
**THE INFLUENCE OF PRIOR VISUAL
GENDER AND ACTION CUES VERSUS
LONG-TERM KNOWLEDGE IN
(SITUATED) LANGUAGE PROCESSING**

Alba Rodríguez Llamazares

Thesis submitted in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

Faculty of Linguistics and Literary Studies
Universität Bielefeld

Bielefeld, 2018

Evaluation Committee:

Prof. Dr. Pia Knoeferle (Humboldt Universität zu Berlin)

Dr. Joana Cholin (Universität Bielefeld)

Prof. Dr. Petra Wagner (Universität Bielefeld)

Dr. Annett Jorschick (Universität Bielefeld)

*Sobre todo creo que
no todo está perdido.
Tanta lágrima, tanta lágrima y yo
soy un vaso vacío.
Oigo una voz que me llama,
casi un suspiro:
¡Rema, rema, rema!
(Jorge Drexler, Al Otro Lado del Río, 2004)*

Contents

Abstract	ix
Acknowledgments	xiii
List of Tables	xvi
List of Figures	xix
List of abbreviations	xxi
1. Introduction	1
1.1. Motivation and aims	2
1.2. Thesis outline	3
2. Situated language processing	9
2.1. Long-term experience	10
2.2. The concurrent visual context	15
2.3. Prior visual cues	20
2.4. Visuolinguistic mismatches	23
3. The influence of gender on language processing	31
3.1. Grammatical gender	32
3.2. Conceptual gender	35
3.2.1. Biological gender	37
3.2.2. Gender stereotypes	41

4. Accounts and models of situated language comprehension	49
5. Gendered expectations: mismatches in situated language comprehension	57
5.1. Experiments 1 and 2	58
5.1.1. Methods and Design	61
5.1.2. Analysis and Results	65
5.1.3. Discussion	72
6. The concurrent visual context: constraining participants' expectations	77
6.1. Experiment 3	79
6.1.1. Methods and Design	82
6.1.2. Analysis and Results	84
6.1.3. Discussion	91
7. The electrophysiological correlates of visual gender verification in language comprehension	95
7.1. Experiment 4	98
7.1.1. Methods and Design	99
7.1.2. Recording, Analysis and Results	100
7.1.3. Discussion	104
8. General discussion	109
8.1. Preference for prior visual cues	111
8.2. Mismatch effects	113
8.3. Contribution of stereotypical gender knowledge	118
8.4. Implications for accounts of situated language comprehension	121
8.4.1. Example: Gender information	125
8.5. Conclusions	130
9. German summary	135

Appendices	141
A. Experimental materials (Experiments 1 to 4)	143
A.1. Experimental sentences	144
A.2. Onsets and offsets of experimental sentence regions	149
A.3. Visual materials	152
A.3.1. Snapshots of the agents' faces and hands with Consent to Use of Image forms	152
A.3.2. Snapshots of the objects from the experimental videos	158
A.4. Example of two filler trials	167
B. Additional statistical analyses (Experiments 1 to 4)	169
B.1. Accuracy analyses using GLME (Experiments 1 to 4)	169
B.2. Alternative reaction-time analyses using LME (Experiments 1 to 3)	172
B.3. Statistical tests for the intercept per sentence region (Experiments 1 to 3)	176
B.4. Alternative eye-movement analyses using LME (Experiments 1 to 3)	177
B.5. Time-course graphs: percentage of looks, Experiment 3	187

Abstract

Studies on situated language comprehension (i.e., comprehension in rich visual contexts), have shown that the comprehender makes use of different information sources in order to establish visual reference and to visually anticipate entities in a scene while understanding language (reflecting expectations on what might be mentioned next). Semantics and world-knowledge (i.e., experiential, long-term knowledge) are among these sources. For instance, when listening to a sentence like *The girl will ride...*, the comprehender will likely anticipate an object that a girl could ride, e.g., a carousel, rather than other objects, such as a motorbike (Kamide, Altmann, & Haywood, 2003). However, following the inspection of events (featuring agents acting upon objects or patients), comprehenders have so far shown a preference to visually anticipate the agents or objects that have been seen as part of those prior events (i.e., *recent event preference* or the preference for event-based representations; Abashidze, Carminati, & Knoeferle, 2014; Knoeferle, Carminati, Abashidze, & Essig, 2011). This preference emerged even when other plausible objects or better stereotypically fitting agents were present. Although the preference for event-based information over other sources (e.g., plausibility or stereotypicality) seems to be strong and has been accommodated in accounts of situated language comprehension (Knoeferle & Crocker, 2006, 2007), its nature when comprehenders generate expectations is still unspecified. Crucially, the preference for recent events has not been generalized from action events to other types of information in the visual and linguistic contexts.

To further examine this issue, this thesis investigated the role of a particular type of information during situated language comprehension under the influence of prior events,

namely, visual gender and action cues and knowledge about gender stereotypes. As many studies in the field of psycholinguistics have highlighted, gender (both a biological and a social feature of human beings) is relevant in language comprehension (e.g., grammatical gender can serve to track reference in discourse, and gender-stereotype knowledge can bias our interpretation of a sentence). However, little psycholinguistic research has examined the comprehension of gender information in a visual context. We argue that gender is worth exploring in a paradigm where prior event representations can be pitted against long-term knowledge. Not only that, inspired by experiments using mismatch designs, we wanted to see how the visual attention of the comprehender might be affected as a function of referential incongruencies (i.e., mismatches between visual events and linguistic information, e.g., Knoeferle, Urbach, & Kutas, 2014; Vissers, Kolk, Van de Meerendonk, & Chwilla, 2008; Wassenaar & Hagoort, 2007) and incongruences at the level of world-knowledge (i.e., gender stereotypes; e.g., Duffy & Keir, 2004; Kreiner, Sturt, & Garrod, 2008). By doing so, we could get insights into how both types of sources (event-based information and gender-stereotype knowledge from language) are used, i.e., whether one is more important than the other or if both are equally exploited in situated language comprehension.

We conducted three eye-tracking, *visual-world* experiments and one EEG experiment. In all of these experiments, participants saw events taking place prior to sentence comprehension, i.e., videos of (female or male) hands acting upon objects. In the eye-tracking experiments, following the videos, a visual scene appeared with the faces of two potential agents: one male and one female¹. While the agent matching the gender features from prior events (i.e., the hands) was considered as the target agent, the other potential agent, whose gender was not cued in previous events, was the competitor agent. The visual scene in Experiment 3 further included the images of two objects; one was the target object (i.e., the object that appeared in prior events), while the other was a competitor object with opposite stereotypical valence. During the presentation of this scene,

¹Experiment 4 had no visual scene displayed during comprehension, but had a cross that participants had to fixate instead.

an OVS sentence was presented (e.g., translation from German: ‘The cake_{NP1/obj} bakes_V soon_{ADV} Susanna_{NP2/subj}’). We used the non-canonical OVS word order as opposed to SVO (more commonly used in prior research, e.g., Knoeferle, Carminati, et al., 2011) precisely to examine participants’ expectations towards the agent, who was mentioned at final position. We manipulated two factors. One factor was the match between prior visual events and language: there were action-verb(-phrase) mismatches in Experiments 1 and 3, and mismatches between the gender of the hands and the final subject (i.e., the proper name) in Experiments 2 and 4. The second manipulation, present in Experiments 1 to 3, was the match between the stereotypical valence of the described actions/events in the sentence and the target agent’s gender. In the eye-tracking experiments, we measured participants’ visual attention towards the agents’ faces during sentence comprehension. In the EEG experiment, we measured ERP responses time-locked to the final, proper name region (i.e., Susanna). Participants’ task was to verify via button press whether the sentence matched the events they just saw.

In line with prior research, our results support the idea that the preference for event-based representations generalizes to another cue, i.e., gender features from the hands of an agent during prior events. Participants generally preferred to look at the target agent compared to the competitor. These results also suggest that the recent-event preference does not just rely on representations of full objects, agents and events, but also subtler (gender) features that serve to identify feature-matching targets during comprehension (i.e., faces of agents are inspected based on the gender features from hands seen in prior events). This preference is however modulated by mismatches in language, i.e., whenever the actions described or the gender implied by the final noun in the sentence were at odds with prior events, attention towards the target agent was reduced. In addition, the scene configuration of Experiment 3 gave rise to gender stereotypicality effects, which had not yet been found in prior studies using a similar design. Participants looked at the target agent (vs. the competitor) to a greater extent when the action described by the sentence stereotypically matched (vs. mismatched) them. As for the electrophysiological response towards mismatches between event-based gender cues and language,

we found a biphasic ERP response, which suggests that this type of verification requires two semantically-induced stages of processing. This response had commonalities both with some effects found in strictly linguistic/discourse contexts but also with previously observed mismatch effects in picture-sentence verification studies (i.e., role relation and action mismatches; Knoeferle et al., 2014), which suggests that a similar (perhaps a single) processing mechanism might be involved in several visiolinguistic relations.

In sum, our results using gender and action cues from prior events and long-term knowledge call for a more refined consideration of the different aspects involved in (situated) language comprehension. On the one hand, existing accounts need to accommodate further reconciliations/verifications of visiolinguistic relations (e.g., roles, actions, gender features, etc.). When it comes to listeners generating expectations during comprehension while inspecting the visual world, we further suggest that a weighted system (i.e., a system indexing the strength of the expectation and how different information sources contribute to it; also suggested in Münster, 2016), applies for gender of information. Not only event-based representations, but also different discrepancies between these representations and language and, depending on the concurrent visual scene configuration, long-term knowledge (e.g., pertaining to gender stereotypes), can affect weighted expectations. Biosocial aspects such as gender may be of particular interest to answer some of the open questions in how situated language comprehension works, as these aspects can be found and manipulated at different levels of communication (e.g., the comprehender, the speaker, the linguistic content, etc.).

Acknowledgments

This thesis is an incomplete reflection of a once-in-a-lifetime experience. Only by reading between the lines can one partly grasp what its production implied. The things I have learned, the places I have been, the people I have met. Regardless of where my future endeavours take me, I feel nothing but gratitude for what has taken place over the course of these years. This gratitude is of course directed to a considerable amount of people, without whom none of this would have been possible.

First and foremost, I would like to express my most sincere gratitude to Prof. Dr. Pia Knoeferle, for giving me the chance to embark on a PhD. Your support and comprehension, which were always there even during complicated times, go beyond my understanding. Thank you for guiding me, and always smiling back at me every time I entered your office feeling overwhelmed. You deserve all my respect and admiration.

