; the 'patient' is the person receiving the action and is represented by the direct object of the sentence. The study introduces the concept of prominence, according to which specific linguistic features make an item more likely to be interpreted as an agent. For example, *animacy* is a prominence feature, with animated items being more prominent, that is, more likely to be interpreted as agents performing an action, as compared to inanimate items (see Fillmore, 1968). The present study postulates that gender may also represent a prominence feature, with male items being more prominent, thus more likely to be interpreted as agents/subjects of the sentence as compared to female items. This hypothesis is based on socio-psychological theories, which describe the perception of male roles in association with characteristics such as agency, assertion, competence in acting,

while female roles are related to communion, warmth and empathy (see for example Fiske, Cuddy, & Glick, 2007; Koenig, Mitchell, Eagly, & Ristikari, 2011). To assess the eventual linguistic bias of gender, sentences are presented in an ambiguous syntactic formulation, which is solved only at the end of the relative sentence, as in the example: *Die Flugbegleiterin, die viele Touristen beobachtet hat/haben, ist aufmerksam* ‘The flight attendant, who has observed many tourists whom many tourists have observed, is attentive’. Only the auxiliary verb *hat/haben* (has/have) allows to establish if the role noun *Touristen* is the subject or the object of the relative clause. In the first experiment the role noun of the relative clause is presented either in masculine (e.g., *Touristen*, tourists<sub>masculine</sub>) or in feminine (e.g., *Touristinnen*, tourists<sub>feminine</sub>). If male cues are more prominent, that is more likely associated to the agent role, then it would be easier for the reader to interpret the masculine form as the subject of the relative clause, as compared to the feminine form, which should be more easily interpreted as the object. Results partially confirmed this hypothesis, showing that in subject-extracted relative clauses (i.e., with the auxiliary verb in singular form, *hat*) the verb was fixated longer after feminine rather than masculine role nouns. This finding suggests that it was more difficult for readers to interpret feminine role nouns as subjects, in comparison to masculine role nouns.

The second experiment introduced a manipulation of typical gender cues. While role nouns of the relative clause were presented in masculine or feminine form, and not marked for gender typicality, the role nouns of the main clause were presented in feminine form and could be neutral or typically female. The grammatical form of the main role noun was feminine, in order to exploit the ambiguous function of the German pronoun *die* (who/whom), which refers both to the singular and plural feminine form. The typical gender manipulation compared female and neutral roles, in order to avoid a possible gender mismatch that could have been produced by male typicality in association with a

feminine suffix on the initial role noun. Results showed that typical gender biased the assignment of agent/patient thematic roles in the expected direction.

In conclusion, the eye-tracking studies of Paper 4 support the hypothesis of an automatic bias produced by grammatical and typical gender cues in sentence comprehension. Results suggest that it was easier for readers to spontaneously interpret feminine/female roles as *patients* receiving an action rather than *agents* performing the action. The opposite result was found for masculine and neutral roles. These findings complement the results of social psychology studies, which document a condition of imbalance in the cognitive representation of male and female roles (e.g., Koenig et al., 2011).

#### **4.5. Paper 5: Gender Hierarchies in the Processing of Ambiguous French Anaphors**

The hypothesis of a linguistic bias elicited by gender cues was explored in a further grammatical gender language, French, in Paper 5. In French the pronoun *lui* (to him/her) is ambiguous with respect to gender reference. This feature offered the opportunity to construct an experimental design with the pronoun in cataphoric position, referring to a following role noun either in masculine or feminine form. This role noun represented the patient thematic role (the person receiving the action). A different role noun at the beginning of the sentence represented the subject thematic role (the person performing the action) and could vary in typical gender (female or neutral in experiment 1 and male or neutral in experiment 2) as in the example: *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoureux* ‘In fact, the dietician<sub>Female+fem</sub> recommended to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>, a strict plan’.



As in the previous German study, the hypothesis concerned grammatical and stereotypical gender cues and postulated that masculine and typically male cues would be more easily interpreted as agents/subjects, while feminine and typically female cues would be more easily interpreted as patients/objects. The results of the first experiment supported the hypothesis in the expected direction, for both grammatical and stereotypical gender cues. However, it could be claimed that the feminine grammatical gender of the first role noun may have primed the representation of a female referent for the interpretation of the pronoun *lui* and the corresponding second role noun. To test this alternative explanation the second experiment was designed, employing masculine role nouns as initial roles. The results of experiment 2 allowed the exclusion of the alternative hypothesis and confirmed the model of a linguistic bias triggered by gender cues.

As a whole, the present study represents a cross-linguistic validation of the hypothesis of gender as a prominence feature, with male cues ranking over female cues on the agency dimension.

### **3. Conclusions**

The present research contributed to further understand the impact of cognitive gender representations on language processing and comprehension on several aspects. From a psycholinguistic perspective, the research findings reported a fine-grained analysis of the anaphor resolution processes, documenting the time-course of gender stereotype activation and the mismatch effect modulation according to different paradigms (role nouns antecedents vs. grammatical-gender-free descriptions; pronominal vs. noun phrase anaphors). Moreover, the research explored the gender mismatch on a cross-linguistic level, drawing a comparison between natural and grammatical gender systems.

From a socio-cognitive perspective, the experiments proved that expectations about gender roles in professions can emerge in automatized behavior such as eye movements during reading, even when they are not present in more explicit measures such as classical questionnaires. The cognitive format of gender representations in professional roles is discussed; we propose a conceptual distinction between gender typicality, intended as the perception of the gender rates in the different roles, and gender stereotype, namely a more stable cognitive structure which may be largely implicit and automatically activated.

The impact of cognitive gender representation emerged also in the form of a linguistic bias in sentence comprehension. Studies showed that the syntactical process of thematic role assignment is affected by cognitive representations related to gender stereotypical beliefs, specifically the implicit association of 'man = agent'.

To conclude, language and cognitive representations have an indissoluble, reciprocal influence; at the same time, both factors can influence and are influenced by the surrounding social environment. Language may either contribute to the maintenance of existing stereotypes, or foster potential change (Maass & Arcuri, 1996). Based on the present findings, future studies may focus on possible cognitive trainings exploiting the specific characteristics of the language under study, to promote cognitive availability of counter-stereotypical gender role representations.

## References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Beukeboom, C. J., Finkenauer, C., & Wigboldus, D. H. J. (2010). The negation bias: When negations signal stereotypic expectancies. *Journal of Personality and Social Psychology*, *99*(6), 978-992.
- Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton, Jr. (Eds.), *The online study of sentence comprehension* (pp. 51-76). New York, NY: Psychology Press.
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, *49A*(3), 639-663.
- Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335-359.
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye-movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*(4), 551-559.
- Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, *29*, 781-803.
- Fillmore, J. (1968). The case for case. In Bach, E. & Harms, R. T. (Eds.) *Universals in linguistic theory* (pp. 1-88). New York: Holt, Rinehart, and Winston.

- Fiske, S.T., Cuddy, A., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Science, 11*, 77-83.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.
- Garnham, A., Doehren, S., & Gygax, P. (2015). True gender ratios and stereotype rating norms. *Frontiers in Psychology, 6*, 1023.
- Garrod, S. C., & Sanford, A. J. (1995). Incrementality in discourse understanding. In D. Milward & P. Sturt (Eds.), *Incremental interpretation: Edinburgh working papers in cognitive science (vol.11)*. Edinburgh: Centre for Cognitive Science, The University of Edinburgh.
- Garrod, S. C., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42*, 526-544.
- Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research, 36*, 431-456.
- Irmen, L., Holt, D. V. & Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research, 1353*, 133-144.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Allyn & Bacon.
- Kennison, S. M., Fernandez, E. C., & Bowers, J. M. (2009). Processing differences for anaphoric and cataphoric pronouns: Implications for theories of discourse processing. *Discourse Processes, 46*(1), 25-45.

- Kennison, S. M. & Trofe, J. L. (2003). Comprehending Pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32(3), 355-378.
- Koenig, A. M., Mitchell, A. A., Eagly, A. H., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616-642.
- Kreiner, H., Sturt, P., & Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58, 239-261.
- Maass, A., & Arcuri, L. (1996). Language and stereotyping. In N. Macrae, M. Hewstone, & C. Stangor (Eds.), *The foundations of stereotypes and stereotyping*. New York: Guilford.
- Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, 57(6).
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). Semantic significance: a new measure of feature salience. *Memory & Cognition*, 42(3), 355-369.
- Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, 33, 972-983.
- Pyykkönen, P., Hyönä, J., & van Gompel, R. P. (2010). Activating Gender Stereotypes During Online Spoken Language Processing. *Experimental Psychology*. 57(2), 126-133.
- Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Reali C., Esaulova, Y., & Von Stockhausen, L. (2014). Isolating stereotypical gender in

- a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics*, 36, 977–1006.
- Reali, C., Esaulova, Y., Öttl, A., & Von Stockhausen, L. (in press). Role descriptions induce gender mismatch effects in eye movements during reading. *Frontiers in Psychology*, 6, 1607.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, 105(1), 125.
- Sanford, A. J. (1985). *Cognition and cognitive psychology*. London: Weidenfeld & Nicolson.
- Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social communication* (pp. 163-187). New York: Psychology Press.
- Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PloS one*, 7(12).
- Siyanova-Chanturia, A., Warren, P., Pesciarelli, F., & Cacciari, C. (2015). Gender stereotypes across the ages: On-line processing in school-age children, young and older adults. *Frontiers in psychology*, 6, 1388.
- Vervecken, D. and Hannover, B. (2012), Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts perception of the speaker. *European Journal of Social Psychology*, 42, 754–762.

# Paper 1

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

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

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

## ADDRESS FOR CORRESPONDENCE

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

## ABSTRACT

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



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

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

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

## EXPERIMENT 1

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

### Method

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

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

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

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

### *Procedure*

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

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

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

### *Design and analysis*

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

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

Table 1. *Experiment 1 factorial structure and results*

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



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

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

## Results

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

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

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

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

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

### *Discussion*

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

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

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

## EXPERIMENT 2

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

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

### *Method*

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

### *Materials.*

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

(1) Description:

*M. F. repariert und stellt Möbel her, arbeitet mit Holz.*

“M. F. repairs and produces pieces of furniture, works with wood.”

Target sentence:

*Gewöhnlich hat er/sie ein ausreichendes Einkommen.*

“Usually he/she has a sufficient income.”

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

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

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

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

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

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

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

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

### *Procedure*

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

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

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



### *Design and hypotheses*

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

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

### *Results*

#### *Eye-tracking data.*

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

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

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

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

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

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

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

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

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

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



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

Measure	Region of Analysis	Effect	F1	F2
First fix. time	Pronoun	Typicality	$F = 2.026, p = .141$	$F = 1.943, p = .159$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F_{2,60} = 3.879, MSE = 466.06, p = .026$	$F_{2,33} = 3.526, MSE = 232.34, p = .041$
	Spillover	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F < 1$	$F < 1$
First pass time	Pronoun	Typicality	$F = 2.193, p = .120$	$F_{2,33} = 2.746, MSE = 1219.25, p = .079$
		Pronoun	$F = 2.607, p = .117$	$F = 1.378, p = .249$
		Typicality $\times$ Pronoun	$F_{2,60} = 2.68, MSE = 1709.95, p = .078$	$F = 1.256, p = .298$
	Spillover	Typicality	$F_{2,60} = 2.566, MSE = 2213.24, p = .085$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F < 1$	$F < 1$
Regression path	Pronoun	Typicality	$F = 1.162, p = .320$	$F < 1$
		Pronoun	$F < 1$	$F = 1.062, p = .310$
		Typicality $\times$ Pronoun	$F_{2,60} = 3.126, MSE = 3681.28, p = .051$	$F = 1.658, p = .206$
	Spillover	Typicality	$F_{2,60} = 3.532, MSE = 22700.63, p = .035$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F < 1$	$F < 1$
Total time	Pronoun	Typicality	$F = 1.293, p = .282$	$F = 1.014, p = .374$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F_{2,60} = 6.925, MSE = 4199.17, p = .002$	$F_{2,33} = 5.880, MSE = 1904.26, p = .007$
	Spillover	Typicality	$F = 1.782, p = .177$	$F < 1$
		Pronoun	$F = 2.780, p = .106$	$F_{2,33} = 4.518, MSE = 2011.32, p = .041$
		Typicality $\times$ Pronoun	$F < 1$	$F = 1.113, p = .341$
Regressions into region	Pronoun	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F_{2,60} = 3.006, MSE = 301.57, p = .057$	$F_{2,33} = 3.017, MSE = 119.37, p = .063$
	Spillover	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		Typicality $\times$ Pronoun	$F < 1$	$F < 1$

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

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

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

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

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

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

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

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

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

*Relating eye movements to individual measures.*

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

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

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

The IAT index showed that 29 participants out of 31 had a positive index, which indicates a stronger implicit association between the concepts of *Men* and *Career*, and between *Women* and *Family*. Two participants had a negative score, indicating the counterstereotypical tendency (stronger association between *Men* and *Family*, and *Women* and *Career*). For our sample, the mean IAT index (0.59,  $SD = 0.39$ ) was higher than the mean index reported by Nosek et al. (0.39,  $SD = 0.36$ ), which was averaged on a sample of 83,084 Gender–Career IATs collected on a publicly available website between 2002 and 2006 (Nosek et al., 2007). We analyzed possible covariation between the IAT index and eye-movement measures. As outlined above, the IAT index results from the subtraction of reaction times for the congruent block from reaction times for the incongruent block. For our study,

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

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

### *Discussion*

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

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



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

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

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

## GENERAL DISCUSSION

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

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

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

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

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

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

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

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

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



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

## APPENDIX A

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

### *Typically male roles*

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

### *Typically female roles*

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

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

### *Typically neutral roles*

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

### ACKNOWLEDGMENTS

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

### NOTES

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

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

## REFERENCES

- Bates, H., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics, 58*, 992–1004.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155–162.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*, 354–364.
- Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton Jr. (Eds.), *The online study of sentence comprehension* (pp. 51–76). New York: Psychology Press.
- Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64*, 153–170.
- Bußmann, H., & Hellinger, M. (2003). German—Engendering female visibility in German. In M. Hellinger & H. Bußmann (Eds.), *Gender across languages: The linguistic representation of women and men* (Vol. 3, pp. 141–174). Amsterdam: Benjamins.
- Cacciari, C., Carreiras, M., & Barbolini Cionini, C. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language, 37*, 517–532.
- Cacciari, C., & Cubelli, R. (2003). The neuropsychology of grammatical gender: An introduction. *Cortex, 39*, 377–382.
- Cacciari, C., & Padovani, R. (2007). Further evidence on gender stereotype priming in language: Semantic facilitation and inhibition on Italian role nouns. *Applied Psycholinguistics, 28*, 277–293.
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology, 49A*, 639–663.
- Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture-word interference paradigm. *Journal of Memory and Language, 53*, 42–59.
- Dieckman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin, 26*, 1171–1188.
- Duden, Volume 4 (1995). *Die Grammatik* [Grammar]. Mannheim: Dudenverlag.
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye-movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32*, 551–559.
- Eckes, T., & Six-Materna, I. (1998). Leugnung von Diskriminierung: Eine Skala zur Erfassung des modernen Sexismus [Denial of discrimination: A scale measuring modern sexism]. *Zeitschrift für Sozialpsychologie, 29*, 224–238.

- Eckes, T., & Six-Materna, I. (1999). Hostilität und Benevolenz: Eine Skala zur Erfassung des ambivalenten Sexismus [Hostility and benevolence: A scale for assessing ambivalent sexism]. *Zeitschrift für Sozialpsychologie*, *30*, 211–228.
- Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, *40*, 206–212.
- Glick, P., & Fiske, S. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, *23*, 464–485.
- Hamilton, D. L., & Trolie, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In S. L. Gaertner & J. F. Dovidio (Eds), *Prejudice, discrimination, and racism* (p. 133). New York: Academic Press.
- Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 393–398.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). Meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.
- International Labour Organization of the United Nations. (2000). *Employment for detailed occupational groups by sex, Germany*. Retrieved from <http://laborsta.ilo.org>
- Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*, 431–456.
- Irmen, L., Holt, D. V., & Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research*, *1353*, 133–144.
- Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498–509.
- Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *32*, 355–378.
- Kreiner, H., Sturt, P., & Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*, 239–261.
- Lattner, S., & Friederici, A. (2003). Talker's voice and gender stereotype in human auditory sentence processing—Evidence from event-related brain potentials. *Neuroscience Letters*, *339*, 191–194.
- Most, S. B., Verbeck Sorber, A., & Cunningham, J. G. (2005). Auditory Stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology*, *43*, 287–294.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*, 273–285.
- Pyykkönen, P., Hyönä, J., & van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye tracking. *Experimental Psychology*, *57*, 126–133.
- Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*, 121–149.

- Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26, 159–190.
- Schneider-Düker, M., & Kohler, A. (1988). Assessment of sex roles: Results of a German version of the Bem Sex-Role Inventory. *Diagnostica*, 34, 256–270.
- Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social communication: Frontiers of Social Psychology* (pp. 163–187). New York: Psychology Press.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68, 199–214.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.
- Van Gompel, R. P. G., & Livesedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 128–139.

# Paper 2



# Role descriptions induce gender mismatch effects in eye movements during reading

Chiara Reali<sup>1\*</sup>, Yulia Esaulova<sup>1</sup>, Anton Öttl<sup>2</sup> and Lisa von Stockhausen<sup>1</sup>

<sup>1</sup> Department of Psychology, University of Duisburg-Essen, Essen, Germany, <sup>2</sup> 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.

## OPEN ACCESS

### Edited by:

Antonino Vallesi,  
University of Padova, Italy

### Reviewed by:

Maria Montefinese,  
University of Padova, Italy  
Chiara Zanini,  
University of Padova, Italy

### \*Correspondence:

Chiara Reali  
chiara.reali@uni-due.de

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 31 May 2015

**Accepted:** 05 October 2015

**Published:** xx October 2015

### Citation:

Reali C, Esaulova Y, Öttl A and von  
Stockhausen L (2015) Role  
descriptions induce gender mismatch  
effects in eye movements during  
reading. *Front. Psychol.* 6:1607.  
doi: 10.3389/fpsyg.2015.01607

**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<sub>feminine</sub>). 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<sub>masculine/feminine</sub> 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<sub>feminine</sub>) 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<sup>1</sup> 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<sup>2</sup>.

<sup>1</sup>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).

<sup>2</sup>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 [ $M_{\text{male}} = 74.00$ ;  $M_{\text{female}} = 74.19$ ;  $M_{\text{neutral}} = 69.06$ ,  $F_{(2, 45)} = 0.86$ , *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<sup>3</sup>. 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)<sup>4</sup>. Contrast analyses revealed that the effect was statistically significant only when the priming sentence was female, with congruent trials being read faster,  $M_{\text{femaleHE}} = 302$ ,  $M_{\text{femaleSHE}} = 263$ ,  $t_{(948)} = 2.55$ ,  $p = 0.01$ ;  $M_{\text{maleHE}} = 257$ ,  $M_{\text{maleSHE}} = 269$ , *ns*;  $M_{\text{neutralHE}} = 269$ ,  $M_{\text{neutralSHE}} = 288$ , *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,  $M_{\text{HE}} = 295$ ,  $M_{\text{SHE}} = 269$ ,  $t_{(514)} = 2.35$ ,  $p = 0.002$  (pronoun region);  $M_{\text{HE}} = 457$ ,  $M_{\text{SHE}} = 407$ ,  $t_{(941)} = 2.14$ ,  $p = 0.03$  (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,  $M_{\text{neutralHE}} = 8.1$ ,  $M_{\text{neutralSHE}} = 13.2$ ,  $t_{(947)} = -2.26$ ,  $p = 0.02$ ;  $M_{\text{maleHE}} = 8.9$ ,  $M_{\text{maleSHE}} = 11.7$ , *ns*;  $M_{\text{femaleHE}} = 14.8$ ,  $M_{\text{femaleSHE}} = 11.2$ , *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,  $M_{\text{femaleSHE}} = 380$ ,  $M_{\text{femaleHE}} = 427$ ,  $t_{(998)} = 2.14$ ,  $p = 0.03$ ;  $M_{\text{maleHE}} = 363$ ,  $M_{\text{maleSHE}} = 355$ , *ns*;  $M_{\text{neutralHE}} = 437$ ,  $M_{\text{neutralSHE}} = 437$ , *ns*. Furthermore, a main effect of *participant gender* emerged on the pronoun region. Contrasts revealed a tendency for female participants to read faster,  $M_{\text{men}} = 355$ ,  $M_{\text{women}} = 316$ ,  $t_{(30)} = 1.86$ ,  $p = 0.073$ .

## 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),

<sup>3</sup>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.

<sup>4</sup>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)
<b>FIRST REGION</b>						
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
<b>PRONOUN REGION</b>						
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
<b>SPILLOVER REGION</b>						
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*
<b>FINAL REGION</b>						
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	
<b>FIRST REGION</b>						
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
<b>PRONOUN REGION</b>						
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
<b>SPILLOVER REGION</b>						
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
<b>FINAL REGION</b>						
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  $\beta$  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,  $M_{\text{male}}$ , pretest = 1.87,  $M_{\text{male}}$ , follow-up = 2.32,  $t_{(22)} = 2.88, p = 0.009$ ;  $M_{\text{female}}$ , pretest = 5.98,  $M_{\text{female}}$ , follow-up = 5.20,  $t_{(22)} = 4.20, p < 0.001$ ;  $M_{\text{neutral}}$ , pretest = 4.04,  $M_{\text{neutral}}$ , 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*,  $\text{Model} < \text{lmer}(\text{fixation\_time} \sim \text{typicality} * \text{pronoun} * \text{participant\_gender} * \text{region\_length} + (1 | \text{participants}) + (1 | \text{items}))$ .

## 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)
<b>FIRST REGION</b>						
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
<b>PRONOUN REGION</b>						
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
<b>SPILLOVER REGION</b>						
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
<b>FINAL REGION</b>						
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
	<b>Regression path</b>		<b>Regressions in</b>		<b>Regressions out</b>	
<b>FIRST REGION</b>						
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
<b>PRONOUN REGION</b>						
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
<b>SPILLOVER REGION</b>						
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
<b>FINAL REGION</b>						
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*<sub>pretest</sub> = 4.0, *M*<sub>follow-up</sub> = 4.1, *t*<sub>(70)</sub> = 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*<sub>s,male, pretest</sub> = 2.99, *M*<sub>s,male, follow-up</sub> = 3.34, *t*<sub>(22)</sub> = -2.86, *p* = 0.009; *M*<sub>s,female, pretest</sub> = 4.98, *M*<sub>s,female, follow-up</sub> = 4.68, *t*<sub>(22)</sub> = 2.20, *p* = 0.039; *M*<sub>neutral, pretest</sub> = 4.04, *M*<sub>neutral, follow-up</sub> = 4.16, *t*<sub>(22)</sub> = 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

The research was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under

grant agreement 237907. We are grateful to Alan Garnham and Jane Oakhill for their support at the time of data collection at the University of Sussex, UK. We also would like to thank David Tobinski for his help and valuable suggestions.

## 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/S014271641400010
- 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
- Sivanova-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

## SUPPLEMENTARY MATERIAL

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

**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.

Copyright © 2015 Reali, Esaulova, Öttl and von Stockhausen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Paper 3

## Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution

Yulia Esaulova\*, Chiara Reali and Lisa von Stockhausen

Department of Psychology, University of Duisburg-Essen, Essen, Germany

(Received 2 May 2012; final version received 30 March 2013)

Two eye-tracking studies addressed the processing of grammatical and stereotypical gender cues in anaphor resolution in German. The authors investigated pronominal (*er* ‘he’/ *sie* ‘she’) and noun phrase (*dieser Mann* ‘this man’/ *diese Frau* ‘this woman’) anaphors in sentences containing stereotypical role nouns as antecedents (Example: *Oft hatte der Elektriker gute Einfälle, regelmässig plante er/ dieser Mann neue Projekte* ‘Often had the electrician good ideas, regularly planned he/this man new projects’). Participants were native speakers of German ( $N = 40$  and  $N = 24$  in Experiments 1 and 2, respectively). Results show that influences of grammatical gender occur in early stages of processing, whereas the influences of stereotypical gender appear only in later measures. Both effects, however, strongly depend on the type of anaphor. Furthermore, the results provide evidence for asymmetries in processing feminine and masculine grammatical gender and are discussed with reference to two-stage models of anaphor resolution.

**Keywords:** anaphor resolution; grammatical gender; stereotypical gender; sentence processing

The central question of this paper concerns the processes involved in the comprehension of gender information encoded in German language. As in other grammatical gender languages, gender information can be conveyed both grammatically and conceptually (e.g., through stereotypical knowledge). Basic comprehension requires an integration of grammatical and conceptual gender cues and yet the mechanisms of this integration are not fully understood. Referential structures such as anaphors, which are commonly used in everyday utterances, illustrate the integration required. Consider, for instance, the following German sentence: *Ständig besuchte der Flugbegleiter<sub>FemaleMasc</sub> verschiedene Länder, vor allem bevorzugte er<sub>masc</sub> exotische Ziele* (The flight attendant<sub>FemaleMasc</sub> visited diverse countries all the time, most of all he<sub>masc</sub> preferred exotic destinations).<sup>1</sup> Understanding this sentence involves the integration of a conceptual component (stereotypically female role noun ‘flight attendant’) and a grammatical component (masculine grammatical gender of the role noun and the pronoun) in order to establish a link between the first and the second clause. Different combinations of such gender cues in a sentence may produce incongruities that, in turn, may cause comprehension difficulties and slow down reading. For example, the combination *der Flugbegleiter<sub>FemaleMasc</sub> – er<sub>masc</sub>* ‘The flight attendant<sub>FemaleMasc</sub> – he<sub>masc</sub>’ presents a reader with an incongruity between the stereotypically

female ‘flight attendant’ and the grammatically masculine pronoun ‘he’, even though these forms agree grammatically. Moreover, comprehension difficulties may already occur in the first clause upon reading the role noun ‘flight attendant’, due to the incongruity between female stereotypicality and the grammatically masculine gender of the role noun.

The interplay of stereotypical gender, grammatical forms and inferred biological gender makes person reference an especially interesting case for research. This is reflected in the variety of methods applied in previous research on personal nouns (e.g., reading times in Carreiras, Garnham, Oakhill, & Cain, 1996; ERP in Irmen, Holt, & Weisbrod, 2010; reaction times in priming tasks in Cacciari & Padovani, 2007; sentence evaluation in Gygax, Gabriel, Sarrasin, Garnham, & Oakhill, 2008) and the wide range of experimental materials used (text passages, isolated sentences, referential and non-referential contexts, etc.). Among other paradigms, earlier research has exploited anaphor resolution as a tool to reveal the mechanisms underlying language comprehension and has demonstrated the influence of grammatical and conceptual information on processing (Cacciari, Corradini, Padovani, & Carreiras, 2011; Carreiras, Garnham, & Oakhill, 1993; Duffy & Keir, 2004; Irmen, 2007; Kreiner, Sturt, & Garrod, 2008; Sturt, 2003). Thus, Duffy and Keir (2004) monitored participants’ eye movements

\*Corresponding author. E-mail: [yulia.esaulova@uni-due.de](mailto:yulia.esaulova@uni-due.de)



while they read sentences containing reflexive pronouns that referred to stereotypically male and female role nouns with or without preceding paragraph context. In Experiment 1, experimental conditions differed in using either feminine or masculine reflexive pronouns in sentences like *the babysitter found herself/himself humming while walking up to the door* and found a mismatch effect reflected in longer reading times, when reflexive pronouns were incongruent with the gender stereotype. In Experiment 2, they used the same target sentences but introduced either a disambiguating context (which explicitly stated that the character was a woman or a man) or a neutral context before them. The mismatch effect between reflexive pronouns and role nouns disappeared after a disambiguating context. The authors applied the lexical interpretation model (Foss & Speer, 1991; Hess, Foss, & Carroll, 1995) to explain the elimination of mismatch effects by the context, arguing that the pronoun is readily integrated into the discourse, despite the mismatch with gender stereotypes, when gender is already clearly instantiated.

A different theoretical perspective was applied by Sturt (2003), who reports the results of two eye-tracking studies investigating anaphoric references with reflexive pronouns in terms of Chomsky's binding theory (Chomsky, 1981). His Experiment 1 was based on paragraphs in which two characters were introduced as potential antecedents for the reflexive pronoun ('himself' or 'herself'). According to syntactic constraints identified by the binding theory, however, only the second character – a stereotypically male or female role noun – was a (grammatically) possible antecedent (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*). The study demonstrated an early effect of incongruity between the stereotypical gender of the grammatical antecedent and the anaphor. This finding supports Principle A of the binding theory, which explains constraints on the reference of reflexive and reciprocal anaphors by the same local domain of an anaphor and an antecedent (e.g., a clause) and their syntactic prominence. Even so, ungrammatical antecedents also affected processing, but only at a relatively later stage.

While Sturt (2003) suggests a specific time-course in the processing of various types of person-related information, other studies seek to define this differential processing further, based on the spatial location of occurring mismatches. Thus, Irmen (2007) used role nouns in the plural as a non-referential form of personal reference, denoting abstract generic categories, as opposed to references to specific persons ('Teachers often say that...' vs. 'The teacher often said that...'). First sentences in text passages introduced

non-referential role nouns of male, female and neutral conceptual gender (Experiment 1), while second sentences contained the anaphoric expression *diese Männer/diese Frauen* ('these men/these women'). Eye-tracking data showed that a conceptual mismatch, which was reflected in a slowdown in reading times, occurred immediately before and after the anaphor, whereas a grammatical mismatch occurred on the anaphoric noun itself.

The studies described above show that the discussion about conceptual vs. grammar-based influences is an important component in most of the research on stereotypical gender and anaphor resolution. Two aspects can be highlighted in this discussion. The first one regards seeing conceptual and grammatical influences as properties of anaphors and antecedents. While grammatical gender is obviously a grammatical property of a word, considering stereotypical gender a word property may require more of a theoretical reasoning. Kreiner et al. (2008) contrast theoretical accounts that assume stereotypical gender to be a lexical feature (Banaji & Hardin, 1996; Osterhout, Bersick, & McLaughlin, 1997) with others that propose considering stereotypical gender a conceptual feature, with stereotypes as a form of general world knowledge (Aitchison, 1994; Garnham, 2001). In their eye-tracking experiments, they used reflexive pronouns in anaphoric and cataphoric constructions to compare nouns characterised by stereotypical gender (e.g., *minister*) with nouns where gender information is part of the word definition – definitional gender nouns (e.g., *king*). They found similar mismatch costs for both types of nouns in anaphoric sentences, but in cataphoric sentences mismatch costs were found only for definitional gender nouns and not for stereotypical gender nouns. They conclude, therefore, that definitional gender is represented lexically, while stereotypical gender is not, a difference, which is reflected in effect strengths of syntactic constraints on these two types of gender.

The second major aspect in the discussion on conceptual vs. grammatical influences concerns the identification of processes involved in anaphor resolution, which are argued to be grammatical or conceptual in nature. According to the unification model proposed by Hammer, Jansma, Lamers, and Münte (2008), anaphors are resolved on the basis of either syntactic or semantic rules, depending on the specific constellation of antecedent characteristics (animate/inanimate) and the distance between antecedent and anaphor. Other models of anaphor resolution go yet further and attribute a specific time-course to conceptual and grammar-based rules involved in anaphor resolution (e.g., Garrod & Sanford, 1990; Marslen-Wilson, Tyler, & Koster, 1993; Trueswell, Tanenhaus, & Garnsey, 1994). Up until now, there seemed to be little



agreement on how early each of these processes takes place and how the processes are related to each other. Some of the models propose that grammatical effects take place early in the processing (Garrod & Sanford, 1990; Nicol & Swinney, 1989; Stewart, Pickering, & Sanford, 2000), and other models predict not only early but also simultaneous effects of both grammar-based and conceptual information (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell et al., 1994).

The experiments reported in this paper were designed to assess the respective influences of grammatical gender and stereotypical gender in the process of anaphor resolution. Since earlier research has consistently shown that gender mismatches lead to inflated reading times, due to a disruption of the reading process and the tendency to re-read (e.g., Frazier & Rayner, 1982; Staub, 2010), we chose eye-tracking as a method which offers detailed temporal and spatial information on the influence of different types of gender cues on language comprehension. In past, psycholinguistic research eye-tracking measures were found to reveal information associated with moment-to-moment cognitive processes, which offers a way of clarifying mechanisms that underlie language comprehension (for more technical details and background information on eye movements, see Rayner, 2009).<sup>2</sup>

The eye-tracking studies on anaphor resolution reported above mostly addressed gender processing in English and used reflexive pronouns as anaphors (Duffy & Keir, 2004; Kreiner et al., 2008; Sturt, 2003). Both of our experiments expand this area of research in that they address the processing of different gender cues involved in anaphor resolution in German (a grammatical gender language) and in that they investigate personal pronouns. In both experiments we used isolated sentences, as opposed to sentences embedded in a context (Duffy & Keir, 2004) or text passages (Sturt, 2003). In contrast to Irmen's (2007) study on gender cues in German, which used non-referential, generic role nouns in the plural, both of our studies contain role nouns which are used referentially in the singular. Moreover, the materials in both Experiments 1 and 2 are identical and differ only in the type of anaphor (personal pronoun vs. noun phrase). This provides an opportunity to manipulate grammatical and stereotypical gender cues and to observe gender effects as directly as possible while excluding possible confounding influences caused by differences in methodologies applied or significant variations in the materials.

The eye-tracking methodology reveals the following aspects of gender processing involved in reading comprehension: the exact timing (onset and duration) of the effects in the process of reading, spatial location of these effects (on a word-by-word basis),

differentiation between the processing of specific gender cues (i.e., grammatical vs. stereotypical, masculine vs. feminine), and the time course of integration of grammatical and stereotypical gender cues.

