Inauguraldissertation
zur Erlangung des akademischen Doktorgrades (Dr. phil.)
im Fach Psychologie
an der Fakultät für Verhaltens- und
Empirische Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg

Titel der publikationsbasierten Dissertation:

The Prominence of Gender Information in On-line Language Processing: Cross-Linguistic Evidence of Implicit Gender Hierarchies

vorgelegt von
Dipl. Psych. Yulia Esaulova

Jahr der Einreichung
2014

Dekan: Prof. Dr. Klaus Fielder
Berater: Prof. Dr. Lisa von Stockhausen

Dedication

I dedicate this dissertation to my father, Prof. Dr. Ivan Esaulov, who gave me an early example of an excellent scholar.

Contents

Acknowledgements

I would like to express my deep appreciation and gratitude to my supervisor, Prof. Dr. Lisa von Stockhausen, who inspired and encouraged my enthousiasm about research from when I first considered applying for a PhD program until the present day. Her patient guidance and commitment to the highest standards motivated me, making the past years a thoughtful and rewarding journey. On a more personal note, I would like to thank her for being my second family when I needed it the most.

I would also like to thank the *Initial Training Network: Language, Cognition and Gender* for giving me an exclusive opportunity to forget about country borders conducting my research and surrounding me with colleagues in the field already at the early stage of my career, whose expertise and thought-provoking suggestions were critical at different stages of this project. An especial thank you goes to my friends from the network and elsewhere (Amal Khaleefa, Phillip Thomas, and Viviane Brunel to name just a few) who contributed to this work in many direct and indirect ways.

I am grateful to my parents for doing the impossible and their continuous support throughout my life and to my elder brother, Dr. Andrey Esaulov, who still picks me up when I fall.

Finally, I would like to thank my fiancé, Dr. Ivan Marozau, for tirelessly listening to me talk about my research, his practical advice, and his unfailing readiness to always be there for me.

Publication list

Paper 1:
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 2:
Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). *Prominence of gender cues in the assignment of agent and patient roles in German.* Manuscript submitted for publication.

Paper 3:
Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). *Prominence hierarchies in the processing of gender-ambiguous anaphors in French.* Manuscript submitted for publication.

Paper 4:
Reali, C., Esaulova, Y., & von Stockhausen, L. (in press). Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics.*†

Paper 5:
Reali, C., Esaulova, Y., & von Stockhausen, L. (2014). *Gender typicality effects on eye movements in sentence reading*. Manuscript submitted for publication.

* Reprinted with permission from Taylor & Francis Ltd.
† Reprinted with permission from Cambridge University Press.

Je suis obligé de toujours choisir entre le masculin et le féminin… La langue… est tout simplement : fasciste ; car le fascisme, ce n'est pas d'empêcher de dire, c'est d'obliger à dire.[*]

–Roland Barthes, *Leçon*, 1978

1. Introduction

A natural language regularly presents its users with a number of referential and ambiguous structures that could hinder the comprehension. However, it has long been a well-know fact that people are still able to understand language with only little delay if any at all (Marslen-Wilson, 1975). Whether and how particular types of information are used to efficiently resolve and interpret ambiguities are central questions in the research on on-line language processing. The present work contributes to this research by focusing on the use of gender information and evaluating its contribution to the interpretation of referential and ambiguous sentences.

As one of the categories essential for social interaction (Fiske, 1998), gender is represented in language on formal and conceptual levels. A line of research in social cognition (e.g., Stahlberg, Braun, Irmen, & Sczesny, 2007) indicates the association between the formal level of grammatical gender features and the conceptual level of gender-related connotations (e.g., gender stereotypes). Representations of gender in language act upon the cognitive representation of men and women, often leading to stereotyping and biases that have important economic and political consequences in social life. Literature on linguistic biases has demonstrated that (gender-)stereotypical information is reflected, transmitted and maintained by language in a number of ways (the use of abstractions – Semin & Fiedler, 1988; explanations - Sekaquaptewa, Expinoza, Thompson, Vargas, & von Hippel, 2003;

[*] I am obliged to always choose from masculine and feminine… The language... is plainly fascist; because fascism is not about preventing from saying but about obliging to say.

negations - Beukeboom, Finkenauer, Wigboldus, 2010). While stereotyping may be due to the repeated activation of cognitive schemata, the exact cognitive mechanisms of linguistic biases and the extent to which the interpretation of language relies on gender-related information are not yet fully understood. In social cognition (e.g., Koenig, Mitchell, Eagly, & Ristikari, 2011), the categorization of individuals according to characteristics associated with male and female roles often results in hierarchies (higher or lower status, more or less power, etc.). The hierarchical organization is considered inherent to most linguistic structures (e.g., subject/object hierarchy) and certain information types (e.g., animacy hierarchy as one of the so-called prominence features) in language (e.g., Aissen, 2003; Lamers & de Swart, 2012). While social gender hierarchies and formal linguistic hierarchies are typically treated as distinct, the present work combines these theoretical notions offering a novel approach to the study of mechanisms underlying gender biases in language.

Chapter 1.1 of the Introduction addresses the research on the on-line processing of gender information that is present in language. Chapter 1.2 introduces interdisciplinary aspects of the problem, relating the social cognitive notion of gender hierarchies to the linguistic and prominence hierarchies in language, and thus offers the theoretical grounding of the present research. Chapter 1.3 states the main research questions of this work and the research program. Chapter 1.4 provides the general information on the method and data analyses that were common for the reported experiments. The Overview of studies is based on a series of experiments and describes materials and main findings of Papers 1-5, as well as my contribution to each of them. The first study focuses on the recruitment of gender information for the resolution of anaphoric structures and is mainly discussed in terms of the time-course of gender processing in language. The next two studies focus on the implicit influences of gender cues in the interpretation of ambiguous structures in German and French and highlight the understanding of gender as a prominence feature and its cross-linguistic validity. The last

two studies explore other representational formats of gender information and their influences. Finally, Conclusions summarize the main findings regarding explicit and implicit influences of gender information, the perspective on gender as a prominence feature, as well as limitations of the present studies and directions for future research.

1.1. Gender in on-line language processing

A large part of the research on the on-line processing of gender information has focused on the cognitive processes related to the use of role noun denotations, such as *electrician* or *soccer fan*. As an instance of person reference, these nouns are subjects to gender stereotypes (Baudino, 2001) and may entail grammatical gender information marked morphologically or by the determiner. Since basic comprehension requires the integration of such gender cues when biological sex of the referent is inferred (Cacciari, Corradini, Padovani, & Carreiras, 2011), role nouns present a special interest for the study of gender processing in language. Among a range of various paradigms employed in the field, reference resolution paradigms occupy a central place, since they can be used to reveal difficulties in the integration processes when linguistically presented gender information is incongruent (e.g., a feminine pronominal anaphor referring to a stereotypically male role noun antecedent, as in "*electrician ...she*"). These incongruities are often discussed in terms of mismatch effects reflected by enhanced neural activity, slowed down reaction times or inflated reading times as behavioral indicators of underlying cognitive processes.

In research on the processing of grammatical gender, the processing of generic masculine has received a lot of attention because of the ambiguity of interpretation that it implies. The probability of its possible interpretation as gender-specific (masculine) rather than generic form that includes both masculine and feminine readings has led to ongoing discussions (e.g., in French – Colé & Segui, 1994; in German – Stahlberg et al., 2007).

However, gender ambiguity is not limited to the formal grammatical encoding of gender. It also occurs in languages where role nouns do not have grammatical gender markings, such as in English. In theses cases, conceptual gender-related information associated with a given role noun (stereotypical gender) surfaces in the cognitive processes involved in the resolution of ambiguities. As cognitive structures that contain perceivers' knowledge, beliefs, and expectancies about a given group of persons (Hamilton & Troiler, 1986, p. 133), gender stereotypes influence the processing of language when presented linguistically in the form of a role noun. These influences were empirically demonstrated in a number of studies. In a self-paced reading study, Carreiras, Garnham, Oakhill and Cain (1996) examined the influence of stereotypical information in English and Spanish sentences, where a sentence containing a role noun antecedent was followed by a pronominal anaphor in the next sentence (e.g., *The footballer wanted to play in the match. He/she had been training very hard during the week*). The results are discussed in terms of a mental model (Garnham, 1987) that incorporates stereotypical gender information into the representation of a character named by the role noun. In the mismatch condition where the stereotypical gender is inconsistent with the pronoun, the model needs to be updated. The mismatch effect was observed in slowed down reading and interpreted as the time needed for the updating of the model. Similarly, in an eye-tracking study, Kennison and Trofe (2003) demonstrated longer processing times for sentences where the stereotypical gender of the role noun did not correspond to the grammatical gender of an anaphoric pronoun that referred back to it (e.g., *The executive… she…*). Duffy and Keir (2004) examined a similar mismatch effect for reflexive pronouns (*himself/herself*) that referred to stereotypically incongruent role nouns. Furthermore, they showed the elimination of the mismatch effect in cases where a disambiguating context preceding the target sentence clearly instantiated the gender of the character in the discourse. The authors argue that in such cases the pronoun is readily integrated into the discourse

despite the mismatch with stereotypes, which is in line with the lexical interpretation model (Hess, Foss, & Carroll, 1995). The mental model and lexical interpretation model offer different theoretical approaches as to the nature of stereotypical gender. According to the mental model approach, stereotypical gender is inferred from world knowledge, while lexical approach takes it to be a part of lexical representation of a word. Kreiner, Sturt, and Garrod (2008) investigated this controversy contrasting the processing of stereotypical (e.g., minister) and definitional (e.g., king) nouns in sentences with cataphoric references (e.g., *After reminding himself/herself about the letter, the king/minister immediately went to the meeting at the office*). The authors found that a mismatch-effect was only observed in cataphors with definitional nouns in which gender information was part of the definition of the word, while stereotypical nouns did not evoke such an effect. They conclude that, unlike stereotypical gender, definitional gender is represented lexically, which is reflected in processing differences, such as the ease of modulation by syntactic constraints provided in the prior discourse.

The differentiation between the processing of distinct types of gender information can be further approached in terms of a time course of processes related to the comprehension of each of those gender types. Understanding at which point in time specific gender information affects comprehension is crucial to account for the recruitment of gender information in the on-line language processing. Sturt (2003) interprets the results of his eye-tracking study with texts including anaphoric references within a framework of a two-stage model of reference resolution (Garrod & Terras, 2000). These texts consisted of two sentences with the first sentence including a named female/male character (e.g., *Jennifer/Jonathan*). In the second sentence a second character was introduced through a stereotypically female or male role noun (e.g., $surgeon_{Male}$) and a feminine/masculine reflexive pronoun (e.g., *himself/herself*). The results show a mismatch effect between the role noun and the reflexive (e.g., *surgeon…*

herself) in measures reflecting very early stages of processing. However, when the role noun and the reflexive were congruent (e.g., *surgeon… himself*), a mismatch between the named antecedent and the reflexive (e.g., *Jennifer… the surgeon… himself*) only occurred at a relatively later stage. According to the two-stage model, two processes can be identified in a reference resolution: bonding and the resolution itself. The bonding process consists in establishing a link made between the anaphor and a possible antecedent on a relatively superficial level, whereas the resolution process involves the deeper evaluation of information and its integration into the semantic interpretation. Sturt suggests that while stereotypical gender information seems to affect the processing immediately, the subsequent interpretation of gender information provided in the text may become available only at later stages.

Specific processing patterns evoked by gender information presented in language motivate cognitive psychological research to examine different types of gender information (grammatical, stereotypical, definitional), variables that may modulate the processing (e.g., context, syntax), and the time course of on-line processes related to gender comprehension. Papers 1, 4 and 5 extend this research by analyzing the contribution of gender information to the interpretation of referential structures with different types of anaphors and antecedents.

1.2. Interdisciplinary perspective on gender and language hierarchies

While the research described above provides an essential knowledge on the topic of on-line gender processing, this chapter contributes to the theoretical basis of the present work by addressing and bringing together interdisciplinary perspectives on the notions of gender and language comprehension.

Gender hierarchies. Even though not directly related to the on-line language processing, a number of gender-related issues in the focus of social psychology offer a different but conceptually related perspective on the understanding of the problem. Within

this area of research, gender is discussed in contexts of sex roles and their characteristics, sex-role socialization, self-concept formation, and social status evaluations (Diehl, Owen, & Youngblade, 2004). The cognitive basis for the differences in sex roles consists in the extent to which the processing of information relies on an individuals' cognitive gender schema. This theoretical grounding underlies the Bem Sex Role Inventory (Bem, 1981), which assesses sex roles through a number of characteristics. Based on these characteristics, the BSRI results in a classification of an individual as masculine (when scored high on masculine and low on feminine scales), feminine (high on feminine and low on masculine), androgynous or undifferentiated type (when scores on masculine and feminine scales are both high or low respectively). According to the theory, sex-typed (masculine and feminine) individuals should demonstrate behavior that is guided by rigid gender schemas, while the behavior of androgynous individuals should display greater flexibility across situations. Even though there is an ongoing debate concerning theoretical and empirical aspects of BSRI (e.g., Spence & Helmreich, 1981; Pedhazur & Tetenbaum, 1979), this line of research is interesting for us as it emphasizes the link between the social construct of gender and its influences on cognition.

The characteristics of sex roles defined by gender schemas also guide a person's behavior in terms of socialization. Men are socialized to be self-sufficient, achievement oriented, and independent, whereas women are socialized to be relationship oriented, nurturing and sensitive (Cross & Madson, 1997). The literature on sex-role socialization provides consistent evidence for the close association between gender differences on one side and agency/communion on the other (Helgeson, 1994). Bakan (1966) describes the notion of agency as a modality of human behavior that is expressed in an individual's desire to experience competence, power, achievement, and master the environment. The cooperation with and close relation to others, on the other hand, characterize the concept of communion.

As can be inferred from the descriptions above, masculine and feminine characteristics are at the core of social psychological concepts of agency and communion respectively. Furthermore, masculinity and femininity are also discussed in relation to such constructs as status and power (Koenig et al., 2011; Spence & Buckner, 2000). These constructs are represented as hierarchies ranging from high to low extremes (e.g., high vs. low status), with the strong tendency to attribute higher rankings on the hierarchy to masculinity.

Taken together, the areas of research described above demonstrate how characteristics that define masculine and feminine sex roles and are reflected in socialization patterns relate to the concepts of agency/communion and hierarchical constructs of power and status. In this way, the results of social cognitive research imply a hierarchical organization within the construct of gender itself, with masculine gender associated with higher rankings on various scales and more active roles (e.g., desire to master the environment) and feminine with lower rankings and less active roles.

Hierarchies in language. Interestingly, a number of formal properties in language are also regarded as hierarchies. Grammatical functions are considered one such hierarchical property, where subjects have higher rankings than objects. Another hierarchy that is widely discussed concerns thematic roles, which represent the roles played in a sentence by entities in events described by the verb (McRae, Ferretti, & Amyote, 1997). Agent and patient roles represent the typical focus of research, even though labels and precise definitions of thematic roles differ widely across existing taxonomies and researchers (e.g. Cook, 1979; Dowty, 1991; Fillmore, 1968). In terms of a hierarchy, agent thematic roles refer to entities performing the event and rank higher than patient roles that have the event performed on them.

Together with the information from word order and case marking, linguistic hierarchies of grammatical functions and thematic roles are essential cues for the adequate

identification of arguments in a sentence and thus its interpretation. Although the underlying processes and the time course of these language properties remain subjects of debate (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2009), theoretical and empirical psycholinguistic research clearly shows that their precise interaction (which ones agree and which provide conflicting information) determines the readers' ability to comprehend sentences and resolve ambiguities when more than one interpretation is available (for an overview, see Lamers & de Swart, 2012).

In addition to the hierarchies mentioned above, cross-linguistic research identified and conceptualized several semantic/pragmatic features of arguments (e.g., animacy, person, definiteness) in terms of hierarchies (e.g., Siewierska, 2004; Wang, 2012). These cross-linguistically motivated hierarchically ordered information types are also called ***prominence*** features, with elements ranking higher on the hierarchy regarded as more prominent (or higher in prominence). Thus, animates are considered more prominent than inanimates, first and second person more prominent than the third, and so on.