Next, I would like to extend my gratitude to Dr. Maria Nella Carminati and Dr. Michele Burigo, who gave me a warm welcome and plenty of advice and support. Also thanks to my student colleagues from the Language and Cognition Group (some of them Drs. already, *phew!*): Ernesto Guerra, Dato Abashidze, Katja Münster, Katharina Wendler, Thomas Kluth, Eva Nunneman and Julia Kröger. Not only did I have interesting discussions with you guys, but also the chance to share my illusions and fears. Thank you for helping me out in so many situations, and for letting me be part of your lives. It has been a real pleasure.

Apart from having an amazing group nearby over these years, I also had the privilege of being part of the EU-funded Initial Training Network LanPercept, which has been an exciting and enriching experience. To all the PIs and students with whom I met during our various workshops and conferences all over Europe, thank you. Special thanks to Prof. Dr. Matthew Crocker for giving me access to his labs at Saarland University, and to Yoana Vergilova, Torsten Jachmann and Jesús Calvillo, for your valuable assistance and for making me feel comfortable over the months spent in Saarbrücken as part of my secondment.

For economically supporting me during the PhD and in my last stages, I would like to thank the European Union's Seventh Framework Programme for research, technological development and demonstration (grant agreement no 316748) and the *Gleichstellungsfonds der Humboldt-Universität*, respectively.

Outside academia, I have also had all sorts of support from people I love deeply. I would like to thank my parents, María Llamazares and Jesús Rodríguez: *gracias por vuestro apoyo y comprensión, sin el cual no hubiera reunido las fuerzas para completar esta aventura*. Thanks to my sister, Isabel Rodríguez, for providing me with challenges during my conference trips (*sí, en breve te llevo los dedos y souvenirs que he ido recolectando*). Also thanks to my cousin, Javier Rodríguez, who also happened to be completing his PhD at the same time as me: thanks for your company during the seasons of laughter and despair. To all my friends, here in Bielefeld, back in Spain and beyond, thank you for visiting me in my imaginary bunker once in a while, either with messages or calls, or taking me out for dinner and drinks. It was all an indispensable part of the process; you gave me confidence and showed me that not everything in life is work.

Last but not least, I would like to thank my *partner in crime*, Tristan Ugarte. You have sacrificed much to join me in this adventure, and I am not sure if I will ever be able to pay you back however much I try. You took care of keeping my mind healthy and my spirits high, and supported me even when I was unable to believe in myself. This piece of work is as much yours as it is mine. *Maitte zaitut.*

List of Tables

5.1. Example item for Experiment 1	64
5.2. Example item for Experiment 2	64
7.1. Example item for Experiment 4	100
A.1. Sentences for the experimental items	144
A.2. Onsets and offsets of sentence regions (in msec)	149
A.3. Onsets and offsets of sentence regions (in msec)	150
B.1. Accuracy analysis, Experiment 1	170
B.2. Accuracy analysis, Experiment 2	170
B.3. Accuracy analysis, Experiment 3	171
B.4. Accuracy analysis, Experiment 4	171
B.5. Reaction-time analysis, Experiment 1	173
B.6. Reaction-time analysis, Experiment 2	174
B.7. Reaction-time analysis, Experiment 3	175
B.8. Statistical tests for the intercept (grand average per subjects) in the log-probability ratios per sentence region (Experiments 1 to 3)	176
B.9. Eye-movement analysis, Experiment 1, verb region	177
B.10. Eye-movement analysis, Experiment 1, adverb region	178
B.11. Eye-movement analysis, Experiment 1, NP2 region	179
B.12. Eye-movement analysis, Experiment 2, NP2 region	180
B.13. Eye-movement analysis, Experiment 3 (agents), NP1 region	181

B.14. Eye-movement analysis, Experiment 3 (agents), verb region	182
B.15. Eye-movement analysis, Experiment 3 (agents), adverb region	183
B.16. Eye-movement analysis, Experiment 3 (agents), NP2 region	184
B.17. Eye-movement analysis, Experiment 3 (objects), NP1 region	185
B.18. Eye-movement analysis, Experiment 3 (objects), verb region	186

List of Figures

2.1. Example image (middle scene) from Knoeferle and Crocker (2006).	21
4.1. Representation of a simple recurrent network. The light grey units are original from Elman (1990); the darker units form the implemented version by Dienes, Altmann, and Gao (1999), p.58	50
4.2. The Revised Coordinated Interplay Account (Knoeferle et al., 2014)	52
5.1. Example of an experimental trial, Experiments 1 and 2.	66
5.2. Time-course graph for Experiment 1.	68
5.3. By-subject mean log-probability ratios at the verb region, Experiment 1 (error bars indicate 95% confidence intervals).	69
5.4. By-subject mean log-probability ratios at the verb region per condition, Experiment 1 (error bars indicate 95% confidence intervals).	69
5.5. By-subject mean log-probability ratios at the adverb region, Experiment 1 (error bars indicate 95% confidence intervals).	70
5.6. By-subject mean log-probability ratios at the final noun (NP2) region, Experiment 1 (error bars indicate 95% confidence intervals).	70
5.7. Time-course graph for Experiment 2.	71
5.8. By-subject mean log-probability ratios at the final noun (NP2) region, Experiment 2 (error bars indicate 95% confidence intervals).	71
6.1. Example of an experimental trial in Experiment 3.	83
6.2. Time-course graph for the agents , Experiment 3.	85

6.3. By-subject mean log-probability ratios for the agents at the NP1 region, Experiment 3 (error bars indicate 95% confidence intervals).	86
6.4. By-subject mean log-probability ratios for the agents in the action-verb match condition (a) and the stereotypicality match condition (b) at the verb region, Experiment 3 (error bars indicate 95% confidence intervals).	87
6.5. By-subject mean log-probability ratios for the agents at the verb region per condition, Experiment 3 (error bars indicate 95% confidence intervals).	87
6.6. By-subject mean log-probability ratios for the agents in the action-verb match condition (a) and the stereotypicality match condition (b) at the adverb region, Experiment 3 (error bars indicate 95% confidence intervals).	88
6.7. By-subject mean log-probability ratios for the agents at the final noun (NP2) region, Experiment 3 (error bars indicate 95% confidence intervals).	88
6.8. By-subject mean log-probability ratios for the agents at the final noun (NP2) region per condition, Experiment 3 (error bars indicate 95% confidence intervals).	89
6.9. Time-course graph for the objects in Experiment 3.	90
6.10. By-subject mean log-probability ratios for the objects at the NP1 region per condition, Experiment 3 (error bars indicate 95% confidence intervals).	90
6.11. By-subject mean log-probability ratios for the objects in the action-verb match condition (a) and the stereotypicality match condition (b) at the verb region, Experiment 3 (error bars indicate 95% confidence intervals).	91
7.1. Example of an experimental trial in Experiment 4.	101
7.2. Electrode configuration, using Acticap 32-channel active electrode system (Brain Products). Two electrodes were moved to the outer canthi (T7 and T8), two to the left eye (PO9 and PO10) and two to the left and right mastoids (TP9 and TP10).	102
7.3. Grand average ERPs (mean amplitude) for 9 electrodes (3 frontal, 3 middle and 3 posterior) time-locked to the final noun (NP2).	103

7.4. Grand average ERPs (mean amplitude) across the scalp at the final noun region (300-500 and 500-900 time windows), obtained by subtracting the matching condition from the mismatching condition.	104
A.1. Snapshot of the agents' faces	152
A.3. Snapshot of the agents' hands from the experimental videos	153
A.5. Snapshots of the objects from the experimental videos	158
A.15. Filler trial with two pairs of hands	167
A.16. Filler trial with an object picture	167
B.1. Time-course graphs per condition with percentages of looks, Experiment 3	187

List of abbreviations

ABE: Anticipatory Baseline Effects

ANOVA: Analysis of Variance

CIA: Coordinated Interplay Account

ERP: Event-Related Potentials

NP1: First Noun-Phrase

NP2: Second Noun-Phrase

(G)LME: (Generalized) Linear Mixed Effects

OVS: Object-Verb-Subject (word order)

RT: Reaction Time

sCIA: social Coordinated Interplay Account

SRN: Serial Recurrent Network

SVO: Subject-Verb-Object (word order)

VWP: Visual-World Paradigm

WM: Working Memory

1 | Introduction

Language is a complex system, not only on its own (e.g., syntax, semantics and pragmatics) but also in its interaction with aspects of our (visual) environment. When we talk about people and things, we sometimes refer to our immediate perceptual world. In doing so, we can actively exploit sources of information around us, while also making use of our experiential (long-term) knowledge (e.g., about how events typically develop in real life, i.e., who tends to do what). These sources of contextual information have an important impact on how fast and efficiently we understand language. When different sources of information are incomplete or conflicting, this could lead to an incorrect interpretation if we decide to rely on a source in the absence of unambiguous information, and we might need to choose which source we want to rely on. Imagine yourself at a supermarket where, a person that you identify as a little boy (e.g., small, short hair, sportswear,...) is about to choose between a blue and a pink bike as a birthday present. The boy might start saying *I would rather get the...* and unintentionally, you might look at the blue bike very early on, waiting for the boy to refer to it. In that case, you would be surprised if the boy ended the sentence with *...pink bike*, even though it is perfectly possible for him to do so.

In this particular example, our knowledge of gender stereotypes (i.e., long-term knowledge about gender) might bias our expectations and the comprehension of the sentence in context. Perhaps the boy's utterance would be less surprising if a few minutes prior to it, we saw the boy trying out the pink bike, or maybe, we would not take that visual experience as relevant if we strongly trusted our beliefs about gender. What do these

types of situations tell us about how we understand sentences in non-linguistic contexts? How do the different information sources (one from long-term experience and a more recent, visually grounded experience) interact or compete? And does the particular type of knowledge we tap into (a general topic, like what type of food a person orders at a restaurant vs. a more *socially relevant* topic like the gender of a person who has a particular behaviour) differentially influence these interactions?

In the study of *situated language comprehension*, i.e., language in relation to a visual, non-linguistic context, different aspects regarding our knowledge of the world could give us insights into how a situation like the one above would develop. Such aspects could be related to, for instance, age (e.g., Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008), social class (e.g., Squires, 2013) or gender (e.g., Hanulíková & Carreiras, 2015; Pyykkönen, Hyönä, & van Gompel, 2010). Our work will however focus on the latter type of information. Gender, both from a biological¹ and a social perspective, will be viewed as a binary dimension². It is an inherent set of features to humans as well as to other animals; it is one of our most salient perceptual characteristics; and it also has an impact on our social, long-term knowledge (i.e., at the time of making inferences about people's traits, behaviour, etc.).

1.1 | Motivation and aims

In this thesis, we will compare the influence of visual gender cues from prior action events (i.e., events taking place prior to the presentation of a target scene and sentence) with knowledge on gender stereotypes during situated language comprehension, by measuring participants' (anticipatory) eye movements and ERP responses. We will see how the comprehender behaves when processing these types of information, and compare the

¹When talking about *biological gender*, we refer to *sex*, and both of these terms will be used as synonyms throughout the present thesis.