Anaphoric expressions of the type presented in both of our experiments require readers to integrate grammatical and conceptual features of anaphors and antecedents in order to allow a sensible interpretation of the sentences. Since pronoun anaphors only contain grammatical gender information, we expected the rules of grammatical gender agreement to dominate possible effects of antecedent stereotypicality. This would be expressed in the earlier onset of grammatical gender effects compared to stereotypical gender effects, which would be reflected in measures of early processing (i.e., first fixation durations and first pass), and their presence until measures of late processing (i.e., regressions in and out of regions, total fixation times) in Experiment 1. Considering that in sentences like *Oft hatte der Elektriker gute Einfälle, regelmässig plante dieser Mann/diese Frau neue Projekte* 'Often had the electrician good ideas, regularly planned this man/this woman new projects', noun phrases 'this man'/'this woman' represent semantically rich anaphors that comprise both conceptual and grammatical gender cues, we expected a search for an antecedent to be based on both types of gender cues. In this case, we would see stereotypical gender effects together with grammatical gender effects already in measures of early processing (i.e., first fixation durations and first pass) in Experiment 2. The results of the two eye-tracking experiments may inform models of reference resolution and prove new evidence for cognitive mechanisms surfacing in eye movement patterns and reading times.

### Experiment 1

In the first experiment we investigated the influence of grammatical and stereotypical gender cues on the process of anaphor resolution in sentences containing role noun antecedents and pronominal anaphors (*er* 'he' or *sie* 'she').

### Method

#### Participants

Forty native speakers of German (20 male and 20 female, mean age 24.8 years,  $SD = 3.9$ ) were paid to participate in Experiment 1. Most of them were students at the University of Heidelberg.

#### Materials

*Experimental stimuli.* Thirty-six experimental sentences were constructed using 12 stereotypically male, 12 stereotypically female and 12 neutral role nouns in

pretested neutral contexts (see Table A1). All of the sentences consisted of two clauses. Role nouns were introduced as antecedents in the first clause and were followed by anaphoric personal pronouns (*er*<sub>masc</sub> ‘he’ or *sie*<sub>fem</sub> ‘she’) in the second. Both clauses were presented simultaneously, with each clause occupying a separate line and line breaks after commas, as shown in Example (1).

(1)

*Oft hatte der Elektriker gute Einfälle,  
regelmässig plante er neue Projekte.*

‘Often had the electrician good ideas,  
regularly planned he new projects.’

(Word-by-word translation is given to render the word order of the original sentence).

All of the sentences had the following fixed structure: adverb + verb + role noun + adjective + noun, adverb + verb + personal pronoun + adjective + noun. Role nouns were presented either in masculine or feminine form. Their stereotypical and grammatical gender could be congruent or incongruent (e.g., *Elektriker*<sub>MaleMasc</sub> ‘[masculine] electrician’ vs. *Elektrikerin*<sub>MaleFem</sub> ‘[feminine] electrician’, *Kosmetiker*<sub>FemaleMasc</sub> ‘[masculine] beautician’ vs. *Kosmetikerin*<sub>FemaleFem</sub> ‘[feminine] beautician’). The pronoun in turn could agree with both the grammatical and the stereotypical gender of the antecedent, or with only one of the two gender cues.

Thus, the experimental design was fully crossed and included the two within-subjects and within-items factors of grammatical gender of the role noun (masculine or feminine) and pronoun gender (masculine or feminine) and one within-subjects but between-items factor of role noun stereotypicality (male, female, neutral). All experimental items were compiled in four randomised lists, which presented each item in one of the four conditions: (1) masculine antecedent + masculine anaphor, (2) masculine antecedent + feminine anaphor, (3) feminine antecedent + masculine anaphor and (4) feminine antecedent + feminine anaphor. Across lists, each item occurred equally often in each condition. Participants were presented with all four conditions and encountered each experimental item only once. To make sure that participants read the sentences carefully, about one third of the sentences (including fillers) were followed by comprehension questions with two alternative answer choices.

*Gender stereotypicality pretest.* Seventy-seven role nouns were selected on the basis of previously published stereotypicality ratings (Gabriel, Gyax, Sarra-sin, Garnham, & Oakhill, 2008). Fifty participants (all native speakers of German) were asked to rate these role nouns for gender stereotypicality on a seven-point scale (1 = stereotypically male, 7 = stereotypically fe-

male). To assess stereotypicality irrespective of grammatical gender, both grammatical forms were provided (i.e., masculine stems with a slash and hyphenated feminine endings: ‘Elektriker/-in’). Epicenes and role nouns with suffices indicating feminine gender (i.e., ‘-ess’, ‘-amme’) were excluded from the list. Twelve role nouns rated as stereotypically male (e.g., ‘astronaut’, ‘carpenter’; rating score of 2.5 or lower), 12 role nouns rated as stereotypically female (e.g., ‘beautician’, ‘babysitter’; rating score of 5.5 or higher) and 12 role nouns rated as neutral (e.g., ‘musician’, ‘writer’; rating score of 3.8–4.3) were used to construct the experimental items of Experiment 1. All of the 36 selected role nouns had received similar ratings from male and female participants and did not differ significantly in either length or frequency within male, female and neutral stereotypicality groups (see Table A2).<sup>3</sup>

*Context pretest.* We conducted a series of pretests that were designed to ensure that the context of the experimental sentences was neutral and did not suggest any gender stereotypicality in the absence of role nouns. This was important in order to exclude potential confounding effects that might result from the context and not the role noun. In the pretest, an X replaced role nouns and pronouns that served as subjects in each of the two clauses of the experimental sentences. These sentences were presented together with others, which had stereotypically male and stereotypically female contexts and served as fillers in this pretest. The pretest materials were compiled in two lists to prevent effects of item order. Thirty participants (all native speakers of German) were asked to rate these sentences on a scale from 1 to 7 (1 = stereotypically male, 7 = stereotypically female). They received course credit or a candy bar for their participation. Contexts with ratings from 3.5 to 4.5 were selected for the main study.

*Fillers.* To prevent participants from developing expectations of gender-related incongruities, we constructed 72 filler sentences. In addition to the role nouns used in the experimental items, 24 slightly male (rating score: 2.5–3.4), 24 slightly female (rating score: 4.6–5.5) and 24 neutral (rating score: 3.5–4.5) role nouns were selected to create filler sentences. Like experimental sentences, fillers consisted of two clauses and had a fixed structure similar to that of the experimental sentences. The role noun was introduced in the first clause: in the second clause, there was either a demonstrative pronoun which referred to the object of the first clause (50% of the sentences, e.g., *Häufig stellte der Psychiater tiefgehende Fragen, meistens brachten diese wichtige Erkenntnisse* ‘Often posed the psychiatrist profound questions, mostly yielded these important knowledge’) or a second subject (e.g., *Fast immer hielt der Politiker fabelhafte Reden, daher gewann*

*die Partei viele Wahlen* ‘Almost always gave the politician fabulous speeches, as a result won the party many elections’). Twenty-five per cent of the filler items contained grammatical violations that imitated incongruities in the experimental sentences.

### Procedure

Eye movements were monitored by a video-based head-mounted eye-tracking system (EyeLink II) that sampled pupil location with a sampling rate of 250 ms. The experiment was implemented using the Eye-Track software provided by the Department of Psychology at the University of Massachusetts Amherst.<sup>4</sup> The experimental stimuli were presented in Lucida Console 22 point font on the screen, which was located at a distance of 70 cm from the participants’ eyes. Viewing was binocular with eye location being recorded from the dominant eye.

Participants were tested individually and used a chinrest during the whole experiment. Before the experimental session began, participants were instructed to read the sentences for comprehension in their normal reading speed. To move to the next sentence and to answer content-related questions, participants had to press corresponding buttons on the keypad. The first three filler sentences served as practice trials. Each session started with a calibration and validation procedure after the eye-tracker was adjusted to the head and eyes of the participants. At the beginning of each trial the participants had to focus on a black rectangle. The sentence appeared only after the rectangle was fixated accurately enough, starting at the exact point of the rectangle location. Whenever fixations were too inaccurate to trigger the next item, calibration and validation were repeated. The experiment lasted approximately 30 minutes.

## Results

### Data analysis

All of the experimental sentences consisted of two clauses and had the following fixed structure (analysed regions of interest are marked with <brackets>): *Oft hatte <der Elektriker or die Elektrikerin> <gute Einfälle>, regelmäßig <plante> <er or sie neue> <Projekte>* ‘Often had <the electrician, masc. or fem.> <good ideas>, regularly <planned> <he or she new> <projects>’. In the first clause, the determiner plus *role noun* as well as the following adjective and noun (*role noun spillover*) served as regions of analysis. In the second clause, the regions of analysis were the *verb* preceding the pronoun (as a possible launching position before skipping the pronoun), the pronoun together with the following adjective (*pronoun region*), and the noun following the pronoun (*pronoun*

*spillover*). Initial stages of data analysis were carried out using EyeDoctor and EyeDry software provided by the Department of Psychology at the University of Massachusetts Amherst. Short fixations (below 70 ms) were merged with neighbouring fixations within three characters. Following Rayner, Sereno, Morris, Schmauder, and Clifton (1989) and the current practice in eye-tracking research (e.g., Breen & Clifton, 2011), we assume fixations below 70 ms and above 600 ms in the regions of interest not representative of normal acquisition of information by the reader. These fixations had been removed (3.7% of the data) before further statistical analyses were performed. Computations reported below are based on the data averaged across participants ( $F_1$ ) and items ( $F_2$ ). The analyses were based on residual fixation times after correction for region lengths (Trueswell et al., 1994).

Fixation times are reported for five measures that reflect early, late and intermediate stages of processing. *First fixation durations* reflect the durations of the very first fixation on a region of interest that is entered from the left. *First-pass reading time* is the sum of all fixations from first entering a region from the left until leaving it for the first time, either to the left or to the right. First fixation durations and first-pass reading time are identical in case of a single fixation on a region during the first reading. *Regression path* time reflects fixation time from first entering a region until leaving it to the right including regressions to previous regions. *Total fixation times* reflect the time spent on a region including re-reading and excluding regressions from this region. *Regressions into a region* are defined as the probability of regressing into a region of interest (i.e. entering from the right) (see Boland, 2004; Mitchell, Shen, Green, & Hodgson, 2008).

The basic design of the reported analyses is a 2 (grammatical gender of the role noun: masculine/feminine)  $\times$  2 (grammatical gender of the pronoun: masculine/feminine)  $\times$  3 (role name stereotypicality: male/female/neutral) analysis of variance (ANOVA) with grammatical gender of both the role noun and the pronoun as within-subjects factors and stereotypicality as a within-subjects and a between-items factor. Means and standard deviations of residual fixation times and probabilities of regressions for all measures and all regions of interest are given in Table 1. Results of analyses of variance are shown in Table 2.

Below we report and interpret results that were reliable in both the analysis by subjects ( $F_1$ ) and the analysis by items ( $F_2$ ) or reliable in one ( $p \leq .05$ ) and marginally reliable in the other analysis ( $p \leq .10$ ) with similar patterns of mean differences. Results of the reported *t*-tests are based on data averaged across participants and were significant at  $p < .05$  unless otherwise stated. Analyses of variance with the

Table 1. Means (and standard deviations) of residual fixation times and probabilities of regressions (Experiment 1).

Region	Factor			Measure <sup>a</sup>					
	Typicality (RN)	Grammatical gender (RN)	Pronoun	FF	FP	RP	RI	TT	
Role noun	Male	Masculine	Masculine	2.46 (46.33)	15.95 (152.05)	-14.81 (166.36)	33.33 (30.19)	-15.60 (326.73)	
			Feminine	-3.22 (36.45)	15.47 (168.48)	-34.75 (175.34)	36.67 (31.85)	-44.50 (281.19)	
	Female	Masculine	Masculine	-2.22 (42.89)	-1.85 (158.41)	27.21 (244.50)	39.17 (31.02)	122.90 (408.51)	
			Feminine	0.32 (33.64)	10.78 (179.05)	4.10 (180.97)	39.17 (31.93)	-25.69 (251.30)	
		Feminine	Masculine	3.47 (41.43)	15.95 (152.05)	35.37 (262.56)	29.17 (32.19)	33.10 (420.69)	
			Feminine	10.08 (39.62)	15.47 (168.48)	39.57 (232.51)	41.67 (33.55)	109.42 (435.93)	
	Neutral	Masculine	Masculine	-0.78 (32.35)	-1.85 (158.41)	11.05 (209.00)	40.83 (36.58)	39.66 (298.83)	
			Feminine	5.18 (43.34)	10.78 (179.05)	29.92 (258.44)	20.00 (27.01)	-44.27 (378.39)	
		Feminine	Masculine	-0.07 (38.86)	-30.88 (127.43)	-39.08 (143.65)	25.83 (30.65)	-95.12 (201.96)	
			Feminine	-4.87 (31.20)	-21.30 (120.00)	-39.24 (124.35)	33.33 (30.19)	-32.55 (231.96)	
	RN spillover	Male	Masculine	Masculine	-2.57 (31.29)	-3.00 (152.48)	-20.70 (178.79)	43.33 (37.13)	26.37 (337.69)
				Feminine	-8.21 (36.84)	-27.41 (119.45)	-11.17 (228.66)	27.50 (33.66)	-73.90 (280.86)
Female		Masculine	Masculine	-11.64 (45.91)	-37.61 (198.80)	-5.35 (260.10)	7.50 (15.99)	-31.48 (266.21)	
			Feminine	-7.89 (38.55)	-16.77 (161.75)	-64.90 (238.00)	15.00 (23.81)	-14.04 (292.37)	
		Feminine	Masculine	1.16 (40.15)	4.65 (151.38)	24.78 (356.63)	10.83 (17.52)	28.26 (372.16)	
			Feminine	-8.13 (38.95)	23.26 (188.92)	39.82 (327.79)	9.17 (18.47)	-2.71 (320.96)	
Neutral		Masculine	Masculine	-1.10 (38.37)	-4.44 (183.37)	-5.95 (274.94)	6.67 (15.47)	19.52 (420.87)	
			Feminine	3.78 (58.85)	18.07 (186.43)	105.78 (585.30)	12.50 (20.93)	18.46 (357.26)	
		Feminine	Masculine	-5.77 (42.76)	14.03 (199.40)	-30.65 (214.28)	10.00 (18.80)	-2.88 (327.43)	
			Feminine	11.33 (56.29)	17.33 (207.05)	-4.18 (402.88)	11.67 (19.32)	-10.22 (393.93)	
Verb		Male	Masculine	Masculine	5.67 (56.37)	16.32 (188.42)	44.26 (386.28)	10.00 (20.25)	1.24 (307.88)
				Feminine	6.79 (45.03)	-11.39 (181.31)	-13.24 (287.07)	13.33 (23.63)	4.91 (305.95)
	Female	Masculine	Masculine	-4.64 (37.88)	-4.71 (166.63)	-42.70 (273.80)	10.00 (15.47)	4.48 (282.87)	
			Feminine	12.01 (51.53)	-0.45 (175.51)	-1.81 (295.12)	5.83 (16.69)	-19.66 (296.83)	
		Feminine	Masculine	13.46 (35.18)	27.27 (66.08)	29.54 (201.18)	15.83 (22.63)	-10.91 (112.74)	
			Feminine	15.09 (48.48)	16.85 (69.27)	21.25 (103.66)	26.67 (25.26)	31.76 (112.74)	
	Neutral	Masculine	Masculine	12.51 (49.70)	15.91 (71.54)	8.52 (72.17)	23.33 (24.11)	21.27 (165.86)	
			Feminine	15.73 (43.74)	4.93 (52.78)	90.94 (250.09)	15.00 (21.28)	-20.55 (114.32)	
		Feminine	Masculine	-9.77 (36.65)	-15.98 (48.50)	-33.19 (100.45)	9.17 (16.86)	-67.32 (107.28)	
			Feminine	6.70 (57.64)	3.38 (66.97)	3.25 (113.23)	27.50 (26.03)	7.36 (115.81)	
	Pronoun	Male	Masculine	Masculine	-6.51 (38.87)	-16.22 (54.23)	-4.74 (89.52)	24.17 (26.14)	17.97 (145.41)
				Feminine	-6.76 (40.19)	-13.70 (48.03)	-13.91 (93.85)	19.17 (26.03)	-21.09 (145.41)
Female		Masculine	Masculine	-5.76 (34.26)	2.33 (65.72)	-27.23 (53.00)	17.50 (23.86)	-15.95 (147.57)	
			Feminine	0.93 (46.84)	1.61 (72.39)	-41.58 (54.90)	23.33 (28.44)	6.72 (138.55)	
		Feminine	Masculine	-8.94 (40.98)	-12.22 (61.24)	-39.93 (66.43)	23.33 (22.90)	12.25 (199.14)	
			Feminine	8.08 (51.31)	18.04 (70.99)	-38.51 (68.33)	16.67 (23.87)	-1.18 (138.85)	
Neutral		Masculine	Masculine	3.22 (31.62)	-6.29 (104.20)	-188.32 (199.97)	21.67 (24.52)	-75.51 (159.85)	
			Feminine	4.33 (40.24)	33.38 (151.57)	165.25 (791.56)	22.50 (24.33)	89.76 (312.09)	
		Feminine	Masculine	7.10 (51.06)	15.86 (155.75)	256.93 (951.57)	18.33 (21.28)	43.72 (243.32)	
			Feminine	0.52 (44.96)	-10.51 (139.24)	-117.42 (400.40)	25.00 (31.80)	-44.54 (215.62)	
Female		Masculine	Masculine	3.05 (38.90)	-14.72 (112.58)	-163.49 (286.84)	28.33 (26.74)	-72.55 (225.86)	
			Feminine	3.39 (60.51)	-26.23 (145.84)	-9.73 (459.16)	22.50 (27.62)	4.25 (288.33)	
	Feminine	Masculine	3.41 (50.40)	11.47 (151.75)	49.45 (775.38)	26.67 (30.38)	41.02 (239.13)		
		Feminine	5.39 (51.12)	-40.30 (119.66)	-83.37 (577.18)	23.33 (26.37)	-80.97 (206.80)		



Table 1 (Continued)

Region	Factor		Measure <sup>a</sup>						
	Typicality (RN)	Grammatical gender (RN)	Pronoun	FF	FP	RP	RI	TT	
Neutral		Masculine	Masculine	-8.31 (34.12)	-24.91 (114.60)	-107.90 (390.40)	29.17 (27.41)	-42.75 (184.79)	
		Feminine	Feminine	-9.13 (36.87)	32.54 (188.13)	116.53 (510.03)	19.17 (24.91)	109.84 (363.40)	
Pronoun spillover	Male	Masculine	Masculine	-5.87 (44.54)	-19.50 (119.43)	-29.56 (666.76)	26.67 (25.26)	36.63 (229.61)	
		Feminine	Feminine	-10.54 (38.32)	-19.50 (119.43)	-29.56 (666.76)	28.33 (26.74)	-20.61 (224.79)	
Female		Masculine	Masculine	-7.76 (49.41)	0.57 (170.37)	-88.63 (673.34)	/	-26.25 (168.77)	
		Feminine	Feminine	5.69 (70.21)	-7.69 (142.31)	84.27 (953.24)	/	3.70 (165.79)	
Neutral		Masculine	Masculine	7.08 (85.50)	20.37 (136.87)	8.67 (770.95)	/	20.96 (197.06)	
		Feminine	Feminine	-22.52 (58.47)	-25.86 (137.48)	62.21 (1328.49)	/	-11.99 (216.55)	
Neutral		Masculine	Masculine	7.57 (70.61)	-25.45 (134.32)	-37.52 (945.84)	/	-50.38 (138.89)	
		Feminine	Feminine	-2.71 (65.27)	10.16 (162.88)	149.02 (945.06)	/	49.34 (192.40)	
Neutral		Masculine	Masculine	6.72 (87.37)	14.05 (182.16)	133.61 (682.94)	/	35.51 (205.29)	
		Feminine	Feminine	-3.62 (62.24)	-15.61 (165.28)	-78.97 (761.40)	/	-43.85 (238.31)	
Neutral		Masculine	Masculine	-3.43 (69.47)	-16.85 (149.05)	-97.60 (629.34)	/	-26.67 (166.91)	
		Feminine	Feminine	-1.10 (65.58)	3.44 (156.67)	42.85 (733.91)	/	24.66 (199.84)	
Neutral		Masculine	Masculine	3.45 (63.06)	0.80 (143.54)	45.01 (604.96)	/	24.49 (204.97)	
		Feminine	Feminine	-10.39 (67.48)	-6.95 (157.36)	-82.46 (949.72)	/	-34.59 (174.28)	

<sup>a</sup>FF: first fixation durations, FP: first-pass reading times, RP: regression path times, RI: regressions into the region, TT: total fixation times.

experimental factors mentioned above were performed separately for each of the regions of interest.

*First fixation durations.* The first relevant effect was found on the *verb* region.<sup>5</sup> The ANOVA revealed a main effect of pronoun gender, with shorter fixations on the verb preceding masculine than feminine pronouns,  $M_{\text{Masc}} = -4.36$ ,  $M_{\text{Fem}} = 2.23$ ,  $t(39) = -1.98$ ,  $SEM = 3.32$ ,  $p = .054$ .

*First-pass reading time.* A reliable interaction effect on the *pronoun* was found between the grammatical gender of the role noun and pronoun gender, with shorter fixations on pronouns following grammatically congruent compared to grammatically incongruent role nouns. Masculine pronouns were fixated shorter than feminine pronouns after masculine role nouns,  $M_{\text{Masc/masc}} = -15.04$ ,  $M_{\text{Masc/fem}} = 14.95$ ,  $t(39) = -2.09$ ,  $SEM = 14.39$ . Feminine pronouns were fixated shorter than masculine pronouns after feminine role nouns,  $M_{\text{Fem/masc}} = 15.43$ ,  $M_{\text{Fem/fem}} = -25.05$ ,  $t(39) = 3.16$ ,  $SEM = 12.80$ .

The same interaction – between grammatical gender of the role noun and pronoun gender – was found in the *pronoun spillover* region. After masculine role nouns, *pronoun spillover* was fixated equally long irrespective of pronoun gender,  $M_{\text{Masc/masc}} = -11.04$ ,  $M_{\text{Masc/fem}} = 4.56$ ,  $t(37) = -0.96$ , ns. Following feminine role nouns, however, the spillover was fixated shorter after feminine compared to masculine pronouns,  $M_{\text{Fem/masc}} = 12.57$ ,  $M_{\text{Fem/fem}} = -17.62$ ,  $t(37) = 1.76$ ,  $SEM = 17.18$ ,  $p = .087$ .

*Regression path time.* The first reliable effects were found in the *pronoun* region. Again the ANOVA revealed an interaction between the grammatical gender of the role noun and pronoun gender. As with first-pass reading times, both masculine and feminine pronouns were fixated shorter after a grammatically congruent than a grammatically incongruent antecedent,  $M_{\text{Masc/masc}} = -153.39$ ,  $M_{\text{Masc/fem}} = 84.25$ ,  $t(39) = -4.93$ ,  $SEM = 48.19$ ;  $M_{\text{Fem/masc}} = 138.83$ ,  $M_{\text{Fem/fem}} = -76.06$ ,  $t(39) = 3.67$ ,  $SEM = 58.53$ .

The same interaction – between grammatical gender of the role noun and pronoun – was revealed in the *pronoun spillover* region. The spillover region was fixated shorter when the role noun antecedent was grammatically congruent with the pronoun than when the two were incongruent,  $M_{\text{Masc/masc}} = -66.2$ ,  $M_{\text{Masc/fem}} = 99.78$ ,  $t(39) = -2.01$ ,  $SEM = 82.64$ ;  $M_{\text{Fem/masc}} = 85.77$ ,  $M_{\text{Fem/fem}} = -79.39$ ,  $t(39) = 2.11$ ,  $SEM = 78.22$ .

*Total fixation times.* Once more the ANOVA revealed an interaction effect between the grammatical gender of the role noun and the pronoun on the *role noun* region. Masculine role nouns were fixated equally long

Table 2. Results of analyses of variance for all regions of interest (Experiment 1).

Measure <sup>a</sup>	Region	Effect	$F_1$	df <sub>1, 2</sub>	$F_2$	df <sub>1, 2</sub>	
FF	Role noun	RN typicality (T)	2.88*	2, 78	1.82	2, 33	
		RN grammatical gender (GG)	1.54	1, 39	1.24	1, 33	
		Pronoun (P)	<1		<1		
		T × GG	<1		<1		
		T × P	1.48	2, 78	1.70	2, 33	
		GG × P	<1		<1		
	Role noun spillover	T × GG × P	<1		<1		
		RN typicality (T)	4.50***	2, 78	1.76	2, 33	
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	4.92**	1, 39	2.44	1, 33	
		T × GG	<1		<1		
		T × P	1.43	2, 78	1.15	2, 33	
	Verb	GG × P	<1		<1		
		T × GG × P	1.57	2, 78	1.76	2, 33	
		RN typicality (T)	10.50***	2, 70	5.35***	2, 33	
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	3.95*	1, 35	5.38**	1, 33	
		T × GG	<1		<1		
	Pronoun	T × P	<1		<1		
		GG × P	<1		<1		
		T × GG × P	<1		1.17	2, 33	
		RN typicality (T)	7.86***	2, 78	5.78***	2, 33	
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	<1		<1		
	Pronoun spillover	T × GG	<1		<1		
		T × P	<1		<1		
		GG × P	<1		<1		
		T × GG × P	<1		<1		
		RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	<1		<1		
	FP	Role noun	Pronoun (P)	2.35	1, 37	1.18	1, 33
			T × GG	<1		<1	
			T × P	<1		<1	
			GG × P	2.35	1, 37	1.81	1, 33
			T × GG × P	<1		1.25	2, 33
			RN typicality (T)	6.42***	2, 78	1.92	2, 33
	Role noun spillover	RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	<1		<1		
		T × GG	1.25	2, 78	1.48	2, 33	
		T × P	<1		<1		
		GG × P	<1		<1		
		T × GG × P	1.17	2, 78	<1		
	Verb	RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	2.17	1, 39	3.02*	1, 33	
		Pronoun (P)	<1		<1		
T × GG		1.86	2, 78	1.77	2, 33		
T × P		<1		<1			
GG × P		<1		<1			
Pronoun	T × GG × P	<1		<1			
	RN typicality (T)	8.12***	2, 70	5.64***	2, 33		
	RN grammatical gender (GG)	1.50	1, 35	1.36	1, 33		
	Pronoun (P)	<1		1.43	1, 33		
	T × GG	<1		<1			
	T × P	2.94*	2, 70	1.97	2, 33		
Verb	GG × P	<1		<1			
	T × GG × P	1.03	2, 70	2.34	2, 33		
	RN typicality (T)	2.32	2, 78	<1			
	RN grammatical gender (GG)	<1		<1			
	Pronoun (P)	<1		<1			
	T × GG	<1		<1			
Pronoun	T × P	1.38	2, 78	1.68	2, 33		
	GG × P	11.89***	1, 39	11.50***	1, 33		
	T × GG × P	<1		<1			

Table 2 (Continued)

Measure <sup>a</sup>	Region	Effect	$F_1$	$df_{1, 2}$	$F_2$	$df_{1, 2}$
FP	Pronoun spillover	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	<1		<1	
		T × GG	<1		<1	
		T × P	<1		<1	
		GG × P	3.10*	1, 37	4.26**	1, 33
		T × GG × P	<1		<1	
RP	Role noun	RN typicality (T)	9.03***	2, 78	2.42	2, 33
		RN grammatical gender (GG)	1.78	1, 39	1.50	1, 33
		Pronoun (P)	<1		<1	
		T × GG	1.85	2, 78	1.77	2, 33
		T × P	<1		<1	
		GG × P	<1		<1	
		T × GG × P	<1		<1	
	Role noun spillover	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	<1		<1	
		T × GG	1.80	2, 78	3.69**	2, 33
		T × P	1.12	2, 78	1.32	2, 33
		GG × P	<1		<1	
		T × GG × P	1.74	2, 78	1.56	2, 33
	Verb	RN typicality (T)	9.71***	2, 46	<1	
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	1.31	1, 23	1.18	1, 33
		T × GG	<1		<1	
		T × P	1.04	2, 46	4.14**	2, 33
		GG × P	<1		<1	
		T × GG × P	2.92	2, 46	<1	
Pronoun	RN typicality (T)	2.54*	2, 78	<1		
	RN grammatical gender (GG)	<1		1.96	1, 33	
	Pronoun (P)	<1		<1		
	T × GG	<1		<1		
	T × P	<1		<1		
	GG × P	27.56***	1, 39	15.05***	1, 33	
	T × GG × P	3.77**	2, 78	1.92	2, 33	
Pronoun spillover	RN typicality (T)	<1		1.57	2, 33	
	RN grammatical gender (GG)	<1		<1		
	Pronoun (P)	<1		<1		
	T × GG	<1		<1		
	T × P	<1		<1		
	GG × P	4.07*	1, 36	5.60**	1, 33	
	T × GG × P	<1		<1		
TT	Role noun	RN typicality (T)	5.99***	2, 78	1.95	2, 33
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	2.71	1, 39	2.28	1, 33
		T × GG	6.77***	2, 78	3.19*	2, 33
		T × P	2.15	2, 78	1.21	2, 33
		GG × P	11.65***	1, 39	8.47***	1, 33
		T × GG × P	<1		<1	
	Role noun spillover	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	<1		<1	
		T × GG	1.02	2, 78	<1	
		T × P	<1		<1	
		GG × P	<1		<1	
		T × GG × P	<1		<1	
	Verb	RN typicality (T)	2.15	2, 76	<1	
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	<1		<1	
		T × GG	1.05	2, 76	<1	
		T × P	<1		<1	
		GG × P	17.53***	1, 38	10.40***	1, 33
		T × GG × P	1.26	2, 76	1.17	2, 33



Table 2 (Continued)

Measure <sup>a</sup>	Region	Effect	$F_1$	df <sub>1, 2</sub>	$F_2$	df <sub>1, 2</sub>	
TT	Pronoun	RN typicality (T)	3.06*	2, 78	1.07	2, 33	
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	1.10	1, 39	<1		
		T × GG	<1		<1		
		T × P	2.03	2, 78	<1		
		GG × P	29.03***	1, 39	26.63***	1, 33	
		T × GG × P	<1		<1		
	Pronoun spillover	RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	<1		<1		
		T × GG	<1		<1		
		T × P	<1		<1		
		GG × P	13.84***	1, 37	14.74***	1, 33	
		T × GG × P	1.61	2, 74	<1		
RI	Role noun	RN typicality (T)	1.69	2, 78	<1		
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	<1		<1		
		T × GG	2.44*	2, 78	2.09	2, 33	
		T × P	<1		<1		
		GG × P	14.85***	1, 39	20.28***	1, 33	
		T × GG × P	4.94***	2, 78	3.97**	2, 33	
		Role noun spillover	RN typicality (T)	<1		<1	
	RN grammatical gender (GG)		<1		<1		
	Pronoun (P)		1.49	1, 39	1.34	1, 33	
	T × GG		<1		1.13	2, 33	
	T × P		<1		<1		
	GG × P		3.22*	1, 39	6.28**	1, 33	
	T × GG × P		<1		<1		
	Verb		RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	<1		<1		
		Pronoun (P)	1.66	1, 39	1.12	1, 33	
		T × GG	<1		<1		
		T × P	1.38	2, 78	<1		
		GG × P	19.56***	1, 39	23.29***	1, 33	
		T × GG × P	<1		<1		
		Pronoun	RN typicality (T)	1.24	2, 78	<1	
	RN grammatical gender (GG)		<1		<1		
	Pronoun (P)		<1		<1		
	T × GG		<1		<1		
	T × P		1.13	2, 78	1.32	2, 33	
	GG × P		1.94	1, 39	1.96	1, 33	
	T × GG × P		<1		<1		
RI the role noun region	Role noun spillover		RN typicality (T)	1.18	2, 78	<1	
			RN grammatical gender (GG)	<1		<1	
			Pronoun (P)	2.40	1, 39	2.20	1, 33
			T × GG	1.22	2, 78	<1	
			T × P	<1		<1	
		GG × P	1.49	1, 39	1.08	1, 33	
	Pronoun	T × GG × P	2.26	2, 78	2.17	2, 33	
		RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	1.35	1, 39	1.40	1, 33	
		Pronoun (P)	<1		<1		
		T × GG	1.14	2, 78	1.35	2, 33	
		T × P	<1		<1		
Pronoun spillover	GG × P	7.61***	1, 39	5.18**	1, 33		
	T × GG × P	<1		<1			
	RN typicality (T)	<1		<1			
	RN grammatical gender (GG)	<1		<1			
	Pronoun (P)	2.29	1, 39	2.30	1, 33		
	T × GG	<1		<1			
	T × P	<1		<1			
	GG × P	8.00***	1, 39	6.63**	1, 33		
	T × GG × P	4.49**	2, 78	1.92	2, 33		

<sup>a</sup>FF: first fixation durations, FP: first-pass reading times, RP: regression path times, RI: regressions into the region, TT: total fixation times; \* $p \leq .10$ , \*\* $p \leq .05$ , \*\*\* $p \leq .01$ .

irrespective of pronoun gender in the second clause,  $M_{\text{Masc/masc}} = -25.74$ ,  $M_{\text{Masc/fem}} = 11.72$ ,  $t(39) = -1.14$ , ns. However, feminine role nouns were fixated shorter when followed by congruent (feminine) pronouns compared to incongruent (masculine) ones,  $M_{\text{Fem/masc}} = 62.2$ ,  $M_{\text{Fem/fem}} = -48.73$ ,  $t(39) = 3.76$ ,  $SEM = 29.47$ .