The interaction of prominence features with other formal properties of language (e.g., grammatical functions) often results in mapping (or aligning) of their hierarchical structure onto that of grammatical functions or thematic roles. Such alignments match highly ranked animates with highly ranked subjects/agents and lower ranked inanimates with objects/patients (also referred to as harmonic alignment - Aissen, 2003). In this way, prominence features can be viewed as mediators that provide a link between conceptual and linguistic factors that are at play during language comprehension. This characteristic of prominence features brings up a controversial theoretical perspective on semantics and syntax, known as two-stage and constraint-based approaches. The two-stage approach stems from the garden-path model (Frazier, 1987) and isolates the influence of syntactic information from the semantic content (e.g., thematic roles are considered syntactic slots void of

conceptual content; Caplan, Hildebrandt, & Waters, 1994). According to this approach, the initial stage of language processing involves determining the syntactic structure, while semantic information may be used to revise that structure at a relatively later stage. Contrary to this theoretical perspective, the constraint-based approach explores the connection between syntactic and conceptual information in language (e.g., thematic roles are seen as concepts reflecting world knowledge; McRae et al., 1997). Constraint-based models of language comprehension claim that a number of constraints (e.g., syntactic principles – Frazier & Rayner, 1982; frequency – MacDonald, Pearlmutter, & Seidenberg, 1994; thematic roles – Clifton, 1993), including conceptual ones, is evaluated resulting in the interpretation that satisfies them best.

McRae, Spivey-Knowlton and Tanenhaus (1998) modeled the influences of different constraints on the resolution of syntactic ambiguity and referred to them in terms of biases towards a certain interpretation. In the light of constraint-based approach, the model of Incremental Optimization of Interpretation (de Hoop & Lamers, 2006) considers prominence information as one of such constraints. As such, prominence information can be seen as the source of a bias itself. Research on animacy provides some empirical support fro this understanding of prominence information. Mak, Vonk, and Schriefers (2002, 2006) and Traxler, Morris, and Seely (2002) demonstrated the modulation of the default preference in the interpretation of relative clauses (subject- over object-extracted) as a function of animate/inanimate sentence heads. Object-extracted relative clauses with inanimate heads (*The movie that the director watched received the prize*) were almost as easy to comprehend as subject-extracted ones (*The director that watched the movie received the prize*). In this case, the prominence feature of animacy can be said to bias the interpretation of relative clauses towards object-extracted after inanimate and subject-extracted after animate heads.

While the use of inanimate entity as a head of the sentence increased the probability of its reading as an object-extracted relative clause, other linguistic biases reflect the probability of stereotype violation. Beukeboom et al. (2010) analyzed the use of negations in descriptions of behavior that either confirmed or went against existing stereotypes. The results show a tendency to use negations (negation bias) when behavior is inconsistent with stereotypical expectations, such as in saying *not stupid* rather than *smart* in a situation when a blond girl solves a complex math problem. Similarly, the use of different levels of abstraction (distinguished by Semin & Fiedler, 1988) to describe behavior has been shown to indicate the probability to which it is expected or unexpected, again in terms of its consistency with a stereotype (expectancy bias – Wigboldus, Semin, & Spears, 2000). Expected behaviors (e.g., crying women) were encoded by abstract adjectives (e.g., *emotional*, *aggressive*), while unexpected behaviors (e.g., crying men) were encoded by verbs referring to concrete events (e.g., *cry*, *hit*). In this way, biases in language can be understood as linguistic choices made during the encoding of information or tendencies to use a particular structure/information type that reflect the probability of its interpretation in a particular way.

To summarize, the research on gender as a subject of social cognition focuses on its conceptual (agency/communion) and hierarchical (status, power) content that contributes to the definition of cognitive gender schemas that guide a person's behavior. On the other hand, linguistic behavior is also guided by conceptual (prominence) and hierarchical (grammatical functions, thematic roles) information in language and is discussed in theories on language comprehension. By relating the two areas of research that have so far been considered distinct, the present work offers a novel approach for the processing of gender information in language. This approach broadens the traditional view on the processing of gender information in language by considering gender a hierarchically organized prominence feature which influence can be revealed through (biases in) the processing of formal linguistic

properties, such as thematic roles and grammatical functions. Papers 2 and 3 explore the potential advantages of this approach by using empirical techniques to study the processing of gender as a prominence feature in locally ambiguous sentences.

1.3. Research questions and program

The research program of this work focuses on clarifying the influences of gender information on the sentence comprehension starting with the analysis of explicit mismatch effects of gender cues in referential structures and proceeding to more subtle gender influences, such as in the assignment of thematic roles. In sentences with referential structures, where the integration of information is required in order to interpret the reference, gender cues represented on formal (grammatical gender) and conceptual (stereotypical gender) levels may differ in their relevance for the understanding of a sentence. Paper 1 addressed the question about processes involved in the comprehension of gender information encoded on different levels and their time course. While the resolution of an anaphoric structure in a sentence required the integration of gender cues rather explicitly, Paper 2 examined whether gender information influences language comprehension in a more implicit way, such as by having an effect on the assignment of thematic roles in sentences with local ambiguities. Paper 3 further explored the question about subtle influences of gender by examining whether gender information is used in the identification of thematic roles/grammatical functions cross-linguistically, thus validating the notion of gender information as a prominence feature. Papers 1-3 represent the core of the present work. As the principal investigator in these papers, I was responsible for the experimental design, stimuli, data collection and analyses, as well as writing and preparation of the manuscripts for publication. Furthermore, I present Papers 4 and 5 to which I contributed as a second co-author by participating in the discussions related to the design, assisting in the construction of

stimuli, data collection and initial analyses of eye-tracking data, and revising manuscript drafts before their submission for publication. Papers 4 and 5 examined the extent, to which gender stereotypes represented in descriptions of gender-typical occupational activities influence anaphor resolution in a grammatical gender language (German, Paper 4) and a natural gender language (English, Paper 5). These papers contribute to the overarching question of how gender information influences sentence comprehension during reading.

1.4. Method

In order to capture a moment-to-moment ease (or difficulty) during reading that would reflect mechanisms of language comprehension, a method that allows on-line registration was required. Eye-tracking was chosen as a method that offers high spatial (where readers fixate in a sentence) and temporal (how long they look at a specific part of a sentence) resolution, which provides reliable data for the analysis of comprehension processes involved in reading. While saccade programming (readers' decisions about where to move the eyes) is largely determined by low-level perceptual processing of visual information, the time spent focused on a given point in the sentence reflects higher-level linguistic or conceptual processing. Generally, the effects of linguistic manipulations can be seen in increased fixation times (slowed down processing) indicating the influence of higher-level variables in the reading comprehension. At the same time, time points at which these effects occur can be highly informative about the course and nature of the underlying cognitive processes. This is why different eye-tracking measures are often referred to as "early" or "late", as they are assumed to reflect early or delayed processing. In the present study we report eye-tracking measures commonly used in psycholinguistic research that address temporal and probabilistic information about eye movements during reading. The reported measures include *first fixation duration* (the duration of the first fixation on a critical region), *first-pass reading time* (the

sum of all fixation durations on the region before leaving it for the first time), *regression path* (the time from first fixating the region to first leaving it to the right, including rereading of previous regions), *total fixation times* (the time spent on the region including re-reading), and *regressions into a region* (the probability of regressive eye movements into a region on the first pass through the region).

The characteristics of the chosen method determined a number of necessary requirements for the construction of stimuli materials. In order to make the experimental items comparable between conditions, all sentences within an experiment had to have the same basic structure. The regions of interest could not be placed at the very beginning or the end of a sentence in order to avoid the effects of additional processes (e.g., wrap-up processes). Multiple pre-tests were conducted in order to control for gender influences other than the ones provided by the experimental manipulation and thus could be regarded as a confounding variable (e.g., gender neutrality of the context, adjective neutrality). Filler items constituted an important part of materials construction and served to prevent participants from developing reading strategies that could interfere with the expected effects of linguistic manipulation.

Two eye-tracking systems – EyeLink II and EyeLink 1000 – were used to monitor participants' eye movements during different experiments of the present study. Eye-tracking software programs used to set up an experiment and collect and analyze data included EyeTrack, EyeDoctor, EyeDry, ExperimentBuilder, and DataViewer. In all cases, the procedure of an eye-tracking session necessarily involved the calibration at the beginning of the recording and re-calibration at different points during the session that served to improve the accuracy of the collected data.

Following the current practice in eye-tracking research, data analysis included several stages. During the first stage, fixations shorter than 70 ms were merged with neighboring fixations located within one character, after which procedure fixations not representative of

normal information extraction during reading (below 70 ms and above 600 ms) were excluded. Statistical analyses report residual reading times after the correction for differences due to the length of the analyzed regions (Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994). The correction included computing the best linear fit between the length of the region and unadjusted reading times for each participant, followed by the subtraction of the predicted times from the original reading times thus removing linear variance related to the length of the region.

2. Overview of studies

2.1. Paper 1: Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution

Paper 1 focused on the processes involved in the comprehension of grammatical and stereotypical gender information in German referential structures. This study included two eye-tracking experiments that investigated how grammatical and stereotypical gender information encoded in German role nouns and anaphors referring to them is used in the on-line sentence processing. Both experiments employed an anaphor resolution paradigm, which required the integration of both grammatical and conceptual information in order to establish the link between the anaphor and the antecedent. In Experiment 1, anaphors were denoted by personal pronouns (*er* 'he'/*sie* 'she') and referred to the role nouns mentioned earlier in the sentence (e.g., *Oft hatte der Elektriker gute Einfälle, regelmässig plante er/sie neue Projekte* 'Often had the electrician good ideas, regularlz planned he/she new projects'). In Experiment 2, anaphors were denoted by semantically rich noun phrases (NP) *dieser Mann* 'this man'/*diese Frau* 'this woman'. The 3 x 2 x 2 factorial design of both experiments comprised the factors of stereotypical gender of role nouns (male/female/neutral), their grammatical gender (masculine/feminine), and the anaphor gender (masculine/feminine). The study

evaluates the relevance of gender information for the interpretation of the linguistic input based on the timing, the location in the sentence, and the time course of the observed effects.

The results suggest that anaphor resolution is based first of all on the rules of grammatical agreement. The analysis of eye movement patterns revealed the time-course of influences from grammatical and conceptual information: grammatical gender seemed to affect eye-tracking measures reflecting the initial stages of sentence processing, while stereotypical gender influenced measures associated with later processing stages. The differences in the time-course of gender influences are further discussed in terms of a semantic content of each anaphor type. Grammatical gender influences were delayed in case of fuller noun phrase anaphors compared to pronominal anaphors, as indicated by the measures and locations of their occurrence. Stereotypical gender affected the processing of noun phrase anaphors, but influenced the processing of the role noun itself in the case of a pronominal anaphor.

Paper 1 demonstrates that different gender information types are recruited at different time points during the on-line processing of referential structures, such as anaphors. Grammatical gender influences earlier stages of processing, while stereotypical gender affects the processing at relatively later stages. Moreover, the role of stereotypical gender information for the sentence comprehension appears to depend on the anaphor type. In sentences with semantically rich noun phrase anaphors, the conceptual knowledge from stereotypical gender is used for the resolution of the referential structure. In case of pronominal references, this information is integrated with grammatical gender of the role noun and serves for the interpretation of the role noun itself.

2.2. Paper 2: Prominence of gender cues in the assignment of agent and patient roles in German

Paper 2 explored the semantic content of cognitive gender representations yet further

by addressing the relevance of gender information for the assignment of agent and patient thematic roles in locally ambiguous sentences. Two eye-tracking experiments presented German sentences with subject- and object-extracted relative clauses (SRC and ORC) like *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet hat/haben, ist aufmerksam* 'The flight attendant who has observed many tourists / whom many tourists have observed, is attentive'. In these sentences, the auxiliary verb *hat/haben* 'has/have' disambiguated thematic agent/patient roles of the two role nouns but only at the end of the relative clause, making readers create expectations about agent (performing an action) and patient (having the action performed on it) role assignments based on other cues. At the same time, gender cues of the two role nouns were systematically varied resulting in two experimental designs. A 2 x 2 factorial design in Experiment 1 included RN1 grammatical gender (masculine/feminine) and the relative clause type (SRC/ORC). In Experiment 2, stereotypical gender constituted an additional factor resulting in a 2 x 2 x 2 design, with the following structure: RN1 stereotypical gender (neutral/female), RN2 grammatical gender (masculine/feminine), and the relative clause type (SRC/ORC).

The results consistently showed longer processing times on the auxiliary verb when it disambiguated grammatically masculine role nouns as patients and feminine role nouns as agents. Longer reading times and regressions into the action verb (preceding the critical for the interpretation of the sentence auxiliary verb) indicated comprehension difficulties when stereotypically female role nouns were disambiguated as agents and neutral role nouns as patients. Longer processing times are interpreted as the violation of readers' expectations regarding the assignment of thematic agent roles to feminine/female and patient roles to masculine/neutral ones.

Paper 2 provides the first empirical evidence for the assignment of thematic agent/patient roles based on the gender information provided in the sentence. Grammatically

masculine and neutral role nouns rather than feminine/female ones are interpreted as instigators of an action, or agents, whereas grammatically feminine and female role nouns are expected to have that action performed on them, serving as patients. Based on these differences in processing, gender is viewed as another prominence feature that biases the interpretation of a role noun towards an agent or a patient thematic role. Considering the hierarchy of thematic roles (agents over patients), the hierarchical organization of gender as a prominence feature places grammatically masculine role nouns over feminine and neutral role nouns over stereotypically female. This hierarchical structure relates linguistic and social psychological notions of gender and agency and encourages to consider the two research areas in a broader context of linguistic biases.

2.3. Paper 3: Prominence hierarchies in the processing of gender-ambiguous anaphors in French

Paper 3 addressed the cross-linguistic validity of gender as a prominence feature by examining readers' expectations about gender information based on the information about grammatical functions/thematic roles in French anaphoric sentences. In two eye-tracking experiments, a gender-ambiguous indirect object pronoun *lui* 'him/her' served as a backwards anaphor and referred to a role noun that served as an object/patient: *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoreux* 'In fact, the dietician recommended to him/her, so to this pharmacist, a strict plan'. The first role noun that served as a subject/agent varied in stereotypical gender (female/neutral in Experiment 1 and male/neutral in Experiment 2) and the second role noun varied in grammatical gender (masculine/feminine). The gender-ambiguous pronoun clearly indicated the upcoming personal reference as an object/patient leaving its gender unspecified. The specification of the referent gender later in the sentence either matched or mismatched with the reader's

expectation about it and resulted in processing times reflecting relative comprehension difficulties.

The results showed longer fixation times on the second role noun and its gender-marked determiner when they were grammatically masculine rather than feminine already during the first reading of the sentences. As to stereotypical gender influences, stereotypical (both male and female) rather than neutral subjects/agents produced longer fixations of different regions in measures reflecting early processing.

Paper 3 establishes the processing of gender information as a prominence feature in the resolution of gender-ambiguous anaphors. Reading patterns demonstrate a systematic expectation of a grammatically feminine rather than masculine antecedent based on its grammatical function (object)/thematic role (patient). These findings correspond to those reported for German language in Paper 2 and suggest that grammatical gender characteristics can be conceptualized in terms of a prominence hierarchy with masculine gender ranking higher than feminine. At the same time, further research is needed to define the place of stereotypical gender information on the gender prominence hierarchy.

2.4. Paper 4: Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements; and Paper 5: Gender typicality effects on eye movements in sentence reading

Papers 4 and 5 investigated the cognitive representation of gender on a conceptual level going beyond gender stereotypes represented in role nouns. These papers examined the relevance of stereotypical gender information for the processing of referential structures in the absence of grammatical gender influences. Based on our contribution to these studies, we focus on Experiment 2 (in German) reported in Paper 4 and Experiment 1 (in English) reported in Paper 5. Like in Paper 1, these eye-tracking experiments employed an anaphor

resolution paradigm. This time, however, the stereotypical gender information was conveyed by descriptions of typically male, typically female and neutral occupational activities instead of role nouns. These descriptions served as antecedents and were followed by sentences containing a masculine or feminine pronominal anaphor (e.g., *M. F. repariert und stellt Möbel her, arbeitet mit Holz. Gewöhnlich hat er/sie ein ausreichendes Einkommen.* 'M. F. repairs and produces pieces of furniture, works with wood. Usually he/she has a sufficient income').

The results of both experiments demonstrate mismatch effects in cases when stereotypical gender of the description and the anaphoric pronoun referring to it were incongruent. These effects occurred immediately upon the encounter of the anaphor in Experiment 2 and on the region following the anaphoric pronoun in Experiment 1. In both cases, mismatch effects were remained throughout later measures of language processing.

Papers 4 and 5 show that descriptions of occupational activities effectively convey stereotypical gender information. This gender information is activated during anaphor resolution affecting both early and late processing. The findings suggest that linguistic formats for cognitive representations of gender information are not limited to explicit cues, such as those present in grammatical gender suffixes or stereotypical role nouns, but go beyond them relying on the conceptual content of gender information. The cross-linguistic validity of the results in English (a language without grammatical gender) indicates that the stereotypical gender information does not depend on the gender system of a language but is represented on a semantic level.