²By using gender as a binary dimension (i.e., female vs. male opposition), we do not mean to deny the existence of some evidence in favour of a more "colourful" gender spectrum (e.g., Ainsworth, 2015). These theories are however beyond the scope of this thesis.

findings to other sources of information that have been tested in the literature (e.g., depicted events vs. occupational stereotypes, Knoeferle & Crocker, 2007, past vs. future event plausibility; Knoeferle, Carminati, et al., 2011, or action and emotion cues; Münster, Carminati, & Knoeferle, 2014). The questions addressed in this thesis are: 1) How rapidly are visual gender and action cues integrated, and how do they guide attention towards (female vs. male) agents during comprehension? 2) How do different types of incongruencies, a) referential mismatches between prior events and language, and b), mismatches in the stereotypical congruence between gender cues and language, modulate this attention? With this investigation, we aim at further informing current processing accounts that accommodate the real-time interplay between visual and linguistic cues in comprehension.

1.2 | Thesis outline

In the second chapter, we will first introduce relevant background literature about the different aspects that are involved in the study of situated language comprehension and which establish the linking hypotheses between eye movements in the visual world and language comprehension. We will highlight the ability of the comprehender to create mental representations from language alone and to generate expectations about the linguistic input that may come next. We will then continue to explain the interactions that can occur between language-based representations and representations from the outside visual world, and what these interactions can tell us about our cognitive capacities and the strategies³ that come into play during comprehension. Crucially, we will also discuss how the temporal dynamics of these different sources of information (whether both visual and linguistic information come into play simultaneously or one after the other) make a difference in the way our attention is guided and our comprehension processes affected. We will argue that so far in the literature on situated language comprehension, semantics

³When using the term *strategy*, we don't necessarily mean that a conscious effort is taking place. We also use this term to refer to the unconscious cognitive biases of the comprehender, some of which may facilitate comprehension processes.

and plausibility have been shown to guide participants' attention over a concurrent scene (e.g., when listening to a sentence like *The boy will eat...* in front of a scene where a boy and several objects are present, the comprehender will likely direct their attention towards an edible object during the verb, such as a cake; Altmann & Kamide, 1999; Altmann & Mirković, 2009). However, when available, prior visual information about recent action events seems to enjoy priority over our knowledge about other plausible events (Knoeferle, Carminati, et al., 2011) or knowledge about (occupational) stereotypes (Knoeferle & Crocker, 2007) when visually anticipating entities during comprehension. Imagine hearing a subject-verb-object (SVO) sentence like 'The experimenter just sugared the...' (translated from German) in front of a scene where a plate with strawberries and a plate with pancakes are visually available. If a prior strawberry-sugaring action has been presented, the comprehender will preferentially look at the strawberries rather than the pancakes, even though these are equally plausible candidates for a sugaring action. This preference remains even though the strawberry-sugaring event is in the past and the sentence uses future tense (e.g., 'The experimenter sugars soon...'; Knoeferle, Carminati, et al., 2011). The term *recent-event preference* has been coined to refer to the preference for visually anticipating entities from event-based representations over other plausible candidates based on the linguistic input during situated language comprehension. In this chapter, we will also address the topic of visiolinguistic mismatches (i.e., when language is at odds with different aspects of visual information), their motivation in psycholinguistic research and how they further inform us about which processing mechanisms may be involved in connecting language with the visual world.

In the third chapter, we will move towards the psycholinguistics of gender, in order to outline its importance in sentence processing for strictly linguistic contexts, and the necessity to explore its effects further in situated language comprehension. We will review studies involving different dimensions of gender, from grammatical to conceptual gender. The latter dimension would encompass biological gender knowledge and knowledge about gender stereotypes, and this is what we will focus on. We will discuss the experimental methods that have been used in order to investigate the influence of gender information

and the conclusions that have been drawn about how this information modulates language processing.

Based on the previous chapters, in chapter 4 we will discuss relevant accounts and models of situated language comprehension that have been put forward in order to underline the processes implicated during comprehension in a rich contextual, visual world. We will have a special focus on the Coordinated Interplay Account (CIA; Knoeferle & Crocker, 2006, 2007; Knoeferle et al., 2014) and we will identify the potential aspects of the account that could be further informed.

The fifth and sixth chapters are an extensive description of our eye-tracking experiments using the *visual-world paradigm* (i.e., a paradigm where participants' eye movements are measured while understanding language and looking at a relevant visual display; Huettig, Rommers, & Meyer, 2011). In these experiments, we studied the influence of action and gender cues from prior events during situated language comprehension, and we pitted this information against knowledge about gender stereotypes (Experiments 1 and 2 also to be seen in Rodríguez, Burigo, & Knoeferle, 2015; Experiment 3, Rodríguez, Burigo, & Knoeferle, 2016).

Participants first saw a particular event (i.e., a video of female/male hands acting upon an object). Then a visual scene appeared with the faces of two potential agents: one male and one female. While the agent matching the gender features from the prior event was considered as the target agent, the other character, whose gender was not cued in previous events, was the competitor agent. The eye movements of the participants towards the agents were measured during the comprehension of German OVS sentences (e.g., *Den Kuchen backt gleich Susanna*; 'The cake_{NP1/obj} bakes_V soon_{ADV} Susanna_{NP2/subj}'). We used the non-canonical OVS word order as opposed to SVO (more commonly used in prior research, e.g., Knoeferle, Carminati, et al., 2011) precisely to examine participants' expectations towards the agent, who was mentioned at final position. Participants' task was to verify via button press whether the sentence matched the events they just saw (e.g.,

see chapter 5, Figure 5.1). To test the relative strength/weight of event-based representations (i.e., gender and action representations based on prior events) on the one hand and stereotypical gender knowledge on the other during comprehension, we adopted a mismatch design, something that has not yet been extensively tested in situated language comprehension studies (e.g., Knoeferle, Carminati, et al., 2011; Knoeferle & Crocker, 2006, 2007, but see Knoeferle et al. 2014), and we manipulated two factors. One factor was the referential congruence between prior events and language: there were action-verb(-phrase) mismatches in Experiments 1 and 3, and mismatches between the gender of the hands and the final subject (i.e., the proper name) in Experiment 2. The second manipulation was the stereotypicality match between the actions/events described in the sentence and the target agent's gender (i.e., the agent whose gender features matched the hands seen in the previous action video).

In these three experiments, we saw that action videos in which an agent (implicitly identified as male or female via the hand gender in the video) acts upon objects can rapidly affect subsequent interactions between an utterance and a visual scene. Gender cues from prior events preferentially guide attention towards one potential agent over another (of the opposite gender) in a scene during language comprehension. We also observed that mismatches between prior events and the linguistic input at different points in a sentence likewise affect this preference (mismatches reduce the preference for inspecting the agent whose gender was cued in prior events, i.e., the target agent). Experiment 3 additionally showed that, provided that the visual scene concurrent with language contains sufficient constraints (where not just characters/agents, but also objects are present), not just event-based information, but also stereotypical gender knowledge is used in order to preferentially inspect one agent as opposed to another, something that had not yet been found in research on situated language comprehension.

The seventh chapter describes Experiment 4, which was conducted using event-related brain potentials (ERPs), and the hand-subject gender match manipulation from Experiment 2. This experiment aimed at identifying some of the underlying mechanisms of

sentence comprehension as a function of prior visual gender cues. On the one hand, we aimed at exploring the commonalities between the flow of information in discourse comprehension studies conveying gender information and the current experiments. On the other hand, we wanted to see how responses to this type of verification (i.e., the reconciliation of visual and language-based gender information) compare to already existing evidence for other picture-sentence relations, i.e., thematic role relations and action-verb relations. Our results suggest that two semantically induced processing stages can be identified in the verification of linguistic gender cues with prior visual gender cues. This is in some respects similar to what has been observed with strictly linguistic stimuli as well as with other picture-sentence relations during processing. The result indicate that this methodology can provide us with further information about functional mechanisms (at present not identifiable in the eye-tracking data) involved in situated language comprehension in general and in the processing of gender information in particular.

Taking our contribution into account, in chapter eight we will discuss and interpret the findings, and will explain how these findings inform state-of-the-art accounts on situated language comprehension, including an illustrative example on how gender information would be handled according to the latest version of the CIA (Knoeferle et al., 2014). We will briefly outline future directions that could extend the present line of research.

2 | Situated language processing

Across several experimental paradigms it has been shown that comprehenders actively build mental representations based on the events described by the linguistic input and our knowledge about how those events usually take place, i.e., long-term experience with objects, people, actions, and so on. However, language comprehension does not usually take place in isolation. Rather, it takes place in rich contexts, i.e., in which linguistic and other perceptual information (e.g., visual) is present.

Prior to the extensive use of visually situated contexts in language comprehension studies there have been recurrent debates between those in favor of syntax-first (e.g., Frazier & Rayner, 1982; Friederici, 2002) and those in favor of interactive and parallel constraint-based models of language comprehension (where semantic, pragmatic and other sources like prior discourse context immediately interact with syntax; e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). In this debate, the *visual-world paradigm* has turned the scale in favor of the latter (Huettig, Rommers, & Meyer, 2011; Spivey & Huette, 2016). This turn has enriched later accounts of sentence comprehension, which have started to include the information from the visual context in which comprehension takes place (e.g., Altmann & Mirković, 2009; Knoeferle & Crocker, 2007).

As we will see in the following sections, the study of visually situated language comprehension as a real time process has provided evidence for a very tight coupling between the comprehension of language and visual attention within our visual environment

(e.g., Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Kamide, Altmann, & Hayward, 2003; Knoeferle & Crocker, 2006; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). At the same time, the visual world itself may constrain or narrow the representations a person might consider during comprehension.

2.1 | Long-term experience

As mentioned above, to achieve a successful understanding of language, comprehenders need to have the capacity of creating mental representations of the situations being described, also known as *situation* or *mental models* (Garnham, 1981; Gernsbacher, 1991; McKoon & Ratcliff, 1992; Sanford & Garrod, 1981; Zwaan & Radvansky, 1998). These representations include tokens that stand for individuals language is referring to, as well as the events such individuals are involved in (Garnham, 1981). Linguistic cues give us hints to construct the different aspects of a mental model, from causal relations, intentionality, time and space to properties or traits (Garnham, 1981; Zwaan & Radvansky, 1998).