Furthermore, there was an interaction effect between the grammatical gender of the role noun and its stereotypicality on the *role noun* region. For both stereotypically male and stereotypically female role nouns, fixation times were shorter when grammatical gender was congruent with stereotypicality (e.g. *Elektriker*<sub>MaleMasc</sub> vs. *Elektrikerin*<sub>MaleFem</sub> or *Kosmetikerin*<sub>FemaleFem</sub> vs. *Kosmetiker*<sub>FemaleMasc</sub>),  $M_{\text{MaleMasc}} = -29.09$ ,  $M_{\text{MaleFem}} = 48.48$ ,  $t(39) = -2.40$ ,  $SEM = 32.37$ ;  $M_{\text{FemaleMasc}} = 73.86$ ,  $M_{\text{FemaleFem}} = -0.83$ ,  $t(39) = 2.27$ ,  $SEM = 32.86$ . For the neutral role nouns there was no difference in fixation times depending on their grammatical gender,  $M_{\text{NeutMasc}} = -63.72$ ,  $M_{\text{NeutFem}} = -23.75$ ,  $t(39) = -1.53$ , ns.

For the *verb* region, the ANOVA revealed an interaction between the grammatical gender of the role noun and pronoun gender. Verbs preceding the pronouns were fixated shorter when grammatically masculine role nouns were followed by masculine rather than feminine pronouns,  $M_{\text{Masc/masc}} = -31.14$ ,  $M_{\text{Masc/fem}} = 15.85$ ,  $t(38) = -3.30$ ,  $SEM = 14.26$ . There was no difference in verb fixation times for grammatically feminine role nouns followed by masculine and feminine pronouns,  $M_{\text{Fem/masc}} = 18.24$ ,  $M_{\text{Fem/fem}} = -11.97$ ,  $t(38) = 1.89$ , ns.

In the *pronoun* region and in the *pronoun spillover*, the ANOVA revealed the same interaction between grammatical gender of the role noun and pronoun gender. The total fixation times on these regions were shorter when grammatical gender of the role nouns and pronoun gender matched and longer when there was grammatical disagreement between the two, *pronoun* region:  $M_{\text{Masc/masc}} = -65.16$ ,  $M_{\text{Masc/fem}} = 67.56$ ,  $t(39) = -4.18$ ,  $SEM = 31.68$ ;  $M_{\text{Fem/masc}} = 41.06$ ,  $M_{\text{Fem/fem}} = -51.08$ ,  $t(39) = 3.78$ ,  $SEM = 24.41$ ; *pronoun spillover*:  $M_{\text{Masc/masc}} = -33.72$ ,  $M_{\text{Masc/fem}} = 24.90$ ,  $t(37) = -3.75$ ,  $SEM = 15.62$ ;  $M_{\text{Fem/masc}} = 27.84$ ,  $M_{\text{Fem/fem}} = -34.24$ ,  $t(37) = 3.38$ ,  $SEM = 18.34$ .

*Regressions into a region.* The analysis of the regressions into the *role noun* region revealed two types of interaction. First, an interaction occurred between the grammatical gender of the role noun and that of the pronoun. There were fewer regressions when role nouns and pronouns were grammatically congruent than when they were incongruent,  $M_{\text{Masc/masc}} = 29.44$ ,  $M_{\text{Masc/fem}} = 37.26$ ,  $t(39) = -2.83$ ,  $SEM = 2.76$ ;

$M_{\text{Fem/masc}} = 41.11$ ,  $M_{\text{Fem/fem}} = 28.89$ ,  $t(39) = 3.03$ ,  $SEM = 4.03$ . Secondly, the relation between role nouns and pronouns was qualified by a three-way interaction between the grammatical gender of the role noun, its stereotypicality and the grammatical gender of the pronoun. There was no reliable difference in regressions into stereotypically male role nouns,  $M_{\text{MaleMasc/masc}} = 33.33$ ,  $M_{\text{MaleMasc/fem}} = 36.67$ ,  $t(39) = -0.61$ , ns;  $M_{\text{MaleFem/masc}} = 39.17$ ,  $M_{\text{MaleFem/fem}} = 39.17$ ,  $t(39) = 0$ , ns. But there were fewer regressions into stereotypically female role nouns when pronouns were grammatically congruent with role nouns than when they were incongruent,  $M_{\text{FemaleMasc/masc}} = 29.17$ ,  $M_{\text{FemaleMasc/fem}} = 41.67$ ,  $t(39) = -2.73$ ,  $SEM = 4.57$ ;  $M_{\text{FemaleFem/masc}} = 40.83$ ,  $M_{\text{FemaleFem/fem}} = 20$ ,  $t(39) = 3.26$ ,  $SEM = 6.39$ . Finally, there were fewer regressions into neutral role nouns with feminine grammatical gender when they were followed by feminine compared to masculine pronouns,  $M_{\text{NeutFem/masc}} = 43.33$ ,  $M_{\text{NeutFem/fem}} = 27.5$ ,  $t(39) = 2.46$ ,  $SEM = 6.43$ . The percentage of regressions into neutral role nouns with masculine grammatical gender did not differ according to the subsequent pronoun,  $M_{\text{NeutMasc/masc}} = 25.83$ ,  $M_{\text{NeutMasc/fem}} = 33.33$ ,  $t(39) = -1.33$ , ns.

Analysis of the regressions into the *verb* region showed another interaction between the grammatical gender of the role noun and that of the pronoun. There were fewer regressions into the *verb* when role noun antecedents and pronouns agreed in grammatical gender than when they were incongruent,  $M_{\text{Masc/masc}} = 14.17$ ,  $M_{\text{Masc/fem}} = 25.83$ ,  $t(39) = -4.58$ ,  $SEM = 2.55$ ;  $M_{\text{Fem/masc}} = 23.61$ ,  $M_{\text{Fem/fem}} = 16.94$ ,  $t(39) = 2.15$ ,  $SEM = 3.10$ .

To specify the exact source of regressions into the role noun, we conditionalised regressions into the *role noun* region by launching region. Regressions from the *pronoun* into the *role noun* showed an interaction between the grammatical gender of the role noun and the pronoun. The probability of regressions into masculine role nouns after masculine or feminine pronouns did not differ,  $M_{\text{Masc/masc}} = 6.94$ ,  $M_{\text{Masc/fem}} = 9.44$ ,  $t(39) = -1.27$ , ns. However, there were more regressions into feminine role nouns after incongruent masculine pronouns than after congruent feminine pronouns,  $M_{\text{Fem/masc}} = 12.50$ ,  $M_{\text{Fem/fem}} = 7.26$ ,  $t(39) = 2.29$ ,  $SEM = 2.29$ .

Regressions from the *pronoun spillover* into the *role noun* also showed an interaction between the grammatical gender of the role noun and of the pronoun. Again, the probability of regressions into masculine role nouns after masculine or feminine pronouns did not differ,  $M_{\text{Masc/masc}} = 11.11$ ,  $M_{\text{Masc/fem}} = 14.17$ ,  $t(39) = -1.15$ , ns. But as before, there were more regressions into feminine role nouns after incongruent (masculine) pronouns than after congruent (feminine)

pronouns,  $M_{\text{Fem/masc}} = 16.67$ ,  $M_{\text{Fem/fem}} = 8.37$ ,  $t(39) = 3.11$ ,  $SEM = 2.67$ .

### Discussion

One effect occurred reliably in all measures except the first fixation durations: the interaction between the grammatical gender of the role noun and that of the pronoun. Generally, fixations were shorter and probabilities of regressions were lower when the grammatical gender of a role noun was congruent with the pronoun. In some cases, however, this pattern applied only to feminine role nouns (first-pass reading times of the *pronoun spillover*, regressions into the *role noun* from the *pronoun* and *pronoun spillover*, total fixation times of the *role noun* region), which suggests that there is an asymmetry in the processes involved in coreference establishment regarding masculine and feminine grammatical gender. The details of this asymmetry will be discussed in the General Discussion below. Note that the influence of grammatical gender demonstrated by the interaction between the grammatical gender of the role noun and of the pronoun appears already during first-pass reading times, a measure reflecting early processing, and lasts until the final stages of processing.

Sentence processing was further influenced by role noun stereotypicality, as reflected in regression path on the *pronoun* region. Only stereotypically male role nouns and grammatically masculine neutral role nouns required longer processing when their grammatical gender was incongruent with pronoun gender. Interestingly, these difficulties did not arise earlier, which indicates the activation of stereotypical gender information only at a later stage. Moreover, the processing of role nouns was not slowed down by an incongruity between stereotypicality and role noun gender until the very last stage, as reflected in total fixation times (as the effect was not present in any of the earlier measures, it must be due to the repeated reading of the region). These effects show that the influence of stereotypical gender, compared to grammatical gender, appears relatively late in sentences with anaphoric pronouns referring back to the first clause.

### Experiment 2

In the second experiment, we examined if grammatical and stereotypical gender influences observed in Experiment 1 affect processing in different ways when the pronominal anaphor is replaced with a noun phrase. Namely, we examined if different time-course patterns of processing these gender cues emerge in sentences containing the semantically rich anaphors *this man/this*

*woman* ‘dieser Mann/diese Frau’ compared to the personal pronouns *he/she* ‘er/sie’ used in Experiment 1.

### Method

#### Participants

Twenty-four native speakers of German (9 male and 15 female, mean age 23.3 years,  $SD = 2.5$ ) were paid to participate in the study. Most of them were students at the University of Heidelberg.

#### Materials and procedure

The anaphoric pronouns in the 36 experimental sentences used in Experiment 1 were replaced by the noun phrase *dieser Mann* ‘this man’ or *diese Frau* ‘this woman’. These experimental items were presented on the screen in the form shown in Example (2).

- (2)  
*Oft hatte der Elektriker gute Einfälle,  
 regelmässig plante dieser Mann neue Projekte.*  
 ‘Often had the electrician good ideas,  
 regularly planned this man new projects’  
 (Word-by-word translation is given to render the word order of the original sentence).

All procedural details as well as the experimental design were the same as in Experiment 1.

### Results

#### Data analysis

While the structure of the sentences was the same as in Experiment 1, the anaphoric pronouns used in the first experiment were replaced by the noun phrases *dieser Mann* ‘this man’ and *diese Frau* ‘this woman’. The regions of analysis in the first clause were identical to the ones in Experiment 1. In the second clause, the regions of analysis were the anaphor *determiner* (this-masc/fem), the *noun* itself (man or woman) and the following adjective and noun (*anaphor spillover*). Fixations below 70 ms and above 600 ms were removed (3.5% of the data) before the statistical analyses were performed.

Means and standard deviations of residual fixation times and probabilities of regressions for all measures and all regions of interest are given in Table 3. Results of analyses of variance are shown Table 4.

The same strategies of reporting and interpreting results apply as in Experiment 1.

The ANOVA did not reveal any reliable main effects or interactions in either the first fixation durations or in first-pass reading time measures.

*Regression path time.* In the *anaphor spillover*, an interaction was found between the grammatical gender

Table 3. Means (and standard deviations) of residual fixation times and probabilities of regressions (Experiment 2).

Region	Factor			Measure <sup>a</sup>				
	Typicality (RN)	Grammatical gender (RN)	Pronoun	FF	FP	RP	RI	TT
Role noun	Male	Masculine	Masculine	-2.06 (28.34)	2.06 (167.27)	1.36 (211.11)	27.78 (30.56)	-31.33 (311.79)
			Feminine	-1.79 (42.28)	-26.11 (149.65)	-28.20 (187.90)	40.28 (29.45)	-12.33 (316.53)
	Female	Feminine	Masculine	-13.96 (31.98)	-14.17 (170.72)	-11.19 (236.42)	45.83 (39.09)	95.48 (416.32)
			Feminine	1.43 (38.53)	2.72 (155.36)	-25.61 (178.02)	33.33 (29.49)	-42.40 (266.90)
		Masculine	Masculine	0.41 (43.49)	20.96 (207.57)	37.67 (241.63)	31.94 (26.88)	-11.42 (329.90)
			Feminine	6.85 (32.03)	63.91 (173.54)	36.28 (211.95)	30.56 (29.35)	61.86 (305.10)
	Neutral	Masculine	Masculine	0.07 (34.62)	-15.38 (167.63)	-6.16 (188.09)	45.83 (32.32)	56.45 (406.89)
			Feminine	-2.49 (42.26)	34.46 (156.18)	45.40 (264.79)	31.94 (31.82)	-14.38 (356.09)
			Masculine	3.59 (35.54)	-12.68 (120.82)	-28.33 (131.30)	34.72 (31.82)	-33.24 (238.51)
		Feminine	Masculine	1.14 (43.07)	-40.61 (119.31)	-45.07 (132.74)	30.56 (32.48)	-47.04 (221.80)
			Masculine	7.14 (37.61)	-8.62 (128.67)	-6.23 (167.01)	38.89 (27.22)	33.66 (322.04)
			Feminine	-0.14 (36.34)	-5.70 (143.74)	22.94 (256.61)	31.94 (30.26)	-56.88 (251.65)
RN spillover	Male	Masculine	Masculine	-3.79 (40.97)	-13.48 (189.42)	-24.85 (231.97)	12.50 (19.19)	-47.24 (266.51)
			Feminine	3.80 (45.25)	-63.57 (129.94)	-18.28 (285.82)	11.11 (16.05)	-19.53 (309.72)
	Female	Feminine	Masculine	-10.07 (38.88)	-32.86 (178.88)	-38.68 (258.74)	18.06 (24.04)	3.48 (417.91)
			Feminine	4.15 (56.15)	15.23 (217.24)	-12.56 (322.73)	16.67 (26.01)	11.20 (350.24)
		Masculine	Masculine	-7.04 (41.59)	5.51 (232.12)	16.69 (379.64)	6.94 (13.83)	-42.67 (352.83)
			Feminine	-3.22 (51.26)	-6.11 (208.81)	56.87 (362.29)	9.72 (18.33)	51.54 (442.43)
	Neutral	Masculine	Masculine	4.40 (43.58)	-10.16 (222.61)	48.75 (377.69)	15.28 (19.61)	16.17 (368.65)
			Feminine	3.17 (38.51)	6.02 (188.95)	39.63 (427.26)	6.94 (13.83)	-16.52 (369.83)
			Masculine	-0.43 (46.80)	-14.02 (197.89)	-36.06 (252.39)	11.11 (18.82)	6.65 (364.60)
		Feminine	Masculine	3.88 (46.58)	24.71 (252.98)	-4.02 (276.35)	9.72 (15.48)	46.56 (419.27)
			Masculine	8.03 (35.52)	35.40 (194.70)	-16.06 (248.70)	16.67 (19.66)	17.44 (289.49)
			Feminine	-4.00 (41.43)	18.08 (197.97)	-30.72 (223.39)	8.33 (17.72)	-61.03 (260.96)
Determiner	Male	Masculine	Masculine	12.39 (67.95)	4.04 (77.61)	-12.74 (100.95)	8.33 (17.72)	-47.28 (101.07)
			Feminine	16.00 (55.57)	21.31 (70.18)	3.85 (72.55)	22.22 (28.94)	19.09 (147.79)
	Female	Feminine	Masculine	0.89 (44.80)	26.04 (49.19)	41.18 (134.60)	16.67 (26.01)	79.72 (192.21)
			Feminine	7.88 (39.78)	8.69 (51.63)	-22.24 (60.19)	18.06 (25.97)	16.25 (162.92)
		Masculine	Masculine	5.68 (41.47)	1.83 (53.63)	-12.53 (134.19)	9.72 (15.48)	-48.58 (108.29)
			Feminine	-5.08 (53.47)	0.36 (77.63)	-21.24 (134.40)	13.89 (21.80)	-44.12 (136.55)
	Neutral	Feminine	Masculine	-0.44 (34.77)	3.38 (78.08)	-9.53 (85.78)	30.56 (32.48)	34.18 (162.75)
			Feminine	-11.08 (38.27)	5.73 (72.64)	44.06 (223.38)	12.50 (21.56)	-30.45 (91.02)
			Masculine	0.02 (28.90)	12.92 (40.40)	-40.70 (43.04)	19.44 (25.85)	-53.76 (83.98)
		Masculine	Feminine	6.38 (73.42)	3.81 (87.71)	-16.53 (129.23)	19.44 (19.45)	19.46 (133.95)
			Masculine	2.23 (51.36)	-2.83 (50.10)	10.47 (132.49)	29.17 (28.34)	27.27 (98.50)
			Feminine	15.42 (47.16)	8.27 (54.21)	-28.40 (80.67)	16.67 (26.01)	-3.45 (119.17)
Noun	Male	Masculine	Masculine	203.50 (50.18)	218.50 (38.28)	266.27 (82.18)	6.94 (13.83)	216.53 (77.24)
			Feminine	192.00 (70.05)	192.00 (70.05)	290.27 (136.34)	11.11 (16.05)	246.67 (83.46)
	Female	Feminine	Masculine	197.83 (95.45)	197.83 (95.45)	436.53 (448.86)	19.44 (23.91)	292.13 (126.07)
			Feminine	162.67 (57.00)	211.33 (151.43)	345.87 (193.95)	9.72 (20.80)	269.20 (114.19)
		Masculine	Masculine	228.67 (15.94)	252.33 (53.60)	223.47 (79.51)	8.33 (17.72)	249.20 (89.27)
			Feminine	190.50 (54.60)	201.50 (62.61)	359.73 (308.58)	9.72 (18.33)	232.00 (90.55)
	Neutral	Masculine	Masculine	203.50 (48.29)	203.50 (48.29)	220.80 (61.80)	22.22 (23.40)	304.33 (95.22)
			Feminine	189.50 (36.78)	229.00 (60.96)	247.60 (30.87)	6.94 (13.83)	214.00 (68.60)

Table 3 (Continued)

Region	Factor		Measure <sup>a</sup>						
	Typicality (RN)	Grammatical gender (RN)	Pronoun	FF	FP	RP	RI	TT	
Neutral		Masculine	Masculine	168.00 (62.35)	184.00 (65.69)	307.60 (191.07)	9.72 (18.33)	232.07 (63.48)	
		Feminine	Feminine	199.00 (28.21)	209.50 (33.12)	233.20 (60.26)	16.67 (21.98)	249.87 (42.33)	
NP spillover	Male	Masculine	Masculine	215.50 (72.79)	222.83 (81.19)	549.20 (374.90)	16.67 (24.08)	258.27 (149.11)	
		Feminine	Feminine	207.00 (41.23)	207.00 (41.23)	270.00 (121.53)	15.28 (21.93)	285.20 (138.77)	
Female		Masculine	Masculine	3.83 (70.33)	-28.27 (204.83)	-281.83 (655.05)	/	-88.89 (277.02)	
		Feminine	Feminine	-2.64 (61.81)	2.68 (275.18)	88.12 (1039.86)	/	20.89 (369.27)	
Neutral		Masculine	Masculine	3.26 (46.70)	4.74 (215.19)	417.27 (1393.60)	/	93.62 (401.66)	
		Feminine	Feminine	-7.17 (43.00)	-23.78 (232.06)	-70.93 (803.99)	/	-38.28 (314.99)	
Female		Masculine	Masculine	-2.42 (45.25)	35.45 (224.21)	-359.03 (566.48)	/	-60.88 (267.66)	
		Feminine	Feminine	9.59 (76.74)	-9.10 (274.09)	-0.92 (1329.78)	/	-49.49 (394.59)	
Neutral		Masculine	Masculine	12.41 (49.76)	-103.79 (210.60)	181.54 (1127.61)	/	-1.94 (322.81)	
		Feminine	Feminine	-6.64 (50.84)	-22.76 (247.70)	-156.80 (1069.00)	/	-78.34 (253.46)	
Female		Masculine	Masculine	6.84 (52.37)	11.84 (225.95)	-150.29 (921.13)	/	29.83 (425.82)	
		Feminine	Feminine	-2.99 (49.12)	0.84 (257.29)	157.35 (1091.65)	/	45.46 (333.21)	
Neutral		Masculine	Masculine	-5.64 (59.35)	4.33 (257.99)	179.63 (1129.87)	/	109.64 (381.36)	
		Feminine	Feminine	-5.21 (59.22)	13.32 (247.20)	-164.51 (762.00)	/	-13.36 (306.60)	

<sup>a</sup>FF: first fixation durations, FP: first-pass reading times, RP: regression path times, RI: regressions into the region, TT: total fixation times.

of the role noun and that of the anaphor gender. When role nouns and anaphors were grammatically congruent, the fixation times on the anaphor *spillover* were shorter than when they were incongruent,  $M_{\text{Masc/masc}} = -259.24$ ,  $M_{\text{Masc/fem}} = 71.49$ ,  $t(22) = 2.75$ ,  $SEM = 120.42$ ;  $M_{\text{Fem/masc}} = 255.06$ ,  $M_{\text{Fem/fem}} = -125.57$ ,  $t(22) = 3.34$ ,  $SEM = 113.93$ .

*Total fixation times.* In the *role noun* region, the ANOVA revealed another interaction between the grammatical gender of the role noun and anaphor gender. Masculine role nouns were fixated equally long irrespective of the gender of the anaphor,  $M_{\text{Masc/masc}} = -26.57$ ,  $M_{\text{Masc/fem}} = 0.83$ ,  $t(23) = -0.79$ , ns. Feminine role nouns, however, were fixated shorter when the anaphor was feminine rather than masculine,  $M_{\text{Fem/masc}} = 61.87$ ,  $M_{\text{Fem/fem}} = -37.38$ ,  $t(23) = 2.34$ ,  $SEM = 42.49$ .

In the *determiner* region, the ANOVA revealed a main effect of the grammatical gender of the role noun. The determiner was fixated shorter when the role noun was in the masculine form than when it was feminine,  $M_{\text{Masc}} = -27.89$ ,  $M_{\text{Fem}} = 16.22$ ,  $t(23) = -3.51$ ,  $SEM = 12.57$ . There was also an interaction between the grammatical gender of the role noun and that of the anaphor. The determiner was fixated shorter when role nouns were grammatically congruent with anaphors than when they were incongruent,  $M_{\text{Masc/masc}} = -50.77$ ,  $M_{\text{Masc/fem}} = -2.27$ ,  $t(23) = -2.99$ ,  $SEM = 16.22$ ;  $M_{\text{Fem/masc}} = 39.49$ ,  $M_{\text{Fem/fem}} = -10.87$ ,  $t(23) = 2.36$ ,  $SEM = 21.38$ .

Another interaction between the grammatical gender of the role noun and the anaphor gender emerged in the *anaphor spillover* region. There was no difference in total fixation times when feminine and masculine anaphors followed masculine role nouns,  $M_{\text{Masc/masc}} = -39.65$ ,  $M_{\text{Masc/fem}} = 4.38$ ,  $t(23) = -1.35$ , ns. At the same time, fixations were shorter when feminine role nouns were followed by congruent (feminine) anaphors compared to incongruent (masculine) ones,  $M_{\text{Fem/masc}} = 66.84$ ,  $M_{\text{Fem/fem}} = -41.46$ ,  $t(23) = 3.01$ ,  $SEM = 36.03$ .

*Regressions into a region.* In the *role noun* region, the main effect of the grammatical gender of the role noun manifested itself in more regressions into feminine compared to masculine role nouns,  $M_{\text{Masc}} = 32.64$ ,  $M_{\text{Fem}} = 37.96$ ,  $t(23) = -1.78$ ,  $SEM = 3.00$ .

For the anaphor *determiner*, the ANOVA revealed two interactions. First, an interaction emerged between role noun stereotypicality and anaphor gender. In sentences with stereotypically male role nouns, there were fewer regressions into the determiner when the anaphor was masculine than when it was feminine,  $M_{\text{Male/masc}} = 12.5$ ,  $M_{\text{Male/fem}} = 20.14$ ,  $t(23) = -2.2$ ,  $SEM = 3.47$ . In sentences with stereotypically female



Table 4. Results of analyses of variance for all regions of interest (Experiment 2).

Measure <sup>a</sup>	Region	Effect	F <sub>1</sub>	df <sub>1, 2</sub>	F <sub>2</sub>	df <sub>1, 2</sub>	
FF	Role noun	RN typicality (T)	1.30	2, 46	1.19	2, 33	
		RN grammatical gender (GG)	<1		<1		
		Noun phrase (NP)	<1		<1		
		T × GG	<1		<1		
		T × NP	1.17	2, 46	<1		
		GG × NP	<1		<1		
		T × GG × NP	<1		1.49	2, 33	
	Role noun spillover	RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	<1		<1		
		Noun phrase (NP)	<1		<1		
		T × GG	<1		<1		
		T × NP	1.16	2, 46	<1		
		GG × NP	<1		<1		
		T × GG × NP	<1		<1		
	Determiner	RN typicality (T)	1.77	2, 30	3.16*	2, 33	
		RN grammatical gender (GG)	<1		<1		
		Noun phrase (NP)	<1		<1		
		T × GG	<1		1.70	2, 33	
		T × NP	<1		<1		
		GG × NP	<1		<1		
		T × GG × NP	<1		<1		
	Noun	RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	<1		<1		
		Noun phrase (NP)	<1		<1		
		T × GG	<1		1.25	2, 31	
		T × NP	<1		2.41	2, 31	
		GG × NP	<1		2.31	1, 31	
		T × GG × NP	<1		<1		
	Noun phrase spillover	RN typicality (T)	<1		<1		
		RN grammatical gender (GG)	<1		<1		
		Noun phrase (NP)	1.08	1, 23	1.16	1, 33	
		T × GG	<1		<1		
		T × NP	<1		<1		
		GG × NP	<1		<1		
		T × GG × NP	1.63	2, 46	1.90	2, 33	
	FP	Role noun	RN typicality (T)	3.72**	2, 46	1.90	2, 33
			RN grammatical gender (GG)	<1		<1	
			Noun phrase (NP)	<1		<1	
			T × GG	1.84	2, 46	<1	
			T × NP	3.42**	2, 46	<1	
GG × NP			<1		<1		
T × GG × NP			<1		<1		
Role noun spillover		RN typicality (T)	2.36	2, 46	<1		
		RN grammatical gender (GG)	1.62	1, 23	<1		
		Noun phrase (NP)	<1		<1		
		T × GG	<1		<1		
		T × NP	<1		<1		
		GG × NP	<1		<1		
Determiner	T × GG × NP	1.96	2, 46	<1			
	RN typicality (T)	1.13	2, 30	1.54	2, 33		
	RN grammatical gender (GG)	<1		<1			
	Noun phrase (NP)	<1		<1			
	T × GG	<1		1.04	2, 33		
	T × NP	<1		<1			
	GG × NP	<1		<1			
Noun	T × GG × NP	1.03	2, 30	1.24	2, 33		
	RN typicality (T)	<1		<1			
	RN grammatical gender (GG)	<1		<1			
	Noun phrase (NP)	<1		<1			
	T × GG	<1		<1			
	T × NP	<1		1.73	2, 31		
	GG × NP	28.24**	1, 3	1.69	1, 31		
T × GG × NP	<1		2.09	2, 31			



Table 4 (Continued)

Measure <sup>a</sup>	Region	Effect	$F_1$	$df_{1,2}$	$F_2$	$df_{1,2}$		
FP	Noun phrase spillover	RN typicality (T)	1.06	2, 44	2.71*	2, 33		
		RN grammatical gender (GG)	1.44	1, 22	<1			
		Noun phrase (NP)	<1		<1			
		T × GG	2.65*	2, 44	<1			
		T × NP	<1		<1			
		GG × NP	<1		<1			
		T × GG × NP	2.42	2, 44	<1			
		RP	Role noun	RN typicality (T)	3.72**	2, 46	1.55	2, 33
				RN grammatical gender (GG)	<1		<1	
				Noun phrase (NP)	<1		<1	
T × GG	1.90			2, 46	<1			
T × NP	<1				<1			
Role noun spillover	GG × NP		1.45	1, 23	<1			
	T × GG × NP		<1		<1			
	RN typicality (T)		3.67**	2, 46	1.62	2, 33		
	RN grammatical gender (GG)		<1		<1			
	Noun phrase (NP)		<1		<1			
Determiner	Determiner	T × GG	<1		<1			
		T × NP	<1		<1			
		GG × NP	<1		<1			
		T × GG × NP	<1		<1			
		RN typicality (T)	<1		<1			
	Noun	RN grammatical gender (GG)	5.58**	1, 15	1.82	1, 33		
		Noun phrase (NP)	<1		<1			
		T × GG	<1		<1			
		T × NP	1.04	2, 30	<1			
		GG × NP	1.72	1, 15	<1			
Noun phrase spillover	Noun	T × GG × NP	1.81	2, 30	1.06	2, 33		
		RN typicality (T)	1.51	2, 8	<1			
		RN grammatical gender (GG)	2.35	1, 4	1.33	1, 31		
		Noun phrase (NP)	1.12	1, 4	<1			
		T × GG	<1		<1			
	Noun phrase spillover	T × NP	1.39	2, 8	<1			
		GG × NP	1.81	1, 4	3.06*	1, 31		
		T × GG × NP	<1		<1			
		RN typicality (T)	1.79	2, 44	1.83	2, 33		
		RN grammatical gender (GG)	8.30***	1, 22	2.55	1, 33		
TT	Role noun	Noun phrase (NP)	<1		<1			
		T × GG	1.83	2, 44	<1			
		T × NP	<1		<1			
		GG × NP	13.14***	1, 22	15.86***	1, 33		
		T × GG × NP	<1		<1			
	Role noun spillover	RN typicality (T)	2.03	2, 46	<1			
		RN grammatical gender (GG)	1.18	1, 23	<1			
		Noun phrase (NP)	2.07	1, 23	<1			
		T × GG	<1		<1			
		T × NP	<1		<1			
Determiner	Role noun	GG × NP	4.63**	1, 23	3.33*	1, 33		
		T × GG × NP	<1		<1			
		RN typicality (T)	<1		<1			
		RN grammatical gender (GG)	<1		<1			
		Noun phrase (NP)	<1		<1			
	Role noun spillover	T × GG	1.27	2, 46	<1			
		T × NP	<1		<1			
		GG × NP	4.15*	1, 23	1.31	1, 33		
		T × GG × NP	<1		<1			
		RN typicality (T)	2.00	2, 42	1.20	2, 33		
Determiner	RN grammatical gender (GG)	12.11***	1, 21	9.66***	1, 33			
	Noun phrase (NP)	<1		<1				
	T × GG	<1		<1				
	T × NP	2.21	2, 42	<1				
	GG × NP	10.55***	1, 21	17.98***	1, 33			
		T × GG × NP	<1		<1			

Table 4 (Continued)

Measure <sup>a</sup>	Region	Effect	$F_1$	$df_{1,2}$	$F_2$	$df_{1,2}$
TT	Noun	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	2.25	1, 9	4.43**	1, 33
		Noun phrase (NP)	<1		<1	
		T × GG	<1		<1	
		T × NP	5.13**	2, 18	1.59	2, 33
		GG × NP	1.68	1, 9	6.67**	1, 33
		T × GG × NP	<1		<1	
	Noun phrase spillover	RN typicality (T)	5.08***	2, 46	4.85**	2, 33
		RN grammatical gender (GG)	2.03	1, 23	<1	
		Noun phrase (NP)	2.28	1, 23	<1	
		T × GG	<1		<1	
		T × NP	<1		<1	
		GG × NP	8.19***	1, 23	3.97*	1, 33
		T × GG × NP	1.20	2, 46	<1	
RI	Role noun	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	3.15*	1, 23	4.32**	1, 33
		Noun phrase (NP)	2.01	1, 23	1.21	1, 33
		T × GG	<1		<1	
		T × NP	<1		<1	
		GG × NP	4.58**	1, 23	5.71**	1, 33
		T × GG × NP	1.35	2, 46	1.31	2, 33
	Role noun spillover	RN typicality (T)	1.77	2, 46	1.51	2, 33
		RN grammatical gender (GG)	3.19*	1, 23	2.47	1, 33
		Noun phrase (NP)	2.46	1, 23	1.82	1, 33
		T × GG	<1		<1	
		T × NP	<1		<1	
		GG × NP	2.46	1, 23	1.56	1, 33
		T × GG × NP	<1		<1	
	Determiner	RN typicality (T)	2.08	2, 46	1.22	2, 33
		RN grammatical gender (GG)	2.39	1, 23	3.24*	1, 33
		Noun phrase (NP)	<1		<1	
		T × GG	1.03	2, 46	<1	
		T × NP	4.10**	2, 46	3.14*	2, 33
		GG × NP	6.23**	1, 23	13.73***	1, 33
		T × GG × NP	<1		<1	
	Noun	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	2.75	1, 23	3.46*	1, 33
		Noun phrase (NP)	<1		<1	
T × GG		<1		<1		
T × NP		2.19	2, 46	1.32	2, 33	
GG × NP		10.53***	1, 23	11.47***	1, 33	
T × GG × NP		<1		<1		

<sup>a</sup>FF: first fixation durations, FP: first-pass reading times, RP: regression path times, RI: regressions into the region, TT: total fixation times; \* $p \leq .10$ , \*\* $p \leq .05$ , \*\*\* $p \leq .01$ .

role nouns, there was a tendency of fewer regressions into the determiner for congruent (feminine) anaphors than for incongruent (masculine) ones,  $M_{\text{Female/masc}} = 20.14$ ,  $M_{\text{Female/fem}} = 13.19$ ,  $t(23) = 1.93$ ,  $SEM = 3.61$ ,  $p = .067$ . In sentences with neutral role nouns, there was no difference in the probability of regressions into the determiner for masculine or feminine anaphors,  $M_{\text{Neut/masc}} = 24.31$ ,  $M_{\text{Neut/fem}} = 18.06$ ,  $t(23) = 1.23$ , ns.

Second, there was once again an interaction between the grammatical gender of the role noun and that of the anaphor. After masculine role nouns, there was no difference in the probability of regressions into the

determiner depending on the gender of the following noun,  $M_{\text{Masc/masc}} = 12.4$ ,  $M_{\text{Masc/fem}} = 17.23$ ,  $t(23) = -1.6$ , ns. After feminine role nouns, there were fewer regressions into the anaphor determiner when anaphors were also feminine than when they were masculine,  $M_{\text{Fem/masc}} = 25.46$ ,  $M_{\text{Fem/fem}} = 15.74$ ,  $t(23) = 2.29$ ,  $SEM = 4.25$ .