3. Summary and conclusions

Bridging the gap between research on language processing and social cognition, the present work is the first, to our knowledge, to provide an empirical evidence for the understanding of gender as a prominence feature. In a series of eye-tracking experiments, we

used experimental paradigms based on the resolution of referential and ambiguous structures that made it possible to demonstrate measurable differences in the on-line processing of gender information. Papers 1, 4 and 5 focused on the time course and the relevance of gender cues for the interpretation of anaphoric structures depending on the representational formats of gender information in language (grammatical gender, stereotypical gender of role nouns, descriptions of gender-typical activities). Papers 2 and 3 introduced a novel approach to the study of gender influences in language that applied experimental eye-tracking techniques to the theoretical notions about gender and language hierarchies merging together research areas traditionally considered distinct. These papers addressed the interpretation of grammatical functions (subject/object) and the assignment of thematic roles (agent/patient) in a sentence as a function of gender characteristics of the corresponding entities, thus instantiating gender as a prominence feature.

Taken together, the findings of our studies indicate both explicit and implicit influences of gender information on on-line language processing. On the one hand, these influences can be seen on a rather explicit level in the resolution of referential structures, where the integration of gender characteristics of anaphors and antecedents is necessary for the interpretation of the sentence (e.g., based on the rules of grammatical agreement). Our studies demonstrate a specific time-course of cognitive processes related to the integration of gender information during sentence comprehension. Namely, the processing is characterized by an earlier onset of grammatical gender influences and is followed by the integration of stereotypical gender information during later stages. On the other hand, gender information represented in language guides cognitive processes involved in language comprehension in a more subtle, or implicit, way. In sentences where the argument structure is ambiguous, gender characteristics of arguments influence their expected interpretation (e.g., their probability to serve as agents/patients), making them relevant event in contexts where other features are

crucial for the adequate comprehension of the sentence (e.g., singular/plural form of auxiliary verbs disambiguating relative clauses as SRCs/ORCs). Our work suggests that readers' expectations about the assignment of thematic roles to one or another noun and the identification of their grammatical functions are gender-biased in that they are driven by gender information represented in those nouns.

The implicit ways in which the interpretation of formal linguistic structures is influenced by gender information is the area that has so far remained disregarded by the research on on-line gender processing in language. At the same time, research on linguistic biases shows that implicit social cognitive processes systematically influence language use. The biased language use is typically represented by specific linguistic choices people make (e.g., to use negations in a description of a behavior, as in negation bias) based on personal assumptions, expectations, and general stereotypical knowledge. In this sense, the interpretation of linguistic structures (e.g., agents/patients, subjects/objects) based on gender-related expectations about them can be regarded as an implicit gender bias, where gender information constrains the sentence comprehension.

While testing constraint-based (e.g., MacDonald et al., 1994) and other models (e.g., two-stage models – Frazier, 1987; extended Argument Dependency Model – Bornkessel-Schlesewsky & Schlesewsky, 2009) was beyond the scope of our research questions at this point, we would like to highlight the clear interaction between the conceptual content of gender information (semantics) and formal linguistic (syntactic) properties of language without further claims about the time-course or mechanisms of this interaction. Instead, we propose to consider gender information as a prominence feature, along with animacy, definiteness, and person (e.g., Lamers & de Swan, 2012). Our findings suggest a clear prominence scale for grammatical gender, with masculine entities ranking over feminine, whereas the rankings of stereotypical gender entities are yet to be clarified. According to the

principle of harmonic alignment (Aissen, 2003), masculine role nouns are perceived as higher-ranking subjects/agents and feminine role nouns as lower-ranking objects/patients. The idea of gender as a prominence feature broadens the understanding of cognitive representations of gender information, viewing them as inferences about the formal linguistic properties of a particular entity (e.g., masculine subjects/feminine patients) as opposed to limiting them to the explicit knowledge about gender stereotypes of the use of gender-specific suffixes.

Overall, the present work is an example of the potential advantages of an interdisciplinary approach to a systematic study of gender influences involved in the on-line language comprehension. Gaining a better understanding of explicit and implicit ways in which gender information shapes our understanding of language, is the first step towards the evidence-based evaluation of language as a powerful means for maintaining certain beliefs and expectations about men and women. Future research should address the questions regarding gender as a prominence feature that the present work left open. These questions mainly concern the rankings of entities on a prominence scale within stereotypical gender and the theoretical framework that would describe the interaction between the semantic gender information and the syntactic properties of language in terms of the time-course of involved processes. Other directions should address the interaction of gender prominence as a semantic feature with other structural dimensions of language, such as word order and case marking, and its cross-linguistic validity.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, *21*, 435-483.

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, *21*, 435-483.

Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Chicago: Rand McNally.

Baudino, C. (2001). *Politique de la langue et différence sexuelle: La politisation du genre des noms de métier* [Language policies and sexual differences: The political awareness of the gender of professional occupations]. Paris: L'Harmattan.

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 SocialPsychology*, *99*(6), 978-992.

Bornkessel-Schlesewsky, I., & Schlesewsky, 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.

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.

Caplan, D., Hildebrandt, H., & Waters, G. S. (1994). Interaction of verb selectional restrictions, noun animacy and syntactic form in sentence processing. *Language and Cognitive Processes*, *9*, 549-585.

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.

Clifton, C. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology*, *47*, 222-246.

Colé, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory & Cognition*, *22*, 387-394.

Cook, W. A. (1979). *Case grammar: The development of the matrix model (1970-1978).* Washington: Georgetown University Press.

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

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.

Diehl, M., Owen, S. K., & Youngblade, L. M. (2004). Agency and communion attributes in adults' spontaneous self-representations. *International Journal of Behavioral Development*, *28*, 1-15.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, *67*, 547-619.

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.

Ferreira, F., & Clifton, C. (1968). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348-368.

Fillmore, C. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1-90). New York, Holt, Rinehart and Winston.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (357-411). New York: McGraw-Hill.

Frazier, L. (1987). Theories of syntactic processing. In J. Garfield (Ed.), *Modularity in knowledge representation and natural language processing* (pp.559-586). Cambridge, MA: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Garnham, A. (1987). *Mental models as representations of discourse and text*. Chichester, UK: Ellis Horwood.

Garnham, A. (1987). *Mental models as representations of discourse and text.* Chichester, UK: Ellis Horwood.

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.

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

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

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.

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

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.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

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

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.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226-228.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing*, *12*, 137-176.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283-312.

Pedhazur, E. J., & Tetenbaum, T. J. (1979). Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology*, *37(6)*, 996-1016.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, *39*, 75-82.

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.

Siewierska, A. (2004). *Person*. Cambridge: Cambridge University Press.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes: what do they signify? *Psychology of Women Quartelry*, *24*, 44-62.

Spence, J. T., & Helmreich, R. L. (1981). Androgyny versus gender schema: A comment on Bem's gender schema theory. *Psychological Review*, *88*, 365-368.

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.

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

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.

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.

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.

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*, *78*, 5-18.

Paper 1

Language, Cognition and Neuroscience, 2014
Vol. 29, No. 7, 781–803, http://dx.doi.org/10.1080/01690965.2013.794295

Routledge
Taylor & Francis Group

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*$_{\text{FemaleMasc}}$ *verschiedene Länder, vor allem bevorzugte er*$_{masc}$ *exotische Ziele* (The flight attendant$_{\text{FemaleMasc}}$ visited diverse countries all the time, most of all he$_{\text{masc}}$ preferred exotic destinations).[1] 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*$_{\text{FemaleMasc}}$ – *er*$_{\text{masc}}$' The flight attendant$_{\text{FemaleMasc}}$ – he$_{\text{masc}}$' 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

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).[2]

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*$_{masc}$ 'he' or *sie*$_{fem}$ '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*$_{MaleMasc}$ '[masculine] electrician' vs. *Elektrikerin*$_{MaleFem}$ '[feminine] electrician', *Kosmetiker*$_{FemaleMasc}$ '[masculine] beautician' vs. *Kosmetikerin*$_{FemaleFem}$ '[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, Gygax, Sarrasin, 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 female). 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).[3]

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.[4] 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) × 2 (grammatical gender of the pronoun: masculine/feminine) × 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).

	Factor			Measure[a]									
Region	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)
		Feminine	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)
	Female	Masculine	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)
		Feminine	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)
	Neutral	Masculine	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)
		Feminine	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)
RN spillover	Male	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)
	Female	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)
	Neutral	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)
		Feminine	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)
Verb	Male	Masculine	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	(11274)
		Feminine	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)
	Female	Masculine	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)
		Feminine	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)
	Neutral	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)
Pronoun	Male	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	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	−9.13 (36.87)	32.54 (188.13)	116.53 (510.03)	19.17 (24.91)	109.84 (363.40)
		Feminine	Masculine	−5.87 (44.54)	15.71 (136.95)	78.15 (801.35)	26.67 (25.26)	36.63 (229.61)
			Feminine	−10.54 (38.32)	−19.50 (119.43)	−29.56 (666.76)	28.33 (26.74)	−20.61 (224.79)
Pronoun spillover	Male	Masculine	Masculine	−7.76 (49.41)	0.57 (170.37)	−88.63 (673.34)	/	−26.25 (168.77)
			Feminine	5.69 (70.21)	−7.69 (142.31)	84.27 (953.24)	/	3.70 (165.79)
		Feminine	Masculine	7.08 (85.50)	20.37 (136.87)	8.67 (770.95)	/	20.96 (197.06)
			Feminine	−22.52 (58.47)	−25.86 (137.48)	62.21 (1328.49)	/	−11.99 (216.55)
	Female	Masculine	Masculine	7.57 (70.61)	−25.45 (134.32)	−37.52 (945.84)	/	−50.38 (138.89)
			Feminine	−2.71 (65.27)	10.16 (162.88)	149.02 (945.06)	/	49.34 (192.40)
		Feminine	Masculine	6.72 (87.37)	14.05 (182.16)	133.61 (682.94)	/	35.51 (205.29)
			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	−1.10 (65.58)	3.44 (156.67)	42.85 (733.91)	/	24.66 (199.84)
		Feminine	Masculine	3.45 (63.06)	0.80 (143.54)	45.01 (604.96)	/	24.49 (204.97)
			Feminine	−10.39 (67.48)	−6.95 (157.36)	−82.46 (949.72)	/	−34.59 (174.28)

(Factor: Region, Typicality (RN), Grammatical gender (RN), Pronoun; Measure[a]: FF, FP, RP, RI, TT)

[a]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.[5] 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[a]	Region	Effect	F_1	$df_{1,\ 2}$	F_2	$df_{1,\ 2}$
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	
		T × GG × P	<1		<1	
	Role noun spillover	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
		GG × P	<1		<1	
		T × GG × P	1.57	2, 78	1.76	2, 33
	Verb	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	
		T × P	<1		<1	
		GG × P	<1		<1	
		T × GG × P	<1		1.17	2, 33
	Pronoun	RN typicality (T)	7.86***	2, 78	5.78***	2, 33
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	<1		<1	
		T × GG	<1		<1	
		T × P	<1		<1	
		GG × P	<1		<1	
		T × GG × P	<1		<1	
	Pronoun spillover	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	<1		<1	
		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
FP	Role noun	RN typicality (T)	6.42***	2, 78	1.92	2, 33
		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	
	Role noun spillover	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	
		T × GG × P	<1		<1	
	Verb	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
		GG × P	<1		<1	
		T × GG × P	1.03	2, 70	2.34	2, 33
	Pronoun	RN typicality (T)	2.32	2, 78	<1	
		RN grammatical gender (GG)	<1		<1	
		Pronoun (P)	<1		<1	
		T × GG	<1		<1	
		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[a]	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[a]	Region	Effect	F_1	$df_{1,\ 2}$	F_2	$df_{1,\ 2}$
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
		T × GG × P	2.26	2, 78	2.17	2, 33
	Pronoun	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	
		GG × P	7.61***	1, 39	5.18**	1, 33
		T × GG × P	<1		<1	
	Pronoun spillover	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

[a]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. $Elektriker_{\text{MaleMasc}}$ vs. $Elektrikerin_{\text{MaleFem}}$ or $Kosmetikerin_{\text{FemaleFem}}$ vs. $Kosmetiker_{\text{FemaleMasc}}$), $M_{\text{MaleMasc}} = -29.09$, $M_{\text{MaleFem}} = 48.48$, $t(39) = -2.40$, $SEM = 32.37$; $M_{\text{FemaleMasc}} = 73.86$, $M_{\text{FemaleFem}} = -0.83$, $t(39) = 2.27$, $SEM = 32.86$. 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, $M_{Fem/masc} = 16.67$, $M_{Fem/fem} = 8.37$, $t(39) = 3.11$, $SEM = 2.67$.

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 co-reference 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	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)
		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)
	Female	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)
		Feminine	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)
	Neutral	Masculine	Masculine	3.59	(35.54)	−12.68	(120.82)	−28.33	(131.30)	34.72	(31.82)	−33.24	(238.51)
			Feminine	1.14	(43.07)	−40.61	(119.31)	−45.07	(132.74)	30.56	(32.48)	−47.04	(221.80)
		Feminine	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)
		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)
	Female	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)
		Feminine	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)
	Neutral	Masculine	Masculine	−0.43	(46.80)	−14.02	(197.89)	−36.06	(252.39)	11.11	(18.82)	6.65	(364.60)
			Feminine	3.88	(46.58)	24.71	(252.98)	−4.02	(276.35)	9.72	(15.48)	46.56	(419.27)
		Feminine	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)
		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)
	Female	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)
		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)
	Neutral	Masculine	Masculine	0.02	(28.90)	12.92	(40.40)	−40.70	(43.04)	19.44	(25.85)	−53.76	(83.98)
			Feminine	6.38	(73.42)	3.81	(87.71)	−16.53	(129.23)	19.44	(19.45)	19.46	(133.95)
		Feminine	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)
		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)
	Female	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)
		Feminine	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)

[a] Measure

Table 3 (*Continued*)

Factor				Measure[a]				
Region	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	199.00 (28.21)	209.50 (33.12)	233.20 (60.26)	16.67 (21.98)	249.87 (42.33)
		Feminine	Masculine	215.50 (72.79)	222.83 (81.19)	549.20 (374.90)	16.67 (24.08)	258.27 (149.11)
			Feminine	207.00 (41.23)	207.00 (41.23)	270.00 (121.53)	15.28 (21.93)	285.20 (138.77)
NP spillover	Male	Masculine	Masculine	3.83 (70.33)	−28.27 (204.83)	−281.83 (655.05)	/ /	−88.89 (277.02)
			Feminine	−2.64 (61.81)	2.68 (275.18)	88.12 (1039.86)	/ /	20.89 (369.27)
		Feminine	Masculine	3.26 (46.70)	4.74 (215.19)	417.27 (1393.60)	/ /	93.62 (401.66)
			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	9.59 (76.74)	−9.10 (274.09)	−0.92 (1329.78)	/ /	−49.49 (394.59)
		Feminine	Masculine	12.41 (49.76)	−103.79 (210.60)	181.54 (1127.61)	/ /	−1.94 (322.81)
			Feminine	−6.64 (50.84)	−22.76 (247.70)	−156.80 (1069.00)	/ /	−78.34 (253.46)
	Neutral	Masculine	Masculine	6.84 (52.37)	11.84 (225.95)	−150.29 (921.13)	/ /	29.83 (425.82)
			Feminine	−2.99 (49.12)	0.84 (257.29)	157.35 (1091.65)	/ /	45.46 (333.21)
		Feminine	Masculine	−5.64 (59.35)	4.33 (257.99)	179.63 (1129.87)	/ /	109.64 (381.36)
			Feminine	−5.21 (59.22)	13.32 (247.20)	−164.51 (762.00)	/ /	−13.36 (306.60)

[a]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[a]	Region	Effect	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
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	
		T × GG × NP	1.96	2, 46	<1	
	Determiner	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	
		T × GG × NP	1.03	2, 30	1.24	2, 33
	Noun	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[a]	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	
		GG × NP	1.45	1, 23	<1	
		T × GG × NP	<1		<1	
	Role noun spillover	RN typicality (T)	3.67**	2, 46	1.62	2, 33
		RN grammatical gender (GG)	<1		<1	
		Noun phrase (NP)	<1		<1	
		T × GG	<1		<1	
		T × NP	<1		<1	
		GG × NP	<1		<1	
		T × GG × NP	<1		<1	
	Determiner	RN typicality (T)	<1		<1	
		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	
		T × GG × NP	1.81	2, 30	1.06	2, 33
	Noun	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	
		T × NP	1.39	2, 8	<1	
		GG × NP	1.81	1, 4	3.06*	1, 31
		T × GG × NP	<1		<1	
	Noun phrase spillover	RN typicality (T)	1.79	2, 44	1.83	2, 33
		RN grammatical gender (GG)	8.30***	1, 22	2.55	1, 33
		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	
TT	Role noun	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	
		GG × NP	4.63**	1, 23	3.33*	1, 33
		T × GG × NP	<1		<1	
	Role noun spillover	RN typicality (T)	<1		<1	
		RN grammatical gender (GG)	<1		<1	
		Noun phrase (NP)	<1		<1	
		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	
	Determiner	RN typicality (T)	2.00	2, 42	1.20	2, 33
		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[a]	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	

[a]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'.[6]

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_{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_{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/10.1080/713755647

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*/*Grammar*/. 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 letzen 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.