As many authors have pointed out, how mental representations are built seems to be not just related to linguistic knowledge; it is also closely related to our knowledge of the world. Long-term experiential knowledge from memory is part and parcel of the construction of mental situations from language (Carreiras, Garnham, Oakhill, & Cain, 1996; Garnham, 1981). Making use of our general knowledge about situations similar to the ones described by language allows us to establish relations across sentences (i.e., discourse coherence, when knowledge from one sentence prepares us for the next; e.g., Carreiras et al., 1996; McKoon & Ratcliff, 1992), as well as to make inferences during sentence processing, i.e., extract information that is not made explicit in the text:

- (1) a. The diplomat was lying dead on the floor.

- b. Meanwhile the servant wiped the blood off the knife.

When presented with a sentence pair like in (1), for example, several inferences could be drawn. The most obvious ones are, first, that the diplomat was murdered and second, that he was (most probably) murdered by the servant. Therefore in this process, a) we establish relations between the representation of one utterance and that of the following one and b) we fill in the gaps in the representations that the propositions of such utterances convey by making inferences. Comprehenders put their inferencing abilities to work not just with complete sentences, but also with smaller sentence elements, like constituents (e.g., a verb phrase like *fly a kite* already allows us to build up an event representation where a kite is being flown) or even single lexical items (e.g., when hearing a noun like *the surgeon*, we might automatically make the inference, and therefore build a representation, in which the referent of that noun is a male person. Such an inference might be confirmed or dismissed based on later discourse. We will discuss similar cases in later sections, e.g., 2.2 and 3.2).

More often than not, we have the impression that the mental representations we construct from a sentence are built in an *incremental* way: meaning seems to build up on a word-by-word basis (but see Jackendoff, 2007), and considering several constraints at a time: phonological, syntactic, semantic and pragmatic (Huettig, Rommers, & Meyer, 2011; MacDonald et al., 1994; Trueswell et al., 1994). This leads us to the intuition that during language understanding, comprehension also seems to involve a constant assimilation of new incoming linguistic elements to enrich the mental representations that are under construction (Kamide, 2008). However, given that comprehenders have a tendency to make inferences, they are not just passively dealing with incoming information. We tend to go beyond the assimilation process, that is, we moreover form expectations. Many authors have explored the ability of comprehenders to elaborate predictions about upcoming information, both in the form of linguistic input (i.e., whether a noun, a verb, an adjective, etc. is expected) and based on an abstraction process of how the events described should unfold (i.e., based on our long-term knowledge, we might picture in our

heads what could happen next during the unfolding of discourse; Altmann & Mirković, 2009; Kamide, 2008; Sanford & Garrod, 1981).

As more information is disclosed by language, we can also anticipate thematic information, i.e., which type of object will most likely be involved in the generated mental representation of the events (e.g., whether it is an agent, a patient, or a theme). In one study, Altmann (1999) used a *stop-making sense* judgment task to test predictive processes based on prior context and verb information. Participants were presented with two sentences as prior context (e.g., (2a) and (2b)), and they then had to press ‘yes’ to reveal each word of the final sentence, and press ‘no’ whenever the sentence stopped making sense:

- (2) a. A car was driving downhill when it suddenly veered out of control.
- b. In its path were some dustbins and a row of bollards.
- c. It *injured/missed* several bollards that came close to being destroyed.

In the example (2c) above, participants pressed ‘no’ more often and had longer reading times at the verb (i.e., before *bollards* was read) when it was *injured* compared to when it was *missed*. The author took this as evidence for anticipatory processes: participants used the representation based on prior linguistic context together with the verb information to expect an antecedent that could fulfill its thematic restrictions (in the case of *injured*, i.e., the verb with more selectional restrictions, an animated, patient role).

As suggested by Altmann’s study and many other examples in the psycholinguistic literature, verbs are believed to be the most powerful *predictors* in sentence processing, as they don’t tend to stand alone in language (Altmann & Kamide, 1999; McRae, Hare, Elman, & Ferretti, 2005). On the one hand, verbs impose restrictions on the syntactic structure of a sentence (i.e., the number of arguments it should contain). On the other hand, verbs also place semantic (i.e., thematic) constraints on event structure (i.e., which

type of arguments the sentence should contain). Verbs prime agents, patients and instruments that would typically fill the roles of its arguments (Ferretti, McRae, & Hatherell, 2001; Gentner, 1981; McRae et al., 2005). However, this is not to say that other sentence parts do not have a role in generating predictions. Tanenhaus, Boland, Garnsey, and Carlson (1989) investigated how lexical information was used in long distance dependencies (e.g., questions such as *Which book did the boy read in class?*). Participants performed a stop-making-sense task while they were listening to such sentences. Results showed that they were faster in responding ‘no’ at the verb whenever the initial noun was not a suitable role filler for the verb (e.g., *Which food did the boy read in class?*), compared to when it was suitable. McRae et al. (2005) later found in a priming study that reading times for verbs were shorter when the priming word was a suitable thematic agent, patient or instrument for it (e.g., *nun-praying, guitar-strummed, bedroom-sleeping*) compared to unrelated pairs (e.g., *sniper-praying, musician-petting*). They concluded that the generation of expectancies does not only occur from verbs to nouns, but also vice versa, provided that such nouns are sufficiently strong cues to the generalized events that people store in memory.

So far we have talked about how the understanding of language, aided by our knowledge of the world, contributes to mental representations of events. However, it is sensible to think that participants draw on both linguistic and other non-linguistic, perceptual (e.g., visual) sources for the construction of these representations. Moreover, when conflicts between the two sources arise, this may cause disruptions in participants’ language comprehension processes, and consequently, in their performance in different tasks. In the past, it has been argued that such sources are dealt with separately via encapsulated cognitive systems or modules, where the output of one system is fed into the central cognitive system, and the different modules don’t need to inform one another (Fodor, 1983). However, more recent research has moved to a view in which at the representational level, the different perceptual sources are intertwined and interact at different stages of processing, arguably sharing a common representational substrate (Altmann & Mirković, 2009; Barsalou, 2008; Potter, Kroll, Yachzel, Carpenter, & Sherman, 1986).

In a study investigating the nature of representations of meaning (Potter et al., 1986), participants were presented with written sentences in serial visual presentation, where some objects appeared in their pictorial or lexical form. In the conditions where objects were presented as a picture, no disadvantage in processing was found compared to when written words appeared, supporting the idea of a common conceptual system for both types of input. More recently and in relation to inference-making studies, Zwaan, Stanfield, and Yaxley (2002) presented participants with a written sentence followed by the picture of an object. Participants then had to respond whether the sentence had mentioned the object in the picture. For each of the objects that were tested in the experiment, two different sentence versions were used (e.g., *The egg was in the refrigerator* vs. *The egg was in the pan*) as well as two different pictures from the same object, in which the object was in a different state or shape (e.g., a solid, unprocessed egg vs. a fried egg). Each sentence version implied a different state for the object, therefore corresponding to one of the two pictures. Participants were faster to respond when the perceptual characteristics implied by the sentence matched the following picture (the picture of a full egg after *The egg was in the refrigerator*) compared to when they did not (*The egg was in the pan*), which is consistent with the idea that during language comprehension, people simulate the shapes and states of objects and these dynamic perceptual representations interfere with the processing of the visual information of objects.

To summarize, mental representations can be generated from several sources: from linguistic information combined with the comprehender's knowledge about how events in the world take place, and from information in visual scenes. Arguably, representations from the linguistic and the visual input successfully interact in comprehension, but they can also interfere with one another. This has led some authors to suggest that different types of input may contribute to the construction of the same underlying representation (Altmann & Mirković, 2009; Huettig, Mishra, & Olivers, 2011), which then feeds to several cognitive processes, from inference-making to predictions on how both language and the real world events should unfold. In this respect, studies on language-mediated visual attention have made big steps in providing an insight into such processes and the

kinds of interactions involved.

2.2 | The concurrent visual context

As already mentioned, linguistic exchange between individuals usually takes place in rich contexts, e.g., in conjunction with the visual world. We constantly refer to things (e.g., objects or people) in our visual environment while pursuing different goals (e.g., asking someone to pass you the salt during dinner). At the same time, we also tend to visually search for things referred to by the language we hear. Grounding linguistic expressions in our perceptual world enriches the comprehension process, and allows for fast and successful achievement of communicative goals. In what follows, we will argue that the relation between linguistic and non-linguistic sources is bidirectional: linguistic information aids visual perception (Allopenna, Magnuson, & Tanenhaus, 1998; Eberhard et al., 1995; Marslen-Wilson, 1987), but the visual context also influences language comprehension in real time (Altmann & Kamide, 1999; Anderson, Chiu, Huette, & Spivey, 2011; Chambers, Tanenhaus, & Magnuson, 2004; Knoeferle et al., 2005; Tanenhaus et al., 1995).

How we make use of the visual context to understand language and how we map language into the visual world gives us insights into a central concept in language comprehension, namely, *reference* (see Jackendoff, 2002, for a discussion). We can understand reference as the process of connecting a linguistic entity to the object it denotes, be it perceived or in the mind (Jackendoff, 2002; Knoeferle & Guerra, 2016). In a perceptual, visual context, the establishment of reference would be indexed via eye movements towards the appropriate object. To measure how this process takes place, what types of information (i.e., syntactic, semantic, pragmatic and so on) are implicated in comprehension and when, psycholinguists have developed what is called the *visual-world paradigm*, first used by Cooper (1974) in the context of a narrative¹. The visual-world paradigm is

¹Although it was Cooper (1974) who first established this relation between language and visual attention, it was not until Tanenhaus et al. (1995) that the paradigm became popular.

an experimental setup where participants are seated in front of a display and their eye movements towards elements on that display are measured as language unfolds. Displays can contain semi-realistic scenes (e.g., clip-art images) or real pictures or videos. Language either appears in the form of an instruction, or as a description of what is depicted (see Huettig, Rommers, & Meyer, 2011; Knoeferle & Guerra, 2016, for a review). Tasks used in such studies can be either active or passive: in active tasks, participants may be required to respond to questions about the content or to verify the match between different aspects of the visual and the linguistic inputs. In passive tasks, participants are only asked to listen to the linguistic content while inspecting the scenes (Pyykkönen-Klauck & Crocker, 2016).

Studies on situated language comprehension, i.e., language in relation to a visual, non-linguistic context, have provided evidence for the view that people actively and rapidly exploit the visual environment in order to link both linguistic and visual information (Allopenna et al., 1998; Eberhard et al., 1995; Knoeferle et al., 2005; Knoeferle & Guerra, 2016; Tanenhaus et al., 1995). When listening to a word, we try to establish a connection between that word and the elements in our visual field, already as the word unfolds (Allopenna et al., 1998; Eberhard et al., 1995), which narrows down the set of potential visual referents (i.e., the objects referred to by language) until one instance from the set is selected for attention. When uniquely identified, visual referents tend to be inspected as fast as 200 ms after word onset, although when an object with a phonologically similar realization is present (e.g., *candle* and *candy*), this process is slower, as both entities (i.e., the candle and the candy) undergo a temporal competition (Allopenna et al., 1998; Eberhard et al., 1995).