In the *noun* region ('man' or 'woman'), there was again an interaction between the grammatical gender of the role noun and anaphor gender. After masculine role nouns, there was no difference in the probability of regressions into masculine or feminine anaphors,

$M_{\text{Masc/masc}} = 8.33$ ,  $M_{\text{Masc/fem}} = 12.5$ ,  $t(23) = -1.4$ , ns. After feminine role nouns, there were fewer regressions into feminine anaphors than into masculine ones,  $M_{\text{Fem/masc}} = 19.44$ ,  $M_{\text{Fem/fem}} = 10.65$ ,  $t(23) = 2.74$ ,  $SEM = 3.21$ .

### Discussion

Experiment 2 revealed an interaction between the grammatical gender of the role noun and that of the anaphoric expression. This interaction was reliable in all measures except first fixation durations and first-pass reading times. When comparing anaphors that are grammatically congruent or incongruent with their antecedents, a general pattern with shorter fixations and fewer regressions in congruent cases emerges. A closer look at this interaction, however, reveals an asymmetry in the processing of grammatically masculine and feminine role nouns, similar to the one found in Experiment 1. This will be discussed in more detail in the General discussion below.

The main effect of the grammatical gender of the role noun (in total fixation times on the *determiner*) indicates an asymmetry as well. It suggests that the processing of grammatically feminine gender generally requires more effort than the processing of masculine gender, when anaphoric sentences with noun phrase references to the first clause are being processed.

Furthermore, role noun stereotypicality was found to influence the process of anaphor resolution. Participants regressed more frequently to the anaphor when it was incongruent with the stereotyped role noun, which suggests that expectations of feminine and masculine grammatical gender after stereotypically female and male antecedents, respectively, were violated. After neutral role nouns, no indication of such a violation emerged. This suggests that expectations regarding the grammatical gender of the subject are less specific after reading neutral role nouns than they are in the case of stereotyped role nouns. Note that this influence of stereotypical gender occurred at a relatively late stage of processing, i.e., when participants regressed back to the anaphoric expression.

### General discussion and conclusions

The analysis of our results revealed several aspects concerning gender processing: the timing of the observed effects, their location in a sentence, their nature (grammatical/stereotypical, masculine/feminine) and the time course of the processes involved. Slightly different structuring of regions of interest required by two anaphor types, as well as the uniformity in the general structure of stimuli used in Experiments 1 and 2 and the fact that both samples were drawn from the

same population substantiate qualitative comparison of major findings as more appropriate comparison than statistical one. In this section, the results of both experiments are brought together in order to provide a better picture of anaphor resolution processes in sentences with antecedents containing both grammatical and stereotypical gender cues.

The eye movement patterns of the two experiments have shown reliable influences of grammatical gender both on the resolution of pronominal anaphors and noun phrase anaphors. Furthermore, these effects display interesting differences in timing when compared across experiments. Sentences with role nouns that were grammatically congruent rather than incongruent with anaphors caused less difficulty in processing. The violation of grammatical agreement affected comprehension already upon the first reading of pronominal anaphors, while in the case of noun phrase resolution, the effects of grammatical violations did not appear before regression path times of the region following the anaphor. Interestingly, in the sentences with noun phrase anaphors, this is overall the earliest effect found in the experiment. Anaphor resolution, therefore, seems to depend above all on the rules of grammatical agreement in the context of overlapping gender cues. In sentences with pronominal anaphors, the grammatical analysis starts immediately upon first reading, whereas with noun phrase anaphors the analysis is probably delayed by the additional semantic content which needs to be processed. Garrod and Sanford (1995) offer another possible explanation of this finding arguing that the difference in processing of pronominal and fuller anaphors comes from presupposition of a particular interpretation. Fuller descriptions do not seem to lead to immediate commitment to one particular (anaphoric) interpretation, since sentences containing them would still be possible without antecedents allowing different interpretations. According to Fraurud (1990), over 60% of full definite descriptions are mentioned in written text without discourse antecedents. This could be another reason why the interpretation of definite descriptions *this man/this woman* as anaphors in Experiment 2 was delayed.

An asymmetry in the processing of grammatical gender was observed in both experiments, for there were cases where congruity/incongruity with the anaphor affected either only masculine or only feminine role nouns. Feminine role nouns, particularly in sentences with noun phrase anaphors, made participants revisit antecedent and anaphor regions; the same tendency emerged in sentences with pronominal anaphors. It seems that masculine gender, due to its generic functions (Duden Grammatik, 1995), is more open for different gender interpretations. It may therefore allow an easier integration of masculine role nouns into a

context with other gender cues and make surprise effects less pronounced. This finding can be related to the elimination of gender mismatch effects by disambiguating context in English language reported in previous research (Duffy & Keir, 2004). Feminine role nouns, on the other hand, do not allow generic interpretations and may therefore require more revisiting in the attempt to resolve the anaphor (see Irmen & Schumann, 2011, for a similar asymmetry in the processing of masculine and feminine grammatical gender). This asymmetry is more pronounced in sentences with pronouns than with noun phrase anaphors. The reason may be that grammatical cues are of greater importance for the resolution of pronominal anaphors than for the resolution of noun phrase anaphors because the latter also require an earlier recruitment of conceptual gender cues. Distribution information could also influence the found asymmetry. Frequency analyses of anaphors used in Experiments 1 and 2 showed that pronouns were overall more frequent than noun phrases and differed in frequencies within themselves: masculine pronoun 'he' in German was more frequent than feminine pronoun 'she', and masculine noun phrase 'this man' was more frequent than feminine noun phrase 'this woman'.<sup>6</sup>

In both experiments, the influence of stereotypicality appeared only in measures reflecting later processing. The locations of stereotypicality effects indicate that in sentences with pronominal anaphors (Experiment 1) stereotypical gender information was involved in the processing of the role noun, whereas in Experiment 2 it affected the anaphor itself. Previous research has shown that the effect of stereotypical gender information is weaker than that of biological or definitional gender and can be modulated through a preceding context (e.g., Kreiner et al., 2008). The difference between the two experiments demonstrates the subtle nature of stereotypical gender influences, as well as their sensitivity to changes in the linguistic form of experimental materials. It is quite plausible that the semantically rich noun phrases used as anaphors in Experiment 2 highlighted the importance of stereotypicality information and enhanced its effect compared to pronouns providing little semantic content in Experiment 1. Garnham (2001) argues that most of the constraints on interpretation of anaphoric expressions must come from the context, since many anaphors do not have enough semantic content of their own (see also Duffy & Keir, 2004). The sentences used in both experiments provided very little context, and the absence of additional semantic information in Experiment 1 resulted in a dominance of grammatical gender in the process of anaphor resolution. While stereotypical gender cues are not as useful in identifying the

antecedent of a pronominal anaphor, the semantic content of noun phrase anaphors makes the recruitment of stereotypicality information quite important for the establishment of co-reference.

Our findings can be interpreted within the framework of two-stage models of reference resolution (Cook & Myers, 2004; Garrod & Sanford, 1995; Garrod & Terras, 2000), which claim that the first stage of resolution (*linkage/bonding*) is influenced by lexical information only, whereas the second stage (*verification/resolution*) can also be affected by semantic information already stored in memory. This sequence was indeed found in both experiments. Grammatical features of anaphoric expressions that contained both grammatical and stereotypical gender cues were used first, while stereotypicality information was recruited during later stages. The delay of the grammatical effect in Experiment 2 might be due to the additional semantic content in the anaphor, which required additional processing.

This is, however, in contrast with other studies that reported immediate effects of role stereotypicality on role noun processing as well as on reference resolution. These differences in findings may indicate that the processes under study are sensitive to the exact materials and procedures involved. In Carreiras et al. (1996), experimental passages started with role nouns, which may have emphasised the question of congruity between grammatical and stereotypical gender and may have caused immediate delays in reading the role noun in cases of incongruity. In Irmen and Schumann's (2011) materials, role nouns served as the second of two co-referring expressions within one clause. Here, again, stereotypicality affected the first reading of the role noun. In Irmen (2007), role nouns were used in non-referential, generic ways, thus emphasising semantic aspects of the resolution process and resulting in an effect of stereotypicality on the first reading of the anaphoric expression.

While the stages of anaphor resolution can be defined by the type of information that is being processed, the timing of these stages varies greatly depending on the availability and relevance of the information in each particular case. Non-referential use of role nouns, for example, could make stereotypicality a more relevant cue for resolving anaphors that refer to them than specific grammatical features of the antecedent (e.g., Irmen, 2007). Similarly, the noun phrases in Experiment 2 of the present investigation provide additional semantic information, as opposed to the pronominal anaphors of Experiment 1, which makes recruitment of stereotypicality information more relevant for processing at an earlier stage. This is reflected in regressions back to the anaphor region and not only in later wrap-up processes, as in the case of pronouns.

The findings of our experiments provide evidence that grammatical features lead the early stages of anaphor resolution with an earlier onset for pronominal than for noun phrase anaphors. The background knowledge about stereotypical gender roles, which influences later stages of processing gender-related information, is, in turn, recruited earlier for noun phrase anaphors than for pronominal anaphors. Even though the two-stage model of reference resolution seems to fit our data quite well in a general sense, a more refined model, one which considers gender asymmetries and specifies the timing of stages depending on the relevance of the processing of different types of information, would be needed to cover all the results concerning the processing of gender cues in reference resolution. In addition, it is important to realise that even though the processing of grammatical gender cues seems to start early, it may not be resolved by the time stereotypical gender comes into play (and vice versa; Irmen, 2007), which results in overlapping stages. The processing of gender cues at specific points in time seems to depend on the relevance of recruiting the most useful type of information. So far, we can say that in referential constructions this depends on the type of reference (i.e., anaphora or cataphora), grammatical features of antecedents and anaphors, semantic features of antecedents and anaphors, the distance between them, and context characteristics. Obviously, further research is needed to integrate all these factors and to differentiate the stages in the processing of cues from different grammatical and conceptual sources in anaphor resolution.

### Acknowledgements

This research was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 237907. We thank Friederike Braun for her feedback on earlier versions of the paper. At the time of data collection, all authors were affiliated with the Department of Cognitive and Theoretical Psychology at the University of Heidelberg, Germany.

### Notes

1. Capitalised subscripts refer to role noun properties ( $M_{\text{MaleMasc}}$  = mean value of stereotypically male role nouns in the grammatically masculine form), non-capitalised subscripts refer to the grammatical gender of the anaphor ( $M_{\text{Masc}}$  = mean value of masculine anaphors).
2. Eye movements have also been reported to reflect such cognitive mechanisms as, for instance, an identification of candidate antecedents or verification of those candidates (Duffy & Rayner, 1990), lexical or semantic access to

words (Garrod & Terras, 2000), selective reanalysis in syntactic parsing (Mitchell, Shen, Green, & Hodgson, 2008) and so on.

3. Frequency analyses were based on the corpora from the Archive of written language, Institute for German Language, Mannheim, Germany. Frequencies were collected based on the role noun stems including all inflections. In general, neutral role nouns were more frequent than stereotypically male, which in turn were more frequent than stereotypically female role nouns. However, frequencies did not differ significantly within the groups of stereotypically male, female and neutral role nouns.
4. We would like to thank Chuck Clifton for providing us with software for the analysis of regressions into a region conditionalised by launching region (used in Experiment 1) in addition to other software packages available on the website of the eye-tracking lab at the University of Massachusetts Amherst (<http://www.psych.umass.edu/eyelab/software/>).
5. The main effect of typicality in the *verb* region detected in first fixation durations and first-pass reading times is not relevant for the processes under study and will therefore be included in Table 2 only. It is not reported or interpreted in the text.
6. Frequency analyses were based on the corpora from the Archive of written language, Institute for German Language, Mannheim, Germany. Frequencies were collected for non-capitalised pronouns ('er', 'sie') and noun phrases '*dieser Mann*' and '*diese Frau*' excluding other inflections.

### References

- Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7(3), 136–141. doi:10.1111/j.1467-9280.1996.tb00346.x
- Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton (Eds.), *The online study of sentence comprehension* (pp. 51–76). New York, NY: Psychology Press.
- Breen, M., & Clifton Jr., C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, 64(2), 153–170. doi:10.1016/j.jml.2010.11.001
- Cacciari, C., Corradini, P., Padovani, R., & Carreiras, M. (2011). Pronoun resolution in Italian: The role of grammatical gender and context. *Journal of Cognitive Psychology*, 23, 416–434. doi:10.1080/20445911.2011.526599
- Cacciari, C., & Padovani, R. (2007). Further evidence on gender stereotype priming in language: Semantic facilitation and inhibition on Italian role nouns. *Applied Psycholinguistics*, 28, 277–293. Retrieved from <http://journals.cambridge.org/action/displayFullText?type=1&fid=836536&jid=APS&volumeId=28&issueId=02&aid=836532&bodyId=>
- Carreiras, M., Garnham, A., & Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns: Evidence from Spanish. *European Journal of Cognitive Psychology*, 5(1), 93–116. doi:10.1080/09541449308406516
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A, 639–663. Retrieved from <http://dx.doi.org/10.1080/71375647>



- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influence of context and world knowledge. *Journal of Memory and Language*, 50, 268–288. doi:10.1016/j.jml.2003.11.003
- Duden (1995). *Die Grammatik/Grammarl*. Volume 4. Mannheim: Dudenverlag.
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye-movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, 32, 551–559. doi:10.3758/BF03195846
- Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, 33, 103–119.
- Foss, D. J., & Speer, S. R. (1991). Global and local context effects in sentence processing. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives* (pp. 115–139). Hillsdale, NJ: Erlbaum.
- Fraurud, K. (1990). Definiteness and the processing of noun phrases in natural discourse. *Journal of Semantics*, 7, 395–433. doi:10.1093/jos/7.4.395
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210. doi:10.1016/0010-0285(82)90008-1
- Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au-pairs are rarely male: Role names' gender stereotype information across three languages. *Behavior Research Methods*, 40, 206–212. doi:10.3758/BRM.40.1.206
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.
- Garrod, S. C., & Sanford, A. J. (1990). Referential processing in reading: Focusing on roles and individuals. In D. A. Ballota, G. B. Floresd' Arcais & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 465–486). Hillsdale, NJ: Lawrence Erlbaum.
- Garrod, S. C., & Sanford, A. J. (1995). Incrementality in discourse understanding. In D. Milward & P. Sturt (Eds.), *Incremental interpretation: Edinburgh working papers in cognitive science* (vol. 11, pp. 99–122). Edinburgh: Centre for Cognitive Science, The University of Edinburgh.
- Garrod, S. C., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, 42, 526–544. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0749596X99926943>
- Gygax, P., Gabriel, U., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). There is no generic masculine in French and German: When beauticians, musicians and mechanics are all men. *Language and Cognitive Processes*, 23, 464–485. doi:10.1080/01690960701702035
- Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Research*, 1230, 177–191. doi:10.1016/j.brainres.2008.07.004
- Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124(1), 62–82. doi:10.1037/0096-3445.124.1.62
- Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, 36, 431–456. doi:10.1007/s10936-007-9053-z
- Irmen, L., Holt, D. V., & Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research*, 1353, 133–144. doi:10.1016/j.brainres.2010.07.018
- Irmen, L., & Schumann, E. (2011). Processing grammatical gender of role nouns: Further evidence from eye-movements. *Journal of Cognitive Psychology*, 23, 998–1014. doi:10.1080/20445911.2011.596824
- Kreiner, H., Sturt, P., & Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58, 239–261. doi:10.1016/j.jml.2007.09.003
- MacDonald, J. L., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703. doi:10.1037/0033-295X.101.4.676
- Marslen-Wilson, W. D., Tyler, L. K., & Koster, C. (1993). Integrative processes in utterance resolution. *Journal of Memory and Language*, 32, 647–666. doi:10.1006/jmla.1993.1033
- Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language*, 59, 266–293. doi:10.1016/j.jml.2008.06.002
- Nicol, J. L., & Swinney, D. (1989). The role of structure in co-reference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18(1), 5–19. doi:10.1007/BF01069043
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273–285. doi:10.3758/BF03211283
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62, 1457–1506. doi:10.1080/17470210902816461
- Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton Jr., C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4, 21–49. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/01690968908406362>
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116(1), 71–86. doi:10.1016/j.cognition.2010.04.002
- Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing vs. integration accounts. *Journal of Memory and Language*, 42, 423–443. doi:10.1006/jmla.1999.2691
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562. doi:10.1016/S0749-596X(02)00536-3
- Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 258–318. doi:10.1006/jmla.1994.1014



### Appendix 1

Table A1. Experimental stimuli in the grammatically masculine form (Experiment 1).

#### Male role nouns

Jeden Tag hatte der Dachdecker schöne Aussichten, von oben sah er interessante Dinge.  
 Normalerweise schlief der Nachtwächter jeden Nachmittag, hierdurch hatte er geordnete Tagesrhythmen.  
 Häufig protestierte der Handwerker gegen Schwarzarbeit, natürlich fürchtete er illegale Konkurrenz.  
 Schon immer genoss der Hausmeister großes Vertrauen, infolgedessen besaß er alle Schlüssel.  
 Oft hatte der Elektriker gute Einfälle, regelmäßig plante er neue Projekte.  
 In den letzten Monaten hatte der Tischler viel Arbeit, jetzt brauchte er erholsame Ferien.  
 Offenbar hatte der Mechaniker gute Augen, häufig entdeckte er kleinste Schäden.  
 Oft arbeitete der Informatiker lange Stunden, daher hatte er schmerzende Augen.  
 Saisonbedingt trug der Straßenkehrer regenfeste Kleidung, seit langem hasste er nasses Wetter.  
 Immer bot der Metzger hochwertige Produkte, zuverlässig erfüllte er alle Kundenwünsche.  
 Oft absolvierte der Astronaut besondere Trainingseinheiten, dadurch ertrug er belastende Situationen.  
 Meistens trieb der Mathematiker ausreichend Sport, auf die Dauer brauchte er körperlichen Ausgleich.

#### Female role nouns

Natürlich kannte der Diätberater alle Kassentarife, täglich stellte er mehrere Rechnungen.  
 Natürlich mied der Fußpfleger schlechtes Schuhwerk, schließlich kannte er mögliche Folgeschäden.  
 Inzwischen hatte der Florist schlimmen Heuschnupfen, daher suchte er geeignete Jobalternativen.  
 Routinemäßig besuchte der Flugbegleiter diverse Länder, vor allem bevorzugte er exotische Ziele.  
 Oft erfand der Grundschullehrer kreative Aufgaben, immer lobte er gute Ideen.  
 Öfter las der Arzthelfer aktuelle Fachliteratur, dadurch erhielt er wertvolle Informationen.  
 Oft löste der Erzieher schwere Konflikte, offenbar liebte er soziale Brennpunkte.  
 Tatsächlich besaß der Wahrsager normale Fähigkeiten, deswegen nutzte er schlaue Tricks.  
 Abends bekam der Babysitter immer Langeweile, dann suchte er interessante Fernsehsendungen.  
 Oft erzählte der Kindergärtner spannende Geschichten, damit lieferte er wunderbare Unterhaltung.  
 Oft hatte der Geburtshelfer anstrengende Tage, selten bekam er regelmäßigen Schlaf.  
 Täglich verjüngte der Kosmetiker zahlreiche Gesichter, offenbar hatte er nützliche Fertigkeiten.

#### Neutral role nouns

Morgens lief der Skifahrer einige Kilometer, offensichtlich brauchte er tägliche Trainings.  
 Mühelos ertrug der Schwimmer kaltes Wasser, trotzdem hatte er trockene Haut.  
 Jede Woche besuchte der Praktikant neue Abteilungen, bald kannte er alle Arbeitsbereiche.  
 Regelmäßig kaufte der Geiger neue Saiten, offenbar hatte er hohen Verschleiß.  
 Manchmal hatte der Künstler originelle Ideen, anscheinend dachte er ungewöhnliche Dinge.  
 Regelmäßig gab der Musiker theoretischen Unterricht, offenbar schätzte er stabile Einkünfte.  
 Oft recherchierte der Schriftsteller interessante Geschichten, daher erfand er lebendige Romane.  
 Häufig hatte der Schauspieler starkes Lampenfieber, daher brauchte er viel Ruhe.  
 Lange verdiente der Rentner gutes Geld, schließlich hatte er einige Ersparnisse.  
 Regelmäßig hatte der Student wenig Geld, deswegen bevorzugte er billige Wohnungen.  
 In letzter Zeit gab der Sänger viele Benefizkonzerte, damit unterstützte er mehrere Organisationen.  
 Jeden Tag gruppierte der Apotheker eingehende Pakete, zuerst ordnete er vorbestellte Medikamente.

Table A2. Role nouns used in experiments 1 and 2, with rating scores.

<b>Stereotypical gender</b>	<b>Role noun (German, masc.)</b>	<b>English translation</b>	<b>Rating score (1 = male, 7 = female)</b>	
Male	Dachdecker	Roof tiler	1.6	
	Nachtwächter	Night guard	1.6	
	Handwerker	Craftsman	1.8	
	Hausmeister	Janitor	1.9	
	Elektriker	Electrician	1.9	
	Tischler	Carpenter	2	
	Mechaniker	Mechanic	2	
	Informatiker	Computer scientist	2.1	
	Straßenkehrer	Street sweeper	2.1	
	Metzger	Butcher	2.2	
	Astronaut	Astronaut	2.2	
	Mathematiker	Mathematician	2.5	
	Neutral	Skifahrer	Skier	3.8
		Schwimmer	Swimmer	3.9
		Praktikant	Intern	4
		Geiger	Violinist	4
Künstler		Artist	4	
Musiker		Musician	4	
Schriftsteller		Writer	4	
Schauspieler		Actor	4.1	
Rentner		Pensioner	4.1	
Student		Student	4.1	
Sänger		Singer	4.2	
Apotheker		Pharmacist	4.3	
Female		Diätberater	Dietician	5.5
		Fußpfleger	Pedicurist	5.7
		Florist	Florist	5.8
		Flugbegleiter	Flight attendant	5.8
	Grundschullehrer	Primary school teacher	5.8	
	Arzthelfer	Doctor's assistant	5.9	
	Erzieher	Educator	5.9	
	Wahrsager	Fortuneteller	5.9	
	Babysitter	Babysitter	5.9	
	Kindergärtner	Kindergarten teacher	6.1	
Geburtshelfer	Obstetrician	6.3		
Kosmetiker	Beautician	6.5		

# Paper 4

Prominence of Gender Cues in the Assignment of Thematic Roles in German

Yulia Esaulova, Chiara Reali, and Lisa von Stockhausen

University of Duisburg-Essen

Author Note

Yulia Esaulova, Department of Psychology, University of Duisburg-Essen, Essen, Germany; Chiara Reali, Department of Psychology, University of Duisburg-Essen, Essen, Germany; Lisa von Stockhausen (formerly Irmen), Department of Psychology, University of Duisburg-Essen, Essen, Germany.

Correspondence concerning this article should be addressed to Yulia Esaulova, Department of Psychology, Berliner Platz 6-8, Essen 45127, Germany. Tel.: +49 201 1837024; E-mail: [yulia.esaulova@uni-due.de](mailto:yulia.esaulova@uni-due.de).

This research was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant number 237907.

## Abstract

Two eye-tracking experiments examined influences of grammatical and stereotypical gender of role nouns on the assignment of agent and patient roles in locally ambiguous subject- and object-extracted relative clauses in German. Participants ( $N_1 = 32$ ;  $N_2 = 40$ ) read sentences like *Die Flugbegleiterin, die viele Touristen/-innen beobachtet hat/haben, ist aufmerksam* ‘The flight attendant<sub>Female+feminine</sub>, who has observed many tourists<sub>Neutral+feminine/masculine</sub> / whom many tourists<sub>Neutral+feminine/masculine</sub> have observed, is attentive’, where only the auxiliary verb at the end of the relative clause disambiguated each of the two role nouns as a thematic agent or patient. The results reveal a linguistic gender bias: agent roles are assigned easier to grammatically masculine than feminine role nouns and stereotypically neutral than female ones. The opposite pattern is observed in the assignment of patient roles for stereotypical but not grammatical gender. The findings are discussed within the framework of situation model theories, as well as constraint-based and similarity-based interference accounts, while gender is viewed as a dimension of prominence.

*Keywords:* grammatical gender, stereotypical gender, thematic roles, relative clauses, prominence

Prominence of gender cues in the assignment of thematic roles in German

Resolving linguistic ambiguities can reveal tendencies in the use of provided information – or biases towards one of the possible interpretations – that can be easily overlooked otherwise. In this paper we examine whether gender markings may function as cues moderating the assignment of thematic roles in complex relative clause constructions. The processing of sentence (1) may appear as difficult as that of sentence (2), as they both contain object-extracted relative clauses (ORC). Sentences (3) and (4) contain subject-extracted relative clauses (SRC) and therefore (3) may seem equally difficult to comprehend as (4).

(1) *The beautician, whom both designers recognized, is experienced.*

(2) *The artist, whom both designers recognized, is experienced.*

(3) *The beautician, who recognized both designers, is experienced.*

(4) *The artist, who recognized both designers, is experienced.*

However, there is evidence that certain features shared by nouns or noun phrases (e.g., animacy) facilitate the assignment of specific thematic roles, such as when agent roles are assigned easier to animate and patient roles to inanimate nouns (e.g., Bornkessel-Schlesewsky & Schlewsky, 2009; MacDonald, 1994; Trueswell, Tanenhaus, Kello, 1993). If gender is one of such features, then stereotypically female *beautician* may in fact be perceived as a better patient compared to neutral *artist* (i.e., receiving an action) in the ORC sentences, making the comprehension of (1) easier than (2). Similarly, neutral *artist* may be perceived as a better agent compared to stereotypically female *beautician* (i.e., producing an action) in the SRC sentences, making (4) easier than (3). While claims about animate nouns as better agents (Gennari & MacDonald, 2008) and inanimate nouns as poorer agents (Clifton et al., 2003; Just & Carpenter, 1992) have already been supported by empirical evidence, the claim about gender cues as predictors of good or poor agents has



not yet been demonstrated experimentally. The experiments reported in the present paper use relative clause structures in German as a tool to address this issue.

### **Thematic Structure and Prominence Hierarchies**

A number of studies invoked thematic structure to explain biases that influence the interpretation of complex linguistic constructions, such as relative clauses (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; Pickering & Traxler, 1998; Pickering, Traxler & Crocker, 2000; Traxler, Morris, & Seely, 2002). Evaluating the role of various factors in the comprehension of relative clauses, previous research has repeatedly shown that ORCs are more difficult to process than SRCs (Gordon, Hendrick, & Johnson, 2001; Traxler, Williams, Blozis, & Morris, 2005; Staub, 2010). However, Mak, Vonk, and Schriefers (2002, 2006) and Traxler et al. (2002) demonstrated that ORCs with inanimate heads, such as *The movie that the director watched received the prize*, were almost as easy to comprehend as SRCs of the type *The director that watched the movie received a prize*. Thus, the feature of animacy has been shown to modulate the difficulty in the interpretation of relative clauses, showing that linguistic characteristics are a significant factor modulating the likelihood of the assignment of an agent role to one of the two nouns in a sentence.

The expectations readers have about entities possessing certain characteristics to occupy syntactically prominent positions in a sentence can be seen within the framework of the thematic hierarchy hypothesis (e.g., Grimshaw, 1990; Jackendoff, 1987). This hypothesis states the ordering of thematic roles by prominence, with the agent role ranking the highest on the hierarchy of semantic features. At the same time, prominence can be assessed along several dimensions other than thematic agency, such as animacy, definiteness or person, with animate entities ranking over inanimate, definite over indefinite, and first and second person over third (Lamers & de Swart, 2012). Tripartite

animacy hierarchy proposed by Croft (1990) offers a similar ordering by person (first and second over third), NP-type (pronouns over common nouns), and animacy itself (human over non-human animate over inanimate). The definition of the agentive case given by Fillmore as “the typically animate perceived instigator of the action identified by the verb” (Fillmore, 1968, p. 24) indicates the relatedness of the two concepts: agency and animacy. In line with this definition, Yamamoto (1991) suggests that agency presupposes animacy, considering that previous research has named such conceptual properties of agency as intentionality (Davidson, 1971), dynamicity and control (Dik, 1989). These properties are not purely linguistic, which contributes to Yamamoto’s understanding of animacy as an “extra-” or “supra-linguistic” concept, which nevertheless relates to such linguistic phenomena as case-marking, word order, subject selection, and gender.

Even though the interaction of different prominence dimensions remains a subject of debate (e.g., Klein, Guntsetseg & von Heusinger, 2012; Primus, 2012), the principle of harmonic alignment suggests that hierarchies within separate dimensions map onto one another, so that hierarchy within the dimension of animacy, for instance, correlate with that of thematic roles (Lamers, 2012). The processing is facilitated when rankings on different hierarchies point to the same argument in a sentence as being more prominent (Bornkessel-Schlesewsky & Schlewsky, 2008). As a result of such alignment, animacy information can be used in the assignment of thematic roles during language comprehension. Thus, readers seem to have expectations about high-ranked animate entities to rather produce actions represented by the verb (i.e., serve as agents that are high-ranked on a thematic role hierarchy), while low-ranked inanimate entities are expected to rather receive those actions (i.e., serve as patients that are low-ranked on a thematic role hierarchy). This tendency can be regarded as a bias moderating difficulties in the interpretation of syntactically complex sentences.

## **Gender Processing and Agency**

The present investigation extends the current knowledge about biases in linguistic structures by examining the role of grammatical and stereotypical gender in the resolution of relative clauses. In the literature on reference resolution, the integration of grammatical and stereotypical gender cues is widely discussed in terms of mismatch effects which are reflected in longer processing times when stereotypically male (e.g., electrician) or female (e.g., beautician) role nouns co-refer with mismatching information, such as gender suffixes, gender-specific pronouns or noun phrases (e.g., Cacciari, Corradini, Padovani, & Carreiras, 2011; Esaulova, Reali, & von Stockhausen, 2014; Kreiner, Sturt, & Garrod, 2008; Irmen, 2007; Reali, Esaulova, & von Stockhausen, in press). Grammatical and stereotypical gender cues have been shown to affect readers' interpretation of role nouns in highly automatized ways and to strongly influence the comprehension of sentences (e.g., Banaji & Hardin, 1996; Cacciari & Padovani, 2007; Carreiras, Garnham, Oakhill, & Cain, 1996; Esaulova et al., 2014; Irmen, Holt, & Weisbrod, 2010).

In the literature mentioned above, the effects of both grammatical features and stereotype-based connotations of roles and contexts are analysed. On the one hand, this research clearly points at the fact that both gender representations affect reference resolution and, on the other hand, it relates the linguistic and the social psychological understanding of the term gender on a conceptual level. This is in line with social cognition research (e.g., Stahlberg, Braun, Irmen, & Sczesny, 2007), which indicates the association between conceptual and formal gender representations, where the former are expressed through gender stereotypes and the latter through grammatical features, such as gender suffixes. However, neither research on thematic structures (e.g., Clifton et al., 2003; Trueswell & Tanenhaus, 1994), nor research on gender processing (e.g., Carreiras et al., 1996; Cacciari & Padovani, 2007; Irmen, 2007; Esaulova, Reali & von Stockhausen, 2014)

and social cognition (Stahlberg et al., 2007) have ever supposed the link between gender and agency or considered gender a relevant factor in the assignment of thematic roles. Based on these three areas of research, we supposed that gender characteristics of nouns should be examined as constituting another dimension along which prominence of thematic role nouns can be assessed. This perspective draws largely on McRae's account of thematic roles, which views them as concepts formed through everyday experience rather than simply lexical information (McRae, Hare, Elman & Ferretti, 2005). This account, in turn, is based on the situation model theories (Sanford & Garrod, 1981), which emphasize the role of long-term memory in the representation of a linguistically described event. Following Yamamoto, we suggest that, similar to agency, gender presupposes animacy, most certainly in those cases when it points at the sex of the referent. The evidence of animacy-based role assignments (Wang, Schlesewsky, Philipp, Bornkessel-Schlesewsky, 2012) encourages the assumption that gender cues may influence readers' implicit beliefs about good or poor agents in a similar way. This should be reflected through readers' expectations about nouns of certain gender to produce or receive actions represented by the verb in a sentence. In social psychology, male roles are associated with higher status and power and are described as more agentic and less communal than female ones (e.g., Koenig, Mitchell, Eagly, & Ristikari, 2011). Agency and communion are fundamental dimensions of social categorization (e.g., Fiske, Cuddy, & Glick, 2007), in which the former comprises such characteristics as assertion, competence and independence, while the latter is associated with cooperation, warmth and empathy. An extensive research in this field indicates a possible association between agency and gender representations (e.g., masculine and feminine sex roles – Bem, 1981; sex-role socialization – Cross & Madson, 1997; Helgeson, 1994; masculinity – Spence & Buckner, 2000; Koenig et al., 2011).

## Hypotheses

In this study we examine gender represented through grammatical and conceptual characteristics, since research on language-based representation of women and men points at the commonalities in their processing. If gender constitutes an important factor in the assessment of agency, as we would like to argue, we should be able to observe its influences on agency both when it is expressed through gender-marking suffixes or through gender stereotypes, such as in typically male (/neutral)/female occupational role nouns. Predictions made about each of these two types of gender representations are described below in two hypotheses.

The phenomenon of differential object marking described in functional/typological literature (e.g., Aissen, 2003) offers a theoretical frame considering grammatical gender in relation to agency. Differential object marking defines the likelihood of an object to be overtly case-marked as a function of prominence ranking: the higher the prominence, the more likely is an overt case-marking. In German, the case-marking of singular masculine determiners is expressed overtly (*der*<sub>Nominative</sub>; *den*<sub>Accusative</sub>), while feminine determiners in some cases remain unmarked (*die*<sub>Nominative/Accusative</sub>). According to differential object marking, such differentiation of case-marking suggests that masculine gender is more prominent than feminine in German. Since prominence hierarchies underlie grammatical functions according to the concept of harmonic alignment, more prominent subjects should align with masculine role nouns and less prominent objects with feminine ones. This prediction is expressed in *Hypothesis I*, which concerns grammatical gender: If grammatical gender constitutes a prominence dimension and feminine is ranked lower on the hierarchy than masculine, then readers should have expectations about grammatically feminine role nouns to rather function as patients than agents (to receive rather than

produce actions) compared to masculine role nouns, which would be reflected in longer processing times for feminine agents than masculine ones.