Stereotypical gender	Role noun (German, masc.)	English translation	Rating score (1 = male, 7 = female)
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 2

Prominence of Gender Cues in the Assignment of Agent and Patient 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.

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$_{\text{Female+feminine}}$, who has observed many tourists$_{\text{Neutral+feminine/masculine}}$ / whom many tourists$_{\text{Neutral+feminine/masculine}}$ 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, while the opposite is observed in the assignment of patient roles. The findings are discussed within the framework of constraint-based accounts and situation model theories, 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 agent and patient roles in German

Biases in linguistic structures are often implicit and can be easily overlooked. 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 & Schlesewsky, 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 lexical information is 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 animate). 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 & Schlesewsky, 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 analyzed. 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. Following Yamamoto, we suggest that, similar to agency, gender presupposes animacy, most certainly in those cases when it is used in reference to persons. 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 masculine determiners is expressed overtly (*der*$_{Nominative}$*; den*$_{\text{Accusative}}$), while feminine determiners in some cases remain unmarked (*die*$_{Nominative/Accusative}$). 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/female role nouns are agents and shorter fixations and fewer regressions in sentences where masculine/neutral[1] 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). Our research question consisted in clarifying whether gender cues are relevant indicators of prominence in

readers' expectations about thematic agents and patients. 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$_{masculine}$ who have overlooked the cyclist$_{feminine}$ are hurt.'

(6) (ORC) *Die Studenten, die die Fahrradfahrerin übersehen hat, sind verletzt.* 'The students$_{masculine}$ whom the cyclist$_{feminine}$ 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.

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$_{\text{feminine sg/masculine or feminine plural}}$' 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$_{\text{Neutral+masculine/feminine}}$'). 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). 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$_{\text{Neutral+masculine/feminine}}$, who have overlooked the cyclist$_{\text{Neutral+feminine}}$, are hurt.'

(8) *Die Student-en/-innen, die die Fahrradfahrerin übersehen hat, sind verletzt.* 'The students$_{\text{Neutral+masculine/feminine}}$, whom the cyclist$_{\text{Neutral+feminine}}$ 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 pretests. To exclude potential confounding effects resulting from the context, a series of pretests were conducted to ensure that verbs and adjectives used in experimental items did not contain any gender information. The *verb pretest* 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 pretest fillers. The *adjective pretest* 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 pretest fillers. A total of 37 participants were asked to rate pretest 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$_{\text{Neutral+masculine/feminine}}$, whom the cyclist$_{\text{Neutral+feminine}}$ has overlooked / who have overlooked the cyclist$_{\text{Neutral+feminine}}$, are hurt'. We refer to the analyzed regions of the relative clause as *relative pronoun* (the relative pronoun with the following determiner), *RN2*, *action verb*, and *auxiliary verb.*[2]

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 neighboring 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[3].

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_{masc} = -11.40; M_{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_{masc} = -475.30; M_{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, $M_{mascSRC}$ = -198.99; M_{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_{masc} = 22.23; M_{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_{masc} = .60; M_{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.

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.

In Experiment 2 we extended the research question to gender typicality of role nouns as a cue to agency.

Experiment 2

In Experiment 2, we examined to which extent 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. Fourty 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$_{\text{feminine sg/masculine or feminine plural}}$'. 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, see Table A2).

(9) *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet hat, ist aufmerksam.* 'The flight attendant$_{\text{Female+feminine}}$, who has observed many tourists$_{\text{Neutral+masculine/feminine}}$, is attentive.'

(10) *Die Studentin, die zwei Fahrradfahrer/-innen übersehen hat, ist verletzt.* 'The student$_{\text{Neutral+feminine}}$, who has overlooked two cyclists$_{\text{Neutral+masculine/feminine}}$, is hurt.'

(11) *Die Flugbegleiterin, die viele Tourist-en/-innen beobachtet haben, ist aufmerksam.* 'The flight attendant$_{\text{Female+feminine}}$, whom many tourists$_{\text{Neutral+feminine/masculine}}$ have observed, is attentive.'

(12) *Die Studentin, die zwei Fahrradfahrer/-innen übersehen haben, ist verletzt.* 'The student$_{\text{Neutral+feminine}}$, whom two cyclists$_{\text{Neutral+feminine/masculine}}$ 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. According to *Hypothesis I*, we expected grammatically feminine patients and grammatically masculine agents to facilitate comprehension compared to masculine patients and feminine agents. *Hypothesis II*

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[4] + 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$_{masculine}$' (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$_{\text{Female+feminine}}$, who has observed many tourists$_{\text{Neutral+ masculine /feminine}}$ / whom many tourists$_{\text{Neutral+ masculine/feminine}}$ have observed, is attentive.' We refer to the analyzed 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 neighboring fixations within the neighboring 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.[5] 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_{masc} = -40.91, M_{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, $M_{NeutralSRC}$ = -52.13, $M_{FemaleSRC}$ = 2.15, $t(39) = -2.16$, $SEM = 25.18$, $p = .037$. In ORCs, there were shorter fixations after female than neutral RN1 patients, $M_{NeutralORC}$ = 56.50, $M_{FemaleORC}$ = -52.80, $t(38) = 3.37$, $SEM = 32.40$, $p = .002$. An interaction between the RN2 agents grammatical gender and the relative clause type on the *action verb* showed shorter

fixations after masculine than feminine RN2 in ORCs, $M_{mascORC}$ = -53.56, M_{femORC} = 57.25, $t(38)$ = -3.32, SEM = 33.34, 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*, $M_{mascORC}$ = -53.56, M_{femORC} = 57.25, $t(38)$ = -3.32, SEM = 33.34, p = .002; and on the *auxiliary verb*, M_{mascORC} = -61.15, M_{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 RN2 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, $M_{Female/mascORC}$ = 0.63, $M_{Female/femORC}$ = 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, $M_{NeutralSRC}$ = .23, $M_{FemaleSRC}$ = .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, $M_{mascORC}$ = .31, M_{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 facilitation effect 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.

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 Experiment 1 and 2 support *Hypothesis I* about grammatical gender influences demonstrating that grammatically masculine role nouns are rather expected to refer to agents and grammatically feminine role nouns to patients. Similarly, Experiment 2 confirms *Hypothesis II* about stereotypical gender influences showing that stereotypically female role nouns are rather expected to serve as patients and neutral role nouns 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) than neutral ones. This is in line with previous research demonstrating influences of other prominence hierarchies on reading (e.g., animacy – Mak et al., 2006; definiteness/specificity – Kretzschmar, Bornkessel-Schlesewsky, Staub, Roehm, & Schlesewsky, 2012). Research on ambiguity resolution 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 (Gennari & MacDonald, 2008; Reali & Christiansen, 2007). 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.

Another aspect that our results point to is that 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 behavior, research on the linguistic intergroup bias (Maass, Salvi, Arcuri, & Semin, 1989) and expectancy bias (Wigboldus, Semin, & Spears, 2000) demonstrated that expected behaviors were encoded at higher levels of abstraction (using adjectives that are detached from specific behaviors, 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 and patient roles to feminine/female role nouns.

Interestingly, the influence of grammatical gender 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 our study, grammatical gender could be perceived as reflecting linguistic information and stereotypical gender as reflecting background knowledge. As stereotypical gender information was not varied in Experiment 1, the contribution of background knowledge to ascribing agency was reduced. Thus, in Experiment 1 – compared to Experiment 2 – the comprehension was reduced to linguistic information only (grammatical gender markings), which resulted in gender cues affecting the resolution of ambiguous SRCs but not ORCs.

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.

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 and patient roles to feminine/female role nouns. Even though grammatical gender perfectly corresponds to biological sex in case of role nouns 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 be especially 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 11th 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 23rd 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 SocialPsychology*, *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., & Schlesewsky, 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., & Schlesewsky, 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., Modamed, 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*. Dodrecht: 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 wuing 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

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., & Schlesewsky, 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

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. P*sychology 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

---

[1] Stereotypically male role nouns were not used in the study in order to avoid the incongruity between gender cues within feminine/male role nouns which could

cause confounding processing difficulties (see, e.g. Carreiras et al., 1996; Irmen & Schumann, 2011).

[2] 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.

[3] Due to the 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 analyzed 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.

[4] Quantifiers were used instead of determiners in order to unambiguously refer to plural RN2 avoiding the misinterpretation of the determiner *die* 'the$_{\text{feminine sg/masculine or feminine plural}}$' 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$_{\text{feminine sg/plural}}$ joggers $_{\text{masculine sg/masculine plural}}$').

[5] 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.

Table 1

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

Region	Factor		Measure[a]									
	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)

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

Table 2

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

Measure[a]	Region	Effect[b]	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
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	

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

[b] GG: RN1 grammatical gender, RC: relative clause type; * $p \leq .1$, ** $p \leq .05$, *** $p \leq .01$.

Table 3

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

Region	Factor			Measure[a]									
	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)
		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)
		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)	/	/
		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)	/	/

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

Table 4

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

Measure[a]	Region	Effect[b]	F_1	$df_{1,2}$	F_2	$df_{1,2}$
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	<1		<1	
		GG	<1		<1	
		RC	<1		<1	
		SG * GG	1.49	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, 22
		GG	2.48	1, 38	1.03	1, 22
		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	

Table 4 (Continued)

Measure[a]	Region	Effect[b]	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
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

Table 4 (Continued)

Measure[a]	Region	Effect[b]	F_1	$df_{1,2}$	F_2	$df_{1,2}$
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	
		SG * GG * RC	4.94**	1, 38	4.62**	1, 22
	Role Noun 2	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	
		GG* RC	1.29	1, 38	<1	
		SG * GG * RC	<1		<1	
	Action Verb	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	
		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	
	Auxiliary Verb	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

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region
[b] SG: RN1 stereotypical gender, GG: RN2 grammatical gender, RC: relative clause type; * $p \leq .1$, ** $p \leq .05$, *** $p \leq .01$.

Appendix 1

Table A1. Experimental stimuli with masculine RN1 (Experiment 1).

Die Nachbarn, die die Spaziergängerin beobachtet hat, sind träge.
Die Psychiater, die die Kinderärztin befragt haben, sind alarmiert.
Die Designer, die die Kollegin gewürdigt hat, sind extravagant.
Die Schwimmer, die die Athletin gelobt haben, sind motiviert.
Die Patienten, die die Masseurin benachrichtigt hat, sind besorgt.
Die Künstler, die die Astrologin beauftragt haben, sind dankbar.
Die Praktikanten, die die Kunstmalerin beeindruckt hat, sind engagiert.
Die Skifahrer, die die Joggerin getroffen haben, sind müde.
Die Passagiere, die die Barkeeperin eingeladen hat, sind erfreut.
Die Tennisspieler, die die Partnerin ermutigt haben, sind sympathisch.
Die Dozenten, die die Forscherin unterstützt hat, sind beunruhigt.
Die Sportler, die die Schulanfängerin angeschaut haben, sind aufmerksam.
Die Camper, die die Kleingärtnerin beschuldigt hat, sind gereizt.
Die Zuschauer, die die Poetin begeistert haben, sind raffiniert.
Die Gynäkologen, die die Freundin informiert hat, sind gesellig.
Die Touristen, die die Autofahrerin aufgeheitert haben, sind zufrieden.
Die Geiger, die die Zuhörerin erschreckt hat, sind empört.
Die Sänger, die die Autorin gelangweilt haben, sind vulgär.
Die Musiker, die die Gastgeberin begrüßt hat, sind talentiert.
Die Apotheker, die die Ärztin empfohlen haben, sind ehrgeizig.
Die Schriftsteller, die die Archivarin angeschrieben hat, sind interessiert.
Die Schauspieler, die die Theaterbesucherin irritiert haben, sind exzentrisch.
Die Rentner, die die Mieterin ignoriert hat, sind arrogant.
Die Studenten, die die Fahrradfahrerin übersehen haben, sind verletzt.

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

Female RN1

Die Arzthelferin, die beide Patienten angerufen hat, ist beschäftigt.
Die Babysitterin, die viele Spaziergänger angeschaut haben, ist müde.
Die Diätberaterin, die beide Tennisspieler beschuldigt hat, ist unerbittlich.
Die Erzieherin, die einige Schulanfänger aufgeheitert haben, ist humorvoll.
Die Floristin, die einige Kleingärtner beraten hat, ist zufrieden.
Die Flugbegleiterin, die viele Touristen beobachtet haben, ist aufmerksam.
Die Fußpflegerin, die beide Masseure informiert hat, ist alarmiert.
Die Geburtshelferin, die viele Nachbarn geschätzt haben, ist beunruhigt.
Die Grundschullehrerin, die zwei Kinderärzte befragt hat, ist warmherzig.
Die Kindergärtnerin, die viele Freunde eingeladen haben, ist beliebt.
Die Kosmetikerin, die beide Designer erkannt hat, ist erfahren.
Die Wahrsagerin, die einige Astrologen begeistert haben, ist fasziniert.

Neutral RN1

Die Skifahrerin, die zwei Jogger ermutigt hat, ist sympathisch.
Die Praktikantin, die einige Kunstmaler überrascht haben, ist engagiert.
Die Geigerin, die viele Kollegen gewürdigt hat, ist extravagant.
Die Künstlerin, die einige Autoren gelangweilt haben, ist mittelmäßig.
Die Musikerin, die einige Zuhörer erschreckt hat, ist eigensinnig.
Die Schwimmerin, die einige Athleten provoziert haben, ist ärgerlich.
Die Sängerin, die beide Gastgeber belästigt hat, ist rücksichtslos.
Die Apothekerin, die viele Ärzte kontaktiert haben, ist ungeschickt.
Die Schriftstellerin, die einige Archivare inspiriert hat, ist ordentlich.
Die Schauspielerin, die einige Theaterbesucher irritiert haben, ist egozentrisch.
Die Rentnerin, die viele Mieter ignoriert hat, ist einsam.
Die Studentin, die zwei Fahrradfahrer übersehen haben, ist verletzt.

Paper 3

Prominence Hierarchies in the Processing of Gender-Ambiguous Anaphors in French

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.

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 rigoreux* 'In fact, the dietician$_{\text{Female+fem}}$ recommended to him/her$_{\text{gender-ambiguous}}$, so to this$_{\text{masc/fem}}$ pharmacist$_{\text{Neutral+masc/fem}}$, 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 neutral subjects/agents are easier than stereotypically female or male ones. The findings suggest that gender characteristics can be conceptualized as prominence hierarchies.

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

Prominence Hierarchies in the Processing Of Gender-Ambiguous Anaphors In French

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, 2014a). 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 (2014b), 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$_{\text{Female+feminine}}$, who has observed many tourists$_{\text{Neutral+ masculine/feminine}}$ / whom many tourists$_{\text{Neutral+ masculine/feminine}}$ have observed, is attentive'). It was observed that agent roles were assigned easier to grammatically masculine (e.g., *Touristen* 'tourists$_{\text{masculine}}$') than feminine (e.g., *Touristinnen* 'tourists$_{\text{feminine}}$') 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 & Schlesewsky, 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 rigoreux* 'In fact, the dietician$_{\text{Female+fem}}$ recommended to him/her$_{\text{gender-ambiguous}}$, so to this$_{\text{masc/fem}}$ pharmacist$_{\text{Neutral+masc/fem}}$, 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., 2014b) 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 I***). 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., 2014b) 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*$_{masc}$ / *pharmacienne*$_{fem}$ 'pharmacist$_{masc/fem}$'). 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$_{masc/fem}$'(see Examples (1) and (2).

(1) *En vérité, la diététicienne lui a recommandé, donc à ce/cette pharmacien/pharmacienne, un plan rigoreux.* 'In fact, the dietician$_{Female+fem}$ recommended to him/her$_{gender-ambiguous}$, so to this$_{masc/fem}$ pharmacist$_{Neutral+masc/fem}$, a strict plan'.

(2) *Toutefois, la vétérinaire lui a apporté, donc à ce/cette pharmacien/pharmacienne, un nouveau livre.* 'Anyways, the veterinarian$_{Neutral+fem}$ brought to him/her$_{gender-ambiguous}$, so to this$_{masc/fem}$ pharmacist$_{Neutral+masc/fem}$, 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 méchanicien* 'the mechanic$_{Male+masc}$') and 10 neutral RN1 (e.g., *le joggeur* 'the jogger$_{Neutral+masc}$'), 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 méchanicien 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$_{\text{Neutral+fem}}$> <to him/her$_{\text{gender-ambiguous}}$ has> <brought>, <so to this$_{\text{masc/fem}}$> <pharmacist$_{\text{Neutral+masc/fem}}$>, <a new book>.' The regions are referred to as *RN1*, *indirect object pronoun*, *action verb*, *demonstrative adjective*, *RN2*, and *spillover regions* in the order of their appearance in the original experimental materials.