Incremental processing of lexical information in sentence contexts has also been examined using the visual-world paradigm. In one of the first studies investigating the role of the incremental disambiguation of referents (Eberhard et al., 1995), participants were presented with displays showing four blocks, which differed in marking (e.g., starred vs. plain), color and shape. Participants were required to select a block based on spoken

instructions like *Touch the starred yellow square*. Disambiguating information was provided at different points in the sentence, i.e., by the marking adjective (early), by the color adjective (mid) or by the final shape noun (late). The authors found that participants established reference with the target objects on average 75 milliseconds after the offset of disambiguating words for early and mid conditions and about 200 milliseconds after the onset of the disambiguating word for the late condition. The results of the experiment supported a view where language is processed incrementally and non-linguistic information is rapidly integrated during that process.

The visual context has been shown to help resolve temporary ambiguities at the syntactic level (Chambers et al., 2004; Tanenhaus et al., 1995), and it also permits successful thematic role assignment during incremental sentence comprehension (Altmann & Kamide, 1999; Knoeferle et al., 2005). One influential study tested how changes in the concurrent visual context affected the resolution of syntactic ambiguity in sentences like *Put the apple on the towel in the box* (Tanenhaus et al., 1995). An example scene contained either just one referent for an apple (i.e., an apple on a towel) and an empty towel, or it contained a second apple (i.e., on a napkin). When only one apple was present ('one referent' condition), participants tended to incorrectly fixate the empty towel as the goal for the apple after hearing the modifier *on the towel*. However, when two apples were present ('two referent' condition) participants rarely looked at the empty towel (the incorrect goal). Thus, the presence of two referents in the scene prompted participants to interpret *on the towel* as the modifier of the noun *apple*, resolving the syntactic ambiguity against their preferred analysis (attachment into the verb phrase). Therefore, differences in the configuration of the concurrent scene can change the type of inferences and interpretations that language may convey.

Sometimes, if the linguistic and the visual context allow for it, visual reference might be established in an *anticipatory* manner, i.e., we even guide our attention towards entities before they are mentioned. This tends to happen, for example, when the action a verb denotes identifies an object in the scene (e.g., Altmann & Kamide, 1999, 2007; Kamide,

Altmann, & Haywood, 2003; Weber, Grice, & Crocker, 2006). In that sense, the visual context has a similar function to that fulfilled by prior sentential contexts, although it adds a cross-modal dimension (i.e., from linguistic to visual; e.g., Kamide, Scheepers, & Altmann, 2003). Further studies using clip-art pictures (where individuals and objects are depicted) have shown how important verb-mediated information is for establishing visual reference with objects or characters. Altmann and Kamide (1999) used visual scenes depicting a young boy sitting on the floor, surrounded by a toy train set, a toy car, a balloon, and a birthday cake. For the same scene, they used two types of sentences (as in (3)):

- (3) a. The boy will eat the cake.
b. The boy will move the cake.

While the verb *eat* clearly restricted the number of referents to one (i.e., the cake, the only edible entity), the verb *move* could be used for all of the items in the scene. The probability of looking towards the target object (the cake in both cases), was significantly higher when participants heard *The boy will eat* compared to *The boy will move*. The authors concluded that much like a noun preceded by some modifying adjectives, verb-mediated information (i.e., its selectional restrictions), can rapidly trigger saccades towards objects, even before these objects are mentioned.

Verb tense is another way in which restrictions can be placed on comprehension even when the visual referent is never explicitly mentioned in language, as evidenced by a later study (Altmann & Kamide, 2007). When shown a scenario with a man, a full glass of beer and an empty glass of wine, participants directed more anticipatory looks to the full glass of beer when listening to *The man will drink* compared to *The man has drunk* at the onset of the final referring expression (i.e., the beer). The opposite happened for the empty glass of wine; participants directed more anticipatory looks towards it when listening to *The man has drunk* compared to *The man will drink* at the onset of *the wine*. The authors interpreted this as evidence for anticipatory looks taking place not

exclusively based on the linguistic input, but guided by the *affordances* of the objects that are in the scene, i.e., the non-linguistic knowledge based on our experience with those objects and their interactions. In the absence of actual drinkable wine in the scene, the affordances of the empty glass of wine indicate that the object might have previously contained wine, triggering the anticipatory looks towards it when the past tense was being used. Thus, the tense of the verb had an effect on anticipatory looks towards the appropriate object, even when the object itself (i.e., an empty glass of wine), say the authors, violated the selectional restrictions of the verb (i.e., to drink) and was not even mentioned by language.

Another aspect of world-knowledge, such as plausibility, adds to the different cues that trigger visual anticipation in situated language comprehension. For example, when presented with a context depicting a little girl, a man, a motorbike and a carousel, participants direct more anticipatory looks to the motorbike after hearing *The man will ride* compared to *The girl will ride* (Kamide, Altmann, & Haywood, 2003). Arguably, given the restrictions of the verb *ride*, both the man and the girl could ride the motorbike. However, the results can be explained on the basis of knowledge about real-world plausibility. This knowledge tells us that little girls do not tend to ride motorbikes. First, the results support the idea that combinatorial information (i.e., the combination of the initial noun followed by the verb) successfully drives anticipatory eye movements. Second, real world plausibility is used together with the restrictions of prior linguistic information to predict the role filler in a sentence.

In sum, we have seen that reference between linguistic and visual entities can be established in different ways, sometimes immediate as in the case of noun-object relations, sometimes in an anticipatory manner, as in the case of verb-noun relations, as long as enough cues from either the linguistic (e.g., direct reference, disambiguating adjectives, semantic restrictions and world-knowledge) or the visual context are available (i.e., how the visual context is configured or how much visual information is available).

2.3 | Prior visual cues

Although in some cases linguistic and visual information may be simultaneously available for the comprehender, in many situations, either one or the other source of information is available first. This temporal asynchrony between the different sources has important influences in the comprehension process. For example, actions in an event might be short-lived, and after such actions, maybe only the people or the objects involved remain. In those cases, rather than current events, language may describe something no longer present (or only partially available) in the current visual scene. Yet the prior visual context (i.e., recent events) seems to exert a strong influence on how comprehension takes place and how our visual attention is guided in the scene.

In a study by Knoeferle and Crocker (2006), depicted events were pitted against world-knowledge (about occupational stereotypes) during comprehension. For example, participants were first presented a dynamic clip-art scene with three characters (e.g., a wizard, a pilot and a detective, see Figure 2.1). The central character (the pilot in this example) was the patient of the depicted actions the other two characters performed (e.g., the wizard would appear as spying on the pilot and the detective would appear as serving him food). After the actions took place, the three characters (but no object related to the actions) remained on screen for inspection. During the comprehension of non-canonical German OVS sentences like *Den Piloten bespitzelt gleich der...* ('The pilot_{acc} spies on soon the...') participants looked more often to the agent of the depicted event (i.e., the wizard) compared to the stereotypically matching character (i.e., the detective). That is, when the verb identified both an agent that was seen as performing the action described prior to sentence comprehension and a stereotypically more appropriate agent in the scene, participants preferred to look at the agent of the depicted action. These results suggest that representations from depicted events during the interpretation of a sentence describing those events preferentially guide visual attention over the entities in a scene during comprehension.



Figure 2.1.: Example image (middle scene) from Knoeferle and Crocker (2006).

Even manipulating verb tense in the context of prior visual cues has yielded similar results. In another study (Knoeferle & Crocker, 2007), participants saw another dynamic clip-art scene where an agent performed an action over one out of two available objects (e.g., a waiter polishing candelabra). Participants were then presented with a sentence like *Der Kellner poliert...* ('The waiter polish...'). Crucially, both depicted objects (i.e., the candelabra and some glasses) were plausible thematic role (i.e., theme) fillers for the action indicated by the verb. The verb was ambiguous until tense was revealed as either in the past tense (i.e., *...polierte...* 'polished') or in the futuristic present tense by means of an adverb (i.e., *...poliert demnächst...* 'polishes soon'). The past tense verb and ensuing adverb were followed by the mention of the recently acted-upon object (i.e., the candelabra) and the present tense verb and the futuristic adverb were followed by the mention of the other object, which was a potential target for an unseen, future event (i.e., the glasses). Although both tenses appeared equally often during the experiment, eye-movement patterns at the verb and the adverb regions showed an overall preference for the recent-event target over the plausible future event target.

In a following study using real videos (Knoeferle, Carminati, et al., 2011), participants saw an experimenter acting upon one out of two available objects, e.g., a plate

with strawberries and a plate with pancakes. Participants then listened to a German SVO sentence like *Der Versuchsleiter zuckert sogleich/zuckerte soeben...* ('The experimenter sugars soon/just sugared...') while inspecting a still frame with the two objects and the experimenter in the middle. Similar to the findings from Knoeferle and Crocker (2007), results revealed that if a prior strawberry-sugaring action had been presented, the comprehender preferentially looked at the strawberries rather than the pancakes, even though the pancakes are equally plausible candidates for a sugaring action. This preference was persistent, even though the strawberry-sugaring event took place before sentence comprehension, and the sentence was in the futuristic present form. A within-experiment frequency bias towards the future events (by introducing filler trials showing more frequent post-sentential/'future' event videos) did elicit an earlier rise of looks to the plausible 'future event' object when the futuristic present was used, although the overall preference for the target that had been acted upon prior to language comprehension remained (Abashidze et al., 2014). The term *recent-event preference* has been used to designate this preference for (visual) event-based representations over other types of knowledge, such as plausibility or stereotypical knowledge, during situated language comprehension.

In relation to this priority of event-based information versus other potential outcomes based on merely language-based, long-term knowledge, Altmann and Kamide (2009) put into test the ability of comprehenders to update internalized mental representations from the visual scene in the presence of an unchanging (therefore, to some extent, also prior to key linguistic components) visual environment. They presented participants with scenes in which, according to the linguistic input, certain objects were going to experience a change in location. For instance, there was a scene with a woman, an empty glass of wine and a bottle on the floor and a table. In a concurrent manner, participants could either listen to a sentence describing a situation in which the glass of wine is unmoved (e.g., *The woman is too lazy to put the glass onto the table*) or moved (e.g., *The woman will put the glass on the table*). The visual scene did not change and the glass always remained on the floor. A second sentence introduced the same event for both conditions (i.e., ... *she will pick up the bottle, and pour the wine carefully into the glass*). Eye movements at the

final regions (i.e., *pour the wine / carefully into / the glass*) revealed more looks towards the table when the first sentence had indicated that the woman had put the glass on the table before pouring wine onto it, compared to the ‘unmoved’ condition. The authors took this as evidence for the existence of a dynamic mental representation of the object location as mediated by language (a representation that goes beyond what is depicted in a scene). However, the relative difference in fixations to the table between the moved and the unmoved conditions was obscured by the fact that the glass (which was in the same location for both described scenarios) was fixated significantly more than the table in both conditions. The actual position in which the glass was depicted for both conditions (i.e., on the floor) seemed to interfere with the language-mediated representations updated in memory.