Social cognition research indicates theoretical grounds to relate agency and stereotypical gender. If the association between agency and masculinity (e.g., Koenig et al., 2011) described above can be carried over to linguistic terms, then stereotypically male (/neutral) nouns should be good agents and poor patients, while stereotypically female nouns should be good patients and poor agents. *Hypothesis II* is based on these considerations and regards stereotypical gender: If stereotypical gender constitutes a prominence dimension and female is ranked lower on the hierarchy than neutral, then readers should have expectations about stereotypically female role nouns to rather function as patients than agents compared to neutral role nouns, which would become evident through longer processing times for female agents than neutral ones.

In terms of eye-tracking measures, both hypotheses translate into the prediction that longer fixation times and more regressions should occur in sentences where feminine or female role nouns are agents and shorter fixations and fewer regressions in sentences where masculine or neutral role nouns are agents.

### **Overview of the Present Research**

The influence of thematic role characteristics on syntactic variations in language production and comprehension does not seem to be restricted to a particular language (e.g., English – McDonald, Bock, & Kelly, 1993; Spanish – Prat-Sala, 1997; German – Van Nice & Dietrich, 2003). The standard finding that SRCs are interpreted with greater difficulty than ORCs mentioned above also extends to the case of German (e.g., Friederici, Steinhauer, Mecklinger, & Meyer, 1998). Since German is one of the languages in which the verb often appears in clause-final position, the incremental integration of information is more likely to occur rather than a computational mechanism deferring hypotheses about



structure and meaning until the end of the clause (Kamide, Scheepers, & Altmann, 2003). According to McRae's account on thematic roles, conceptual/world knowledge information from thematic role characteristics is computed and used immediately in online language processing (McRae et al., 2005). Our research question consisted in clarifying whether gender cues are relevant indicators of prominence in readers' expectations about thematic agents and patients, which contribute to the rapid integration of information in a sentence. In German, certain combinations of gender and number in nouns of the main and the relative clause make it possible to construct sentences where ORCs and SRCs can only be identified as such by the form of the auxiliary verb at the end of the relative clause but are otherwise identical in structure, as in the following examples (5) and (6):

(5) (SRC) *Die Studenten, die die Fahrradfahrerin übersehen haben, sind verletzt.*

'The students<sub>masculine</sub> who have overlooked the cyclist<sub>feminine</sub> are hurt.'

(6) (ORC) *Die Studenten, die die Fahrradfahrerin übersehen hat, sind verletzt.*

'The students<sub>masculine</sub> whom the cyclist<sub>feminine</sub> overlooked are hurt.'

Such ambiguity in the thematic structure allowed us to vary grammatical and stereotypical gender of role nouns in German sentences containing ORCs and SRCs to test whether gender information is used in the assignment of thematic agent and patient roles. Previous research has shown that the analysis of subject-object ambiguity is influenced by the relative ranking of the arguments on prominence hierarchies (Haupt, 2008). If gender information constitutes a relevant dimension that indeed contributes to the prominence of thematic roles in a sentence, then it should be reflected in the processing of syntactically ambiguous structures, such as German sentences containing SRCs and ORCs mentioned above.

In both experiments reported in this paper, we examined the empirical validity of our theoretical assumptions about gender as a dimension of prominence. The hypotheses were

tested in two experiments designed using locally ambiguous sentences containing SRC and ORC structures, as provided in examples (5) and (6). The identification of role nouns as agents and patients in these sentences was not possible until the auxiliary verb of relative clauses had been reached. Experiment 1 examined the effects of grammatical gender by varying grammatical cues (masculine and feminine) in role nouns that were neutral with regard to stereotypical gender. Experiment 2 extended the focus to stereotypical gender influences and included the variation of grammatical (masculine and feminine) and stereotypical (neutral and female) gender cues of involved role nouns<sup>1</sup>.

### Experiment 1

In Experiment 1 we investigated to which extent grammatical gender functions as a cue to agency and affects the resolution of ambiguous relative clauses.

#### Method

**Participants.** Thirty-two students at the University of Duisburg-Essen (15 male, 17 female, mean age 26.3 years,  $SD = 4.7$ ) were paid to participate. All of them were native speakers of German and had normal or corrected-to-normal vision.

#### Materials.

**Experimental stimuli.** Twenty-four experimental sentences consisted of a main and a relative clause connected by the relative pronoun *die* ‘who/whom<sub>feminine sg/masculine or feminine plural</sub>’ that can be interpreted either as feminine singular or as masculine or feminine plural. Main clauses contained plural forms of 24 role nouns (RN1) which varied in grammatical gender (feminine and masculine, feminine marked by the feminine plural suffix *-innen*) but were neutral with regard to stereotypical gender (e.g., *Student/-innen* ‘students<sub>Neutral+masculine/feminine</sub>’). Relative clauses contained singular forms of 24 role nouns (RN2), all of which were grammatically feminine and neutral with regard to stereotypical gender. The verb in the relative clause was an action verb and necessarily involved two

arguments, while the verb of the main clause was a state verb (see Table A1 for experimental stimuli with masculine RN1). Both main and relative clauses of each sentence were presented simultaneously in one line. In sentences with SRCs, such as (7), RN1 served as agents and RN2 served as patients. According to *Hypothesis I*, masculine RN1 agents were expected to facilitate comprehension compared to feminine ones. In sentences with ORCs, such as (8), RN2 served as agents and RN1 served as patients. Therefore feminine RN1 patients were expected to facilitate comprehension compared to masculine ones.

(7) *Die Student-en/-innen, die die Fahrradfahrerin übersehen haben, sind verletzt.*

‘The students<sub>Neutral+masculine/feminine</sub>, who have overlooked the cyclist<sub>Neutral+feminine</sub>, are hurt.’

(8) *Die Student-en/-innen, die die Fahrradfahrerin übersehen hat, sind verletzt.* ‘The

students<sub>Neutral+masculine/feminine</sub>, whom the cyclist<sub>Neutral+feminine</sub> has overlooked, are hurt.’

All of the sentences had the following fixed structure: determiner + RN1 + relative pronoun + determiner + RN2 + action verb + auxiliary verb + main clause verb + adjective. The identification of a relative clause as subject-extracted (*die* = ‘who’) or object-extracted (*die* = ‘whom’) was not possible until its last word – the auxiliary verb *hat* ‘has’ or *haben* ‘have’ – had been reached.

**Verb and adjective pre-tests.** To exclude potential confounding effects resulting from the context, a series of pre-tests were conducted to ensure that verbs and adjectives used in experimental items did not contain any gender information. The *verb pre-test* consisted of transitive verbs requiring a two-argument structure (e.g., *erkennen* ‘to recognize’), while stereotypically female (e.g., *kochen* ‘to cook’), stereotypically male (e.g., *boxen* ‘to box’), and verbs allowing a different number of arguments (e.g., *versprechen* ‘to promise’) served as pre-test fillers. The *adjective pre-test* consisted of items that were structurally similar to the main clause in experimental items, except that

role nouns were replaced with X (e.g., *X ist aufmerksam* ‘X is thoughtful’). Stereotypically male (e.g., *wetteifernd* ‘competitive’) and stereotypically female adjectives (e.g., *liebevoll* ‘affectionate’) were used as pre-test fillers. A total of 37 participants were asked to rate pre-test items on a scale from 1 (stereotypically male) to 7 (stereotypically female). Only verbs and adjectives with ratings from 3.5 to 4.5 were used in the study.

**Fillers.** To prevent participants from developing reading strategies based on the gender characteristics of role nouns and on ambiguous relative clause structures of experimental items, 24 filler items were constructed. Filler sentences consisted of a main clause and a relative clause connected by the relative pronoun *die* which was followed by an unambiguous nominative or accusative masculine determiner *der/den* (each occurring in 50% of all fillers). All fillers had a fixed structure that imitated the experimental sentences. Plural forms of 24 neutral role nouns (rating score between 3.5 and 4.5 on a 7-point scale from 1 = stereotypically male to 7 = stereotypically female) served as RN1 in main clauses; they were either grammatically masculine or nominalized participles, which do not express grammatical gender. Singular forms of another 24 neutral role nouns served as RN2 in relative clauses.

**Design.** The experimental design included two factors: 1. grammatical gender of RN1 (masculine vs. feminine; within-subjects and within-items), 2. type of relative clause (SRC vs. ORC; within-subjects and within-items). Four randomized lists presented each item in one of the four conditions: 1. masculine RN1 + SRC; 2. masculine RN1 + ORC; 3. feminine RN1 + SRC; 4. feminine RN1 + ORC. Across lists, each item occurred equally often in each condition. Participants were presented with one of the lists, i.e. they received all four conditions and encountered each item only once. One fourth of the sentences (including fillers) was followed by a yes/no comprehension question to ensure that participants read materials carefully enough and understood their content.

**Procedure.** Eye movements were recorded using an Eyelink 1000 eye-tracker with a sampling rate of 1000 Hz and angular resolution of 10-30 min of arc (about 0.15° to 0.5°). Participants were seated 70 cm from the computer screen, at which distance 3.0 characters subtended 1° of visual arc. All experimental sentences were presented in 22-point Lucida Console font and displayed on a single line. Viewing was binocular, but only the dominant eye was recorded. A chin rest was used to minimize head movements.

Participants were tested individually. Before the experiment began, they were instructed to read for comprehension in their normal reading speed, pressing corresponding buttons on a response pad to move on to the next sentence, and to answer questions. Then a calibration procedure with a nine-point grid was performed. Each trial started with the presentation of a fixation point located at the beginning of the sentence to be triggered. Whenever the experimenter judged fixation on the point as inaccurate, re-calibration was carried out. The first four sentences with two questions served as practice trials. The eye-tracking session lasted approximately 20 minutes.

## Results

**Data analysis.** For the analysis of the eye movement data, the experimental sentences were divided into the following *regions* (marked with <brackets> in the example below and in *italics* in the following text): *Die Student-en/-innen, <die die> <Fahrradfahrerin> <übersehen> <hat/haben,> <sind verletzt>* ‘The students<sub>Neutral+masculine/feminine</sub>, whom the cyclist<sub>Neutral+feminine</sub> has overlooked / who have overlooked the cyclist<sub>Neutral+feminine</sub>, are hurt’. We refer to the analysed regions of the relative clause as *relative pronoun* (the relative pronoun with the following determiner), *RN2*, *action verb*, and *auxiliary verb*.<sup>2</sup>

For each region, five reading time measures were computed: *first fixation duration* (the duration of the very first eye fixation on a region entered from the left), *first-pass*

*reading time* (the sum of fixation times from first entering a region from the left until leaving it either to the right or to the left), *regression path* (the sum of fixation times from first entering a region from the left until leaving it to the right, including the time spent regressing to the left of the region), *total fixation time* (the sum of all fixation times on a region), and *regressions into a region* (the percentage of regressions crossing the right boundary of a region during the first pass through the sentence) (see Staub & Rayner, 2007).

Initial stages of data analysis consisted in merging fixations shorter than 70 ms with neighbouring fixations within one character and removing fixations below 70 ms and above 600 ms (2.13% of the data), for previous research on reading had shown that such fixations are not representative of normal acquisition of information (Breen & Clifton, 2011; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). Trials that exceeded the reading time range of total reading time mean plus three *SD* were considered outliers (1.30% of all trials) and were excluded from the analyses. The data were subjected to analyses of variance with the RN1 grammatical gender (masculine vs. feminine) and the relative clause type (SRC vs. ORC) treated as within-subjects and within-items factors. Computations based on the data averaged across participants and across items are referred to as  $F_1$  and  $F_2$  analyses respectively. The analyses were based on residual fixation times after region-length correction (Trueswell, Tanenhaus, & Garnsey, 1994). Table 1 provides means and standard deviations for all measures and regions.

(Table 1 about here)

Table 2 presents the results of analyses of variance.

(Table 2 about here)

Pairwise contrast analyses were performed based on  $F_1$  only in cases when patterns of mean differences were similar and significant in either both  $F_1$  and  $F_2$ , or significant in



one ( $p \leq .05$ ) and marginally significant ( $p \leq .1$ ) in the other analysis. Corresponding  $t$ -test results are reported and interpreted below<sup>3</sup>.

**First fixation durations.** The ANOVA revealed a main effect of the RN1 grammatical gender on the RN2 region with shorter fixations after masculine than feminine RN1,  $M_{\text{masc}} = -11.40$ ;  $M_{\text{fem}} = -1.66$ ,  $t(31) = -2.05$ ,  $SEM = 4.76$ ,  $p = .049$ .

**Regression path.** A main effect of grammatical gender emerged on the *action verb* with shorter fixations after masculine than feminine RN1,  $M_{\text{masc}} = -475.30$ ;  $M_{\text{fem}} = -416.95$ ,  $t(31) = -2.15$ ,  $SEM = 27.18$ ,  $p = .040$ . The ANOVA also revealed an interaction between the RN1 grammatical gender and the relative clause type on the *auxiliary verb*. A  $t$ -test showed shorter fixations of the auxiliary verb in SRCs after masculine compared to feminine RN1 agents,  $M_{\text{mascSRC}} = -198.99$ ;  $M_{\text{femSRC}} = -104.80$ ,  $t(31) = -3.00$ ,  $SEM = 31.45$ ,  $p = .005$ .

**Total fixation time.** A main effect of grammatical gender occurred on the RN2 showing shorter fixations after masculine than feminine RN1,  $M_{\text{masc}} = 22.23$ ;  $M_{\text{fem}} = 104.34$ ,  $t(31) = -2.22$ ,  $SEM = 36.96$ ,  $p = .034$ .

**Regression into a region.** A main effect of grammatical gender showed a tendency for fewer regressions into the RN2 after masculine than feminine RN1,  $M_{\text{masc}} = .60$ ;  $M_{\text{fem}} = .72$ ,  $t(31) = -1.88$ ,  $SEM = .06$ ,  $p = .070$ .

**Response accuracy.** The accuracy in answering comprehension questions during the experiment was 95.6%.

## Discussion

As expected, the grammatical gender of RN1 affected the resolution of ambiguous relative clauses in Experiment 1, which was represented by the interaction between the grammatical gender and the relative clause type. In SRCs, feminine RN1 agents caused more difficulties in processing than masculine. This finding suggests that grammatical

gender may function as a cue to agency in that masculine role nouns are more expected to serve as agents in relative clauses than feminine role nouns. However, the extent to which this result can be generalized remains limited at this point, since no such pattern was observed in ORCs and, though reliable in regression path, the effect in SRCs did not reach significance across measures.

As to the main effect of grammatical gender, most of the examined measures reliably showed that feminine RN1 caused more difficulties in processing compared to masculine RN1. Since agents are more likely to precede patients (e.g., Bornkessel et al., 2005), this may indicate a general tendency in readers to expect masculine rather than feminine agents to be mentioned in a sentence first.

## **Experiment 2**

In Experiment 2 we extended the research question to the stereotypical gender of role nouns. Based on our previous reasoning, not only grammatical but also stereotypical gender of role nouns should contribute to thematic role assignment as gender typicality has repeatedly been shown to prompt the representation of gender congruent referents. Theoretically this can be explained by the situation model approach (Sanford & Garrod, 1981; Sanford & Garrod, 1998), which claims that updating the representation of an event that is being described from long-term memory is a constant process running in parallel to the unfolding of linguistic input. Stereotypical gender information represented linguistically through role nouns likely affects the upgrading process, in which it is mapped onto the world knowledge about typical gender roles. Research on gender processing has shown that both stereotypical and grammatical gender information is relevant for the interpretation of sentences, which manifests itself through mismatch effects when gender cues are incongruent (e.g., Carreiras et al., 1996; Esaulova et al., 2014). Furthermore, are even stronger when gender cues are made more salient such as

through gender markings and or through varying both grammatical and stereotypical gender (e.g., Irmen, 2007). Making gender information more salient through the use of grammatical and stereotypical gender in Experiment 2 should result in both functioning as cues to agency, as predicted in *Hypotheses I and II*.

Therefore, in Experiment 2 we examined to which extent both grammatical gender and stereotypical gender function as cues to agency and affect the resolution of ambiguous relative clauses. For this purpose, we varied the stereotypical gender of RN1 and the grammatical gender of RN2.

## Method

**Participants.** Forty students at the University of Duisburg-Essen (15 male, 25 female, mean age 25.2 years,  $SD = 3.6$ ) were paid to participate. All of them were native speakers of German and had normal or corrected-to-normal vision.

### Materials and procedure.

The preparation of materials and procedural details in Experiment 2 were similar to those of Experiment 1; details in which they differed are described below.

**Experimental stimuli.** Twenty-four experimental sentences consisted of a main and a relative clause connected by the relative pronoun *die* ‘who/whom<sub>feminine sg/masculine or feminine plural</sub>’. Main clauses contained singular forms of 12 stereotypically female and 12 neutral role nouns (RN1), all grammatically feminine (see Examples (9), (10), (11) and (12) below). Relative clauses contained plural forms of 24 neutral role nouns (RN2) that varied in grammatical gender (feminine and masculine, feminine marked by the feminine plural suffix *-innen*; see Table A2 for experimental stimuli with grammatically masculine RN2).

(9) *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet hat, ist aufmerksam.*

‘The flight attendant<sub>Female+feminine</sub>, who has observed many tourists<sub>Neutral+masculine/feminine</sub>, is attentive.’

(10) *Die Studentin, die zwei Fahrradfahrer/-innen übersehen hat, ist verletzt.* ‘The student<sub>Neutral+feminine</sub>, who has overlooked two cyclists<sub>Neutral+masculine/feminine</sub>, is hurt.’

(11) *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet haben, ist aufmerksam.* ‘The flight attendant<sub>Female+feminine</sub>, whom many tourists<sub>Neutral+feminine/masculine</sub> have observed, is attentive.’

(12) *Die Studentin, die zwei Fahrradfahrer/-innen übersehen haben, ist verletzt.* ‘The student<sub>Neutral+feminine</sub>, whom two cyclists<sub>Neutral+feminine/masculine</sub> have overlooked, is hurt.’

In sentences with SRCs, such as (9) and (10), RN1 served as agents and RN2 served as patients. In sentences with ORCs, such as (11) and (12), RN2 served as agents and RN1 served as patients. *Hypothesis I* concerned stereotypically neutral RN2 and predicted grammatically feminine patients and grammatically masculine agents to facilitate comprehension compared to masculine patients and feminine agents. *Hypothesis II* concerned RN1 (which grammatical gender was held constant) and predicted longer processing of stereotypically female than neutral agents and after neutral than stereotypically female patients.

All of the sentences had the following fixed structure: determiner + RN1 + relative pronoun + quantifier<sup>4</sup> + RN2 + action verb + auxiliary verb + main clause verb + adjective. As in Experiment 1, the identification of a relative clause as subject-extracted (*die* = ‘who’) or object-extracted (*die* = ‘whom’) was not possible until the last word of the relative clause had been reached.

**Fillers.** Fillers consisted of a main clause and an unambiguous relative clause connected by relative pronouns *der/den* ‘who/whom<sub>masculine</sub>’ (each occurring in 50% of all fillers). Singular forms of 24 slightly male (rating score: 2.5 to 3.4) and grammatically masculine role nouns served as RN1 in main clauses. Plural forms of 12 nominalized

participles and 12 neutral role nouns (6 grammatically feminine and 6 grammatically masculine) appeared as RN2 in relative clauses.

**Design.** The experimental design included the following three factors: 1. RN1 stereotypical gender (neutral vs. female; within-subjects and between-items), 2. RN2 grammatical gender (masculine vs. feminine; within-subjects and within-items), 3. relative clause type (SRC vs. ORC; within-subjects and within-items). Four randomized lists presented each item with either stereotypically female or neutral RN1 in one of the four conditions: 1. masculine RN2 in SRC; 2. masculine RN2 in ORC; 3. feminine RN2 in SRC; 4. feminine RN2 in ORC. Each participant was presented with one list only, where one fourth of the sentences was followed by a yes/no question to ensure an adequate reading comprehension.

## Results

**Data Analysis.** Details of data analysis and reporting of results were similar to those in Experiment 1, differences are described below. Experimental sentences of Experiment 2 were divided into similar regions as in Experiment 1 (marked with <brackets> in the example below and in *italics* in the following text): *Die Flugbegleiterin, <die viele> <Tourist-en/-innen> <beobachtet> <hat/haben,> <ist aufmerksam.>* ‘The flight attendant<sub>Female+feminine</sub>, who has observed many tourists<sub>Neutral+ masculine /feminine</sub> / whom many tourists<sub>Neutral+ masculine/feminine</sub> have observed, is attentive.’ We refer to the analysed regions of the relative clause as *relative pronoun* (the relative pronoun with the following quantifier), *RN2*, *action verb*, *auxiliary verb*, and to the last two words of the main clause as *spillover*.

Initial stages of data analysis consisted in merging fixations shorter than 70 ms with neighbouring fixations within one character and removing fixations below 70 ms and above 600 ms (1.98% of the data). Trials that exceeded the total reading time mean plus 3 *SD* were considered outliers (1.25% of all trials) and were excluded from the analyses. The

data were subjected to analyses of variance with RN1 stereotypical gender (neutral vs. female) treated as a within-subjects and between-items factor and with RN2 grammatical gender (masculine vs. feminine) and relative clause type (SRC vs. ORC) treated as within-subjects and within-items factors. Means and standard deviations for all measures and regions are given in Table 3.

(Table 3 about here)

Results of analyses of variance are shown in Table 4.

(Table 4 about here)

**First fixation durations.**<sup>5</sup> The ANOVA revealed an interaction on the *action verb* between the RN1 stereotypical gender and the relative clause type. It showed shorter fixations in ORCs preceded by stereotypically female compared to neutral RN1 patients,  $M_{FemaleORC} = 14.83$ ,  $M_{NeutralORC} = 32.63$ ,  $t(38) = 2.44$ ,  $SEM = 7.31$ ,  $p = .020$ .

**First-pass reading time.** A main effect of the RN1 stereotypical gender occurred on the RN2 with the consistent pattern of shorter fixations after stereotypically female than neutral RN1 in the first pass,  $M_{Female} = -151.60$ ,  $M_{Neutral} = -28.78$ ,  $t(38) = 10.46$ ,  $SEM = 11.75$ ,  $p < .001$ .

**Regression path.** As before, the main effects of the RN1 stereotypical gender on the RN2 resulted in shorter fixations after stereotypically female compared to neutral RN1,  $M_{Female} = -357.32$ ,  $M_{Neutral} = -236.09$ ,  $t(38) = -4.52$ ,  $SEM = 26.81$ ,  $p < .001$ . A main effect of the RN2 grammatical gender was also observed on the RN2, with shorter fixations on feminine than masculine RN2,  $M_{masc} = -243.94$ ,  $M_{fem} = -349.47$ ,  $t(38) = 3.75$ ,  $SEM = 28.17$ ,  $p = .001$ .

**Total fixation time.** Consistent with earlier occurrences, the main effect of the RN1 stereotypical gender on the RN2 showed shorter fixations after stereotypically female than neutral RN1,  $M_{Female} = -61.33$ ,  $M_{Neutral} = 78.36$ ,  $t(38) = 5.75$ ,  $SEM = 24.30$ ,  $p < .001$ . The



main effect of the RN2 grammatical gender also occurred on the RN2 with shorter fixations on masculine compared to feminine RN2,  $M_{\text{masc}} = -40.91$ ,  $M_{\text{fem}} = 57.94$ ,  $t(38) = -2.64$ ,  $SEM = 37.43$ ,  $p = .012$ .

The ANOVA revealed an interaction between the RN1 stereotypical gender and the relative clause type on the *action verb* region. In SRCs, there were shorter fixations after neutral than stereotypically female RN1 agents,  $M_{\text{NeutralSRC}} = -52.13$ ,  $M_{\text{FemaleSRC}} = 2.15$ ,  $t(39) = -2.16$ ,  $SEM = 25.18$ ,  $p = .037$ . In ORCs, there were shorter fixations after female than neutral RN1 patients,  $M_{\text{NeutralORC}} = 56.50$ ,  $M_{\text{FemaleORC}} = -52.80$ ,  $t(38) = 3.37$ ,  $SEM = 32.40$ ,  $p = .002$ .

The ANOVA revealed another interaction between the RN2 grammatical gender and the relative clause type showing the same pattern of shorter fixations after masculine than feminine RN2 agents in ORCs on the *action verb*,  $M_{\text{mascORC}} = -53.56$ ,  $M_{\text{femORC}} = 57.25$ ,  $t(38) = -3.32$ ,  $SEM = 33.34$ ,  $p = .002$ ; and on the *auxiliary verb*,  $M_{\text{mascORC}} = -61.15$ ,  $M_{\text{femORC}} = -3.16$ ,  $t(38) = -2.58$ ,  $SEM = 22.49$ ,  $p = .014$ .

**Regressions into regions.** The ANOVA revealed a three-way-interaction on the *relative pronoun* between the RN1 stereotypical gender, the grammatical gender of RN2, and the relative clause type. Follow-up comparisons showed fewer regressions into the region after typically female RN1 followed by masculine than feminine RN2 agents in sentences with ORCs,  $M_{\text{Female/mascORC}} = 0.63$ ,  $M_{\text{Female/femORC}} = 1.02$ ,  $t(39) = -3.14$ ,  $SEM = 0.12$ ,  $p = .003$ .

An interaction between the RN1 stereotypical gender and the relative clause type showed fewer regressions into the *action verb* in SRCs after neutral than typically female RN1 agents,  $M_{\text{NeutralSRC}} = .23$ ,  $M_{\text{FemaleSRC}} = .44$ ,  $t(39) = -3.66$ ,  $SEM = .05$ ,  $p = .001$ . An interaction between the RN2 grammatical gender and the relative clause type also emerged in regressions into the *action verb* region and showed fewer regressions into the region in

ORCs after masculine compared to feminine RN2 agents,  $M_{\text{mascORC}} = .31$ ,  $M_{\text{femORC}} = .50$ ,  $t(38) = -4.30$ ,  $SEM = .04$ ,  $p < .001$ .

**Response accuracy.** The accuracy in answering the comprehension questions during the experiment was 82.85%.

## Discussion

Several interactions between gender cues and the relative clause type revealed the relevance of gender information in the assignment of agent and patient roles to role nouns when resolving ambiguous SRCs and ORCs. The interaction between the RN1 stereotypical gender and the relative clause type showed more difficulties after stereotypically female RN1 agents compared to neutral ones in the processing of SRCs. Similarly, it also showed more difficulties after neutral RN1 patients than stereotypically female ones in the processing of ORCs. These findings indicate that stereotypical gender information functions as a cue to agency in that it reflects readers' expectations about neutral role nouns to rather serve as agents and stereotypically female as patients. Furthermore, the interaction between the RN2 grammatical gender and the relative clause type consistently showed more difficulties after feminine RN2 agents than masculine ones in the processing of ORCs. These findings suggest that grammatical gender information also functions as a cue to agency and reflects readers' expectations about masculine rather than feminine role nouns to serve as agents.

Furthermore, the results revealed characteristic patterns in the processing of gender cues in Experiment 2. The processing of grammatical gender differed from earlier to later stages. During earlier stages (regression path on the RN2), feminine RN2 were processed faster than masculine, while later stages showed more difficulties in the processing of feminine than masculine RN2. This may be due to the lexical priming through the feminine RN1, which was read first and could cause the advantage of RN2 during the early stages of

processing. As to the main effect of stereotypical gender, female RN1 required less processing time than neutral RN1. Keeping in mind that the RN1 grammatical gender was always feminine, this may reflect the congruency between stereotypical and grammatical gender cues, which might be higher in the case of female than neutral RN1 (for similar findings see Esaulova et al., 2014).

### **General Discussion**

Taken together, the results of both experiments provide consistent evidence that confirms our hypotheses concerning the relationship between gender markings and readers' expectations about thematic roles in ambiguous relative clauses. The results of both Experiments 1 and 2 partially support *Hypothesis I* about grammatical gender influences demonstrating that grammatically masculine rather than feminine role nouns are expected to refer to agents, while the processing of patient roles does not seem to be affected directly by grammatical gender cues (see discussion below). Experiment 2 confirms *Hypothesis II* about stereotypical gender influences showing that stereotypically female rather than neutral role nouns are expected to serve as patients, while neutral rather than female role nouns are expected to serve as agents. These results are interesting in different ways.

First of all, these findings indicate that gender – along with animacy and definiteness (Silverstein, 1976) – can be regarded as another relevant dimension in the assessment of prominence of arguments in a sentence. Like other dimensions of prominence, gender cues can be ordered in terms of a hierarchy, with masculine/neutral entities ranking higher than feminine/female ones. Following the principle of harmonic alignment, grammatically masculine references to persons are expected to serve as more thematically prominent roles (i.e., agents) in ambiguous sentences than grammatically feminine ones. In terms of stereotypical gender, stereotypically female references seem to be associated with less

prominent thematic roles (i.e., patients) and neutral ones with more prominent roles (i.e., agents). These findings can be related to research demonstrating influences of other prominence hierarchies on reading (e.g., animacy – Mak et al., 2006; definiteness/specificity – Kretzschmar, Bornkessel-Schlesewsky, Staub, Roehm, & Schlewsky, 2012) and research on ambiguity resolution (e.g., Gennari & MacDonald, 2008; Reali & Christiansen, 2007). This research is relevant from a theoretical perspective, because it provides support for constraint-based accounts that view comprehension difficulties as a function of probabilistic constraints provided by certain types of linguistic information. A number of findings demonstrate that noun animacy, voice (active vs. passive), the use of highly frequent pronominal subjects, and certain kinds of verbs are constraints modulating the comprehension difficulty of relative clauses. The results of our study suggest that grammatical and stereotypical gender can also be considered as such constraints affecting the probability that role nouns will function as agents or patients.

Interestingly, the influence of gender information on sentence processing is more apparent in Experiment 2 compared to Experiment 1, since the interaction effects between gender cues and relative clause types are only documented in sentences with SRCs but not ORCs in Experiment 1. On the one hand, there is some evidence that singular personal references facilitate comprehension compared to plural ones (e.g., Müsseler, Hielscher, & Rickheit, 1995). This facilitation may have reduced differences between masculine and feminine RN2 in Experiment 1 but not Experiment 2 due to the differences in the design. On the other hand, situation model theories propose that people use both linguistic cues and background knowledge, mapping one onto another, when comprehending a text (Zwaan & Radvansky, 1998; Sanford & Garrod, 1998). In particular, Sanford and Garrod (1981) suggest that representations of situations stored in long-term memory (background knowledge) are activated as soon as there is *enough* information provided through

linguistic input. Applying this perspective on the present study, gender information (grammatical and stereotypical) represented linguistically may have activated corresponding representations of situations from long-term memory. However, since stereotypical gender information was only varied in Experiment 2, the linguistic input containing gender information was reduced in Experiment 1. Thus, in Experiment 1 – compared to Experiment 2 – the provided linguistic gender information was not sufficient for ascribing agency in ORCs but only affected the resolution of SRCs.

An alternative interpretation of this difference between Experiments 1 and 2 supposes that the relevance of gender information for comprehension appears to be modulated by its salience in the sentence. Earlier research has shown that a person's sex is often made salient (marked or noted) especially when her or his role is inconsistent with the stereotypical one, thus indirectly reinforcing stereotypes (Stahlberg et al., 2007; Romaine, 2001). The salience of gender cues in Experiment 2 could have made readers more attentive to contrasts in gender information (i.e., masculine vs. feminine, neutral vs. female) and therefore gender appeared more relevant for the resolution of relative clauses than in Experiment 1. Our results demonstrate that the number of gender cues present in a sentence and their variety (grammatical only vs. both grammatical and stereotypical) increase the effect of gender biases during comprehension.

It has to be noted that while grammatical gender effects appear in SRCs in Experiment 1 and in ORCs in Experiment 2, they concern agent and not patient thematic roles in both types of ambiguous sentences. This should not be surprising considering the central role of the agent role for comprehension and its prominence compared to other thematic roles. A closer look at the results, however, reveals that the effects observed for agents always occurred when patients were feminine, while the design of both experiments allowed for both masculine and feminine patients. This indicates that specific agent-patient

combinations were particularly prone to the effect, namely, masculine rather than feminine role nouns were expected to be agents when patients were feminine and not masculine.

As an alternative or complementary explanation to the one based on gender roles, research on the role of working memory in the interpretation of complex sentences demonstrated that two NPs of a different type are processed faster than two similar NPs due to the similarity-based interference (Gordon et al., 2001). One important question that similarity-based interference model leaves open concerns critical dimensions that define similarity. If we consider gender such a dimension, similarity-based interference could account for longer processing times we observed for grammatically feminine compared to masculine agents when patients were also grammatically feminine. However, role nouns that we have studied had a number of commonalities (e.g., animacy, person) and differences (e.g., number, stereotypical gender) and it cannot be determined based on our experiments, which of these dimensions constitute possible sources of interference. Furthermore, this approach does not hold for the observed stereotypical gender effects, where neutral RN1 were found to take shorter in agent and longer in patient roles than female RN1 while RN2 were always neutral.

Even though the similarity-based interference approach cannot be applied to all of our results, it points at the possible interaction between the two NPs mentioned in a sentence. This interaction may be important in predicting the thematic role of NPs, which is highly relevant for our study. Kamide, Altmann and Haywood (2003) argue that the goodness of fit between a NP and a thematic role may depend on the other arguments in addition to the constraints afforded by the verb. In their eye-tracking experiments using a ‘visual-world’ paradigm, on hearing *the girl will ride* or *the man will ride* participants looked more often at a picture of a carousel or a motorbike respectively. Since the verb (*ride*) afforded the same constraints for both agents (*the girl* and *the man*), the most likely

NP to be selected as a patient varied as a function of the agent. As opposed to our study, the interpretation of NP agents in Kamide et al.'s experiments was unambiguous and allowed the prediction of another argument as a patient. In our study, both thematic roles remained ambiguous until the auxiliary verb and therefore which thematic role possibly predicted the other cannot be determined. Nevertheless, the observed effects suggest an interaction, in which feminine patients were more likely to enter into a dependency with grammatically masculine rather than feminine agents.