The five 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)*, and regressions into a region* (the probability of regressions crossing the left boundary of a region during the first pass through the sentence) (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.

(Tables 1 and 2 about here)

First fixation durations. The ANOVA revealed a main effect of RN2 grammatical gender on the *RN2*, with masculine role nouns fixated shorter than feminine ones, M_{masc} = 7.61, M_{fem} = 34.89, $t(24) = -4.60$, $SEM = 5.93$, $p < .001$. A main effect of RN1 stereotypical gender emerged on the *spillover* region, with shorter fixations after neutral compared to female RN1, $M_N = 2.64$, $M_F = 18.87$, $t(24) = -3.44$, $SEM = 4.72$, $p = .002$.

First-pass reading time. A main effect of RN2 grammatical gender on the *demonstrative adjective* showed that masculine adjectives were fixated longer than feminine ones, $M_{masc} = 63.46$, $M_{fem} = -23.24$, $t(24) = 3.43$, $SEM = 25.25$, $p = .002$. A main effect of RN1 stereotypical gender emerged on the *RN2* region, again with shorter fixations after neutral compared to female RN1, $M_N = -75.54$, $M_F = -32.58$, $t(24) = -3.35$, $SEM = 12.82$, $p = .003$.

Regression path. The ANOVA revealed a main effect of RN2 grammatical gender on the *demonstrative adjective* showing longer fixations on masculine compared to feminine adjectives, $M_{masc} = 3.78$, $M_{fem} = -217.24$, $t(24) = 3.58$, $SEM = 61.75$, $p = .002$. An interaction

between RN2 grammatical and RN1 stereotypical gender emerged on *demonstrative adjective*. After female RN1, there were longer fixations on masculine than feminine adjectives, M_{Fmasc} = 69.13, M_{Ffem} = -259.73, $t(24) = 5.04$, $SEM = 65.24$, $p < .001$. After neutral RN1, fixations on masculine than feminine adjectives did not differ, M_{Nmasc} = -61.56, M_{Nfem} = -174.75, $t(24) = 1.30$, $SEM = 87.33$, *ns*.

Total fixation time. A main effect of RN2 grammatical gender emerged on *demonstrative adjective* region, with longer fixations on masculine than feminine adjectives, M_{masc} = 164.41, M_{fem} = 62.79, $t(24) = 3.73$, $SEM = 27.19$, $p = .001$.

Regressions into a region. A main effect of RN2 grammatical gender showed more regressions into the *demonstrative adjective* region in case of masculine compared to feminine adjectives, M_{masc} = .72, M_{fem} = .54, $t(24) = 2.09$, $SEM = .09$, $p = .048$. An interaction between RN2 grammatical and RN1 stereotypical gender was revealed on the *demonstrative adjective*. After neutral RN1, there were more regressions into masculine compared to feminine adjectives, M_{Nmasc} = .86, M_{Nfem} = .47, $t(24) = 3.90$, $SEM = .10$, $p = .001$. After female RN1, the probability of regressions into masculine and feminine adjectives did not differ, M_{Fmasc} = .57, M_{Ffem} = .60, $t(24) = -2.09$, $SEM = .11$, *ns*.

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

Discussion

The systematic pattern with longer fixations on and more regressions into masculine compared to feminine antecedents confirms the predicted relative difficulty of masculine patients/objects compared to feminine ones. This difficulty appeared already on the demonstrative adjective, reflecting the violation of readers' expectations upon the encounter of the first gender marking of the antecedent on the demonstrative adjective *ce* 'this$_{masc}$'. 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 as the indication of violated expectations about the antecedent. The reverse pattern in the first fixation durations

on the following *RN2* region could suggest further processing difficulties when readers reach the very next word. At the same time, regressions into the *demonstrative adjective* region seem to indicate persisting attempts to resolve the gender-ambiguous anaphor during later processing stages after the region had been left.

Stereotypical gender information affected the processing relatively later and suggests that patient roles are easier to comprehend after neutral than female agents. As indicated by regressions during reading, the final interpretation of the sentence involved the integration of stereotypical gender information. Based on the eye-tracking measures where the corresponding interactions occurred, it may be noted that female RN1 elicited expectations about RN2 grammatical gender earlier than neutral RN1 (regression path vs. regressions into the *demonstrative adjective*). However, readers' expectations about grammatical gender of the antecedent were not influenced by the stereotypical gender of the RN1, as the same reading patterns (longer reading of masculine vs. feminine antecedents) were observed after female and neutral RN1.

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$_{\text{Male+masc}}$ passed to him/her$_{\text{gender-ambiguous}}$, so to this$_{\text{masc/fem}}$ patient$_{\text{Neutral+masc/fem}}$, an oxygen mask'.

(4) *Toutefois, le vétérinaire lui a apporté, donc à ce/cette pharmacien/pharmacienne, un nouveau livre.* 'Anyways, the veterinarian$_{\text{Neutral+masc}}$ brought to him/her$_{\text{gender-ambiguous}}$, so to this$_{\text{masc/fem}}$ pharmacist$_{\text{Neutral+masc/fem}}$, 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)

First fixation durations. The ANOVA revealed the main effect of RN1 stereotypical gender on the *RN1* region with neutral RN1 fixated shorter than male, M_N = -13.91, M_M = -3.04, $t(32) = 2.22$, $SEM = 4.90$, $p = .034$.

First-pass reading time. The main effect of RN2 grammatical gender first emerged on the *demonstrative adjective* region with longer fixations when the region contained masculine than feminine adjectives, M_{masc} = .20, M_{fem} = -49.83, $t(32) = 2.75$, $SEM = 18.19$, $p = .010$[1].

Regression path. The ANOVA revealed the main effect of RN1 stereotypical gender on the *RN1* and *RN2* regions with shorter fixations on (after) neutral than male RN1, M_N = -285.08, M_M = -90.69, $t(32) = 7.05$, $SEM = 27.59$, $p < .001$; M_N = -85.10, M_M = 63.13, $t(32)$ = 4.88, $SEM = 30.40$, $p < .001$.

The effect of grammatical gender revealed on the *demonstrative adjective* and *RN2* region showed longer fixations on masculine than feminine antecedents, M_{masc} = -93.20, M_{fem}

= -324.67, $t(32) = 7.82$, $SEM = 29.61$, $p < .001$; $M_{masc} = 47.37$, $M_{fem} = -69.34$, $t(32) = 2.99$, $SEM = 39.05$, $p = .005$.

Total fixation time. The ANOVA revealed the main effect of RN2 grammatical gender on the *demonstrative adjective* region with longer fixations on the region containing masculine than feminine demonstratives, $M_{masc} = 58.46$, $M_{fem} = .31$, $t(32) = 2.24$, $SEM = 25.95$, $p = .032$.

Response accuracy. The mean comprehension question accuracy was 92%[2].

Discussion

Like in Experiment 1, masculine antecedents were fixated longer than feminine ones indicating the relative difficulty in comprehending masculine referents as objects/patients compared to feminine ones. This time, however, the difficulty appeared on both the demonstrative adjective and the RN2, reflecting the violation of gender expectations upon the encounter of the grammatically masculine antecedent.

The processing of stereotypical gender information suggests that neutral RN1 were easier to comprehend than male RN1. Readers' expectation about the grammatical gender of gender-ambiguous backwards anaphors did not seem to be influenced by stereotypical gender of RN1, since the analyses revealed no interactions. At the same time, the integration of stereotypical gender information seems to be relevant to complete the processing of the sentence.

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 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 which served as objects/patients in the sentence, while

grammatically masculine objects/patients produced processing difficulties. 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 first role noun. This suggests that such preference was not a result of a lexical priming after reading a grammatically feminine role noun. Our findings extend the results of previous research that showed the association between syntactic preferences and variations in thematic role characteristics – namely, animacy and definiteness (e.g., Mak et al., 2006; Kretzschmar, Bornkessel-Schlesewsky, Staub, Roehm, & Schlesewsky, 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 according to 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.

While grammatical gender mainly affected the *demonstrative adjective* region, stereotypical gender influences of the first role noun were observed on the *RN2* and the following *spillover* region (in addition to the *RN1* itself). This is in line with previous research on gender influences in anaphor resolution showing that stereotypical gender affects the resolution of semantically rich rather than pronominal references (Esaulova et al., 2014a). This perspective could explain the differentiation of gender effects for each of the two parts that constituted antecedents (demonstrative adjectives and RN2): demonstrative adjectives were rather informative in terms of grammatical gender, while RN2 were more semantically rich and reflected stereotypical gender influences. Moreover, since all RN2 were neutral in

regard to gender stereotypes, readers appear to relate them easier to neutral RN1 than to gender typical ones. Specific gender stereotypes associated with male and female role nouns, on the other hand, seem to slow down the sentence processing. Even though these findings speak for the relevance of stereotypical gender information in the comprehension of sentences in both of our experiments, stereotypical gender of RN1 does not seem to modulate readers' expectations regarding RN2 patients. This suggests an interesting contrast with our earlier findings, which demonstrated readers' expectations about gender characteristics of agents and patients in German (Esaulova et al., 2014b). In this earlier study, preferences for female rather than neutral patients in German relative clause resolution were clearly indicated, while the results of the present study did not show such a difference in French backwards anaphor resolution. Since male role nouns were not used due to the design of German experiments, 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 subject/agent role nouns that varied in stereotypical gender. This may have resulted in stereotypical gender affecting object/patient role nouns as part of a more active anaphor resolution process that required the integration of different types of gender information. Moreover, 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. 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, Schlesewsky, Phillipp, and Bornkessel-Schlesewsky (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, neutral role nouns in French seem to rank higher on the prominence hierarchy compared to both stereotypically female and male ones. This finding is in contrast to that in German, which indicates that gender prominence hierarchy in the assignment of thematic patients may differ across languages but certainly is not language-specific. 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 correlation between gender hierarchy and other prominence hierarchies, and the cross-linguistic validity of prominence hierarchies of gender.

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., & Schlesewsky, 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*. Dodrecht: Foris.

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014a). 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. (2014b). *Prominence of gender cues in the assignment of agent and patient 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., & Schlesewsky, 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

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.

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.

---

[1] The main effect of grammatical gender on the *indirect object pronoun* region is a non-systematic effect appearing before the grammatical gender is introduced in a sentence and therefore is considered irrelevant for the report.

[2] Regressions into a region are not reported in this section, since no significant effects relevant for the question under study occurred in the measure.

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

Region	Factors		Measure[a]									
	Stereotypical Gender (RN 1)	Grammatical Gender (RN 2)	FF		FP		RP		TT		RI	
Role Noun 1	Neutral	Masculine	-10.13	(49.18)	-31.77	(157.99)	-262.55	(273.18)	178.70	(414.21)	0.83	(0.65)
		Feminine	-10.81	(43.83)	-54.90	(206.08)	-303.66	(303.55)	185.77	(461.92)	0.94	(0.55)
	Female	Masculine	-4.91	(48.29)	-57.66	(181.25)	-261.60	(235.41)	118.62	(445.45)	0.93	(0.76)
		Feminine	-14.28	(43.02)	-108.49	(103.15)	-283.60	(359.74)	99.94	(522.53)	0.88	(0.63)
Indirect Object Pronoun	Neutral	Masculine	2.34	(44.81)	69.92	(63.89)	269.63	(159.02)	28.70	(245.43)	0.47	(0.46)
		Feminine	-7.88	(41.45)	66.34	(71.22)	214.06	(147.94)	15.92	(227.39)	0.59	(0.46)
	Female	Masculine	-1.74	(30.79)	69.48	(61.04)	254.93	(300.75)	69.64	(293.58)	0.54	(0.57)
		Feminine	22.95	(59.13)	82.78	(84.63)	254.97	(168.06)	17.45	(207.81)	0.42	(0.33)
Action Verb	Neutral	Masculine	3.18	(44.91)	-12.91	(74.21)	105.33	(532.86)	-76.61	(227.86)	0.27	(0.20)
		Feminine	-6.41	(41.73)	-27.85	(77.93)	4.93	(218.08)	-62.29	(256.67)	0.35	(0.36)
	Female	Masculine	-3.52	(42.98)	-22.60	(84.89)	101.06	(418.89)	-68.27	(244.74)	0.33	(0.25)
		Feminine	17.79	(57.66)	1.83	(101.87)	-12.21	(130.29)	-36.64	(258.14)	0.37	(0.30)
Demonstrative Pronoun	Neutral	Masculine	7.10	(37.88)	49.98	(114.71)	-61.56	(196.06)	133.62	(333.88)	0.86	(0.50)
		Feminine	-2.72	(40.61)	-36.31	(153.97)	-174.75	(526.08)	28.93	(377.66)	0.47	(0.39)
	Female	Masculine	0.74	(41.18)	76.93	(154.92)	69.13	(384.30)	195.21	(385.80)	0.57	(0.46)
		Feminine	-2.17	(39.35)	-10.17	(161.45)	-259.73	(294.06)	96.64	(414.57)	0.60	(0.46)
Role Noun 2	Neutral	Masculine	8.20	(48.80)	-95.44	(66.10)	-12.05	(365.13)	-86.70	(232.85)	0.14	(0.20)
		Feminine	38.24	(61.34)	-55.63	(91.94)	-165.18	(302.95)	-80.94	(241.72)	0.16	(0.20)
	Female	Masculine	7.01	(40.86)	-31.68	(94.13)	17.76	(182.36)	-105.70	(176.46)	0.16	(0.19)
		Feminine	31.53	(57.83)	-33.48	(81.39)	-0.45	(300.64)	-83.79	(184.24)	0.11	(0.17)
Spillover	Neutral	Masculine	3.94	(38.97)	90.42	(283.72)	256.45	(1162.72)	-20.25	(412.41)	/	/
		Feminine	1.34	(41.20)	93.82	(268.32)	277.68	(1119.59)	-27.32	(409.94)	/	/
	Female	Masculine	18.46	(39.31)	101.29	(239.93)	442.45	(1360.20)	-0.94	(420.78)	/	/
		Feminine	19.28	(45.41)	47.02	(246.15)	384.28	(1370.20)	-72.42	(460.97)	/	/

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

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

Measure[a]	Region	Effect	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
FF	Role Noun 1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	3.99*	1, 24	<1	
		GG	<1		<1	
		Typicality * GG	5.86**	1, 24	3.46*	1, 18
	Action Verb	Typicality	1.44	1, 24	1.26	1, 18
		GG	1.42	1, 24	<1	
		Typicality * GG	4.10*	1, 24	6.86**	1, 18
	Demonstrative Pronoun	Typicality	<1		<1	
		GG	<1		2.51	1, 18
		Typicality * GG	<1		1.08	1, 18
	Role Noun 2	Typicality	<1		<1	
		GG	21.15***	1, 24	9.79***	1, 18
		Typicality * GG	<1		<1	
	Spillover	Typicality	11.83***	1, 24	4.76**	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
FP	Role Noun 1	Typicality	2.40	1, 24	2.75	1, 18
		GG	2.43	1, 24	2.56	1, 18
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Action Verb	Typicality	1.08	1, 24	<1	
		GG	<1		<1	
		Typicality * GG	1.71	1, 24	2.37	1, 18
	Demonstrative Pronoun	Typicality	1.33	1, 24	1.38	1, 18
		GG	11.79***	1, 24	35.47***	1, 18
		Typicality * GG	<1		<1	
	Role Noun 2	Typicality	11.23***	1, 24	8.43***	1, 18
		GG	2.03	1, 24	1.30	1, 18
		Typicality * GG	2.34	1, 24	2.99	1, 18
	Spillover	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
RP	Role Noun 1	Typicality	<1		<1	
		GG	1.09	1, 24	<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Action Verb	Typicality	<1		<1	
		GG	1.47	1, 24	5.36**	1, 18
		Typicality * GG	<1		<1	
	Demonstrative Pronoun	Typicality	<1		<1	
		GG	12.81***	1, 24	34.38***	1, 18
		Typicality * GG	5.46**	1, 24	7.90**	1, 18
	Role Noun 2	Typicality	5.92**	1, 24	2.94	1, 18
		GG	2.93*	1, 24	3.38*	1, 18
		Typicality * GG	3.38*	1, 24	1.60	1, 18
	Spillover	Typicality	1.64	1, 24	<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
TT	Role Noun 1	Typicality	1.43	1, 24	<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	<1		<1	
		GG	1.62	1, 24	5.95**	1, 18
		Typicality * GG	<1		1.90	1, 18
	Action Verb	Typicality	1.09	1, 24	<1	
		GG	<1		1.65	1, 18
		Typicality * GG	<1		<1	
	Demonstrative Pronoun	Typicality	2.25	1, 24	3.96*	1, 18
		GG	13.97***	1, 24	10.29***	1, 18
		Typicality * GG	<1		<1	
	Role Noun 2	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Spillover	Typicality	<1		<1	
		GG	1.23	1, 24	2.03	1, 18
		Typicality * GG	<1		1.10	1, 18

Table 2 (Continued)

Measure[a]	Region	Effect	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
RI	Role Noun 1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	1.15	1, 24	1.71	1, 18
	Indirect Object Pronoun	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	4.51**	1, 24	4.16*	1, 18
	Action Verb	Typicality	<1		<1	
		GG	1.57	1, 24	2.08	1, 18
		Typicality * GG	<1		<1	
	Demonstrative Pronoun	Typicality	1.13	1, 24	<1	
		GG	4.35**	1, 24	5.30**	1, 18
		Typicality * GG	10.57***	1, 24	7.40**	1, 18
	Role Noun 2	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		1.49	1, 18

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region
* $p \leq .1$, ** $p \leq .05$, *** $p \leq .01$.