Overall, it is apparent that visual cues from prior events do have a strong influence on how our referential strategies develop: we seem to preferentially relate the sentential verb to entities that have been recently depicted as taking part in prior events, rather than entities that might be linked to the verbal input by means of other sources of information (i.e., long-term knowledge). However, as suggested by within experiment manipulations, this preference is not invariant, and depending on how the different aspects of the context are presented to the comprehender, they may interfere with this priority of event-based, visually grounded information (Knoeferle & Guerra, 2016).

2.4 | Visuolinguistic mismatches

Sometimes when trying to reconcile the representations coming from different sources in comprehension (e.g., event- or scene- and language-based information), some parts of either one or the other source might be at odds. This could presumably cause different types of disruptions, or the use of different strategies during comprehension. In this sense, mismatch-based designs can be very informative for the area of psycholinguistics. By creating mismatches during language processing and comparing their (online and

post-comprehension) effects with cases where comprehension and verification processes take place smoothly (i.e., matching context-sentence pairs), we can gain an insight into the mechanisms that are involved in the course of comprehension and would otherwise go unnoticed. One interesting question to address, for instance, is to which extent the preference we have seen to rely on prior visual cues (e.g., recent events) in situated language comprehension studies can be modulated by incongruences between event-based and sentence information (e.g., if language is at odds with recent visual events, will comprehension of a sentence still rely more on event-based representations, or will long-term knowledge from the linguistic input take precedence?).

Picture-sentence verification experiments have a long tradition in psycholinguistics (Gough, 1965; Just & Carpenter, 1971; Wannemacher, 1974) and when combined with continuous measures like eye-tracking or neurophysiological methods they can provide very accurate information about the processes that visiolinguistic interactions require (Knoeferle et al., 2014; Vissers et al., 2008; Wassenaar & Hagoort, 2007). In one of the first studies exploring picture-sentence verification (Gough, 1965), participants' reaction times were measured as they verified the match between sentences with different structures (i.e., active and passive sentences presented in the affirmative or negative form) and pictures that were presented at the end of the sentence; the obtained response latencies were interpreted as the time it took participants to understand the sentence. Gough found that response times were faster for picture-sentence matches compared to mismatches, faster for affirmatives compared to negatives, and faster for actives compared to passives. The truth value of the sentence interacted with the affirmative/negative opposition, which led the author to the conclusion that not only syntactic structure, but also semantic reversal processes (i.e., turning the proposition expressed by the sentence into its negation) are involved in the verification of language with pictorial information.

In a more step-by-step manner, Wannemacher (1974) manipulated the different parts at which a mismatch between the pictorial stimuli and the sentence could be encountered. In Experiment 2, pictures showing different situations were presented together with the

auditory sentences; they used reversible sentences (where the entities mentioned could interchange their roles with regards to the verb; e.g., *The boy is chasing the dog*) and non-reversible sentences (e.g., *The girl is picking the flowers*), both in the active and passive voices. Mismatches could be encountered at the subject, verb, and object positions, and combinations of such mismatch types were also used (s-v, s-o, v-o and s-v-o); in all cases reaction times were measured from the beginning of the sentence. Participants responded whether the sentence they were listening to matched the picture (*same* vs. *different*). RTs for mismatches occurring at the initial noun were the fastest, followed by the verb and then the object. For mismatches at the first noun, RTs for the active, non-reversible sentences were the fastest. The author interpreted participants' behaviour as a serial *self-terminating comparison strategy*; rather than waiting to process the whole sentence in order to verify its meaning with the visual input, participants adopted a strategy where each discrete constituent was processed as a unit.

Wannemacher's results may seem trivial at first sight (i.e., as reaction times were all measured from sentence start, it is not surprising to obtain reaction-time increases the longer it takes to identify a mismatch in a sentence); but the mismatch technique and sentence structures such as the ones from their experiment have served later research well, e.g., in the area of neurophysiological methods like event-related brain potentials (ERPs), which have been very popular in psycholinguistics. ERPs are electrophysiological responses recorded from several electrodes placed over the scalp of a participant which are time-locked to a particular event (e.g., the presentation of a particular word during sentence comprehension). Responses, which are then averaged across participants and conditions, take the form of positive or negative going deflections, also called *components*. Amplitude differences in these components as a function of experimental manipulations (e.g., a baseline condition vs. a mismatch) lead to ERP effects and can index, for instance, the different processes being affected during comprehension.

Two very commonly studied ERP components are the N400 and the P600. The N400 is a negativity peaking at around 400 ms after stimulus onset, linked to the processing

of meaning (Kutas & Federmeier, 2011). Both words and pictures can generate it, and anomalies at the semantic level can lead to more negative going responses compared to more sensible counterparts. Since its initial appearance in the work of Kutas and Hillyard (1980), several functional properties have been proposed for the modulation of the N400. Some authors have argued that it is related to violations of semantic expectations (Kutas & Hillyard, 1980, 1984; Wicha, Moreno, & Kutas, 2003). It has further been associated with semantic integration processes (e.g., Baggio & Hagoort, 2011; Brown & Hagoort, 1993; Osterhout & Holcomb, 1992) and also to processes of memory retrieval (Kutas & Federmeier, 2000, see Kutas and Federmeier 2011 for a review about the various interpretations behind the N400). Another very well-known component is the P600, which takes the form of a positivity peaking between 500 and 700 ms. Differential effects on this component have been traditionally linked to structural disambiguation (e.g., in garden path sentences) and reanalysis upon the encounter of syntactic anomalies (e.g., Hagoort, Brown, & Groothusen, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994) but also more general monitoring processes (i.e., not merely syntactic; Kolk, Chwilla, Van Herten, & Oor, 2003; Vissers et al., 2008).

Using this method, Wassenaar and Hagoort (2007) conducted a picture-sentence verification experiment with both healthy older adults and aphasic participants. Participants inspected line drawings showing either a reversible event (e.g., a man pushing a woman, where both entities could perform the action of *pushing*) or an irreversible event (e.g., a woman reading a book, where only the woman could perform the action of *reading*). After the presentation of those line drawings, participants heard a sentence in either the active (for the semantically reversible and irreversible cases) or the passive voice (only for the semantically reversible cases, e.g., ‘The woman on this picture is pushed by the tall man’, translation from Dutch). Sentences could either match or mismatch the depicted visual information. For healthy older participants (but not the aphasic group), mismatches (vs. matches) elicited larger early negative amplitudes time-locked to the acoustic onset of the verb in reversible active sentences. For the irreversible active sentences and the reversible

passive ones, the early negativity was followed by a late positive shift, peaking at around 600 ms (i.e., a P600 effect). The authors argued that these ERP effects indexed processes of thematic role assignment.

In an attempt to explore the potential processing mechanisms of different picture-sentence relations, Knoeferle et al. (2014) measured ERPs as participants read English SVO sentences and verified whether they matched a recently seen picture depicting an event. Pictures showed two characters: one was the agent and the other one the patient of the event (e.g., a gymnast applauding a journalist). Sentences could either fully match the pictures or contain different types of mismatches, i.e., verb-action (e.g., *The gymnast punches the journalist*), agent-patient role-relations (e.g., the sentence was *The gymnast applauds the journalist* when the pictures showed the reverse), or both (e.g., the sentence was *The gymnast punches the journalist* when an applauding action was depicted and in the reversed direction). These different mismatches elicited distinct ERP responses. Role mismatches (vs. matches) elicited larger anterior mean amplitude negativities (200-400 ms after the onset of the subject noun), while verb-action mismatches (vs. matches) elicited a somewhat later centro-parietal negativity (300-500 ms after verb onset), resembling a typical N400 effect. Additionally, in one of their experiments (with a word onset asynchrony of 500ms and a word duration of 200ms), role mismatches also elicited P600 like effects, which like Wassenaar and Hagoort (2007), they ascribed to thematic role assignment. The authors interpreted the results as implicating functionally distinct cognitive mechanisms for the different picture-sentence relations.

Different types of incongruences between pictorial stimuli and language seem to affect the timing as well as the type of response when trying to reconcile the representations derived from both. As we have seen, effects of mismatches have mainly been studied in the form of reaction-time studies and event-related brain potentials. Looking at the literature, we can see that eye-tracking experiments using this type of design are scarce (e.g., Abashidze & Knoeferle, 2015; Dumitru, Joergensen, Cruickshank, & Altmann, 2013; Wendler, Burigo, Schack, & Knoeferle, 2016). Worth mentioning is one

recent eye-tracking study that has explored eye-movement behaviour towards referents as a function of matches or mismatches in language, to see how referents are inspected in cases where anticipation might lead to failure (Dumitru et al., 2013). They used visual scenes in which two objects were present, and these were accompanied by an auditory sentence with conjunctive and disjunctive constructions (e.g., *Nancy examined an ant and/or a cloud*). They had different sentence manipulations in which either both nouns in the sentence matched their visual referents (MM condition) or the first, the second, or both nouns mismatched the pictures (yielding the conditions mM, Mm and mm, respectively). One of their main findings was that a match between the first noun and its visual referent increased the probability of fixations for the second noun referent, which the authors ascribed to what they called a *referential anchoring hypothesis*, based on the anchoring hypothesis by Tversky and Kahneman (1973). This hypothesis broadly describes an *heuristic strategy* (i.e., a rule of thumb used by individuals to ease a particular task) by which the estimate of a certain piece of information will depend on an estimate previously considered (i.e., a referential ‘anchor’)². In the case of visually guided language comprehension, this translates into anticipatory processes implicating referents being dependent on the stability of the previous information being processed, i.e., the degree of match between the previous linguistic and visual information. In other words, if the linguistic input currently being processed does not match the visual scene, anticipation of what was supposed to be the upcoming entity will be reduced as compared to a case where both linguistic and visual information are matching.

The studies mentioned in this section argue in favor of mismatch designs in order to explore the mechanisms involved during situated language comprehension. Eye-tracking studies in the visual world can nicely contribute to the literature, as they would not only tell us about disruptions in comprehension, but also strategies participants may pursue

²In some studies on anchoring, participants are first presented with a question that poses an initial *anchoring* value (e.g., *Is Harrison Ford younger or older than 10 years?*), then they are asked to give an estimation of the actual value. Results have shown that although people might try to stay away from the anchor at the time of making their estimation, this estimation will be biased by the initial anchor (i.e., participants will likely give a higher value if the anchor is 30 compared to 10 years; Jacowitz & Kahneman, 1995; Tversky & Kahneman, 1973).

or stop pursuing in such contexts, e.g., where participants look at during mismatches, for how long and whether other sources of information competing with prior visual events (e.g., long-term knowledge) gain greater relevance in such situations at the time of establishing reference or anticipating entities.