Research on linguistic biases demonstrates how linguistic structures can help to uncover gender influences that are easily missed otherwise. In this respect, our findings indicate tendencies related to the processing of linguistically represented gender information that can be viewed as linguistic biases. Based on the linguistic category model by Semin and Fiedler (Semin & Fiedler, 1988), who distinguished different levels of abstraction that may be used to describe the same behaviour, research on the linguistic intergroup bias (Maass, Salvi, Arcuri, & Semin, 1989) and expectancy bias (Wigboldus, Semin, & Spears, 2000) demonstrated that expected behaviours were encoded at higher levels of abstraction (using adjectives that are detached from specific behaviours, e.g., *emotional*, *aggressive*) compared to unexpected information, which was encoded at a more concrete level (e.g., via descriptive action verbs that referred to a specific observable event, e.g., *cry*, *hit*). Similarly, the negation bias implies that the use of negations (e.g., *not stupid*, rather than *smart*) is more likely in stereotype-inconsistent compared to stereotype-consistent descriptions (Beukeboom, Finkenauer, & Wigboldus, 2010). In this context, the gender bias revealed in our study can be defined as the tendency to assign thematic agent roles to masculine/neutral rather than feminine/female role nouns.

Finally, these findings demonstrate that the assignment of thematic roles is associated with gender cues in the context of reading comprehension in a similar way as



the concept of agency/communion is associated with gender differences as described in social psychology (e.g., Koenig et al., 2011). While in social psychology masculinity and femininity are considered attributes of agency and communion respectively, linguistic cues marking grammatical and stereotypical gender reveal readers' tendencies to assign agent roles to masculine/neutral rather than feminine/female role nouns and patient roles to female rather than neutral role nouns. This finding is again consistent with constraint-based accounts that view thematic roles as concepts based on world knowledge from everyday experiences rather than syntactic slots void of conceptual content (McRae et al., 2005). Even though grammatical gender may correspond to the biological sex of the referent<sup>6</sup> in case of role nouns used in the present studies while stereotypical gender refers to the probability of distribution of men and women in given occupations, both gender cues influence the assignment of thematic agents/patients during reading.

### **Conclusions**

Our study extends the existing knowledge on gender processing relating it, on the one hand, to research on thematic roles and, on the other hand, to research on linguistic biases (e.g., Maass et al., 1989; Wigboldus et al., 2000). The interpretation of masculine/neutral rather than feminine/female role nouns as instigators of an action is the first evidence of a subtle gender bias surfacing in ambiguous relative clause constructions through gender-based role assignments. The results of both eye-tracking experiments encourage to consider interactions between gender and agency in a broader context, which relates linguistic and social psychological aspects of both concepts. We propose to consider gender as another dimension that can be used when the prominence of thematic roles is assessed to determine their hierarchy. Constraint-based models (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994) can be applied to account for readers' expectation about agents and patients associated with specific gender cues. Implications of grammatical and

stereotypical gender processing described in the framework of situation model theories (e.g., Sanford & Garrod, 1981) can, for example, be relevant in the context of guidelines for gender-fair language which are widely discussed today. Since the relevance of prominence dimensions differs across languages (Aissen, 2003), further directions of the current research aim at establishing the extent of gender-based role assignment across languages.

### **Acknowledgements**

This research was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant number 237907. Various portions of this study were presented at the 11<sup>th</sup> International Symposium of Psycholinguistics (2013, Tenerife, Spain), at the Final Conference of the Initial Training Network on Language, Cognition and Gender (2013, Bern, Switzerland) and at the 23<sup>rd</sup> Annual Meeting of the Society for Text and Discourse (2013, Valencia, Spain). We thank the audience of these conferences for their constructive feedback.

### References

- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21, 435-483.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7(3), 136-141. doi:10.1111/j.14679280.1996.tb00346.x
- Bem, S. L. (1981). *Bem Sex Role Inventory professional manual*. Palo Alto, CA: Consulting Psychologists Press.
- Beukeboom, C. J., Finkenauer, C., & Wigboldus, D. H. J. (2010). The negation bias: When negations signal stereotypic expectancies. *Journal of Personality and Social Psychology*, 99(6), 978-992. doi:10.1037/a0020861
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-Questions. *Journal of Memory and Language*, 34, 774-806. doi:10.1006/jmla.1995.1034
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P 600” effects in language comprehension. *Brain Research Reviews*, 59, 55-73. doi: 10.1016/j.brainresrev.2008.05.003
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistic Compass*, 3, 19-58. doi: 10.1111/j.1749-818X.2008.00099.x
- Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, 64, 153-170. doi:10.1016/j.jml.2010.11.001

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

Cacciari, C., Corradini, P., Padovani, R., & Carreiras, M. (2011). Pronoun resolution in Italian: The role of grammatical gender and context. *Journal of Cognitive Psychology*, 23, 416-434. doi:10.1080/20445911.2011.526599

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A(3), 639-663. doi:10.1080/027249896392531

Clifton, C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49, 317-334. doi:10.1016/S0749-596X(03)00070-6

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, 122, 5-37. doi:10.1037/0033-2909.122.1.5

de Villiers, J. G., & Flusberg, H. B. (1975). Some facts one simply cannot deny. *Journal of Child Language*, 2, 279-286. doi:10.1017/S0305000900001100

Croft, W. (1990). *Typology and Universals*. Cambridge: Cambridge University Press.

Davidson, D. (1971). Agency. In Binkley, R., Bronaugh, R. & Marras, A. (Eds.), *Agent, Action, and Reason* (3-15). Toronto: University of Toronto Press.

Dik, S. C. (1989). *The Theory of Functional Grammar. Part I: The structure of the clause*. Dordrecht: Foris.

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, 29, 781-803. doi: 10.1080/01690965.2013.794295

Fillmore, J. (1968). The case for case. In Bach, E. & Harms, R. T. (Eds.) *Universals in linguistic theory* (pp. 1-88). New York: Holt, Rinehart, and Winston.

Fiske, S.T., Cuddy, A., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Science*, 11, 77-83. doi:10.1016/j.tics.2006.11.005

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

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161-187. doi:10.1016/j.jml.2007.07.004

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1411-1423. doi:10.1037/0278-7393.27.6.1411

Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.

Haupt, F.S. (2008). *The Component Mapping Problem: An Investigation of Grammatical Function Reanalysis in Different Experimental Contexts using event-related potentials*. Leipzig: MPI Series in Human Cognitive and Brain Sciences (Vol. 104).

Helgeson, V. S. (1994). Relation of agency and communion to well-being: Evidence and potential explanations. *Psychological Bulletin*, 116, 412-428. doi:10.1037/0033-2909.116.3.412

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

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

Irmen, L. & Schumann, E. (2011). Processing grammatical gender of role nouns: Further evidence from eye movements. *Journal of Cognitive Psychology*, 23, 998-1014. doi: 10.1080/20445911.2011.596824

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122-149. doi:10.1037/0033-295X.99.1.122

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133-156.

Klein, U., Guntsetseg, D., & von Heusinger, K. (2012). Case in conflict: Embedded Subjects in Mongolian. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (43-64). Dordrecht: Springer.

Koenig, A. M., Mitchell, A. A., Eagly, A. H., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616-642. doi:10.1037/a0023557



Kretzschmar, F., Bornkessel-Schlesewsky, I., Staub, A., Roehm, D., & Schlewsky, M. (2012). Prominence facilitates ambiguity resolution: On the interaction between referentiality, thematic roles and word order in syntactic reanalysis. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (239-271). Dordrecht: Springer.

Kreiner, H., Sturt, P., & Garrod, G. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58, 239-261. doi:10.1016/j.jml.2007.09.003

Lamers, M. J. A. (2012). Argument linearization in Dutch: A multi-factorial approach. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (pp. 121-144). Dordrecht: Springer.

Lamers, M. J. A. & de Swart, P. (2012). The interaction of case, word order and prominence: Language production and comprehension in a cross-linguistic perspective. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (pp. 1-16). Dordrecht: Springer.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, 57(6), 981-993. doi:10.1037/0022-3514.57.6.981

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-201.  
doi:10.1080/01690969408402115

MacDonald, J.L., Bock, J.K., & Kelly, M.H. (1993). Word and word order: semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*, 25, 188-230.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703. doi:10.1037/0033-295X.101.4.676

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*, 50-68. doi:10.1006/jmla.2001.2837

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, *54*, 466-490. doi:10.1016/j.jml.2006.01.001

McRae, K., Hare, M., Elman, J.L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory and Cognition*, *33*, 1174-1184.

Müsseler, J., Hielscher, M., & Rickheit, G. (1995). Focussing in Spatial Models. In G. Rickheit & C. Habel (Eds.), *Focus and coherence in discourse processing* (pp. 35-53). Berlin: de Gruyter.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *101*, 608-631. doi:10.1037/0278-7393.24.4.940

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 447-475. doi:10.1006/jmla.2000.2708

Prat-Sala, M. (1997). The production of different word orders: a psycholinguistic and developmental approach. Ph.D. Dissertation. University of Edinburgh, Edinburgh.

Primus, B. (2012). Animacy, generalized semantic roles, and differential object marking. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence:*

*Interacting Cues in Language Production and Comprehension* (pp. 65-90). Dordrecht: Springer.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4(3-4), SI 21-49. doi:10.1080/01690968908406362

Reali, C., Esaulova, Y., & von Stockhausen, L. (in press). Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics*. doi:10.1017/S0142716414000010

Reali, F., Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57, 1-23. doi:10.1016/j.jml.2006.08.014

Romaine, S. (2001). A corpus-based view of gender in British and American English. In M. Hellinger & H. Bussmann (Eds.), *Gender across languages: The linguistic representation of women and men* (pp. 153-175). Amsterdam: John Benjamins.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language*. Chichester: John Wiley & Sons.

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26(2-3), 159-190. doi:10.1080/01638539809545043

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54, 558-568. doi:10.1037/0022-3514.54.4.558

Silverstein, M. (1976). Hierarchy of features and ergativity. In Dixon, R.M.W. (Ed.), *Grammatical Categories in Australian Languages* (112-171). New Jersey: Humanities Press.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. *Psychology of Women Quarterly*, 24, 44-62.

doi:10.1111/j.1471-6402.2000.tb01021.x

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

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116, 71-86. doi:10.1016/j.cognition.2010.04.002

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327-342). Oxford, UK: Oxford University Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90. doi:10.1006/jmla.2001.2836

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204-224. doi:10.1016/j.jml.2005.02.010

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In Clifton, C., Frazier, L. and Rayner, K. (Eds.), *Perspectives in sentence processing* (pp. 155-179). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trueswell, J. C., Tanenhaus, M. K. & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-318. doi:10.1006/jmla.1994.1014

Trueswell, J. C., Tanenhaus, M. K. & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(3), 528-553.

doi:10.1037/0278-7393.19.3.528

Van Nice, K.Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics*, 41, 825-849.

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The role of animacy in online argument interpretation in Mandarin Chinese. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (91-119). Dordrecht: Springer.

Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology*, 28(1), 5-18. doi:10.1037/0022-3514.78.1.5

Yamamoto, M. (1991). *Animacy and Reference: A Cognitive Approach to Corpus Linguistics*. Amsterdam: John Benjamins.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185. doi:10.1037/0033-2909.123.2.162

---

<sup>1</sup> In order to avoid confounding processing difficulties through direct mismatch of gender cues within a role noun (e.g., *Elektrikerin*<sub>Male+feminine</sub> ‘electrician’; see, e.g. Carreiras et al., 1996; Irmen & Schumann, 2011; Esaulova et al., 2014), stereotypically male role nouns were not used in the study.

<sup>2</sup> All regions of analyses are italicized when they are referred to in the text. The effects on the last two words of the sentence (*spillover*) are not reported, as they did not reach significance in Experiment 1.

---

<sup>3</sup> As can be seen from the total fixation time measure, ORCs were overall more difficult than SRCs (for more details see Tables 1-4). Due to these differences in the processing of ORCs and SRCs, only *t*-test comparisons within each type of the relative clause are considered relevant and reported in the results section of both Experiment 1 and 2.

Detailed information on relative clause processing is provided in Tables 1, 2, 3, and 4.

Note that the reading of relative clause types is neither analysed nor discussed in terms of processing costs or the accessibility of particular structures during reanalysis, since this would go beyond the scope of the research question in this paper.

<sup>4</sup> Quantifiers were used instead of determiners in order to unambiguously refer to plural RN2 avoiding the misinterpretation of the determiner *die* ‘the<sub>feminine sg/masculine or feminine plural</sub>’ as feminine singular instead of plural and the following gender incongruity between the determiner and the role noun when plural forms of masculine role nouns were identical to singular ones (e.g., *die Jogger* ‘the<sub>feminine sg/plural joggers masculine sg/masculine plural</sub>’).

<sup>5</sup> There was no consistent pattern in the interaction between the RN1 stereotypical and RN2 grammatical gender across measures and the main effect of the relative clause type detected on the *relative pronoun* in regression path measure. Therefore, these results are considered irrelevant for the processes under study and are not reported or interpreted in the text but included in Table 4 only.

<sup>6</sup> Due to frequent grammatical gender to sex mapping in role nouns but also considering the generic use of masculine forms in German language, our experimental manipulation should not be regarded as perfectly correlated with or independent of the actual sex of the referent.

## GENDER PROMINENCE IN THEMATIC ROLES

Table 1  
Means (and Standard Deviations) of Residual Fixation Times and Probabilities of Regressions (Experiment 1)

Region	Factor		Measure <sup>a</sup>									
	Grammatical Gender (RN 1)	RC	FF	FP	RP	TT	RI					
Relative Pronoun	Masculine	ORC	-6.64	(32.10)	39.39	(47.24)	90.72	(103.35)	31.41	(226.18)	0.62	(0.38)
		SRC	-16.32	(34.17)	24.48	(40.92)	69.00	(102.70)	55.66	(278.37)	0.69	(0.38)
	Feminine	ORC	-7.56	(35.30)	34.84	(45.12)	66.72	(94.84)	91.35	(290.04)	0.66	(0.54)
		SRC	-7.60	(35.89)	28.17	(43.25)	48.75	(67.04)	67.68	(268.37)	0.63	(0.38)
Role Noun 2	Masculine	ORC	-10.05	(28.50)	-32.64	(150.76)	-430.62	(166.34)	2.84	(325.70)	0.55	(0.40)
		SRC	-12.75	(28.04)	-63.74	(138.64)	-404.34	(193.30)	41.62	(419.55)	0.66	(0.54)
	Feminine	ORC	-1.96	(35.44)	-28.05	(151.77)	-384.70	(255.22)	122.38	(435.94)	0.71	(0.56)
		SRC	-1.35	(36.42)	-31.46	(129.69)	-403.42	(216.15)	86.29	(387.86)	0.74	(0.54)
Action Verb	Masculine	ORC	9.91	(44.25)	-77.63	(89.46)	-482.87	(152.10)	-87.06	(285.56)	0.42	(0.33)
		SRC	30.21	(50.33)	-96.34	(91.14)	-467.74	(219.61)	-126.95	(258.65)	0.36	(0.37)
	Feminine	ORC	11.31	(37.94)	-68.20	(94.24)	-370.92	(268.55)	-20.12	(362.35)	0.39	(0.28)
		SRC	17.23	(38.45)	-99.01	(82.69)	-462.98	(125.58)	-97.64	(321.11)	0.36	(0.33)
Auxiliary Verb	Masculine	ORC	12.84	(56.07)	29.37	(64.01)	74.04	(214.11)	-21.87	(198.10)	0.15	(0.15)
		SRC	7.23	(46.82)	-54.36	(49.58)	-198.99	(120.27)	-117.23	(190.79)	0.19	(0.26)
	Feminine	ORC	-1.70	(44.08)	12.81	(52.05)	26.17	(213.17)	-46.74	(203.25)	0.17	(0.25)
		SRC	8.65	(57.77)	-41.55	(77.46)	-104.80	(208.28)	-110.17	(203.80)	0.15	(0.19)

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region



## GENDER PROMINENCE IN THEMATIC ROLES

Table 2  
Results of Analyses of Variance for All Regions of Interest (Experiment 1)

Measure <sup>a</sup>	Region	Effect <sup>b</sup>	F <sub>1</sub>	df <sub>1,2</sub>	F <sub>2</sub>	df <sub>1,2</sub>	
FF	Relative Pronoun	GG	1.50	1, 30	<1		
		RC	2.70	1, 30	<1		
		GG * RC	1.88	1, 30	2.21	1, 23	
	Role Noun 2	GG	4.20**	1, 31	4.53**	1, 23	
		RC	<1		<1		
		GG * RC	<1		<1		
	Action Verb	GG	1.43	1, 31	1.20	1, 23	
		RC	6.20**	1, 31	7.56**	1, 23	
		GG * RC	3.20*	1, 31	3.47*	1, 23	
	Auxiliary Verb	GG	1.19	1, 31	<1		
		RC	<1		<1		
		GG * RC	1.84	1, 31	1.71	1, 23	
	FP	Relative Pronoun	GG	<1		<1	
			RC	5.75**	1, 30	2.50	1, 23
			GG * RC	<1		<1	
Role Noun 2		GG	1.37	1, 31	2.94*	1, 23	
		RC	1.35	1, 31	<1		
		GG * RC	1.07	1, 31	<1		
Action Verb		GG	<1		<1		
		RC	5.17**	1, 31	3.22*	1, 23	
		GG * RC	<1		<1		
Auxiliary Verb		GG	<1		<1		
		RC	56.96***	1, 31	112.71***	1, 23	
		GG * RC	5.03**	1, 31	2.70	1, 23	
RP		Relative Pronoun	GG	3.31*	1, 30	4.03*	1, 23
			RC	3.35*	1, 30	1.49	1, 23
			GG * RC	<1		<1	
	Role Noun 2	GG	<1		1.61	1, 23	
		RC	<1		<1		
		GG * RC	<1		<1		
	Action Verb	GG	4.61**	1, 31	5.07**	1, 23	
		RC	1.96	1, 31	1.49	1, 23	
		GG * RC	2.68	1, 31	6.62*	1, 23	
	Auxiliary Verb	GG	<1		<1		
		RC	31.23***	1, 31	36.60***	1, 23	
		GG * RC	7.76***	1, 31	3.98*	1, 23	
	TT	Relative Pronoun	GG	1.05	1, 31	<1	
			RC	4.78**	1, 31	3.08*	1, 23
			GG * RC	<1		<1	
Role Noun 2		GG	4.94**	1, 31	4.52**	1, 23	
		RC	<1		<1		
		GG * RC	<1		1.70	1, 23	
Action Verb		GG	5.68**	1, 31	2.12	1, 23	
		RC	2.63	1, 31	1.93	1, 23	
		GG * RC	<1		<1		
Auxiliary Verb		GG	<1		<1		
		RC	15.27***	1, 31	13.25***	1, 23	
		GG * RC	<1		<1		
RI		Relative Pronoun	GG	<1		<1	
			RC	<1		<1	
			GG * RC	1.23	1, 30	<1	
	Role Noun 2	GG	3.52*	1, 31	5.18**	1, 23	
		RC	<1		1.14	1, 23	
		GG * RC	<1		<1		
	Action Verb	GG	<1		<1		
		RC	<1		<1		
		GG * RC	<1		<1		
	Auxiliary Verb	GG	<1		<1		
		RC	<1		<1		
		GG * RC	<1		<1		

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

<sup>b</sup> GG: RN1 grammatical gender, RC: relative clause type; \*  $p \leq .1$ , \*\*  $p \leq .05$ , \*\*\*  $p \leq .01$ .

## GENDER PROMINENCE IN THEMATIC ROLES

Table 3  
Means (and Standard Deviations) of Residual Fixation Times and Probabilities of Regressions (Experiment 2)

Region	Factor		Measure <sup>a</sup>												
	Stereotypical Gender (RN 1)	Grammatical Gender (RN 2)	RC	FF	FP	RP	TT	RI							
Relative Pronoun	Neutral	Masculine	SRC	-6.69	(59.66)	76.91	(120.80)	-119.27	(387.53)	79.66	(377.59)	0.55	(0.59)		
			ORC	-5.18	(63.48)	84.26	(132.06)	-230.05	(178.16)	215.39	(440.93)	0.97	(0.92)		
		Feminine	SRC	-11.86	(53.55)	63.57	(103.96)	-208.03	(179.02)	39.01	(275.32)	0.68	(0.59)		
			ORC	-9.04	(47.19)	77.25	(120.72)	-204.85	(217.84)	175.80	(401.97)	0.82	(0.80)		
		Female	Masculine	SRC	-5.81	(62.12)	68.63	(135.36)	-155.74	(223.48)	69.99	(238.68)	0.64	(0.50)	
				ORC	-12.93	(50.32)	57.52	(127.28)	-207.93	(217.41)	26.38	(316.38)	0.62	(0.47)	
	Female	Feminine	SRC	-2.38	(60.47)	54.39	(117.85)	-199.39	(234.81)	38.53	(316.44)	0.77	(0.61)		
			ORC	-4.39	(49.83)	39.84	(113.25)	-253.88	(200.92)	157.08	(389.22)	1.04	(0.93)		
		Role Noun 2	Neutral	Masculine	SRC	-10.86	(44.28)	15.69	(132.44)	-170.15	(498.75)	-5.65	(347.26)	0.53	(0.53)
					ORC	-3.30	(59.54)	10.29	(120.80)	-201.23	(269.94)	72.18	(382.39)	0.65	(0.55)
				Feminine	SRC	-1.84	(51.66)	-79.30	(174.62)	-263.25	(271.28)	68.58	(416.26)	0.50	(0.38)
			ORC		5.92	(63.58)	-61.79	(211.81)	-309.73	(260.85)	178.33	(502.83)	0.70	(0.58)	
Female	Masculine		SRC	1.51	(48.60)	-56.14	(104.00)	-269.76	(260.10)	-120.20	(280.26)	0.41	(0.45)		
			ORC	0.21	(49.13)	-66.27	(97.87)	-334.60	(170.85)	-109.97	(266.59)	0.42	(0.40)		
	Feminine	SRC	-14.69	(55.86)	-223.80	(200.85)	-391.85	(289.98)	-69.15	(404.61)	0.49	(0.42)			
ORC		-16.18	(44.46)	-260.21	(144.95)	-433.05	(264.75)	54.01	(509.84)	0.67	(0.66)				
Action Verb	Neutral	Masculine	SRC	19.13	(62.05)	1.58	(122.96)	-257.67	(223.06)	-68.81	(294.77)	0.29	(0.40)		
			ORC	37.20	(68.42)	8.55	(114.79)	-248.75	(185.85)	12.51	(276.63)	0.29	(0.31)		
		Feminine	SRC	15.89	(50.93)	32.83	(131.85)	-274.80	(179.46)	-61.06	(276.79)	0.25	(0.30)		
			ORC	28.06	(52.70)	4.38	(98.17)	-223.80	(226.16)	100.48	(391.11)	0.53	(0.46)		
		Female	Masculine	SRC	35.84	(79.07)	54.52	(137.50)	-246.28	(166.91)	-17.59	(229.02)	0.43	(0.46)	
				ORC	14.74	(56.89)	-13.76	(131.79)	-312.53	(176.33)	-119.62	(204.83)	0.33	(0.31)	
	Female	Feminine	SRC	24.96	(67.85)	22.47	(160.96)	-243.94	(241.46)	-11.89	(356.03)	0.43	(0.40)		
			ORC	14.92	(59.72)	-11.97	(122.52)	-308.96	(201.05)	14.02	(335.21)	0.47	(0.49)		
		Auxiliary Verb	Neutral	Masculine	SRC	-5.85	(56.01)	270.31	(67.36)	-12.95	(300.18)	-47.52	(178.56)	0.11	(0.26)
					ORC	4.39	(64.05)	107.65	(71.11)	-153.49	(254.82)	-66.14	(202.01)	0.20	(0.27)
				Feminine	SRC	-22.92	(55.10)	255.29	(55.96)	-19.88	(273.40)	-86.03	(151.87)	0.18	(0.27)
			ORC		27.57	(67.86)	127.45	(70.60)	-134.44	(177.66)	16.64	(228.29)	0.27	(0.33)	
Female	Masculine		SRC	-18.63	(57.22)	265.30	(82.73)	-20.45	(263.66)	-65.86	(170.52)	0.12	(0.23)		
			ORC	24.29	(71.49)	140.81	(102.60)	-75.19	(212.34)	-56.17	(202.17)	0.15	(0.24)		
	Feminine	SRC	-18.18	(42.38)	262.94	(54.05)	-37.98	(208.08)	-65.14	(161.11)	0.08	(0.19)			
ORC		25.43	(57.81)	126.82	(80.96)	20.20	(518.06)	-22.96	(207.62)	0.27	(0.31)				
Spillover	Neutral	Masculine	SRC	16.49	(60.14)	-13.89	(233.38)	1149.69	(1727.75)	-13.70	(561.98)	/	/		
			ORC	29.62	(73.52)	-119.84	(167.59)	1407.02	(1369.92)	-87.10	(345.17)	/	/		
		Feminine	SRC	27.70	(67.80)	-68.33	(196.86)	1057.21	(1304.67)	-81.68	(462.12)	/	/		
			ORC	22.58	(75.44)	-96.37	(177.20)	1737.33	(1691.87)	-63.88	(331.50)	/	/		
		Female	Masculine	SRC	4.54	(51.56)	-35.40	(169.66)	1057.31	(1246.58)	-83.63	(299.91)	/	/	
				ORC	13.09	(69.57)	-45.25	(231.91)	931.58	(1156.08)	-163.34	(308.70)	/	/	
	Female	Feminine	SRC	23.18	(57.20)	-40.21	(206.66)	1266.95	(1563.36)	-52.75	(421.59)	/	/		
			ORC	20.74	(63.14)	-59.26	(270.29)	1646.30	(1542.78)	-20.48	(367.87)	/	/		

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

## GENDER PROMINENCE IN THEMATIC ROLES

Table 4  
*Results of Analyses of Variance for All Regions of Interest (Experiment 2)*

Measure <sup>a</sup>	Region	Effect <sup>b</sup>	F <sub>1</sub>	df <sub>1,2</sub>	F <sub>2</sub>	df <sub>1,2</sub>	
FF	Relative Pronoun	SG	<1		<1		
		GG	<1		<1		
		RC	<1		<1		
		SG * GG	1.42	1, 38	<1		
		SG * RC	<1		<1		
		GG* RC	<1		<1		
		SG * GG * RC	<1		<1		
	Role Noun 2	SG	<1		<1		
		GG	<1		<1		
		RC	<1		1.06	1, 22	
		SG * GG	7.76**	1, 38	6.53**	1, 22	
		SG * RC	<1		1.02	1, 22	
		GG* RC	<1		<1		
		SG * GG * RC	<1		<1		
	Action Verb	SG	<1		<1		
		GG	<1		<1		
		RC	<1		<1		
		SG * GG	<1		<1		
		SG * RC	8.31***	1, 38	5.57**	1, 22	
		GG* RC	<1		<1		
		SG * GG * RC	<1		<1		
	Auxiliary Verb	SG	<1		<1		
		GG	<1		<1		
		RC	36.51***	1, 35	58.02***	1, 22	
		SG * GG	<1		<1		
		SG * RC	1.20	1,35	2.95*	1, 22	
		GG* RC	3.70*	1, 35	2.57	1, 22	
		SG * GG * RC	2.82	1, 35	3.02*	1, 22	
	Spillover	SG	1.64	1, 38	1.50	1, 21	
		GG	2.48	1, 38	1.03	1, 21	
		RC	<1		<1		
		SG * GG	<1		<1		
		SG * RC	<1		<1		
		GG* RC	2.33	1, 38	1.47	1, 22	
		SG * GG * RC	<1		<1		
	FP	Relative Pronoun	SG	4.49**	1, 38	2.07	1, 22
			GG	1.38	1, 38	<1	
			RC	<1		<1	
			SG * GG	<1		<1	
			SG * RC	<1		1.54	1, 22
			GG* RC	<1		<1	
			SG * GG * RC	<1		<1	
		Role Noun 2	SG	109.32***	1, 38	6.65**	1, 22
			GG	44.02***	1, 38	66.01***	1, 22
			RC	<1		<1	
			SG * GG	10.46***	1, 38	9.05**	1, 22
			SG * RC	1.70	1, 38	<1	
GG* RC			<1		<1		
SG * GG * RC			<1		<1		
Action Verb		SG	<1		<1		
		GG	<1		<1		
		RC	3.09*	1, 38	2.14	1, 22	
		SG * GG	1.18	1, 38	<1		
		SG * RC	4.12**	1, 38	1.84	1, 22	
		GG* RC	<1		<1		
		SG * GG * RC	1.60	1, 38	1.39	1, 22	
Auxiliary Verb		SG	1.33	1, 35	2.07	1, 22	
		GG	<1		<1		
		RC	227.08***	1, 35	525.63***	1, 22	
		SG * GG	<1		<1		
		SG * RC	1.30	1, 35	1.37	1, 22	
		GG* RC	<1		<1		
		SG * GG * RC	2.24		1.70	1, 22	
Spillover		SG	2.35	1, 38	<1		
		GG	<1	1, 38	1.30	1, 22	
		RC	6.14**	1, 38	3.10*	1, 22	
		SG * GG	<1		<1		
		SG * RC	2.19	1, 38	1.22	1, 22	
		GG* RC	<1		<1		
		SG * GG * RC	1.38	1, 38	<1		

## GENDER PROMINENCE IN THEMATIC ROLES

Table 4 (Continued)

Measure <sup>a</sup>	Region	Effect <sup>b</sup>	<i>F</i> <sub>1</sub>	df <sub>1,2</sub>	<i>F</i> <sub>2</sub>	df <sub>1,2</sub>	
RP	Relative Pronoun	SG	<1		<1		
		GG	3.39*	1, 38	1.59	1, 22	
		RC	5.72**	1, 38	6.49**	1, 22	
		SG * GG	<1		<1		
		SG * RC	<1		<1		
		GG* RC	<1		1.38	1, 22	
		SG * GG * RC	2.12	1, 38	3.42*	1, 22	
	Role Noun 2	SG	20.45***	1, 38	6.18**	1, 22	
		GG	14.04***	1, 38	5.96**	1, 22	
		RC	2.35	1, 38	1.92	1, 22	
		SG * GG	<1		<1		
		SG * RC	<1		<1		
		GG* RC	<1		<1		
		SG * GG * RC	<1		<1		
	Action Verb	SG	1.83	1, 38	<1		
		GG	<1		<1		
		RC	<1		<1		
		SG * GG	<1		<1		
		SG * RC	6.77**	1, 38	5.03**	1, 22	
		GG* RC	<1		<1		
		SG * GG * RC	<1		<1		
	Auxiliary Verb	SG	2.76	1, 35	<1		
		GG	<1		<1		
		RC	3.01*	1, 35	2.39	1, 22	
		SG * GG	<1		<1		
		SG * RC	6.82**	1, 35	1.76	1, 22	
		GG* RC	1.41	1, 35	1.56	1, 22	
		SG * GG * RC	<1		<1		
	Spillover	SG	1.34	1, 38	<1		
		GG	7.83***	1, 38	6.59**	1, 22	
		RC	8.39***	1, 38	4.23*	1, 22	
		SG * GG	2.74	1, 38	2.22	1, 22	
		SG * RC	3.64*	1, 38	1.75	1, 22	
		GG* RC	5.17**	1, 38	6.01**	1, 22	
		SG * GG * RC	<1		<1		
	TT	Relative Pronoun	SG	4.97**	1, 38	1.66	1, 22
			GG	<1		<1	
			RC	8.66***	1, 38	4.58**	1, 22
			SG * GG	3.57*	1, 38	2.04	1, 22
			SG * RC	5.25**	1, 38	1.68	1, 22
			GG* RC	1.45	1, 38	2.11	1, 22
			SG * GG * RC	1.37	1, 38	1.06	1, 22
		Role Noun 2	SG	33.04***	1, 38	4.58**	1, 22
			GG	6.97**	1, 38	3.36*	1, 22
			RC	4.90**	1, 38	11.65***	1, 22
			SG * GG	<1		<1	
			SG * RC	<1		<1	
			GG* RC	1.73	1, 38	1.12	1, 22
SG * GG * RC			<1		<1		
Action Verb		SG	1.63	1, 38	<1		
		GG	5.62**	1, 38	2.47	1, 22	
		RC	2.99*	1, 38	1.10	1, 22	
		SG * GG	<1		<1		
		SG * RC	20.18***	1, 38	6.26**	1, 22	
		GG* RC	8.63***	1, 38	4.28*	1, 22	
		SG * GG * RC	<1		<1		
Auxiliary Verb		SG	<1		<1		
		GG	2.13	1, 38	<1		
		RC	3.90*	1, 38	<1		
		SG * GG	<1		<1		
		SG * RC	<1		<1		
		GG* RC	5.51**	1, 38	3.80*	1, 22	
		SG * GG * RC	1.24	1, 38	2.16	1, 22	
Spillover		SG	<1		<1		
		GG	1.08	1, 38	<1		
		RC	<1		<1		
		SG * GG	2.17	1, 38	2.93	1, 22	
		SG * RC	<1		<1		
		GG* RC	2.44	1, 38	2.07	1, 22	

## GENDER PROMINENCE IN THEMATIC ROLES

Table 4 (Continued)

Measure <sup>a</sup>	Region	Effect <sup>b</sup>	<i>F</i> <sub>1</sub>	<i>df</i> <sub>1,2</sub>	<i>F</i> <sub>2</sub>	<i>df</i> <sub>1,2</sub>
TT		SG * GG * RC	<1		<1	
RI	Relative Pronoun	SG	<1		<1	
		GG	8.40***	1, 38	2.66	1, 22
		RC	15.32***	1, 38	6.66**	1, 22
		SG * GG	5.70**	1, 38	4.82**	1, 22
		SG * RC	4.03*	1, 38	1.01	1, 22
		GG* RC	<1		<1	
	Role Noun 2	SG * GG * RC	4.94**	1, 38	4.62**	1, 22
		SG	5.66**	1, 38	1.52	1, 22
		GG	3.03*	1, 38	<1	
		RC	8.93***	1, 38	5.94**	1, 22
		SG * GG	3.53**	1, 38	1.37	1, 22
		SG * RC	<1		<1	
	Action Verb	GG* RC	1.29	1, 38	<1	
		SG * GG * RC	<1		<1	
		SG	3.49*	1, 38	2.69	1, 22
		GG	6.66**	1, 38	5.73**	1, 22
		RC	2.19	1, 38	1.17	1, 22
		SG * GG	<1		<1	
Auxiliary Verb	SG * RC	5.20**	1, 38	3.43*	1, 22	
	GG* RC	9.71***	1, 38	5.88**	1, 22	
	SG * GG * RC	<1		<1		
	SG	1.69	1, 35	1.42	1, 22	
	GG	2.22	1, 35	2.99*	1, 22	
	RC	14.85***	1, 35	5.87**	1, 22	
		SG * GG	<1		<1	
		SG * RC	<1		<1	
		GG* RC	1.19	1, 35	1.20	1, 22
		SG * GG * RC	1.67	1, 35	2.15	1, 22

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

<sup>b</sup> SG: RN1 stereotypical gender, GG: RN2 grammatical gender, RC: relative clause type; \*  $p \leq .1$ , \*\*  $p \leq .05$ , \*\*\*  $p \leq .01$ .