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

Region	Factors		Measure[a]									
	Stereotypical Gender (RN 1)	Grammatical Gender (RN 2)	FF		FP		RP		TT		RI	
Role Noun 1	Neutral	Masculine	1.04	(50.98)	3.90	(139.69)	-97.69	(192.47)	160.36	(301.92)	0.86	(0.58)
		Feminine	-7.12	(43.58)	-29.01	(113.48)	-83.69	(198.86)	163.48	(282.15)	0.96	(0.51)
	Female	Masculine	-14.76	(44.13)	-39.31	(148.88)	-282.45	(249.89)	164.80	(369.56)	0.95	(0.67)
		Feminine	-13.05	(38.25)	-11.25	(200.19)	-287.72	(210.86)	175.42	(328.30)	0.95	(0.64)
Indirect Object Pronoun	Neutral	Masculine	-6.94	(36.50)	57.31	(62.81)	260.34	(125.60)	48.55	(221.54)	0.56	(0.50)
		Feminine	1.53	(40.19)	63.99	(64.19)	246.60	(112.64)	65.19	(204.04)	0.65	(0.52)
	Female	Masculine	-12.83	(38.84)	39.38	(62.77)	222.62	(141.87)	23.28	(183.19)	0.56	(0.45)
		Feminine	-3.84	(41.61)	67.23	(83.46)	274.36	(153.81)	22.22	(177.58)	0.50	(0.42)
Action Verb	Neutral	Masculine	4.30	(43.71)	-1.65	(87.73)	35.47	(129.11)	-13.24	(218.76)	0.31	(0.25)
		Feminine	-2.49	(42.58)	2.95	(78.67)	40.61	(167.51)	-16.05	(216.52)	0.34	(0.37)
	Female	Masculine	6.25	(41.28)	-18.54	(85.83)	17.75	(186.15)	-22.33	(191.74)	0.31	(0.28)
		Feminine	9.03	(48.81)	1.57	(107.50)	30.08	(174.21)	-27.44	(223.55)	0.29	(0.28)
Demonstrative Pronoun	Neutral	Masculine	5.24	(37.44)	9.69	(109.19)	-69.62	(228.84)	49.76	(307.03)	0.67	(0.68)
		Feminine	5.85	(59.25)	-40.26	(144.94)	-329.19	(188.79)	22.56	(261.76)	0.70	(0.64)
	Female	Masculine	-1.82	(53.36)	-9.29	(116.21)	-116.79	(129.85)	67.15	(298.50)	0.74	(0.62)
		Feminine	-2.36	(42.54)	-59.39	(133.47)	-320.15	(250.79)	-21.94	(280.57)	0.69	(0.68)
Role Noun 2	Neutral	Masculine	11.72	(46.41)	-16.45	(87.15)	91.01	(396.51)	-28.55	(344.96)	0.20	(0.29)
		Feminine	20.67	(47.24)	-15.37	(103.88)	35.26	(348.79)	-27.66	(208.28)	0.17	(0.16)
	Female	Masculine	14.93	(48.80)	-42.59	(116.89)	3.73	(400.54)	-1.69	(255.06)	0.17	(0.21)
		Feminine	30.08	(54.82)	-31.44	(85.14)	-173.94	(274.21)	-6.60	(238.52)	0.23	(0.20)
Spillover	Neutral	Masculine	5.61	(40.81)	66.18	(245.82)	370.46	(1005.96)	-16.29	(308.88)	/	/
		Feminine	11.78	(49.51)	41.60	(180.74)	327.29	(962.85)	-33.44	(340.01)	/	/
	Female	Masculine	-0.69	(37.92)	96.33	(280.79)	481.73	(1188.95)	107.93	(429.53)	/	/
		Feminine	7.44	(53.25)	50.95	(211.55)	395.16	(1140.32)	28.93	(387.30)	/	/

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region

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

Measure[a]	Region	Effect	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
FF	Role Noun 1	Typicality	4.92**	1, 32	4.23*	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	1.59	1, 32	1.03	1, 18
		GG	3.29*	1, 32	3.44*	1, 18
		Typicality * GG	<1		<1	
	Action Verb	Typicality	2.07	1, 32	2.61	1, 18
		GG	<1		<1	
		Typicality * GG	<1		1.49	1, 18
	Demonstrative Pronoun	Typicality	3.09*	1, 32	1.96	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Role Noun 2	Typicality	1.22	1, 32	1.23	1, 18
		GG	2.61	1, 32	1.93	1, 18
		Typicality * GG	<1		<1	
	Spillover	Typicality	<1		<1	
		GG	1.02	1, 32	1.97	1, 18
		Typicality * GG	<1		<1	
FP	Role Noun 1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	2.31	1, 32	2.26	1, 18
	Indirect Object Pronoun	Typicality	1.13	1, 32	<1	
		GG	5.89**	1, 32	6.04**	1, 18
		Typicality * GG	1.11	1, 32	<1	
	Action Verb	Typicality	1.47	1, 32	<1	
		GG	2.18	1, 32	2.61	1, 18
		Typicality * GG	<1		1.13	1, 18
	Demonstrative Pronoun	Typicality	2.87*	1, 32	1.33	1, 18
		GG	7.56***	1, 32	17.43***	1, 18
		Typicality * GG	<1		<1	
	Role Noun 2	Typicality	4.46**	1, 32	1.19	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Spillover	Typicality	<1		<1	
		GG	1.28	1, 32	1.11	1, 18
		Typicality * GG	<1		<1	
RP	Role Noun 1	Typicality	49.64***	1, 32	8.05**	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	<1		<1	
		GG	<1		2.25	1, 18
		Typicality * GG	2.77	1, 32	1.31	1, 18
	Action Verb	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Demonstrative Pronoun	Typicality	<1		1.32	1, 18
		GG	61.10***	1, 32	169.53***	1, 18
		Typicality * GG	1.18	1, 32	1.23	1, 18
	Role Noun 2	Typicality	23.77***	1, 32	9.99***	1, 18
		GG	8.93***	1, 32	5.59**	1, 18
		Typicality * GG	3.65*	1, 32	2.06	1, 18
	Spillover	Typicality	1.89	1, 32	<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
TT	Role Noun 1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	3.48*	1, 32	2.25	1, 18
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Action Verb	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Demonstrative Pronoun	Typicality	<1		<1	
		GG	5.02**	1, 32	4.17*	1, 18
		Typicality * GG	2.47	1, 32	2.13	1, 18
	Role Noun 2	Typicality	1.32	1, 32	<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Spillover	Typicality	8.01***	1, 32	2.18	1, 18
		GG	1.27	1, 32	1.16	1, 18
		Typicality * GG	<1		<1	

Table 2 (Continued)

Measure[a]	Region	Effect	F_1	$df_{1, 2}$	F_2	$df_{1, 2}$
RI	Role Noun 1	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Indirect Object Pronoun	Typicality	2.44	1, 32	2.39	1, 18
		GG	<1		<1	
		Typicality * GG	2.21	1, 32	1.25	1, 18
	Action Verb	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Demonstrative Pronoun	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	<1		<1	
	Role Noun 2	Typicality	<1		<1	
		GG	<1		<1	
		Typicality * GG	1.64	1, 32	2.09	1, 18

[a] FF: first fixation durations, FP: first-pass reading time, RP: regression path, TT: total fixation time, RI: regressions into the region
*$p \leq .1$, **$p \leq .05$, ***$p \leq .01$.

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 acceuil 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 oxygene.
C'est-à-dire, le couvreur lui a glissé, donc à ce piéton, une phrase bizarre.
Evidemment, 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 prédit, 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.

Paper 4

Applied Psycholinguistics, page 1 of 30, 2014
doi:10.1017/S0142716414000010

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

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

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

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

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

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

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

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

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

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

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

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

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$_{\text{feminine}}$"). In the last sentence, which contained the anaphoric reference, no effect of typicality was found when referent gender was already established via morphological features of the role noun and its preceding article. This study shows that when a role noun is encountered, the gender information provided by stereotypicality is compared with, and if necessary overruled by, gender cues provided by the local morphology. Once the referent gender is signaled through grammatical cues, no typicality effect emerges in the subsequent steps of discourse comprehension.

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

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

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.[2] Previous studies have shown that word recognition can be facilitated by a prime word with matching grammatical gender and inhibited by a prime with mismatching grammatical gender (about the priming effect of grammatical suffixes, see Bates, Devescovi, Hernandez, & Pizzamiglio, 1996; Cubelli, Lotto, Paolieri, Girelli, & Job, 2005). If the descriptions actually activate morphological gender cues, then target items with corresponding grammatical gender are likely to be processed faster than the same items with the opposite grammatical gender. The possible activation of grammatical gender was tested through a priming task, employing job descriptions as a prime and role nouns as a target. To control for the influence of gender typicality, the test was conducted employing gender-typical as well as gender-neutral role nouns.

Method

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

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

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

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

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

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

Results

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

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

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

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.[4] Reading was binocular, and participants' dominant eye was tracked. The experiment began after a calibration procedure. The presentation of sentences started with a small rectangle indicating the position of the first word of the sentence. The item appeared only when this rectangle was fixated accurately. Sentences were displayed in a monospaced 22 point Lucida Console font. After reading a sentence, participants pressed a button on a keypad to prompt the next item or a question. Two buttons of the keypad were used for answering the questions.

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

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

Design and hypotheses

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

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

Results

Eye-tracking data.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Relating eye movements to individual measures.

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

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

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

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

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

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

Discussion

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

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

(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 (Diekman & Eagly, 2000).

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

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

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 Massachussets 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 β are enhanced (first fixations: $\beta = 0.41$, $p = .047$; first pass: $\beta = 0.36$, $p = .081$; and total time: $\beta = 0.57$, $p = .004$).
8. Correlations between individual IAT results and questionnaire scores were also analyzed; no reliable correlation was found (maximum correlation coefficient: $r = -17$, $p > 1$).

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, *23*, 464–485.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Paper 5

Gender Typicality Effects On Pronominal Anaphor Resolution

Chiara Reali, Yulia Esaulova, and Lisa von Stockhausen

University of Duisburg-Essen

Author Note

Chiara Reali, Department of Psychology, University of Duisburg-Essen, Essen, Germany; Yulia Esaulova, 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 Chiara Reali, Department of Psychology, Berliner Platz 6-8, Essen 45127, Germany. Tel.: +49 201 1837024; E-mail: chiara.reali@uni-due.de.

Abstract

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 of the study 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 an early measure 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.

Gender typicality effects on pronominal anaphor resolution

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$_{\text{feminine}}$). The gender typicality of lexemes results from the likelihood of personal nouns to refer to men or women. Thus the noun *nurse* has female typicality and *surgeon* male typicality, because of their likelihood to be associated with a female or a male referent, respectively, as shown in typicality ratings (cf. Kennison & 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. This 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, which does not possess grammatical gender ("natural gender language", see Hellinger & Bussmann, 2001). Previous studies have analyzed the gender typicality effect of strongly stereotyped roles; however, as most professional roles lie in the range of moderate stereotypicality, we also explore the effect of roles with lesser degrees of gender typicality on the gender representation of the referent.

In the present paper 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. Psycholinguistic studies, both on natural gender and grammatical gender languages, have investigated the influence of gender typicality on anaphor resolution using personal role nouns to provide information on gender typicality. Commonly, these studies find a slowdown in resolution of both personal and reflexive pronouns that are incongruent with the

gender typicality of role noun antecedents. For example, in a reading time study in English, Kennison and Trofe (2003) presented gender-typical role nouns 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.

Most psycholinguistic studies investigating gender typicality effects on anaphor resolution in English used reflexive pronouns (*himself/herself*) to reveal referential gender (Sturt, 2003; Duffy & Keir, 2004; Kreiner, Sturt & Garrod, 2008 for eye-tracking methodology; Osterhout, Bersick & McLaughlin, 1997 for ERP methodology). 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.

A gender mismatch effect between anaphor and antecedent is also documented in studies on grammatical gender languages. However, in grammatical 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, Reali and von Stockhausen (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 Elecktrikerin gute Einfälle, regelmäßig plante er/sie neue Projekte*. 'Often had the electrician$_{\text{masculine/feminine}}$ good ideas, regularly planned he/she new projects.'). In the

condition of a mismatch between grammatical gender and gender typicality of the role noun results showed a mismatch effect not only on the anaphor region but also on the role noun region. The antecedent contained grammatical gender markings (either masculine or feminine ones), therefore the effect of the noun's gender typicality on anaphor resolution resulted from a combined processing of grammatical gender cues and typicality (see also Gygax, Gabriel, Sarrasin, Garnham, & Oakhill, 2008; Irmen & Schumann, 2011).

The interplay of grammatical gender and gender typicality was further explored in a reading study on another grammatical gender language (Italian): Cacciari, Corradini, Padovani and Carreiras (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$_{\text{feminine}}$) or a bigender role noun (a noun which can function both as a feminine and a masculine noun, e.g., *assistente,* assistant) was introduced as antecedent for an anaphoric pronoun. The anaphor could match or mismatch the typical context and/or the grammatical gender of the epicene. Results showed that for bigender role nouns, which did not present a defined grammatical gender, the influence of gender typicality was essential to trigger the mismatch effect; however, when the antecedent was an epicene the grammatical

gender of the role noun, even though purely formal, affected the resolution of the anaphor and interfered with the typicality effect.

As the literature shows, role nouns are 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 role most 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 (cf. Reali, Esaulova & von Stockhausen, in press). The 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, Gygax, & Gabriel, 2013). Starting from these considerations, it is the aim of the present study to analyze the processing of gender typicality in a natural gender language and to compare the resolution process with previous studies conducted on a grammatical gender language (cf. Reali et al., in press).

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 two possible hypotheses: The first hypothesis assumes that varying the degree of gender typicality of the role descriptions will cause a modulation of the gender typicality effect, with lesser reading disruption in the condition of a mismatch with low typicality; the second hypothesis assumes that the effect of gender typicality is an all-or-none effect, with only highly stereotyped roles activating the mismatching gender representation and roles with low typicality causing no disruption in the resolution process.

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.

Materials

Materials were created to provide gender-typical information associated with different occupational activities without employing role nouns. The experimental sentences were translated and adapted from a previous experiment which had been conducted in German (for details of material pretesting see Reali et al., in press)[1]. The roles had been selected from a published collection of gender typicality ratings for different languages (Gabriel, Gygax, Sarrasin, Garnham & Oakhill, 2008; Irmen, 2007; Kennison & Trofe, 2003). The experimental sentences were tested for gender typicality with a sample of students from the University of Heidelberg, Germany, who did not participate in the present eye-tracking experiment. These gender typicality classifications were assumed to be valid for the present experiment, given the comparable characteristics of the two samples. In the pretest, participants estimated to which extent a specific profession (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 were rated as clearly typically male, typically female, and neutral (12 items per typicality).

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 subject of the previous sentence,

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 electric 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 in order to ensure reading for comprehension.

Design and hypothesis

The experiment was designed to test the interaction between the gender typicality of the occupational role and the gender of the anaphoric reference. This resulted in a 3 (Gender Typicality: male, female or neutral) x 2 (Pronoun: masculine or feminine) within-subjects design. In accord with the German study (Reali et al., in press) 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. Based on previous results (Cacciari & Padovani, 2007; Reali et al., in press), we assumed the

mismatch effect to be more pronounced in the mismatch condition with the masculine compared to the feminine pronoun.

Participants

Thirty-one students (17 women and 14 men) from the University of Sussex, UK, with normal or corrected-to-normal vision participated in the study. Their mean age was 21 years (*SD* = 3,9). They received monetary compensation or course credit for their participation.