3 | The influence of gender on language processing

In the psycholinguistics literature, *gender* is a broad term that has comprised both linguistic and extra-linguistic factors. From the Latin stem *genus* and later the Old French form *gendre*, it came to mean ‘kind’ and then moved on to represent concepts such as grammatical gender, biological gender (i.e., sex) and social gender¹. While the former concept is related to language in formal terms, the other two are linked to language via conceptual (or semantic) knowledge. Grammatical gender is a morphosyntactic feature that together with person and number is often reflected in agreement processes (e.g., between determiners and nouns, nouns and adjectives,...), while conceptual gender is usually conveyed at a lexical level (e.g., definitional gender or stereotypical gender nouns; Kreiner et al., 2008).

Evidence for the importance of gender in language comes from several studies, including judgment studies (Bassetti, 2014; Flaherty, 2001; Garnham, Oakhill, & Reynolds, 2002), reading time measures (Garnham et al., 2002; Gygax, Gabriel, Sarrasin, Oakhill, & Garnham, 2008), eye-tracking in reading (Duffy & Keir, 2004; Kreiner et al., 2008), event-related brain potentials (Barber & Carreiras, 2005; Hanulíková & Carreiras, 2015; Molinaro, Su, & Carreiras, 2016; Osterhout, Bersick, & McLaughlin, 1997; Siyanova-Chanturia, Pesciarelli, & Cacciari, 2012) and, to a lesser extent, visual-world studies (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Pyykkönen et al., 2010). These

¹Consulted in Oxford Dictionaries online: <https://en.oxforddictionaries.com/definition/gender> (17/II/2017)

studies have obtained very fine-grained insights into how we use different aspects of gender in sentence processing. Throughout this chapter, we will describe the most commonly studied aspects of gender in comprehension. The first distinction is between grammatical gender and conceptual gender. Within conceptual gender, we will further make a distinction between those cases where biological gender is directly expressed in language (e.g., pronouns like *he* or *she*) and those where gender is not directly expressed, but is usually inferred from the use of long-term knowledge on social stereotypes (e.g., role nouns like *minister* or *nurse*). As we will argue, grammatical and conceptual gender are not completely independent from each other (e.g., nouns designating animated entities tend to be gender-marked in the grammar based on biological gender/sex). In fact, in language comprehension studies, grammatical and conceptual gender tend to be intertwined (often times knowledge about biological and stereotypical gender is tested via coindexation with anaphoric pronouns). Furthermore, both levels of gender knowledge seem to support a constant awareness of a feminine versus masculine representational dichotomy, which may sometimes pose an advantage, but it can also hinder comprehension processes where expectations based on (inferred) gender representations are not met.

3.1 | Grammatical gender

Many languages around the world make use of grammatical gender, which is a formal property inherent to nouns that divides them into two or more classes (Corbett, 1991). Grammatical gender is reflected in agreement processes of language (i.e., the systematic covariance among sentence parts; Steele, 1978) and sometimes also in the morphology of nouns themselves. Although the communicative functions of grammatical gender have been under debate for a long time, one of the most convincing hypotheses is that it serves 'reference-tracking': gender markers can serve to keep track of referents in long or complex stories or discourse passages (Bates & MacWhinney, 1989; Van Berkum, 1996).

Grammatical gender has mainly been defined as an arbitrary category in most lan-

guages; a property that belongs to individual nouns and not to the referents it denotes (Corbett, 1991; Köpcke, Panther, & Zubin, 2010; Van Berkum, 1996). However, to what extent is it arbitrary? Some authors have claimed that very often, conceptual factors motivate the use of one grammatical gender versus another (e.g., words of a particular semantic field, like the case of the German words for fruits *Orange*, *Birne*, and *Erdbeere*, tend to share the feminine grammatical gender; Köpcke et al. 2010). Moreover, the relation between biological gender or sex in living entities and grammatical gender has often been shown to be relatively transparent (clear examples in several languages are nouns like *man*, *uncle* and *king*, which are all marked as masculine, while *woman*, *aunt* and *queen* are all marked as feminine). These sometimes unclear boundaries between what is purely grammatical and what is conceptual may have an impact on cognition during language comprehension.

Grammatical gender has sometimes been shown to alter our representations differently from what is intended. Some psycholinguistic studies have found that in gender-marked languages (e.g., French or German, where nouns are marked as either feminine or masculine, and also neutral in the latter case), masculine forms intended as generic tend to bias comprehenders' representations of referents (Cole, Hill, & Dayley, 1983; Gygax et al., 2008; Hanulíková & Carreiras, 2015; Schneider & Hacker, 1973). Gygax et al. (2008) had their participants read sentences containing a masculine plural form intended as generic (e.g., *Die Sozialarbeiter liefen durch den Bahnhof*, 'The social workers were walking through the station'), and then judge the suitability of continuation sentences containing either a man or a woman as a referent. Continuations with women as referents (e.g., *Wegen der schönen Wetterprognose trugen mehrere der Frauen keine Jacke*, 'Since sunny weather was forecast several of the women weren't wearing a coat') turned out to be less acceptable than those with men in the gender-marked languages (French and German), and response times were longer for the former compared to the latter. Given that masculine forms can also be interpreted with a male-specific meaning, mental representations containing only men may be preferred to those containing both male and female referents. Therefore, the authors concluded that *gender-open* (i.e., no

gender activated) or *gender-spread* (i.e., both genders activated) interpretations of the plural masculine/generic are unlikely in these cases.

Furthermore, Flaherty (2001) obtained evidence for the influence of grammatical gender in assigning biological gender (i.e., sex) traits to objects. He presented Spanish (a gender-marked language) and English (a natural gender language) participants of different ages with cartoon images of different objects and asked them to assign a gender and to put a typical male or female name to them. The oldest groups of Spanish participants (8 to 10 year-old children and adults) assigned gender based on the grammatical gender of the referent, while speakers of English and younger children (5 to 7 years) in both Spanish and English seemed to make use of perceived gender attributes to assign a gender. A similar finding emerged when participants were presented with both animate and inanimate entities and were asked to assign female and male attributes to them: English speakers used the same strategy as before, while Spanish speakers assigned gender attributes based on the grammatical gender of the entities. Flaherty (2001) concluded that grammatical gender affects people's perception once it is acquired (i.e., as in the case of Spanish from the age of 8 years). Although this tendency to assign gender traits based on grammatical gender seems to be common, bilingual speakers whose gender-marked languages may have different grammatical genders for a single referent can better grasp the arbitrariness of grammatical gender (Bassetti, 2014).

On a related topic, Vigliocco and Franck (1999) showed that the arbitrariness of grammatical gender might have its disadvantages when it comes to language processing in certain contexts. In a production experiment in Italian (Experiment 1) and French (Experiment 2), participants were presented with an adjective (both in the feminine and masculine forms) and then with a preamble that consisted of a head noun with a prepositional phrase. The local noun (i.e., embedded in the prepositional phrase) had the opposite grammatical gender from that of the head noun (e.g., *L'inquilino della casa*, 'The tenant_{masc} of the house_{fem}'). Participants had to repeat the preamble and then continue the sentence with the adjective they were presented with at the beginning. By

inserting a local noun between the head noun and the elicited adjective, the authors wanted to induce errors in the use of the grammatical form. They manipulated the type of head nouns: they either had *conceptual gender* (like ‘tenant’ in the masculine form)² or a *purely grammatical gender* (like ‘closet’). In addition, the authors manipulated gender type (masculine vs. feminine). Results showed that gender agreement errors in the production of adjectives were more common when the head noun had grammatical gender compared to when it had conceptual gender. The authors concluded that participants take conceptual correlates into account when they compute a morphosyntactic relation like agreement.

In summary, grammatical gender is not an entirely transparent aspect of language, and moreover seems to depend on the language(s) we speak. Processing of a particular grammatical gender form can sometimes bias our representations as well as our perception of objects; the effects of such cross-linguistic studies go in line with Whorfian approaches to language (i.e., how language influences thought; Whorf 1956). However, the choice of the right morphosyntactic element during language processing can also be reversely supported by conceptual factors associated with our knowledge about biological gender.

3.2 | Conceptual gender

The studies described above suggest that even during the processing of grammatical gender information, an interaction occurs between linguistic and semantic or conceptual factors. Without taking any deterministic approach on how these two aspects of gender information influence each other, between the conceptual factors of gender that we identify, namely, biological and stereotypical gender, there is arguably a tight relation.

²In this experimental context, *conceptual gender* could be understood as referring to the way in which the relation between the sex of the entity is transparently expressed via grammatical gender. However, *conceptual gender* can also refer to the conceptual knowledge that allows for inferential processes when the comprehender is presented with words that are not necessarily gender-marked (e.g., stereotypical role nouns, objects and adjectives). The following section uses the term *conceptual* in the latter, broader sense.

Biological gender knowledge in language (i.e., the use of linguistic entities that convey biological gender information) tends to have a very straightforward relation with grammatical gender (i.e., the use of pronouns like *he/she*) and with referring expressions whose intrinsic meaning designates the sex of a particular entity, such as proper nouns (e.g., *John/Mary* or definitional nouns like *king/queen*). Just like we classify certain pronouns and nouns in language, *categorization* (i.e., a process by which we classify individuals based on different physical or social cues into groups) also occurs in the visual domain, even more so. In human to human interaction, for example, gender categorization during person-construal is an almost inevitable process that serves a clear cognitive function: that of organizing our environment, as this facilitates numerous perceptual and comprehension processes (Bodenhausen, Kang, & Peery, 2012; Oakes, 1996; Stangor, Lynch, Duan, & Glas, 1992; Taylor & Hamilton, 1981). When it comes to biological gender, several physical features can be used (i.e., dimorphic gender cues³, like the shape of the face, body size and so on).