# Paper 5

Gender Hierarchies in the Processing of Ambiguous French Anaphors

Yulia Esaulova, Chiara Reali, and Lisa von Stockhausen

University of Duisburg-Essen

Author Note

Yulia Esaulova, Department of Psychology, University of Duisburg-Essen, Essen, Germany; Chiara Reali, Department of Psychology, University of Duisburg-Essen, Essen, Germany; Lisa von Stockhausen (formerly Irmen), Department of Psychology, University of Duisburg-Essen, Essen, Germany.

Correspondence concerning this article should be addressed to Yulia Esaulova, Department of Psychology, Berliner Platz 6-8, Essen 45127, Germany. Tel.: +49 201 1837024; E-mail: [yulia.esaulova@uni-due.de](mailto:yulia.esaulova@uni-due.de).



## Gender Hierarchies in the Processing of Ambiguous French Anaphors

## Abstract

Two eye-tracking experiments investigated whether prominence information about grammatical functions/thematic roles influences readers' expectations about grammatical and stereotypical gender of role nouns in the resolution of French backwards anaphors. Participants ( $N_1 = 25$ ,  $N_2 = 33$ ) read sentences where gender-ambiguous indirect object pronoun *lui* 'him/her' referred to the second role noun that served as an object/patient and varied in grammatical gender (masculine/feminine): *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoureux* 'In fact, the dietician<sub>Female+fem</sub> recommended to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>, a strict plan'. The first role noun served as a subject/agent and varied in stereotypical gender (female/neutral in Experiment 1 and male/neutral in Experiment 2). The results demonstrate that grammatically masculine objects/patients are more difficult for comprehension than feminine ones and male subjects/agents are easier than stereotypically female or neutral ones. The findings suggest that gender characteristics can be conceptualized as prominence hierarchies.

*Keywords:* prominence, grammatical gender, stereotypical gender, grammatical functions, thematic roles

### Gender Hierarchies in the Processing of Ambiguous French Anaphors

The ability to resolve referential structures, such as anaphors, is critical for the comprehension of a natural language. In this paper, we focus on gender cues that modulate the resolution of French backwards anaphors. Previous research has shown that certain linguistic characteristics of thematic roles may influence the production and the comprehension of particular syntactic organizations (e.g., Aissen, 2003). One of such features, which received a great deal of attention in the literature, is animacy. Thus, Ferreira (1994) reported the reduction in default voice selection preferences (active over passive) when the thematic patient in a sentence is animate and/or human (for similar findings in English, see McDonald, Bock & Kelly, 1993; in German, see Van Nice and Dietrich, 2003; in Spanish – Prat-Sala, 1997). Similarly, the feature of animacy was shown to modulate the default preferences in the comprehension of relative clauses (subject- over object-extracted clauses). While object-extracted relative clauses are usually more difficult than their subject-extracted counterparts, Mak, Vonk, and Schriefers (2002, 2006) and Traxler et al. (2002) demonstrated that object-extracted relative clauses with inanimate heads, such as *The movie that the director watched received the prize*, were almost as easy to comprehend as subject-extracted ones of the type *The director that watched the movie received a prize*. These findings suggest that non-syntactic information is a significant factor influencing the processing of thematic roles in syntactically complex sentences.

The empirical evidence described above speaks for the support of the general assumption about close connectedness of animacy and agentivity (e.g., Primus, 2012). Dahl and Fraurud (1996) suggest that the reason for this connectedness between the two lies in the very nature of animacy, which distinguishes between “persons, that is, essentially human beings perceived as agents, and the rest of the universe” (Dahl, 2008, p.

145). The definition of agentive case as “the typically animate perceived instigator of the action identified by the verb” (Fillmore, 1968, p. 24) also points at this relatedness of the two concepts. Based on the properties of agency suggested by previous research, such as intentionality (Davidson, 1971), dynamicity and control (Dik, 1989), Yamamoto (1991) considers that agency presupposes animacy and understands animacy as a “supra-linguistic” concept, which at the same time relates to various linguistic phenomena (e.g., case marking, word order, gender). Such a pervasive nature of animacy, at the same time, goes together with the invisibility of animacy in language grammars (Dahl & Fraurud, 1996), where animate entities often lack a generic way of referring to them (e.g., words for ‘human’ tend to be identical to or derived from words ‘male being’, as in English ‘man’ or French ‘homme’) (Dahl, 2008). Taken for granted and therefore invisible, animacy *per se* is not as crucial as its manifestations in grammars (Primus, 2012). Personhood can be considered as one of such manifestations, following Dahl’s (2008) conclusion that the notion of “personhood” is “quintessential” to animate beings and the agent role. Indeed, research on semantic properties of first, second, and third person in terms of agentivity is well-represented in literature reporting empirical studies that show a hierarchy where first person is more agentive than second and third (e.g., Siewierska, 1993). While personhood is a highly relevant instance of animacy and agency, gender can be seen as the central feature of personhood. In this paper, we would like to take the argumentation of Primus (2012) further by considering another possible manifestation of animacy – namely, gender and its linguistic variations – as a relevant aspect that influences language comprehension.

Together with definiteness and thematic roles, animacy can be considered an inherent property of verbal arguments and characterized as a semantic prominence feature (e.g., Lamers, 2012). As a prominence feature, it is often conceptualized in terms of a hierarchy, in which humans are taken to be higher in prominence than animates and

animates, in turn, rank higher than inanimates (e.g., Aissen, 2002). Even though grammatical functions are not prominence features, they can also be regarded in terms of a hierarchy with subjects outranking objects. Furthermore, grammatical functions can be aligned with prominence hierarchies, where subjects correlate with high-ranked prominence features (e.g., animates) and objects correlate with low-ranked prominence features (e.g., inanimates). Such organization of hierarchies constitutes a so-called harmonic alignment (Aissen, 2003) and can be applied to various prominence features. Thus, Ferreira (1994) investigated thematic roles, where agents rank the highest on the prominence hierarchy, and showed the preference for agents to be placed in the subject position of a sentence. The idea of a connection between semantic prominence features and syntactic grammatical functions is also reflected in the model of Incremental Optimization of Interpretation (de Hoop & Lamers, 2006), which assumes that language users make probabilistic syntactic choices based on several violable constraints. This model describes prominence as one of the constraints that influences the distinction between subjects and objects, with higher probability for subjects to outrank objects in prominence. The violation of the constraints covered by the model (e.g., case, agreement, prominence) occurs when certain information contradicts probabilistic predictions and is reflected through difficulties in language processing.

Until recently, the line of research on gender processing did not regard gender in terms of its prominence. Nevertheless, it has demonstrated the highly automatized way in which gender is processed and the importance of integration of gender information represented in language for an adequate comprehension. Among other paradigms, anaphoric references have often been used to detect processing difficulties when gender cues (suffixes, gender-specific pronouns or gender-marked NPs) of the antecedent and the anaphor do not match, thus producing a so-called mismatch effect (e.g., for evidence in

Spanish and English see Carreiras, Garnham, Oakhill, & Cain, 1996; in Italian – Cacciari, Corradini, Padovani, & Carreiras, 2011; in German – Esaulova, Reali, & von Stockhausen, 2014). Most recently, however, research on gender processing extended its focus to expectations that language users may have about gender-marked entities in terms of thematic roles. In two eye-tracking studies, Esaulova, Reali, and von Stockhausen (2015), examined readers' expectations about agents and patients in sentences with locally ambiguous subject- and object-extracted relative clauses in German (e.g., *Die Flugbegleiterin, die viele Touristen/-innen beobachtet hat/haben, ist aufmerksam* 'The flight attendant<sub>Female+feminine</sub>, who has observed many tourists<sub>Neutral+ masculine/feminine</sub> / whom many tourists<sub>Neutral+ masculine/feminine</sub> have observed, is attentive'). It was observed that agent roles were assigned easier to grammatically masculine (e.g., *Touristen* 'tourists<sub>masculine</sub>') than feminine (e.g., *Touristinnen* 'tourists<sub>feminine</sub>') role nouns and stereotypically neutral (e.g., musician) than female ones (e.g., beautician), while the opposite was true for the assignment of patient roles. The results can be interpreted in terms of a harmonic alignment of two prominence hierarchies – that of thematic roles and gender – that guides readers' expectations and leads to comprehension difficulties when it is violated.

As we have seen earlier, the principle of harmonic alignment predicts the correspondence between the hierarchy of grammatical functions and prominence features (e.g., Bornkessel-Schlesewsky & Schlewsky, 2009), which is reflected in the relative ease or difficulty (if the alignment is violated) of language comprehension. If gender (as a manifestation of animacy) can be considered a prominence feature and its characteristics can be conceptualized in terms of a hierarchy, then we should observe the relative ease in sentence processing where rankings of thematic roles (agents over patients) and grammatical functions of linguistic entities (subjects over objects) correspond to rankings

of their gender cues (grammatically masculine over feminine and stereotypically neutral over female) and the relative difficulty in processing when this correspondence is violated.

In order to examine whether readers make predictions about gender characteristics of entities in terms of a prominence hierarchy, we studied French sentences that contained two role nouns varying in gender characteristics and their grammatical function/thematic role, such as *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoureux* ‘In fact, the dietician<sub>Female+fem</sub> recommended to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>, a strict plan’. In these sentences, the first role noun served as a grammatical subject/thematic agent and the second role noun served as a grammatical object/thematic patient, thus corresponding to the principle of harmonic alignment between grammatical functions and thematic roles. A gender-ambiguous indirect object pronoun *lui* ‘him/her’ referred to the second role noun indicating its grammatical function of an object while leaving the gender specification of the role noun open until it is reached later on during reading. Cataphoric pronouns are reported to initiate an active search for an antecedent (e.g., Cowart & Cairns, 1987; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2006) and therefore should reflect the relevance of the provided information for the resolution of anaphors. In our case, the cataphoric pronoun *lui* provided information about the antecedent as a grammatical object, while expectations regarding the gender of the antecedent could be elicited for it to be aligned with its grammatical function. Taking into consideration previous findings concerning prominence hierarchy of gender (masculine over feminine for grammatical gender – Esaulova et al., 2015) and the principle of harmonic alignment, we should expect a relative facilitation in reading when object/patient antecedents are grammatically feminine rather than masculine if readers process grammatical gender information as relevant in terms of prominence for the resolution of backwards anaphors (*hypothesis 1*).

Theoretical grounds for the relationship between agency and stereotypical gender originating from social cognition (e.g., agency and masculinity – Koenig, Mitchell, Eagly, & Ristikari, 2011), as well as results of previous research (stereotypical gender prominence hierarchy with neutral over female – Esaulova et al., 2015) motivate our *hypothesis II* concerning the prominence of stereotypical gender. The processing of stereotypically male subjects/agents should be easier than that of neutral ones and stereotypically female subjects/agents should be relatively more difficult than neutral ones. Since we used eye-tracking to detect differences in online processing during reading as the methodology offering high spatial and temporal resolution, the predicted relative difficulties would translate into longer fixation times and higher probability of regressions into relevant regions of the sentences under study.

Based on the theoretical assumptions and empirical evidence described above, the two experiments reported below investigate the role of gender in the resolution of backwards anaphors, thus attempting to establish gender as a prominence feature.



## Experiment 1

In Experiment 1 we investigated whether the resolution of gender-ambiguous backwards anaphors can reveal the effects of grammatical gender (masculine/feminine) on thematic patients and the effects of stereotypical gender (female/neutral) on thematic agents.

### Method

**Participants.** Twenty-five students at the University of Fribourg, Switzerland (12 male, 13 female; mean age 22.2 years,  $SD = 1.8$ ), were paid to participate in Experiment 1. All of them were native speakers of French and had normal or corrected-to-normal vision.

### Materials.

**Experimental stimuli.** Twenty experimental sentences contained two role nouns each (see Table A1 in Appendix). First role nouns (RN1) were agents in terms of thematic structure, they served as subjects, were grammatically feminine and varied in stereotypical gender – female (e.g., *diététicien* ‘dietician’) or neutral (e.g., *vétérinaire* ‘veterinarian’). Second role nouns (RN2) were thematic patients, they served as objects, were stereotypically neutral and varied in grammatical gender – masculine or feminine (e.g., *pharmacien*<sub>masc</sub> / *pharmacienne*<sub>fem</sub> ‘pharmacist<sub>masc/fem</sub>’). The gender-ambiguous indirect object pronoun *lui* ‘him/her’ served as a backwards anaphor that referred to the RN2 and its gender-marked demonstrative adjective *ce/cette* ‘this<sub>masc/fem</sub>’ (see Examples (1) and (2)).

(1) *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoureux.* ‘In fact, the dietician<sub>Female+fem</sub> recommended to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>, a strict plan’.

(2) *Toutefois, la vétérinaire lui a apporté, donc à ce/cette pharmacien/pharmacienne, un nouveau livre.* ‘Anyways, the veterinarian<sub>Neutral+fem</sub> brought to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>, a new book.’

The resolution of an ambiguous indirect object pronoun was only possible after the gender-marked demonstrative adjective and RN2 had been reached. All of the sentences had the following fixed structure: adverb, RN1, indirect object pronoun, auxiliary verb, action verb, adverb, demonstrative adjective, RN2, noun phrase. Final noun phrases slightly varied in structure.

**Context neutrality pretest.** A series of pretests were conducted in order to ensure gender neutrality of the context in the experimental stimuli. For the pretest, RN1 and RN2 in items constructed as described above (experimental sentences) were replaced with an X and a Y respectively. Sentences that had the same structure but stereotypically male and female contexts served as fillers. In order to prevent the undesirable effect of item order, two lists were compiled for the presentation of pretest materials. Thirty-six native speakers of French were asked to rate the presented sentences on a scale from 1 to 7 (1 = stereotypically male, 7 = stereotypically female). Only items with ratings from 3.5 to 4.5 were selected for the construction of experimental stimuli of the main study.

**Fillers.** Thirty filler sentences were presented together with the experimental sentences. Ten different filler types were designed to prevent possible reading strategies that could influence reading patterns in experimental sentences. Grammatically feminine and stereotypically female/neutral RN1 in experimental sentences were balanced by filler sentences containing 10 stereotypically male (e.g., *le mécanicien* ‘the mechanic<sub>Male+masc</sub>’) and 10 neutral RN1 (e.g., *le joggeur* ‘the jogger<sub>Neutral+masc</sub>’), all grammatically masculine. Half of these RN1 was followed by neutral RN2 with alternated masculine and feminine grammatical gender, like in experimental sentences (e.g., *Entre autre, le mécanicien lui a*

*passé, donc à cet assistant / cette assistante, la clef de démontage* ‘Besides, the mechanic passed to him/her, so to this assistant, the wrench’). Another half and additional 10 grammatically feminine and neutral RN1 were followed by a reference to masculine and feminine inanimate nouns, which were referring back to a shortened direct object pronoun *l’* that served as a backwards anaphor (e.g., *Du coup, le joggeur, l’a découverte, donc cette route, tout seul* ‘As a result, the jogger discovered it, so this route, all alone’).

**Design.** The experiment had a 2 X 2 design with RN1 stereotypical gender (female or neutral) as a within-subjects but between-items factor and RN2 grammatical gender (feminine or masculine) as a within-subjects and within-items factor. Experimental items were compiled in two randomized lists, which presented each item in one of the two conditions: 1) RN2 feminine or 2) RN2 masculine. Across lists, each item occurred equally often in each condition. Participants were presented with both conditions and encountered each experimental item only once. To ensure that participants read materials carefully and understood their content, one third of all items was followed by a yes/no comprehension question. To avoid emphasizing the experimental manipulation, the questions never directly probed the referent of the anaphor.

**Procedure.** Eye movements were recorded using an Eyelink 1000 eye-tracker with a sampling rate of 1000 Hz and angular resolution of 10-30 min of arc (about 0.15° to 0.5°). Participants were seated 65 cm from the computer screen, at which distance 3.0 characters subtended 1° of visual arc. All experimental sentences were presented in Lucida Console twelve font and displayed on a single line. Viewing was binocular, but only the dominant eye was recorded. A chin rest was used to minimize head movements.

Participants were tested individually. Before the experiment began, they were instructed to read for comprehension in their normal reading speed, pressing corresponding buttons on a response pad to move to the next sentence and to answer questions. Then a

calibration procedure with a nine-point grid was performed. Each trial started with the presentation of a fixation point located at the beginning of the sentence to be triggered. Re-calibration was carried out whenever the experimenter judged fixation on the point as inaccurate. The first four sentences with two questions served as practice trials. The eye-tracking session lasted approximately 20 minutes.

## Results

### Data Analysis.

The analysis of eye movement data was conducted for each of the following *regions* (marked below with <brackets> and in *italics* in the following text): *Toutefois*, <la vétérinaire> <lui a apporté>, <donc à > <ce/cette pharmacien/pharmacienne>, <un nouveau> <livre>. ‘Anyways, <the veterinarian<sub>Neutral+fem</sub>> <to him/her<sub>gender-ambiguous</sub> has brought>, <so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>>, <a new> <book>.’ The regions are referred to as *RN1* (determiner and RN1), *anaphor* (indirect object pronoun plus action verb), *conjunction* (conjunction plus preposition), *RN2* (demonstrative adjective plus RN2), and *spillovers A* (the first two words after the RN2) and *B* (the last word of the sentence).

The six reading measures computed for each region included *first fixation duration* (the duration of the very first eye fixation on a region entered from the left), *first-pass reading time* (the sum of fixation times from first entering a region from the left until leaving it for the first time either to the right or to the left), *regression path* (the sum of fixation times from first entering a region from the left until leaving it to the right, including the time spent regressing to previous regions), *total fixation time* (the sum of all fixation times on a region excluding regressions from this region), *regressions into a region* (the probability of regressions crossing the right boundary of a region during the

first pass through the sentence), *and regressions out of a region* (the probability of regressions crossing the left boundary of a region during the first-pass or full count) (see Staub & Rayner, 2007).

During initial stages of data analysis, fixations shorter than 70 ms were merged with neighboring fixations located within one character. Following Rayner and Pollatsek (1989) and the current practice in eye-tracking research (e.g., Breen & Clifton, 2011), we consider fixations below 70 ms and above 600 ms not representative of normal information extraction during reading. Such fixations were excluded from the analysis (3.12% of data). Finally, trials identified as outliers ( $M + 3 SD$ ) were also excluded (1 % of all trials).

The experiment was analyzed using a 2 X 2 repeated-measures analysis of variance (ANOVA) based on the data averaged across participants ( $F_1$ ) and across items ( $F_2$ ). The analyses of fixation time data are based on residual fixation times after correction for length of regions (Trueswell, Tanenhaus, & Garnsey, 1994). We considered results reliable when they had similar patterns of mean differences and were either significant in both  $F_1$  and  $F_2$  analyses or significant in one ( $p \leq .05$ ) and marginally significant in the other analysis ( $p \leq .1$ ). Follow up analyses of such results are reported as  $t$ -tests based on data averaged across participants. Means and standard deviations of residual fixation times and probabilities of regressions are given in Table 1, results of analyses of variance are given in Table 2<sup>1</sup>.

(Tables 1 and 2 about here)

***First fixation durations.*** The ANOVA revealed a main effect of RN1 stereotypical gender on the *spillover A*, with shorter fixations after neutral compared to female role nouns,  $M_{\text{Neutral}} = 9.58$ ,  $M_{\text{Female}} = 32.66$ ,  $t(24) = -3.33$ ,  $SEM = 6.94$ ,  $p = .003$ .

***First-pass reading time.*** A main effect of RN1 stereotypical gender on the *RNI* showed that neutral role nouns were fixated longer than female ones,  $M_{\text{Neutral}} = 21.95$ ,

$M_{\text{Female}} = -70.04$ ,  $t(24) = 3.02$ ,  $SEM = 30.49$ ,  $p = .006$ . Additionally, a tendency in the preprocessing of RN2 grammatical gender was observed as a main effect on the *conjunction* region, with shorter fixations preceding female compared to masculine objects,  $M_{\text{masc}} = -5.30$ ,  $M_{\text{fem}} = -39.31$ ,  $t(24) = 2.06$ ,  $SEM = 16.53$ ,  $p = .051$ . A main effect of RN2 grammatical gender then also emerged on *spillover* regions. *T*-test contrasts showed no significant differences in the fixations of the *spillover A* region and shorter fixations of *spillover B* after feminine rather than masculine objects,  $M_{\text{masc}} = 11.95$ ,  $M_{\text{fem}} = -23.80$ ,  $t(24) = 1.63$ ,  $SEM = 21.90$ , *ns*;  $M_{\text{masc}} = 107.68$ ,  $M_{\text{fem}} = 10.46$ ,  $t(24) = 2.69$ ,  $SEM = 36.14$ ,  $p = .013$ .

**Total fixation time.** A main effect of RN2 grammatical gender emerged on the *conjunction* region showing longer fixations when it was followed by masculine rather than feminine objects,  $M_{\text{masc}} = 102.95$ ,  $M_{\text{fem}} = -2.41$ ,  $t(24) = 2.87$ ,  $SEM = 36.70$ ,  $p = .008$ .

**Regressions into a region.** A main effect of RN2 grammatical gender showed more regressions into the *conjunction* region in case of masculine compared to feminine objects,  $M_{\text{masc}} = 1.21$ ,  $M_{\text{fem}} = .89$ ,  $t(24) = 2.88$ ,  $SEM = .11$ ,  $p = .008$ .

**Regressions out of a region.** The full count of regressions showed a main effect of RN2 grammatical gender with more regressions out of masculine compared to feminine RN2,  $M_{\text{masc}} = 1.18$ ,  $M_{\text{fem}} = .91$ ,  $t(24) = 2.46$ ,  $SEM = .11$ ,  $p = .022$ .

**Response accuracy.** The mean comprehension question accuracy was 88.75%.

## Discussion

The systematic pattern of longer fixations and more regressions associated with masculine compared to feminine antecedents confirms the predicted relative difficulty of masculine patients/objects compared to feminine ones. When the gender marking of the antecedent specified the gender-ambiguous backwards anaphor *lui* ‘to him/her’ as

masculine, it resulted in slowed down reading and more regressions indicating difficulties in comprehension likely due to the violation of readers' expectations about the patient/object gender. This reading pattern appeared already on the *conjunction* region, possibly reflecting the pre-processing of the first gender marking of the antecedent by the following demonstrative adjective *ce/cette* 'this<sub>masc/fem</sub>'. Regressions into the *conjunction* and out of the *RN2* region, as well as first-pass fixations of *spillover B*, indicate readers' attempts to resolve the gender-ambiguous anaphor that persist into later stages of processing after the gender of the patient/object had been revealed by the demonstrative adjective *ce/cette* 'this<sub>masc/fem</sub>'. Considering the consistency of observed effects across measures and regions, the results provide a reliable evidence for masculine role nouns to be perceived as less likely patients/objects compared to feminine ones and thus outrank feminine role nouns on a prominence hierarchy.

Stereotypical gender information affected comprehension at relatively early stages and appears to show two qualitatively distinct processing patterns during the first reading of the sentence. First, stereotypical gender information seems to be used in the processing of the role noun itself as indicated by differences in the first reading of the *RN1*. Stereotypical gender of a female *RN1* may prepare its interpretation as a reference to a woman and not a man thus facilitating the integration of its feminine grammatical gender. Neutral *RN1*, however, do not possess a cue indicating such an exclusive interpretation and thus take longer to integrate the grammatical feminine markings than stereotypically female *RN1*. Second, shorter first fixations on the *spillover A* after neutral compared to female *RN1* may reflect the integration of stereotypical gender information into the overall structure of the sentence which at this point can be completed. Since neutral *RN1* appear to be integrated easier, the results may suggest that neutral role nouns are perceived as more



likely agents/subjects compared to female ones and thus outrank female role nouns in terms of prominence.

Given the design of Experiment 1, an alternative explanation of the observed grammatical gender effect could be based on the assumption of the pre-activation of feminine grammatical gender (by RN1) before the anaphor and antecedent are reached. Experiment 2 was conducted in order to rule out this explanation.

## Experiment 2

In Experiment 2 we examined the possibility that grammatical gender of RN1 may function as a lexical prime influencing the expectations of the same gender antecedents following a gender-ambiguous backwards anaphor. In this case we should observe the facilitation in the processing of RN2 antecedents with the same gender markings as RN1 (i.e., masculine).

### Method

**Participants.** Thirty-three students at the University of Fribourg, Switzerland (14 male, 19 female; mean age 22.3 years,  $SD = 3.38$ ), were paid to participate in Experiment 2. All of them were native speakers of French and had normal or corrected-to-normal vision.

### Materials.

**Experimental stimuli.** Twenty experimental sentences had the same structure as in Experiment 1 (see Table A2 in Appendix). This time, RN1 were grammatically masculine and varied in stereotypical gender – male or neutral, while RN2 were, as before, stereotypically neutral and varied in grammatical gender – masculine or feminine (see Examples (3) and (4)).

(3) *En vérité, le pompier lui a passé, donc à ce/cette patient/patiente, la masque à oxygène.*

‘In fact, the firefighter<sub>Male+masc</sub> passed to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> patient<sub>Neutral+masc/fem</sub>, an oxygen mask’.

(4) *Toutefois, le vétérinaire lui a apporté, donc à ce/cette pharmacien/pharmacienne, un*

*nouveau livre.* ‘Anyways, the veterinarian<sub>Neutral+masc</sub> brought to him/her<sub>gender-ambiguous</sub>, so to this<sub>masc/fem</sub> pharmacist<sub>Neutral+masc/fem</sub>, a new book.’

**Fillers.** Fillers were constructed similarly to those in Experiment 1. Grammatically masculine and stereotypically male and neutral RN1 in experimental sentences were balanced by filler sentences containing 10 stereotypically female (e.g., *la couturière* ‘the dressmaker’) and 10 neutral RN1 (e.g., *la joggeuse* ‘the jogger’), all grammatically feminine. Half of these RN1 was followed by neutral RN2 with alternated masculine and feminine grammatical gender, like in experimental sentences (e.g., *Enfin, la couturière lui a récité, donc à cet employé / cette employée, l’histoire de l’entreprise* ‘Finally, the dressmaker recited to him/her, so to this employee, the history of the company’). Another half and additional 10 grammatically feminine and neutral RN1 were followed by a reference to masculine and feminine inanimate nouns, which were referring back to a shortened direct object pronoun *l’* that served as a backwards anaphor (e.g., *Du coup, le joggeur, l’a découverte, donc cette route, tout seul* ‘As a result, the jogger discovered it, so this way, all alone’).

**Design and Procedure.** The experiment had a 2 X 2 design that included RN1 stereotypical gender (male or neutral) as a within-subjects but between-items factor and RN2 grammatical gender (feminine or masculine) as a within-subjects and within-items factor. The design and procedure in Experiment 2 were otherwise the same as in Experiment 1.

## Results

**Data Analysis.** The regions of analysis in Experiment 1 and 2 were identical, as the structure of the sentences did not differ. The same criteria applied for the exclusion of non-informative reading data as in Experiment 1 (2.54% of data were removed). Trials that were identified as outliers ( $M + 3 SD$ ) were excluded from the analyses (1.8% of all trials). The same strategies applied to the reporting of results as in Experiment 1. Means and standard deviations of residual fixation times and probabilities of regressions are given in Table 3, results of analyses of variance are given in Table 4.

(Tables 3 and 4 about here)

**Total fixation time.** The first of observed effects was an interaction between RN2 stereotypical gender and RN2 grammatical gender that ANOVA revealed on the *conjunction* region with longer fixations when neutral RN1 were followed by masculine rather than feminine RN2,  $M_{\text{Neutral+masc}} = 23.64$ ,  $M_{\text{Neutral+fem}} = -42.86$ ,  $t(32) = 2.56$ ,  $SEM = 25.99$ ,  $p = .015$ ; and shorter fixations when masculine RN2 followed stereotypically male compared to neutral RN1,  $M_{\text{Male+masc}} = -29.43$ ,  $M_{\text{Neutral+masc}} = 23.64$ ,  $t(32) = -2.55$ ,  $SEM = 20.84$ ,  $p = .016$ .

**Regressions out of a region.** A main effect of RN2 grammatical gender occurred on the *RN2* showing more first-pass regressions out of masculine compared to feminine RN2,  $M_{\text{masc}} = .71$ ,  $M_{\text{fem}} = .49$ ,  $t(32) = 3.30$ ,  $SEM = .07$ ,  $p = .002$ . Furthermore, an interaction between RN1 stereotypical and RN2 grammatical gender also showed more regressions out of masculine compared to feminine RN2 following neutral RN1 both for first-pass and full count regressions,  $M_{\text{Neutral+masc}} = .60$ ,  $M_{\text{Neutral+fem}} = .26$ ,  $t(32) = 2.56$ ,  $SEM = .06$ ,  $p < .001$ ;  $M_{\text{Neutral+masc}} = .86$ ,  $M_{\text{Neutral+fem}} = .56$ ,  $t(32) = 3.32$ ,  $SEM = .09$ ,  $p = .002$ . This interaction also appeared on the *spillover A* region again showing more regressions out of

masculine than feminine RN2 but following male RN1,  $M_{\text{Male+masc}} = .17$ ,  $M_{\text{Male+fem}} = .09$ ,  $t(32) = 2.11$ ,  $SEM = .035$ ,  $p = .043$ .

**Response accuracy.** The mean comprehension question accuracy was 92%.

### Discussion

Like in Experiment 1, longer fixation and more regressions associated with masculine compared to feminine antecedents indicate the relative difficulty in comprehending masculine referents as objects/patients compared to feminine ones. The corresponding reading pattern appears already on the *conjunction* region probably as a result of a preprocessing of the following demonstrative adjective *ce/cette* ‘this<sub>masc/fem</sub>’ that disambiguates the indirect object pronoun *lui* ‘to him/her’. Similar patterns of regressions out of the following RN2 region indicate that readers experience more difficulties resolving the anaphor when the disambiguating patient/object is masculine rather than feminine. These findings may reflect a hierarchy within the grammatical gender where masculine role nouns outrank feminine ones in prominence.

The processing of stereotypical gender information suggests that male RN1 were easier to process as agents than neutral RN1 thus placing stereotypically male role nouns over neutral ones on the prominence hierarchy. While stereotypical gender information appears to be relevant to complete the interpretation of the sentence, readers’ expectations about the grammatical gender of the antecedent were not influenced by RN1 stereotypical gender, since reading patterns were the same after both male and neutral RN1 and indicated more difficulties for masculine than feminine patients/objects.

### General Discussion

Two eye-tracking experiments reported in the present paper provide evidence that the information about grammatical functions (subject/object) and thematic roles (agent/patient) of role nouns systematically elicit readers' expectations about stereotypical and grammatical gender characteristics of these role nouns during the resolution of gender-ambiguous backwards anaphors. The analysis of reading patterns demonstrated a pervasive preference for grammatically feminine antecedents to serve as objects/patients, which was expressed in processing difficulties when antecedents were masculine. Readers expected the antecedents of gender-ambiguous backwards anaphors to be rather grammatically feminine than masculine in both experiments, irrespective of the grammatical gender of the RN1. This rules out the possibility that such a preference could be the result of increased expectations of a feminine rather than masculine RN2 after reading a grammatically feminine RN1. Our findings extend the results of previous research that showed the association between syntactic preferences and thematic roles through prominence features – such as animacy and definiteness (e.g., Mak et al., 2006; Kretzschmar, Bornkessel-Schlesewsky, Staub, Roehm, & Schlewsky, 2012) – to grammatical gender. It should be noted, however, that in both experiments grammatical gender implied the biological sex of human referents, which imposes an important limitation on the generalization of our findings to other grammatical gender instances. At the same time, gender is a common feature underlying all of the three scales in the tripartite animacy hierarchy proposed by Croft (1990) that orders entities by person (first and second over third), NP-type (pronouns over common nouns), and animacy itself (human over non-human animate over inanimate). Thus, in terms of prominence scales, gender hierarchy revealed by our study can at the same time be considered as a subscale of animacy scale, which allows the differentiation between high-ranked masculine and low-ranked feminine human referents.