Procedure

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. Materials were presented with the software Eyetrack[2]. Reading was binocular and the 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 when the rectangle was fixated accurately. 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 in random order. Sentences were displayed in a monospaced 22-point Lucida Console font and consisted of three lines: the first two lines contained the role description, the third line the target sentence with the anaphoric reference. 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 of the job descriptions that were presented in the eye-tracking

session.

Analysis

We analyzed fixation times and regression patterns on different regions of the experimental items. The segmentation of items into regions of analysis is shown in Table 1. The role description was split into two regions: the first region contained the initials referring to the subject plus the following verb, the second region the rest of the role description. The first region of the role description constituted an area of analysis for possible regressions to the subject of the context. The target sentence was divided into four region of analysis: adverb region, anaphor region, spillover region and final region. Following common practice in eye-tracking research, fixations below 70 ms and above 600 ms were removed from the data analysis (4.1% of the data). Since the regions of interest differed in length across items, analyses were based on residual fixation times after a length correction (Trueswell, Tanenhaus, & Garnsey, 1994).

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 regression 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 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. First pass 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; Sturt, 2003). In general, longer fixation times and a higher probability of regressions are indicative of greater difficulty in processing the respective region.

Analyses were computed on the basis of participant means across items (F_1) and item means across participants (F_2) (Clark, 1973). The F_1 ANOVA was performed with Gender Typicality (male, female, neutral) x Pronoun (masculine, feminine) as within-subjects factors. The F_2 ANOVA was performed with Gender Typicality (male, female, neutral) as between-items factor and Pronoun (masculine, feminine) as within-subjects factor.

Results

Eye-tracking results.

In the following we report and interpret results that were consistent in both analyses (i.e., reliable [$p \leq .05$] in both F_1 and F_2, or significant in one analysis and marginally significant [$p \leq .1$] in the other)[3]. Table 2 shows the means of residual fixation times and probabilities of regressions, separated for experimental factors; details of the statistical tests are given in Table 3.

First pass time. The first reliable interaction effect between Gender Typicality and Pronoun was detected in first pass time on the spillover region immediately following the anaphoric pronoun, where the incongruent pronoun was fixated longer than the congruent one. Pairwise comparisons of the pronouns revealed that the effect was statistically significant only for the masculine pronoun, M_{maleHE}=-31,61, $M_{fe\text{maleHE}}$=27,95, t_{30}=-3,73, p=.001; M_{maleSHE}=-14,70, $M_{\text{femaleSHE}}$=-12,37, *ns*.

Regressions out of a region. The interaction effect emerged in the proportion of regressions out of the last region of the target sentence. The proportion of regressions was higher after reading an incongruent compared to a congruent anaphor.

Pairwise comparisons showed that the effect was significant only for the masculine pronoun, M_{maleHE}=47,31, $M_{femaleHE}$=58,60, t_{30}=-2,46 , p=.020 $M_{maleSHE}$=53,23, $M_{femaleSHE}$=52,15, *ns*.

Total fixation time. The interaction between Gender Typicality and Pronoun emerged on the subject region of the role description, which contained the subject (the initials) and the following verb. The region was fixated longer after reading an incongruent rather than congruent anaphoric pronoun. Pairwise comparisons revealed that the effect was statistically significant for the feminine but not for the masculine pronoun, M_{maleHE}=-10,69, $M_{femaleHE}$=18,14, *ns*; $M_{maleSHE}$=68,41, $M_{femaleSHE}$=-29,00, t_{30}= 3,01, p=.005.

To check for a possible influence of participant gender, an additional ANOVA was run on the measures and regions where the gender mismatch effect occurred, with participant gender as a between-subject factor. The analysis showed that participant gender did not affect eye movements, neither as main effect nor in interaction with other factors, $F's_{1,29} \leq 2{,}06$, *ns*.

Gender typicality ratings and eye movements.

Statistical analyses of typicality ratings were based on the data collected before the eye-tracking experiment, from a sample which did not participate in the eye-tracking experiment[4]. 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 mismatch effect consistently occurred in the ANOVA. Since pairwise comparisons revealed an asymmetry between the masculine and the feminine pronoun, we conducted separate analyses for the two anaphoric pronouns. Results revealed that typicality ratings predicted first pass fixation times after a masculine anaphor (β = .35,

$p < .05$), and total fixation times after a feminine anaphor ($\beta = -.33$, $p < .05$). As the scale for typicality ratings presented the poles 1=male and 7=female, the β coefficient showed a direct correlation in the condition masculine pronoun, with lower ratings predicting shorter fixations after the pronoun *he*, and a reverse correlation for the feminine pronoun, namely lower ratings predicting longer fixations after the pronoun *she*. These results indicate that eye movements on the regions where the mismatch effect emerged corresponded to the degree of gender typicality expressed in the explicit typicality ratings of the respective items.

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.

Results showed that a mismatch effect between gender typicality of the description and pronoun gender occurred reliably in a measure of early processing on the region following the anaphoric pronoun. Moreover, this interaction effect was detected consistently in a measure of early to 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 gender-incongruent items, and in one measure of late processing, namely the total amount of time spent on the initial region of the description in a mismatching item. Furthermore, correlational analyses with gender typicality ratings showed that the typicality degree of the different items predicted the mismatch effect revealed by eye movements, confirming the validity of the description paradigm as a tool to investigate gender typicality.

The location of the early effect is consistent with data from reading studies in English which employed role nouns as antecedents and personal pronouns as anaphors (Kennison & Trofe, 2003). However, the effect appears to be delayed in location and time compared to studies employing reflexive pronouns to trigger the mismatch (e.g., Sturt, 2003). This may be due to the fact that reflexives require faster syntactic processing than personal pronoun anaphors.

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., on press). Interestingly, the mismatch effect occurred earlier (in first fixations) in the German study and already on the pronoun region itself. A possible explanation for this difference 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 may 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 effect asymmetry for the masculine and the feminine pronoun which was revealed in the pairwise comparisons. Based on the results of previous studies, our hypothesis predicted that the gender mismatch effect would be triggered by the masculine rather than the feminine pronominal anaphor in the incongruent condition. This hypothesis was partially confirmed: pairwise analyses of the interaction effect on the target sentence revealed that the gender mismatch was reliable only for the masculine pronoun, which produced an impairment in the sentence processing after an incongruent (female) role description. In later stages of processing, however, the feminine anaphor triggered the mismatch effect in incongruent contexts. The effect asymmetry which surfaced on the target sentence may be interpreted as indicative of readers' difficulty to integrate a masculine 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 in an initial stage. In a subsequent stage, during the re-reading of the stereotypical context, the effect was reversed and the feminine anaphor triggered the mismatch effect. This time pattern suggests that the conflict between masculine anaphors and typically female contexts may first be perceived as more surprising and lead to a mismatch effect in earlier stages, whereas feminine anaphors in male contexts are dealt with only in a subsequent stage, where the mismatching masculine pronouns are already resolved.

Experiment 2

Experiment 1 investigated the effect of typicality with the help of highly gender-typical items. However, the selection of such items excluded most occupational roles, as these tend to fall in the range between gender-typical and neutral (see the Material

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? The design of Experiment 2 was identical to Experiment 1, but employed roles that had been rated as lower in gender typicality. The goal of this experiment was to analyze the effect of different degrees of the gender typicality of role descriptions on the resolution of anaphoric personal pronouns.

Material

Item structure was identical to the one used in the previous experiment. In Experiment 2, the antecedents were descriptions of roles which had been rated as slightly male, slightly female, or neutral. Thus, we selected items with 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 gender typicality Likert scale where 1 represented the pole of male and 7 the pole of female typicality (see Experiment 1 for details). (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. serves drinks and food in a café.

Generally he / she receives adequate tips.

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 ensure reading for

comprehension.

Design and hypothesis

The experiment tested the interaction between the gender typicality of the occupational role description and the gender of the anaphoric pronoun, resulting in a 3 (Gender Typicality: slightly male, slightly female or neutral) x 2 (Pronoun: masculine or feminine) within-subjects design. We tested two alternative outcomes: (1) Different degrees of gender typicality could have gradual effects on the representation of referential gender, so that disruptions in the reading process would be less pronounced for slightly typical roles compared to highly typical roles (e.g., manifest in fewer eye-tracking measures), while they would still be distinguishable from the neutral condition (where no mismatch effect was expected). In line with this assumption, we expected gender-incongruent role descriptions to prompt longer fixation times and more frequent regressions than the congruent and neutral ones. (2) Alternatively, the gender typicality effect could be an all-or-none process; in that case a high degree of gender typicality would be required to activate representation of the corresponding referential gender representation. In this case the low-typicality roles of the present experiment would not produce any disruption in the reading process.

In view of the previous results, the mismatch effect was expected to affect the masculine anaphor condition to a greater extent than the feminine anaphor.

Participants

Twenty-nine students (17 women and 12 men) from the University of Sussex, UK, with normal or corrected-to-normal vision participated in the study. Their mean age was 21 years (*SD* = 2,4). None of them had participated in Experiment 1. They received monetary compensation or course credit for their participation.

Procedure and analysis

The experimental procedure with eye-tracking recordings and the analyses was identical to Experiment 1. The F_1 ANOVA was performed with Gender Typicality (slightly male, slightly female, neutral) x Pronoun (masculine, feminine) as within-subjects factors. The F_2 ANOVA was performed with Gender Typicality (slightly male, slightly female, neutral) as between-items factor and Pronoun (masculine, feminine) as within-items factor. The principle of reporting results is the same as in Experiment 1, so that we report only results that were consistent in both analyses, i.e., reliable [$p \leq .05$] in both F_1 and F_2, or significant in one analysis and marginally significant [$p \leq .1$] in the other.

Results

The interaction between Gender Typicality and Pronoun emerged consistently in both F_1 and F_2 analysis in the Regressions into the first region of the target sentence. Participants regressed more frequently after reading an incongruent compared to a congruent pronominal anaphor. Pairwise comparisons showed that the effect was significant for slightly male and slightly female role descriptions. Comparing the two anaphoric pronouns, the effect was significant for the feminine, but not for the masculine pronoun, M_{maleHE}=18,96, M_{femaleHE}=20,11, *ns*; M_{maleSHE}=27,59, $M_{\text{femaleSHE}}$=10,34, t_{28}=4,05, p<.001.

The effect did not emerge reliably in other eye-tracking measures and was not affected by participant gender ($F_{1\text{-}27}$ = 1,41, *ns*).

The mismatch effect found in eye movements did not correlate with explicit typicality ratings (β´s ≤ . 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, for women, or as neutral. The results confirmed the hypothesis that the gender typicality effect on anaphor resolution depends on the degree of gender typicality of the antecedent/context. The mismatch effect was less evident compared to the effect occurring with highly typical roles and occurred only in one measure, while the effect did not emerge for neutral items, thus reflecting the modulation of the typicality degree of descriptions. When description typicality and pronoun gender mismatched, readers regressed to the beginning of the description, in order to re-check the provided information and to resolve the gender conflict. The description-paradigm resulted to be sensitive, showing that low degrees of typicality may trigger an impair in the resolution process. The description-based paradigm may thus be considered an adequate tool for investigating gender typicality, even when typical gender cues are too subtle to belong to the categorization of "stereotypical roles".

Contrary to our hypothesis and to earlier studies, a significant mismatch effect appeared only with the feminine anaphoric pronoun, in the region where the interaction between gender typicality of description and pronoun gender occurred. Given the low typicality of the items, the asymmetry in the mismatch effect may reflect a generally higher expectation of a masculine referent ("male-as-norm", cf. Stahlberg, Braun, Irmen, & Sczesny, 2007; Irmen, 2007, Exp 2), which may have caused a mismatch when the feminine pronoun was encountered. In Experiment 1, on the other hand, the expectation of a masculine pronoun may have been outweighed by the strong typicality of the items which already induced definite gender expectations.

General conclusions

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 and was modulated according to the degree of typicality of the descriptions. Incongruence between gender typicality of the description and pronoun gender produced a mismatch cost, not only with highly typical items (cf. Experiment 1) but also with slightly typical items (cf. Experiment 2). 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 effect of typicality is not an all-or-none phenomenon, but that different degrees of typicality trigger proportional disruptions in the reading process. 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 only an early to intermediate stage of sentence processing was affected with less typical items.

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 eye movements

during reading. The cognitive dissociation between these two factors is evident in the results of Experiment 2, where items were presented which possessed a low degree of gender typicality. Based on explicit ratings, the roles (e.g. manager, politician, doctor) 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.

Finally, a comparison of the present study 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 & Rossberg, 2004, Gygax et al., 2008, on the relation between gender typicality and grammatical gender); further comparative studies 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 influence on the process of anaphor resolution.

References

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.

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.

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*(4), 416-434.

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

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.

Gabriel, U., Gygax, P., Sarrasin, O., & 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*(1), 206-212.

Gygax, P., Gabriel, U., Sarrasin, O., Garnham, A, & Oakhill, J. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and

mechanics are all men. *Language, Cognition and Neuroscience, 23*(3), 464-485.

Hellinger, M., & Bußmann, H. (2001). *Gender across languages: The linguistic representation of women and men* Volume 1. Amsterdam, The Netherlands: Benjamins.

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., & Rossberg, N. (2004). Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology, 23*, 272-307.

Irmen, L., & Schumann, E. (2011). Processing grammatical gender of role nouns: Further evidence from eye movements. *Journal of Cognitive Psychology, 23*(8), 998-1014.

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.

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.

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

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

Sato, S., Gygax, P., & Gabriel, U. (2013). Gender inferences: Grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism: Language and Cognition, 16*(4), 792-807.

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

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.

---

[1] Due to the translation and the different word order in the two languages, the segmentation of the areas of interest was partially different in the two experiments.

[2] 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).

[3] Main effects are not reported in the text since they are not relevant for the purpose of the study. Moreover they did not occur consistently in the regions and measures where the interaction effect emerged. The results of eye-tracking measures not documenting the interaction effect reliably are not reported in the text but only in tables.

[4] Follow-up ratings were also collected from our participants after completing the eye-tracking experiment. The follow-up ratings turned out to be more skewed towards neutrality, so that typically male and particularly typically female occupations received less extreme ratings. However, this data should be interpreted with caution since the previous exposition to the eye-tracking material may have induced counter-stereotypical representations and influenced the ratings.

Table 1. Factorial structure of Experiment 1.

Context	Male role description	C. R. repairs - and produces pieces of 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. R1 R2 R3 R4

Note: The regions of analysis (R) are delimited by a dash.

Table 2. Experiment 1: Means (Standard Deviations) of residual fixation times and probabilities of regressions, differentiated for region and experimental factor.