Stereotypical gender information affected sentence comprehension in both experiments and appears to be used by readers in two different ways. In Experiment 1, readers experienced more difficulties upon the first encounter of neutral than stereotypically female RN1. Once they reached the spillover, however, this pattern reversed and showed more difficulties after female than neutral RN1. The interpretation of this finding may be that stereotypical gender information is first used to process gender characteristics of RN1 and then – once the structure of the sentence is revealed – stereotypical gender of the role noun is integrated within its thematic role and grammatical function. Thus, we first observe an easier integration of grammatically feminine forms with stereotypically female gender than neutral role nouns that are not directly associated with female referents. This is followed by an easier integration of neutral compared to female RN1 with their role as agents/subjects. In Experiment 2, stereotypical gender affects the processing after the indirect object pronoun determines the structure of the sentence, which results in an easier integration of stereotypically male rather than neutral role nouns as agents/subjects. Taken together, our findings suggest a prominence hierarchy within stereotypical gender with male role nouns over neutral and neutral over stereotypically female ones. This is in line with our earlier findings, which demonstrated readers' expectations about gender characteristics of agents and patients in German (Esaulova et al., 2015), which reported a processing advantage for neutral rather than stereotypically female agents in the resolution of relative clauses. Since male role nouns were not used due to the design of experiments in German, their position relative to female and neutral ones in the prominence hierarchy still remains to be clarified. It should also be noted that experimental sentences in the present study did not require the active processing of RN1 stereotypical gender to resolve the anaphor. This may have resulted in stereotypical gender effects occurring both as the integration of RN1 gender cues and as

part of a more general processing of the sentence structure. Furthermore, Aissen (2003) points out that the relevance of prominence dimensions differs across languages. Given otherwise comparable results of the two studies, we consider the extent to which stereotypical gender moderates difficulties in the assignment of thematic roles and whether its relevance as a prominence dimension differs between German and French languages open questions.

While Dahl (2008) attempts to provide a cognitive grounding for the prominence of animate entities using philosophical notions viewing animacy as an ontological type, we would like to offer social cognitive grounds for the prominence of gender. In social cognition, masculinity and femininity are considered attributes of agency and communion respectively (e.g., Koenig et al., 2011), expressing the relation between gender and agentivity. In a similar way, expectation states theory associates gender with social hierarchy and leadership through status beliefs linking greater status and competence with men than women (Ridgeway, 2001). Our results suggest that the same relation can be observed in language through readers' expectations about grammatical functions/thematic roles of linguistic entities to carry specific gender cues. While empirical evidence is needed to further characterize gender influences on agentivity, the characteristics of animacy dimension summarized by Wang, Schlewsky, Philipp, and Bornkessel-Schlewsky (2012) suggest future directions for research when they are applied to gender. In particular, more evidence is needed to clarify whether gender – similar to animacy – can be regarded as a relational feature that comes into play when several arguments must be related to one another. Another characteristic of animacy dimension mentioned by Wang et al. (2012) concerns the strength of applicability cross-linguistically, which varies in case of animacy. Whether the degree to which gender modulates the comprehension depends on the language being processed is another research question yet to be considered.



To conclude, the prominence hierarchy of grammatical gender seems to be organized in the same way in both German and French languages, with masculine role nouns ranking higher on the hierarchy than feminine ones, which results in readers' perception of masculine role nouns as less likely objects/patients compared to feminine ones. As to stereotypical gender, stereotypically male role nouns in French seem to outrank neutral ones and the latter in turn outrank stereotypically female role nouns. This finding is in line with prominence research in German language, which indicates that gender hierarchy in the assignment of thematic roles is certainly not language-specific, even though it may differ across languages. In the research on gender processing, it is common to use grammatical violations to demonstrate gender influences through mismatch effects (e.g., Carreiras et al., 1996). As opposed to such paradigms, our experiments provide evidence for gender-based grammatical function/thematic role assignment in the absence of such violations, showing that the influence of gender information on language processing can be detected during natural language comprehension and should be accounted for in contexts that go beyond that of experimental manipulations. Further research should address the investigation of the ranking of stereotypically male in relation to female gender in terms of a prominence hierarchy, the applicability of gender prominence hierarchy to inanimate or non-human entities, the relation between gender hierarchy and other prominence hierarchies, and the cross-linguistic validity of gender as a prominence feature.

**Acknowledgements**

This research was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 237907. We would like to thank Pascal Gygax for his help with the preparation of experimental materials. We are also grateful for the feedback received from the audience after the presentation of a portion of this study at the AMLaP (2013, Marseille, France) and TAELEC (2013, Madrid, Spain) conferences, which helped us in the preparation of this paper.

### References

- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21, 435-483.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistic Compass*, 3, 19-58. doi: 10.1111/j.1749-818X.2008.00099.x
- Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, 64, 153-170. doi:10.1016/j.jml.2010.11.001
- Cacciari, C., Corradini, P., Padovani, R., & Carreiras, M. (2011). Pronoun resolution in Italian: The role of grammatical gender and context. *Journal of Cognitive Psychology*, 23, 416-434. doi:10.1080/20445911.2011.526599
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A(3), 639-663. doi:10.1080/027249896392531
- Cowart, W., & Cairns, H. (1987). Evidence for an anaphoric mechanism within syntactic processing: some reference relations defy semantic and pragmatic constraints. *Memory and Cognition*, 15, 318-331.
- Croft, W. (1990). *Typology and Universals*. Cambridge: Cambridge University Press.
- Dahl, Ö. (2008). Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua*, 118, 141-150. doi: 10.1016/j.lingua.2007.02.008

Dahl, Ö., & Fraurud, K. (1996). Animacy in grammar and discourse. In Th. Fretheim, & J.-K. Gundel (Eds.), *Reference and Referent Accessibility* (pp. 47–64). Amsterdam: Benjamins.

Davidson, D. (1971). Agency. In Binkley, R., Bronaugh, R. & Marras, A. (Eds.), *Agent, Action, and Reason* (3-15). Toronto: University of Toronto Press.

de Hoop, H., & Lamers, M. J. A. (2006). Incremental distinguishability of subject and object. In L. Kulikov, A. Malchukov, & P. de Swart (Eds.), *Case, valency, and transitivity* (pp. 269-287). Amsterdam: John Benjamins.

Dik, S. C. (1989). *The Theory of Functional Grammar. Part I: The structure of the clause*. Dordrecht: Foris.

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, 29, 781-803. doi: 10.1080/01690965.2013.794295

Esaulova, Y., Reali, C., & von Stockhausen, L. (2015). *Prominence of gender cues in the assignment of thematic roles in German*. Manuscript submitted for publication.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33, 715-736. doi: 10.1006/jmla.1994.1034

Fillmore, C. J. (1968). The Case for Case. In E. Bach, & R.T. Harms (Eds.): *Universals in Linguistic Theory* (pp.1-88). New York: Holt, Rinehart, and Winston.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M. & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, 56, 384-409. doi: 10.1016/j.jml.2006.09.003

Koenig, A. M., Mitchell, A. A., Eagly, A. H., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616-642. doi:10.1037/a0023557

Kretzschmar, F., Bornkessel-Schlesewsky, I., Staub, A., Roehm, D., & Schlewsky, M. (2012). Prominence facilitates ambiguity resolution: On the interaction between referentiality, thematic roles and word order in syntactic reanalysis. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (pp. 239-271). Dordrecht: Springer.

Lamers, M. J. A. (2012). Argument linearization in dutch: A multi-factorial approach. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (pp. 121-144). Dordrecht: Springer.

MacDonald, J.L., Bock, J.K., & Kelly, M.H. (1993). Word and word order: semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*, 25, 188-230.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50-68.  
doi:10.1006/jmla.2001.2837

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, 54, 466-490.  
doi:10.1016/j.jml.2006.01.001

McDonald, J. L., Bock, J. K., Kelly, M. H. (1993). Word and world order: semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*, 25, 188-230.

Prat-Sala, M. (1997). The production of different word orders: a psycholinguistic and developmental approach. Ph.D. Dissertation. University of Edinburgh, Edinburgh.

Primus, B. (2012). Animacy, generalized semantic roles, and differential object marking. In M. J. A. Lamers, & P. de Swart (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (pp. 65-90). Dordrecht: Springer.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Ridgeway, C. L. (2001). Gender, Status and Leadership. *Journal of Social Issues*, 4, 637-655.

Siewierska, A. (1993). On the interplay of factors in the determination of word order. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: An international handbook of contemporary research* (pp.826-846). Berlin: Walter de Gruyter.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327-342). Oxford, UK: Oxford University Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90. doi:10.1006/jmla.2001.2836

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-318. doi:10.1006/jmla.1994.1014

Van Nice, K.Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics*, 41, 825-849.

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The role of animacy in online argument interpretation in Mandarin Chinese. In Lamers, M. J. A. & de Swart, P. (Eds.), *Case, Word Order and Prominence: Interacting Cues in Language Production and Comprehension* (pp. 91-119). Dordrecht: Springer.

Yamamoto, M. (1991). *Animacy and Reference: A Cognitive Approach to Corpus Linguistics*. Amsterdam: John Benjamins.



## Appendix 1

Table A1. Experimental stimuli with grammatically masculine RN2 (Experiment 1).

### Female RN1

A ce moment-là, la banquière lui a indiqué, donc à ce bijoutier, le prix marqué.  
De plus, la caissière lui a fourni, donc à ce voisin, les renseignements requis.  
D'abord, la coiffeuse lui a montré, donc à ce client, des options intéressantes.  
Naturellement, la couturière lui a réservé, donc à ce nageur, un accueil froid.  
Pourtant, la danseuse lui a présenté, donc à ce spectateur, un programme extraordinaire.  
En vérité, la diététicienne lui a recommandé, donc à ce patient, un plan rigoureux.  
De toute façon, la gouvernante lui a glissé, donc à ce piéton, une phrase bizarre.  
Evidemment, la maquilleuse lui a offert, donc à ce chanteur, un service de qualité.  
Cependant, la vendeuse lui a donné, donc à ce comédien, une importance exceptionnelle.  
D'ailleurs, la voyante lui a prédit, donc à ce campeur, un chemin facile.

### Neutral RN1

Bref, la physiothérapeute lui a conseillé, donc à ce cavalier, un minimum d'exercice.  
Toutefois, la vétérinaire lui a apporté, donc à ce pharmacien, un nouveau livre.  
Ainsi, la zoologiste lui a exposé, donc à ce conservateur, les problèmes de la faune.  
Ensuite, la biologiste lui a parlé, donc à ce coureur, des articulations importantes.  
Finalement, la graphiste lui a envoyé, donc à ce greffier, des images pertinentes.  
En effet, la violoniste lui a prêté, donc à ce musicien, un pupitre trop bas.  
En fait, la sténographe lui a expédié, donc à ce correcteur, une copie du discours.  
Par conséquent, la journaliste lui a dédié, donc à ce romancier, un article biographique.  
En somme, la syndicaliste lui a expliqué, donc à ce manifestant, le but de l'évènement.  
Puis, la psychiatre lui a communiqué, donc à ce coordinateur, les difficultés pratiques.

Table A2. Experimental stimuli with grammatically masculine RN2 (Experiment 2).

**Male RN1**

A ce moment-là, le ministre lui a indiqué, donc à ce bijoutier, le prix marqué.  
De plus, le maçon lui a fourni, donc à ce voisin, les renseignements requis.  
D'abord, le barbier lui a proposé, donc à ce client, des options intéressantes.  
Naturellement, le marin lui a réservé, donc à ce nageur, un accueil froid.  
Pourtant, le batteur lui a présenté, donc à ce spectateur, un programme extraordinaire.  
En vérité, le pompier lui a passé, donc à ce patient, la masque à oxygène.  
C'est-à-dire, le couvreur lui a glissé, donc à ce piéton, une phrase bizarre.  
Évidemment, le portier lui a offert, donc à ce chanteur, un service de qualité.  
Eh oui, le gouverneur lui a accordé, donc à ce comédien, une attention exceptionnelle.  
D'ailleurs, le cordonnier lui a prêté, donc à ce campeur, un chemin facile.

**Neutral RN1**

Bref, le physiothérapeute lui a conseillé, donc à ce cavalier, un minimum d'exercice.  
Toutefois, le vétérinaire lui a apporté, donc à ce pharmacien, un nouveau livre.  
Ainsi, le zoologiste lui a exposé, donc à ce conservateur, les problèmes de la faune.  
Ensuite, le biologiste lui a parlé, donc à ce coureur, des articulations importantes.  
Finalement, le graphiste lui a envoyé, donc à ce greffier, des images pertinentes.  
En effet, le violoniste lui a prêté, donc à ce musicien, un pupitre trop bas.  
En fait, le sténographe lui a expédié, donc à ce correcteur, une copie du discours.  
Par conséquent, le journaliste lui a dédié, donc à ce romancier, un article biographique.  
En somme, le syndicaliste lui a expliqué, donc à ce manifestant, le but de l'évènement.  
Puis, le psychiatre lui a communiqué, donc à ce coordinateur, les difficultés pratiques.

---

<sup>1</sup> In both Experiments 1 and 2, the effects on the *anaphor* region are considered irrelevant for the processes under study and therefore these data are neither reported nor interpreted in the text but included in Tables. Similarly, the effects observed in fixations of the *RN2* region are only reported in Tables, as they cannot be interpreted due to their systematic variation in length.

# GENDER PROMINENCE IN FRENCH ANAPHORS

Table 1  
Means (and Standard Deviations) of Residual Fixation Times and Probabilities of Regressions (Experiment 1)

Region	Factors		Measure <sup>a</sup>												
	Stereotypical Gender (RN 1)	Grammatical Gender (RN 2)	FF	FP		RP		TT		RI		RO		ROF	
RN1	Neutral	Masculine	-20.89 (48.53)	26.78 (166.96)	-93.49 (289.93)	181.76 (413.05)	0.92 (0.70)	0.25 (0.26)	0.51 (0.52)						
		Feminine	-16.03 (43.58)	-9.68 (212.76)	-123.41 (330.64)	190.36 (464.41)	0.94 (0.53)	0.28 (0.30)	0.45 (0.39)						
	Female	Masculine	-14.63 (50.83)	-37.83 (138.84)	-174.29 (257.60)	122.57 (458.34)	0.93 (0.73)	0.24 (0.26)	0.52 (0.59)						
		Feminine	-22.62 (41.05)	-64.41 (113.52)	-188.93 (383.97)	112.60 (548.70)	0.92 (0.64)	0.31 (0.36)	0.50 (0.51)						
Anaphor	Neutral	Masculine	-5.89 (51.72)	21.66 (98.37)	-47.02 (562.49)	74.03 (449.39)	0.49 (0.41)	0.24 (0.30)	0.50 (0.44)						
		Feminine	-17.92 (47.02)	30.66 (125.14)	-176.77 (222.63)	54.41 (385.25)	0.59 (0.54)	0.14 (0.16)	0.51 (0.47)						
	Female	Masculine	-10.65 (33.44)	28.69 (126.74)	-82.18 (506.12)	85.67 (463.81)	0.52 (0.42)	0.22 (0.28)	0.50 (0.43)						
		Feminine	10.64 (63.29)	46.08 (135.79)	-176.59 (218.35)	77.05 (390.66)	0.56 (0.42)	0.18 (0.19)	0.58 (0.46)						
Conjunction	Neutral	Masculine	-4.44 (53.28)	-6.77 (77.22)	-231.74 (158.65)	58.33 (272.99)	0.81 (0.43)	0.06 (0.11)	0.25 (0.33)						
		Feminine	-13.23 (45.39)	-29.27 (47.78)	-174.46 (347.47)	-3.23 (224.74)	0.58 (0.46)	0.11 (0.22)	0.26 (0.38)						
	Female	Masculine	-7.52 (36.19)	2.95 (98.04)	-198.67 (243.35)	89.26 (313.53)	0.80 (0.42)	0.10 (0.17)	0.23 (0.30)						
		Feminine	-13.61 (44.50)	-20.08 (69.84)	-253.79 (177.83)	1.63 (264.28)	0.63 (0.40)	0.08 (0.13)	0.20 (0.27)						
RN2	Neutral	Masculine	-1.18 (41.86)	-3.02 (195.94)	21.22 (388.27)	15.21 (364.39)	0.19 (0.22)	0.61 (0.34)	0.80 (0.45)						
		Feminine	-5.27 (40.18)	-27.16 (185.68)	-54.33 (413.81)	-33.83 (407.47)	0.24 (0.21)	0.46 (0.38)	0.58 (0.44)						
	Female	Masculine	-0.86 (44.97)	-28.51 (128.18)	-43.48 (337.83)	21.36 (282.96)	0.20 (0.22)	0.56 (0.30)	0.77 (0.40)						
		Feminine	5.93 (52.41)	10.92 (184.06)	-45.88 (380.95)	25.24 (388.58)	0.17 (0.20)	0.48 (0.36)	0.65 (0.48)						
Spillover A	Neutral	Masculine	6.99 (38.11)	3.09 (90.90)	-252.35 (164.32)	-37.95 (227.96)	0.55 (0.38)	0.09 (0.16)	0.34 (0.24)						
		Feminine	5.18 (41.85)	-12.33 (108.95)	-311.02 (112.89)	-48.17 (196.66)	0.57 (0.45)	0.06 (0.11)	0.34 (0.34)						
	Female	Masculine	22.61 (41.18)	17.72 (103.24)	-250.85 (176.97)	8.13 (250.35)	0.55 (0.34)	0.09 (0.18)	0.43 (0.42)						
		Feminine	20.10 (48.68)	-22.94 (84.48)	-265.08 (255.91)	-2.72 (295.09)	0.55 (0.36)	0.09 (0.16)	0.41 (0.33)						
Spillover B	Neutral	Masculine	53.76 (72.53)	96.41 (171.09)	1098.03 (1193.11)	-36.88 (258.29)	/	/	1.21 (0.55)	1.21 (0.55)					
		Feminine	45.98 (70.88)	9.65 (133.83)	1123.86 (1143.80)	-64.33 (312.65)	/	/	1.32 (0.52)	1.32 (0.52)					
	Female	Masculine	48.93 (67.81)	22.55 (96.51)	1119.68 (1312.81)	-87.95 (198.79)	/	/	1.25 (0.58)	1.25 (0.58)					
		Feminine	40.30 (47.71)	1.63 (96.72)	1151.49 (1461.52)	-94.11 (310.80)	/	/	1.19 (0.56)	1.19 (0.56)					

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into a region, RO: regressions out of a region, ROF: regressions out of a region (full count)

## GENDER PROMINENCE IN FRENCH ANAPHORS

Table 2  
Results of Analyses of Variance for All Regions of Interest (Experiment 1)

Measure <sup>a</sup>	Region	Effect	F <sub>1</sub>	df <sub>1,2</sub>	F <sub>2</sub>	df <sub>1,2</sub>	
FF	RN1	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	1.01		<1		
	Anaphor	Typicality	3.07*	1, 24	3.22*	1, 18	
		GG	<1		<1		
		Typicality * GG	5.08**	1, 24	3.41*	1, 18	
	Conjunction	Typicality	<1		<1		
		GG	1.11	1, 24	<1		
		Typicality * GG	<1		2.24	1, 18	
	RN2	Typicality	1.00	1, 24	<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	Spillover A	Typicality	9.38***	1, 24	5.52**	1, 18	
		GG	<1		<1		
		Typicality * GG	<1		<1		
	Spillover B	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	FP	RN1	Typicality	7.80***	1, 24	5.46**	1, 18
			GG	1.47	1, 24	1.94	1, 18
			Typicality * GG	<1		<1	
		Anaphor	Typicality	<1		<1	
			GG	<1		<1	
			Typicality * GG	<1		<1	
Conjunction		Typicality	<1		<1		
		GG	4.46**	1, 24	7.00**	1, 18	
		Typicality * GG	<1		<1		
RN2		Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	1.49	1, 24	1.61	1, 18	
Spillover A		Typicality	<1		<1		
		GG	4.53**	1, 24	6.84**	1, 18	
		Typicality * GG	<1		<1		
Spillover B		Typicality	5.11**	1, 24	1.83	1, 18	
		GG	5.63**	1, 24	4.53**	1, 18	
		Typicality * GG	6.16**	1, 24	1.98	1, 18	
RP		RN1	Typicality	2.55	1, 24	1.18	1, 18
			GG	<1		<1	
			Typicality * GG	<1		<1	
		Anaphor	Typicality	<1		<1	
			GG	1.78	1, 24	7.12**	1, 18
			Typicality * GG	<1		<1	
	Conjunction	Typicality	1.55	1, 24	<1		
		GG	<1		<1		
		Typicality * GG	2.44	1, 24	3.15*	1, 18	
	RN2	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	Spillover A	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	Spillover B	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	TT	RN1	Typicality	1.23	1, 24	<1	
			GG	<1		<1	
			Typicality * GG	<1		<1	
		Anaphor	Typicality	<1		<1	
			GG	<1		<1	
			Typicality * GG	<1		<1	
Conjunction		Typicality	<1		<1		
		GG	9.89***	1, 24	7.65**	1, 18	
		Typicality * GG	<1		<1		
RN2		Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
Spillover A		Typicality	2.22	1, 24	<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
Spillover B		Typicality	2.47	1, 24	<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		

# GENDER PROMINENCE IN FRENCH ANAPHORS

Table 2 (Continued)

Measure <sup>a</sup>	Region	Effect	F <sub>1</sub>	df <sub>1,2</sub>	F <sub>2</sub>	df <sub>1,2</sub>
RI	RN1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Anaphor	Typicality	<1		<1	
		GG	<1		1.36	1, 18
		Typicality * GG	<1		<1	
	Conjunction	Typicality	<1		<1	
		GG	9.70***	1, 24	8.59***	1, 18
		Typicality * GG	<1		<1	
	RN2	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		1.17	1, 18
	Spillover A	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
RO	RN1	Typicality	<1		<1	
		GG	3.04*	1, 24	1.27	1, 18
		Typicality * GG	<1		<1	
	Anaphor	Typicality	<1		<1	
		GG	2.16	1, 24	3.11*	1, 18
		Typicality * GG	<1		<1	
	Conjunction	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	1.95	1, 24	2.45	1, 18
	RN2	Typicality	<1		<1	
		GG	3.19*	1, 24	3.15*	1, 18
		Typicality * GG	<1		<1	
	Spillover A	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
Spillover B	Typicality	<1		<1		
	GG	<1		<1		
	Typicality * GG	1.50	1, 24	1.42	1, 18	
ROF	RN1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Anaphor	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Conjunction	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	RN2	Typicality	<1		<1	
		GG	6.15**	1, 24	3.86*	1, 18
		Typicality * GG	<1		<1	
	Spillover A	Typicality	1.40	1, 24	1.15	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
Spillover B	Typicality	<1		<1		
	GG	<1		<1		
	Typicality * GG	<1		1.42	1, 18	

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into a region, RO: regressions out of a region, ROF: regressions out of a region (full count)

\*  $p \leq .1$ , \*\* $p \leq .05$ , \*\*\* $p \leq .01$ .

## GENDER PROMINENCE IN FRENCH ANAPHORS

Table 3

Means (and Standard Deviations) of Residual Fixation Times and Probabilities of Regressions (Experiment 2)

Region	Factors		Measure <sup>a</sup>												
	Stereotypical Gender (RN 1)	Grammatical Gender (RN 2)	FF	FP	RP	TT	RI	RO	ROF						
RN1	Male	Masculine	-5.11 (47.97)	15.23 (145.02)	-178.44 (168.69)	171.31 (320.43)	0.89 (0.64)	0.25 (0.22)	0.48 (0.32)						
		Feminine	-14.21 (46.10)	-22.94 (117.96)	-142.80 (230.63)	173.53 (291.26)	0.98 (0.51)	0.36 (0.36)	0.53 (0.45)						
	Neutral	Masculine	-15.39 (42.87)	6.67 (148.68)	-93.83 (259.93)	162.06 (420.31)	1.01 (0.69)	0.26 (0.23)	0.46 (0.41)						
		Feminine	-8.92 (40.88)	30.14 (198.93)	-100.53 (210.27)	162.54 (342.02)	0.93 (0.57)	0.30 (0.28)	0.47 (0.38)						
Anaphor	Male	Masculine	-9.48 (44.67)	-3.85 (117.38)	-130.80 (256.80)	105.37 (425.14)	0.51 (0.37)	0.20 (0.25)	0.56 (0.44)						
		Feminine	-12.59 (36.37)	23.55 (128.77)	-187.87 (174.27)	89.77 (323.59)	0.59 (0.45)	0.17 (0.19)	0.62 (0.45)						
	Neutral	Masculine	-20.87 (35.30)	-4.17 (133.38)	-181.20 (230.71)	52.78 (305.54)	0.64 (0.43)	0.14 (0.17)	0.49 (0.36)						
		Feminine	-12.26 (44.49)	14.79 (147.56)	-138.22 (269.83)	38.68 (310.28)	0.53 (0.39)	0.21 (0.22)	0.61 (0.37)						
Conjunction	Male	Masculine	-14.71 (40.62)	-17.66 (69.28)	-269.38 (182.48)	-29.43 (191.29)	0.66 (0.75)	0.06 (0.13)	0.18 (0.28)						
		Feminine	-8.29 (62.52)	-15.58 (86.16)	-289.23 (115.32)	-29.76 (175.73)	0.59 (0.60)	0.07 (0.14)	0.18 (0.23)						
	Neutral	Masculine	-13.74 (54.44)	-21.44 (74.65)	-289.90 (108.54)	23.64 (199.80)	0.78 (0.70)	0.10 (0.20)	0.23 (0.25)						
		Feminine	-11.16 (39.93)	-20.11 (55.95)	-274.08 (116.74)	-42.86 (152.78)	0.64 (0.74)	0.10 (0.20)	0.25 (0.32)						
RN2	Male	Masculine	3.59 (52.30)	-6.15 (136.25)	-113.11 (385.60)	40.06 (373.98)	0.27 (0.36)	0.42 (0.55)	0.62 (0.66)						
		Feminine	-6.99 (42.09)	26.73 (167.82)	-57.30 (358.90)	17.35 (332.10)	0.32 (0.27)	0.36 (0.39)	0.67 (0.63)						
	Neutral	Masculine	0.05 (50.29)	-23.40 (175.01)	-31.45 (423.82)	50.24 (340.97)	0.24 (0.24)	0.60 (0.57)	0.86 (0.80)						
		Feminine	-6.97 (40.35)	26.98 (159.22)	-113.22 (345.45)	-29.20 (367.07)	0.37 (0.30)	0.26 (0.37)	0.56 (0.71)						
Spillover A	Male	Masculine	18.12 (42.29)	6.04 (137.97)	-173.50 (270.18)	-1.22 (237.60)	0.57 (0.37)	0.17 (0.20)	0.46 (0.32)						
		Feminine	13.30 (54.24)	-15.74 (131.84)	-229.09 (248.09)	-27.81 (256.77)	0.52 (0.37)	0.09 (0.18)	0.36 (0.32)						
	Neutral	Masculine	14.89 (49.79)	9.04 (163.19)	-240.08 (230.99)	38.08 (291.53)	0.65 (0.39)	0.08 (0.21)	0.40 (0.36)						
		Feminine	18.42 (55.42)	-14.85 (109.75)	-204.08 (193.97)	5.20 (245.10)	0.63 (0.47)	0.14 (0.17)	0.48 (0.39)						
Spillover B	Male	Masculine	35.26 (71.32)	12.88 (130.17)	1115.77 (1114.11)	-43.32 (232.66)	/	/	1.28 (0.66)	1.28 (0.66)					
		Feminine	38.86 (80.65)	15.00 (140.70)	1243.03 (1083.11)	-78.26 (246.20)	/	/	1.36 (0.61)	1.36 (0.61)					
	Neutral	Masculine	48.89 (72.72)	33.91 (108.33)	1336.56 (1261.56)	14.40 (275.62)	/	/	1.43 (0.65)	1.43 (0.65)					
		Feminine	30.10 (77.67)	21.73 (129.98)	1213.93 (1140.56)	-37.21 (219.04)	/	/	1.33 (0.71)	1.33 (0.71)					

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into a region, RO: regressions out of a region, ROF: regressions out of a region (full count)



## GENDER PROMINENCE IN FRENCH ANAPHORS

Table 4  
Results of Analyses of Variance for All Regions of Interest (Experiment 2)

Measure <sup>a</sup>	Region	Effect	F <sub>1</sub>	df <sub>1,2</sub>	F <sub>2</sub>	df <sub>1,2</sub>	
FF	RN1	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	2.21	1, 32	1.36	1, 18	
	Anaphor	Typicality	4.14**	1, 32	<1		
		GG	<1		<1		
		Typicality * GG	1.90	1, 32	<1		
	Conjunction	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	RN2	Typicality	<1		<1		
		GG	2.00	1, 32	2.37	1, 18	
		Typicality * GG	<1		<1		
	Spillover A	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
	Spillover B	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	1.36	1, 32	<1		
	FP	RN1	Typicality	1.36	1, 32	<1	
			GG	<1		3.13*	1, 18
			Typicality * GG	2.21	1, 32	<1	
Anaphor		Typicality	<1		<1		
		GG	2.10	1, 32	1.44	1, 18	
		Typicality * GG	<1		<1		
Conjunction		Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
RN2		Typicality	<1		<1		
		GG	5.90**	1, 32	5.14**	1, 18	
		Typicality * GG	<1		<1		
Spillover A		Typicality	<1		<1		
		GG	1.94	1, 32	1.17	1, 18	
		Typicality * GG	<1		<1		
Spillover B		Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
RP		RN1	Typicality	6.47**	1, 32	1.72	1, 18
			GG	<1		<1	
			Typicality * GG	<1		<1	
	Anaphor	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	3.18*	1, 32	2.23	1, 18	
	Conjunction	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	<1		3.79*	1, 18	
	RN2	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	5.79**	1, 32	<1		
	Spillover A	Typicality	<1		<1		
		GG	<1		<1		
		Typicality * GG	2.71	1, 32	1.04	1, 18	
	Spillover B	Typicality	1.36	1, 32	<1		
		GG	<1		<1		
		Typicality * GG	3.57*	1, 32	<1		
	TT	RN1	Typicality	<1		<1	
			GG	<1		<1	
			Typicality * GG	<1		<1	
Anaphor		Typicality	3.32*	1, 32	<1		
		GG	<1		<1		
		Typicality * GG	<1		<1		
Conjunction		Typicality	1.60	1, 32	<1		
		GG	2.97*	1, 32	3.37*	1, 18	
		Typicality * GG	4.66**	1, 32	5.04**	1, 18	
RN2		Typicality	<1		<1		
		GG	2.25	1, 32	1.39	1, 18	
		Typicality * GG	<1		1.10	1, 18	
Spillover A		Typicality	2.31	1, 32	<1		
		GG	1.15	1, 32	<1		
		Typicality * GG	<1		<1		
Spillover B		Typicality	3.50*	1, 32	<1		
		GG	3.33*	1, 32	2.00	1, 18	
		Typicality * GG	<1		<1		

## GENDER PROMINENCE IN FRENCH ANAPHORS

Table 4 (Continued)

Measure <sup>a</sup>	Region	Effect	<i>F</i> <sub>1</sub>	<i>df</i> <sub>1,2</sub>	<i>F</i> <sub>2</sub>	<i>df</i> <sub>1,2</sub>
RI	RN1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	2.40	1, 32	2.54	1, 18
	Anaphor	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	3.44*	1, 32	3.15*	1, 18
	Conjunction	Typicality	1.72	1, 32	<1	
		GG	3.35*	1, 32	1.00	1, 18
		Typicality * GG	<1		<1	
	RN2	Typicality	<1		<1	
		GG	3.23*	1, 32	3.00*	1, 18
		Typicality * GG	<1		1.17	1, 18
	Spillover A	Typicality	4.22**	1, 32	2.07	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
RO	RN1	Typicality	<1		<1	
		GG	3.13*	1, 32	4.04*	1, 18
		Typicality * GG	<1		<1	
	Anaphor	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	1.78	1, 32	1.24	1, 18
	Conjunction	Typicality	1.16	1, 32	<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	RN2	Typicality	<1		1.93	1, 18
		GG	18.53***	1, 32	6.98**	1, 18
		Typicality * GG	15.92***	1, 32	4.97**	1, 18
	Spillover A	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	9.16***	1, 32	4.76**	1, 18
Spillover B	Typicality	1.22	1, 32	<1		
	GG	<1		<1		
	Typicality * GG	3.18*	1, 32	2.45	1, 18	
ROF	RN1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Anaphor	Typicality	<1		<1	
		GG	3.58*	1, 32	6.66**	1, 18
		Typicality * GG	<1		<1	
	Conjunction	Typicality	1.67	1, 32	<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	RN2	Typicality	1.08	1, 32	<1	
		GG	3.77*	1, 32	1.44	1, 18
		Typicality * GG	13.88***	1, 32	4.63**	1, 18
	Spillover A	Typicality	<1		<1	
		GG	<1		2.89	1, 18
		Typicality * GG	2.88*	1, 32	<1	
Spillover B	Typicality	1.22	1, 32	<1		
	GG	<1		<1		
	Typicality * GG	3.18*	1, 32	2.45	1, 18	

<sup>a</sup> FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into a region, RO: regressions out of a region, ROF: regressions out of a region (full count)

\*  $p \leq .1$ , \*\*  $p \leq .05$ , \*\*\*  $p \leq .01$ .

**Erklärung gemäß § 8 Abs. 1 Buchst. b) und c) der Promotionsordnung  
der Fakultät für Verhaltens- und Empirische Kulturwissenschaften**

---

**Promotionsausschuss der Fakultät für Verhaltens- und Empirische  
Kulturwissenschaften  
der Ruprecht-Karls-Universität Heidelberg**  
Doctoral Committee of the Faculty of Behavioural and Cultural Studies, of Heidelberg  
University

**Erklärung gemäß § 8 Abs. 1 Buchst. b) der Promotionsordnung der Universität  
Heidelberg  
für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften**  
Declaration in accordance to § 8 (1) b) and § 8 (1) c) of the doctoral degree regulation of  
Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die  
angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.  
I declare that I have made the submitted dissertation independently, using only the specified tools  
and have correctly marked all quotations.

**Erklärung gemäß § 8 Abs. 1 Buchst. c) der Promotionsordnung  
der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische  
Kulturwissenschaften**

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht  
anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation  
vorgelegt habe.  
I declare that I did not use the submitted dissertation in this or any other form as an examination  
paper until now and that I did not submit it in another faculty.

Vorname Nachname                      Chiara Reali  
First name Family name                      \_\_\_\_\_

Datum, Unterschrift                      28.10.2015  
Date, Signature                                      \_\_\_\_\_