Region	Typicality	Pronoun	First fixation time		First pass time		Regression path		Total fixation time		Regressions into a region		Regressions out of a region	
Context sentence, Initials region	Male	he	-1,81	(44,30)	3,64	(172,91)	2,15	(172,89)	-10,68	(193,40)	47,85	(26,79)	0,00	(0,00)
		she	13,82	(55,84)	31,78	(173,11)	30,27	(173,01)	68,41	(244,75)	48,92	(21,49)	0,00	(0,00)
	Female	he	-6,12	(39,20)	12,55	(175,16)	10,77	(175,14)	18,14	(248,86)	42,47	(27,50)	0,00	(0,00)
		she	0,71	(33,76)	-16,38	(165,92)	-8,98	(189,78)	-29,00	(224,31)	43,01	(29,74)	0,00	(0,00)
	Neutral	he	-0,88	(35,05)	-21,57	(159,27)	-23,29	(158,96)	-9,73	(210,51)	46,24	(23,46)	0,00	(0,00)
		she	-8,26	(41,14)	-46,76	(128,21)	-48,43	(128,37)	-60,26	(170,98)	45,70	(27,54)	0,00	(0,00)
Target sentence R1	Male	he	-2,39	(36,47)	13,86	(71,42)	16,13	(83,33)	2,92	(78,55)	27,96	(23,33)	2,15	(5,68)
		she	12,62	(53,48)	14,62	(60,97)	41,51	(122,97)	12,78	(81,94)	21,51	(16,77)	3,76	(8,29)
	Female	he	2,27	(34,97)	2,47	(65,86)	-13,38	(65,87)	-5,58	(66,21)	26,34	(19,14)	0,00	(0,00)
		she	0,27	(54,48)	-11,92	(55,11)	-14,40	(88,21)	-13,69	(71,86)	30,11	(22,94)	1,61	(6,60)
	Neutral	he	-7,13	(40,97)	-4,98	(54,69)	-11,14	(75,05)	2,84	(88,77)	22,04	(19,90)	1,08	(4,16)
		she	-9,53	(37,68)	-14,49	(42,42)	-19,32	(52,87)	-15,51	(56,15)	20,97	(16,65)	1,61	(5,01)
Target sentence R2	Male	he	-0,12	(38,76)	3,89	(79,70)	30,13	(146,59)	-6,76	(110,38)	30,11	(25,25)	20,97	(22,35)
		she	-0,87	(38,22)	-2,98	(95,74)	19,48	(128,18)	-12,34	(102,04)	31,72	(26,65)	17,20	(13,25)
	Female	he	2,26	(48,02)	5,35	(93,21)	15,40	(131,34)	21,87	(107,17)	34,95	(27,34)	19,35	(17,79)
		she	-9,36	(31,36)	-11,96	(87,40)	-29,08	(107,91)	-26,62	(122,62)	29,57	(25,72)	19,35	(18,81)
	Neutral	he	3,62	(37,18)	-6,27	(60,80)	-12,56	(101,22)	13,42	(85,84)	44,09	(28,40)	12,90	(17,59)
		she	-5,30	(34,43)	-10,50	(58,07)	-52,78	(101,51)	-8,41	(122,36)	41,94	(25,41)	9,14	(13,50)

Table 2. (continued)

Region	Typicality	Pronoun	First fixation time		First pass time		Regression path		Total fixation time		Regressions into a region		Regressions out of a region	
Target sentence R3	Male	he	-3,85	(35,21)	-31,61	(88,41)	-13,40	(343,34)	-49,90	(146,58)	22,04	(23,33)	30,11	(23,34)
		she	-4,96	(37,68)	-14,70	(96,70)	-63,31	(257,32)	-60,95	(113,88)	22,04	(18,45)	24,73	(21,03)
	Female	he	3,59	(43,99)	27,95	(106,28)	-19,74	(148,34)	29,21	(141,83)	25,81	(17,66)	31,72	(22,09)
		she	-4,75	(36,33)	-12,37	(105,99)	-81,19	(192,10)	-17,04	(167,27)	22,04	(15,15)	29,57	(22,65)
	Neutral	he	1,11	(42,68)	2,02	(102,84)	120,44	(281,22)	49,10	(170,67)	19,89	(17,96)	41,94	(24,29)
		she	7,53	(44,36)	18,17	(70,88)	63,97	(297,52)	40,19	(129,44)	20,43	(19,10)	35,48	(25,73)
Target sentence R4	Male	he	-10,72	(54,32)	1,58	(163,67)	43,37	(750,21)	-7,65	(198,96)	0,00	(0,00)	47,31	(23,21)
		she	-3,44	(59,30)	-10,68	(156,36)	39,29	(577,86)	-29,48	(189,12)	0,00	(0,00)	53,23	(22,53)
	Female	he	-0,22	(58,00)	-19,44	(145,33)	71,07	(586,57)	5,80	(187,38)	0,00	(0,00)	58,60	(24,67)
		she	8,26	(61,35)	-6,11	(117,58)	30,47	(707,20)	-16,37	(147,20)	0,00	(0,00)	52,15	(26,79)
	Neutral	he	0,17	(76,20)	-27,68	(122,74)	-89,76	(485,75)	-28,81	(153,15)	0,00	(0,00)	45,16	(24,79)
		she	-14,21	(54,27)	-4,76	(122,55)	-7,60	(692,40)	-14,86	(147,50)	0,00	(0,00)	55,91	(27,06)

Table 3. Results of statistical analyses (ANOVA) of experiment 1, differentiated for eye-tracking measures and regions of analysis, *= $p \leq .1$, **= $p \leq .05$, ***= $p \leq .01$.

Measure	Region[a]	Effect[b]	*F*1	*F*2
First fixations time	Context	Typicality	F =2.78*	F = 2.37
		Pronoun	F =1.15	F = 1.92
		T x P	F =1.52	$F_{2,33}$ =4.16**
	Target R1	Typicality	$F_{2,60}$ =3.38**	F < 1
		Pronoun	F < 1	F < 1
		T x P	F = 1.37	F = 1.15
	Target R2	Typicality	F < 1	F < 1
		Pronoun	$F_{1,30}$ = 4.42**	$F_{2,33}$ = 4.27**
		T x P	F < 1	F = 2.51
	Target R3	Typicality	F = 1.00	F < 1
		Pronoun	F < 1	F < 1
		T x P	F < 1	F < 1
	Target R4	Typicality	F = 1.06	F = 1.64
		Pronoun	F < 1	F < 1
		T x P	F = 1.04	F = 1.18
First pass time	Context	Typicality	$F_{2,60}$ =6.99***	F = 2.12
		Pronoun	F < 1	F < 1
		T x P	F =1.51	F < 1
	Target R1	Typicality	$F_{2,54}$ =*5.65***	F = 1.76
		Pronoun	F =1.39	F < 1
		T x P	F < 1	F < 1
	Target R2	Typicality	F < 1	F < 1
		Pronoun	F < 1	F < 1
		T x P	F < 1	F < 1
	Target R3	Typicality	$F_{2,60}$ =3.77**	F < 1
		Pronoun	F < 1	F < 1
		T x P	$F_{2,60}$ =3.06**	$F_{2,33}$ = 3.58**
	Target R4	Typicality	F < 1	F < 1
		Pronoun	F < 1	F < 1
		T x P	F < 1	F < 1
Regression path time	Context	Typicality	$F_{2,60}$ = 6.69***	F = 2.29
		Pronoun	F < 1	F < 1
		T x P	F = 1.06	F < 1
	Target R1	Typicality	$F_{2,54}$ = 7.86***	$F_{1,33}$ = 4.05**
		Pronoun	F < 1	F < 1
		T x P	F < 1	F < 1
	Target R2	Typicality	$F_{2,60}$ = 4.52**	F < 1
		Pronoun	$F_{1,30}$ = 4.08**	$F_{1,33}$ = 4.96**
		T x P	F < 1	F < 1
	Target R3	Typicality	$F_{2,60}$ = 6.68***	F = 1.17
		Pronoun	$F_{1,30}$ = 5.04**	F = 3.29
		T x P	F < 1	F < 1
	Target R4	Typicality	F = 2.12	F < 1
		Pronoun	F < 1	F = 1.46
		T x P	F < 1	F < 1

Table 3. (continued)

Measure	Region[a]	Effect[b]	*F*1	*F*2
Total time	Context	Typicality	$F_{2,60} = 3.65$**	$F = 1.23$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F_{2,60} = 4.16$**	$F_{2,33} = 3.07$*
	Target R1	Typicality	$F = 1.70$	$F = 1.22$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F = 2.14$
	Target R2	Typicality	$F < 1$	$F < 1$
		Pronoun	$F = 4.00$	$F = 2.47$
		T x P	$F < 1$	$F = 1.04$
	Target R3	Typicality	$F_{2,60} = 19.15$***	$F_{2,33} = 2.95$*
		Pronoun	$F = 1.27$	$F = 1.33$
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
Regressions in the region	Context	Typicality	$F = 1.55$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R1	Typicality	$F = 2.29$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 1.58$	$F = 1.56$
	Target R2	Typicality	$F_{2,60} = 7.71$**	$F_{2,33} = 2.61$*
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R3	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
Regressions out of the region	Target R1	Typicality	$F = 2.48$	$F_{2,33} = 2.82$*
		Pronoun	$F = 3.08$	$F = 2.28$
		T x P	$F < 1$	$F < 1$
	Target R2	Typicality	$F_{2,60} = 6.09$**	$F = 2.09$
		Pronoun	$F = 1.30$	$F = 2.80$
		T x P	$F < 1$	$F < 1$
	Target R3	Typicality	$F_{2,60} = 7.52$**	$F = 1.14$
		Pronoun	$F_{1,30} = 4.60$*	$F_{1,33} = 3.22$*
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F = 1.60$	$F < 1$
		Pronoun	$F = 2.14$	$F = 1.73$
		T x P	$F_{2,60} = 3.12$*	$F_{2,33} = 3.52$**

[a]= Context: first region of the context sentence; Target R1-4: regions of analysis of the target sentence. [b]=T x P: Gender typicality by pronoun.

Table 4. Experiment 2: Means (Standard Deviations) of residual fixation times and probabilities of regressions, differentiated for region and experimental factor.

Region	Typicality	Pronoun	First fixation time		First pass time		Regression path		Total fixation time		Regressions into a region		Regressions out of a region	
Context sentence, Initials region	Neutral	he	-3,97	(41,61)	-13,92	(170,95)	-13,92	(170,95)	10,97	(305,61)	48,85	(27,07)	0,00	(0,00)
		she	3,06	(45,93)	-19,49	(131,69)	-19,49	(131,69)	11,78	(279,94)	45,98	(30,43)	0,00	(0,00)
	Slightly male	he	1,44	(42,77)	2,69	(181,10)	2,69	(181,10)	4,83	(268,96)	43,10	(27,28)	0,00	(0,00)
		she	5,55	(47,66)	38,83	(158,62)	38,83	(158,62)	7,29	(231,49)	45,40	(21,77)	0,00	(0,00)
	Slightly female	he	-6,47	(39,91)	3,33	(168,99)	3,33	(168,99)	-8,59	(241,86)	46,55	(29,34)	0,00	(0,00)
		she	4,05	(46,49)	-3,97	(180,96)	-3,97	(180,96)	-1,94	(198,44)	53,45	(26,49)	0,00	(0,00)
Target sentence R1	Neutral	he	-1,12	(43,46)	2,15	(64,93)	12,07	(101,26)	0,50	(76,63)	16,67	(12,60)	2,87	(6,41)
		she	-13,73	(31,56)	-5,12	(53,57)	4,52	(113,96)	0,13	(88,31)	18,39	(20,09)	2,87	(6,41)
	Slightly male	he	-,98	(43,25)	-8,79	(54,77)	-6,42	(103,10)	-7,27	(93,91)	18,97	(19,78)	1,15	(4,30)
		she	1,87	(50,34)	-7,71	(60,15)	-7,86	(78,92)	20,67	(108,13)	27,59	(20,55)	1,72	(5,17)
	Slightly female	he	-,38	(48,08)	4,83	(72,25)	-2,21	(72,63)	4,69	(102,36)	20,11	(16,89)	1,72	(5,17)
		she	7,57	(41,57)	3,64	(65,80)	-5,19	(88,51)	-26,18	(80,69)	10,34	(13,67)	1,72	(5,17)
Target sentence R2	Neutral	he	13,61	(43,03)	14,46	(85,09)	2,08	(122,16)	22,12	(128,05)	32,18	(21,79)	5,75	(9,21)
		she	-2,82	(35,60)	-6,75	(65,43)	-2,12	(166,46)	0,85	(136,03)	35,63	(20,76)	10,34	(18,59)
	Slightly male	he	-3,50	(31,82)	-20,03	(73,59)	-,61	(140,40)	0,49	(138,43)	37,36	(23,00)	13,22	(13,64)
		she	-5,26	(32,55)	-13,90	(66,94)	13,47	(147,31)	-5,27	(167,70)	35,06	(29,66)	14,37	(15,25)
	Slightly female	he	-3,18	(31,56)	-2,63	(65,87)	-7,00	(135,05)	-1,61	(138,61)	28,16	(22,32)	10,34	(12,92)
		she	-0,85	(33,73)	11,35	(107,98)	-7,74	(156,58)	-19,41	(146,67)	29,31	(22,12)	6,90	(9,47)

Table 4. (continued)

Region	Typicality	Pronoun	First fixation time		First pass time		Regression path		Total fixation time		Regressions into a region		Regressions out of a region	
Target sentence R3	Neutral	he	-2,75	(43,48)	25,71	(93,82)	136,75	(351,17)	42,53	(114,07)	16,09	(19,15)	5,75	(9,21)
		she	5,06	(39,84)	13,95	(87,33)	78,46	(243,73)	26,48	(113,65)	17,24	(19,15)	10,34	(18,59)
	Slightly male	he	-5,30	(40,45)	-15,92	(76,85)	-63,07	(199,34)	-12,44	(107,91)	24,14	(18,68)	13,22	(13,64)
		she	-12,37	(41,84)	-25,29	(70,22)	-79,29	(235,74)	-2,80	(140,63)	25,29	(18,70)	14,37	(15,25)
	Slightly female	he	8,77	(45,29)	-0,30	(77,97)	-12,57	(340,36)	-30,55	(129,24)	20,11	(17,47)	10,34	(12,92)
		she	9,43	(34,53)	-0,15	(64,18)	-23,96	(217,61)	-29,65	(90,91)	19,54	(14,13)	6,90	(9,47)
Target sentence R4	Neutral	he	-7,72	(46,83)	-15,21	(108,01)	-102,60	(401,69)	-32,96	(140,28)	0,00	(0,00)	42,53	(22,97)
		she	-7,04	(46,93)	19,86	(122,20)	-13,58	(425,62)	8,18	(129,97)	0,00	(0,00)	43,10	(27,28)
	Slightly male	he	13,32	(51,39)	-1,86	(126,66)	84,53	(473,88)	5,41	(147,17)	0,00	(0,00)	57,47	(19,20)
		she	4,36	(36,79)	-4,66	(149,73)	30,03	(478,56)	3,62	(164,35)	0,00	(0,00)	56,90	(26,17)
	Slightly female	he	3,60	(44,76)	-25,96	(123,07)	-32,84	(414,48)	-33,32	(126,86)	0,00	(0,00)	51,72	(25,72)
		she	-5,03	(38,56)	-2,46	(155,28)	15,83	(521,74)	-4,88	(159,74)	0,00	(0,00)	53,45	(24,95)

Table 5. Results of statistical analyses (ANOVA) of experiment 2, differentiated for eye-tracking measures and regions of analysis, *= $p \leq .1$, **= $p \leq .05$, ***= $p \leq .01$.

Measure	Region[a]	Effect[b]	*F*1	*F*2
First fixations time	Context	Typicality	$F < 1$	$F < 1$
		Pronoun	$F = 2.85$	$F = 1.96$
		T x P	$F < 1$	$F < 1$
	Target R1	Typicality	$F_{2,56} = 2.59$*	$F = 2.43$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 1.57$	$F = 1.34$
	Target R2	Typicality	$F = 2.36$	$F = 2.13$
		Pronoun	$F = 1.27$	$F_{2,33} = 3.09$*
		T x P	$F = 1.89$	$F_{2,33} = 4.58$**
	Target R3	Typicality	$F_{2,56} = 4.91$***	$F = 1.93$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F_{2,56} = 2.51$*	$F = 1.92$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
First pass time	Context	Typicality	$F_{2,56} = 3.18$**	$F = 1.69$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 1.28$	$F = 1.53$
	Target R1	Typicality	$F = 1.81$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R2	Typicality	$F_{2,56} = 3.56$*	$F = 1.53$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 1.81$	$F = 2.20$
	Target R3	Typicality	$F_{2,56} = 9.01$***	$F = 1.51$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F = 1.77$
		T x P	$F < 1$	$F < 1$
Regression path time	Context	Typicality	$F_{2,56} = 3.18$**	$F = 1.69$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 1.28$	$F = 1.53$
	Target R1	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R2	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R3	Typicality	$F_{2,56} = 14.15$***	$F_{2,33} = 3.05$*
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F = 1.42$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$

Table 5. (continued)

Measure	Region[a]	Effect[b]	*F*1	*F*2
Total time	Context	Typicality	$F < 1$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R1	Typicality	$F = 1.12$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 2.08$	$F = 2.45$
	Target R2	Typicality	$F < 1$	$F < 1$
		Pronoun	$F = 2.09$	$F = 1.22$
		T x P	$F < 1$	$F < 1$
	Target R3	Typicality	$F_{2,56} = 7.39$***	$F = 1.36$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F < 1$	$F < 1$
		Pronoun	$F = 1.26$	$F = 2.82$
		T x P	$F < 1$	$F < 1$
Regressions in the region	Context	Typicality	$F = 1.07$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F = 1.13$
	Target R1	Typicality	$F_{2,56} = 4.31$**	$F = 1.68$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F_{2,56} = 4.79$**	$F_{2,33} = 3.48$**
	Target R2	Typicality	$F_{2,56} = 2.74$*	$F = 1.22$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R3	Typicality	$F_{2,56} = 5.13$***	$F = 1.68$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
Regressions out of the region	Target R1	Typicality	$F = 1.11$	$F < 1$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R2	Typicality	$F_{2,56} = 4.26$**	$F = 1.98$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F = 1.70$	$F = 1.18$
	Target R3	Typicality	$F_{2,56} = 5.52$***	$F = 1.67$
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$
	Target R4	Typicality	$F_{2,56} = 8.23$***	$F = 3.35$**
		Pronoun	$F < 1$	$F < 1$
		T x P	$F < 1$	$F < 1$

[a]= Context: first region of the context sentence; Target R1-4: regions of analysis of the target sentence. [b]=T x P: Gender typicality by pronoun.

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
First name Family name _____

Datum, Unterschrift
Date, Signature _____