The type of gender categorization that stems from linguistic and visual (biological) gender cues, seems to have immediate and persistent effects, but these are not the only cues that can be used. Some words do not unambiguously identify the gender of a referent, yet we tend to assign a gender to them as we hear them, arguably trying to apply the same heuristic strategy as with biological gender. For example, in the case of English, we have the nouns *doctor* and *nurse*, which in principle do not denote the gender of an individual. However, studies have shown that we do not process such words without exploiting the gender information they convey (Carreiras et al., 1996; Clifton & Staub, 2011; Clifton, Staub, & Rayner, 2007; Duffy & Keir, 2004). These inferential processes can happen not just with nouns, but also descriptions of actions (i.e., verb-phrases, which make us infer the gender of the possible agent for such an action; Reali, Esaulova, Öttl, & von Stockhausen, 2015). The type of knowledge that leads us to infer the gender associated with such words is part of a stereotyping process, in which beliefs

³The term *dimorphic* refers to the group of visual characteristics that differentiate between two biological genders (or sexes): female and male.

about the roles, traits and abilities of the different genders are applied (Bussey & Bandura, 1999). The process of categorization based on this type of information tends to be more inferential (probabilistic, therefore indirect), however, it can be very persistent in person perception. Once stereotypes are acquired, which tends to be at a very early age (Serbin, Poulin-Dubois, Colburne, Sen, & Eichstedt, 2001), they become implicit and easy to use, which favours automatic activation in numerous contexts. This activation may sometimes result in a cognitive advantage (e.g., shorter reaction times when categorizing a word or processing a sentence) when stereotypes are primed (e.g., Banaji, Hardin, & Rothman, 1993; Bargh, Chen, & Burrows, 1996). However, gender stereotypes can arguably be challenged by exposing a person to individualizing, counterstereotypical events (e.g., de Lemus, Spears, Bukowski, Moya, & Lupiáñez, 2013).

As we will see in the following two sections, information from biological gender and gender stereotypes is readily used in language comprehension, including (although it has not extensively studied yet) comprehension in the visual world. Given the somewhat different nature of both sources of information, however, the effects of both types of gender knowledge have shown slightly different effects. As we will argue later on (e.g., chapter 4, as well as our experimental chapters) although one type of knowledge (prior visual gender cues) may have a greater influence in comprehension than the other (long-term knowledge about gender stereotypes), this is not to say that these two sources of knowledge cannot interact or compete in certain contexts.

3.2.1 | Biological gender

In psycholinguistic studies, particularly those looking at discourse comprehension, several probabilistic cues have been studied, which participants can use in order to establish reference to a pronoun during language comprehension. These (often heuristic) cues include accessibility (e.g., which character has been mentioned first?), recency (which character has most recently been mentioned?) and the grammatical function of the potential antecedents for the pronoun (e.g., subject vs. object position; Crawley, Stevenson, &

Kleinman, 1990). Gender cues have also been subject to debate. Some studies have argued that information about the sex of the antecedents (whether it is John or Mary) in a sentence, together with the morphosyntactic features of the pronoun, may help to identify the potential referent for that pronoun faster (Arnold et al., 2000; Boland, Acker, & Wagner, 1998; Crawley et al., 1990; Ehrlich, 1980; Garnham, Oakhill, & Cruttenden, 1992).

Although findings in the literature have been mixed, several studies have found that when biological gender information is available, there is a rapid use of this information in comprehension, and it is preferred over other heuristic strategies (Arnold et al., 2000; Boland et al., 1998; Crawley et al., 1990; Ehrlich, 1980). Crawley et al. (1990), for example, used a series of passages manipulating both the position of the potential antecedents for a pronoun (subject vs. object position) as well as their gender (same vs. different). The first sentence introduced two characters, followed by a sentence in which third character was introduced. The third sentence was the target sentence, where the two characters that were mentioned first occupied the subject and object positions.

- (4) Brenda and Harriet were starring in the local musical. Bill was in it too and none of them were very sure of their lines or the dance steps. Brenda copied Harriet and Bill watched *her*.

The main findings were that in cases where the gender of the antecedents was the same (ambiguous condition, as in (4)), pronouns in both subject and object positions were preferentially assigned to an antecedent in subject position (*subject assignment strategy*). However, this preference did not emerge in cases where gender was available as a cue (unambiguous condition, where the two antecedents had different genders). Besides, passages with unambiguous gender cues were also read faster than the ambiguous passages. Thus, whenever available, information about gender alone served as the strongest cue for pronoun resolution. Garnham et al. (1992), on the other hand, found effects of gender cues on pronoun resolution in the presence of sentences with implicit causality, but only

under certain circumstances. They used sentences (as in (5)) where the first clause either had or did not have a gender cue (i.e., different vs. same gender characters) and the subordinate clause (a *because* sentence) was either congruent or incongruent with the bias of the verb in the first clause (e.g., NP1-biased as in *confess*, or NP2-biased as in *punish*). After reading the sentence, participants had to respond to yes/no questions (e.g., *Did Max want a reduced sentence?*). In Experiment 4, fillers were also manipulated, so that some participants were presented with fillers similar to the experimental items and others with fillers in which subordinate clauses did not require pronoun resolution (i.e., no pronoun was present, therefore questions did not focus on the pronoun). In Experiment 5, the authors embedded the target sentence in a passage and asked more than one yes/no question about the passage to divert the attention from the pronouns.

- (5) a. John/Jane punished Bill because he had done wrong.
b. Bill punished John/Jane because he had been wronged.

In Experiment 4, the authors found that while the congruency effect was more or less present across their experiments (i.e., subordinate clauses were read faster when they were congruent vs. incongruent with the bias of the verb in the main clause). However, gender cues (i.e., whether there were same or different gender characters in the sentence) were used most (i.e., subordinate clauses were read faster when the cue was present compared to when it was not) when all trials required pronoun resolution, rendering the use of this type of information strategic. Additionally, in Experiment 5 gender cue effects emerged, but only for the main clause (with faster reading times when the cue was present compared to when it was absent). Although the gender cue did not show effects on pronoun resolution (i.e., the subordinate clause) in this experiment, the authors argued based on the effects for the main clause that it helps construct a more distinct mental representation of the characters involved in the events.

In an attempt to further investigate the use of gender information in pronoun resolution using an online methodology and spoken instead of written language comprehension,

Arnold et al. 2000 conducted a visual-world study. Participants' eye movements were measured as they inspected scenes with the pictures of cartoon characters like Mickey or Minnie Mouse and Donald Duck and were listening to a story about the scene (as in (6)). Each story contained two sentences with two clauses each. They manipulated two factors: a) the gender of the characters (same vs. different) and b) the order of mention (first vs. second):

- (6) Donald is bringing mail to *Mickey/Minnie* while a violent storm is beginning. *He/She's* carrying an umbrella, and it looks like they're both going to need it.

The authors found that both cues (i.e., order of mention and gender), were rapidly used to identify the pronoun referent (i.e., the character who carried an umbrella) in the visual scene. Either when the characters were of different genders or the pronoun referred back to the character that was first mentioned (e.g., Donald Duck in example (6)), or both, the proportion of looks to the target character was greater than for the competitor at the onset of the pronoun. Only when both cues were absent (same gender antecedents and second mention target) did participants experience difficulties in establishing reference (i.e., target and competitor characters were inspected to the same extent). The authors concluded that these cues affected the initial stages of pronoun resolution and gender was not used to a greater extent than a heuristic cue like order of mention. However, the results in hand did also reject the claim that gender is only used in special circumstances or in a strategic manner. Overall, their effects supported a dynamic model of language processing where several constraints guide referential processing and pronoun resolution, one of them being biological gender. Using the same paradigm, only with different characters, Arnold, Brown-Schmidt, and Trueswell (2007) later found that during childhood (i.e., around 5 years of age), gender cues are more readily used than order-of mention, leading to the interpretation that children exploit gender cues before other (potentially less reliable) sources for language processing.

3.2.2 | Gender stereotypes

Another conceptual aspect of gender is that of stereotypes. This aspect concerns people's long-term knowledge; as social constructs, they prescribe a distribution of roles, occupations and abilities (Bussey & Bandura, 1999; Harper & Schoeman, 2003). Several types of words have been typically associated with a certain gender, from stereotypical role nouns (e.g., *electrician* vs. *nurse*) and objects (e.g., *doll* vs. *truck*) to traits (*agentic* vs. *communal*). The gender implied from these words is arguably inferred based on people's beliefs about the amount of men and women showing a certain type of behaviour, or engaged in a particular occupation (Garnham et al., 2002; Gygax, Garnham, & Doehren, 2016).

In this respect, social psychology studies investigating priming effects have observed a rapid integration of gender-stereotype information (e.g., Banaji et al., 1993; Blair & Banaji, 1996; Cacciari & Padovani, 2007). In a study by Blair and Banaji (1996), participants classified female and male names (e.g., *Alice* and *Adam*, respectively) that were followed by trait (e.g., *gentle* or *courageous*) and nontrait words (e.g., *lingerie* or *sports*) via button press (i.e., one button for male names, another one for female names). Although participants could have ignored the trait/nontrait words to classify the names, response times turned out to be faster in classification when the names were consistent with the gender implied by the trait/nontrait word that preceded them, compared to when they were inconsistent (e.g., *gentle-Alice*, *courageous-Adam*). Similar effects have been found with role nouns; Cacciari and Padovani (2007), for example, used role nouns (e.g., *insegnante*, 'teacher') as primes for the pronouns *he* and *she*, which participants had to classify (male vs. female, respectively). The authors also found shorter reaction times when prime and target were congruent (e.g., *she* preceded by *teacher*) compared to when they were incongruent (e.g., *she* preceded by *engineer*).

Studies on anaphor resolution have provided important insights into the effects of gender-stereotype knowledge in discourse (e.g., cases where *she* is preceded by *engineer*

in a text). Carreiras et al. (1996), for example, conducted a self-paced reading study in which an initial sentence introduced a stereotypically feminine, masculine or neutral role noun (e.g., *electrician*, a stereotypically male role noun, as in (7b)). An ensuing sentence introduced a pronoun which could either match or mismatch the gender implied by the previously mentioned stereotype role noun (7a):

- (7) a. The *electrician* examined the light fitting.
b. *He/She* needed a special attachment to fix it.

Participants pressed a button once they read the first sentence, and again after reading the second sentence. Right after the second sentence, a comprehension (yes/no) question appeared. The authors found that reading times for the second sentence were longer (i.e., it took longer for participants to press the button that triggered the comprehension question) when there was a mismatch between the stereotypical gender of the role noun and the pronoun following it (*she* in the example above), compared to matching conditions.

Using eye-tracking, stereotypical gender mismatches have also been shown to cause disruptions in on-line reading. For example, Duffy and Keir (2004) obtained clear gender mismatch effects in early reading measures at the reflexive pronoun site, e.g. in the case of a sentence containing *electrician* followed by the target sentence with the reflexive *herself*. First pass (i.e., the sum of fixations within an area before moving the eyes out of that particular area) and go-past times (i.e., the time spent re-reading previous parts of the sentence before moving beyond that area) were higher for gender mismatches, compared to matching conditions. The effect on these measures shows that the violation of gender expectations based on the role noun (*engineer*) makes it difficult for the reader to integrate the mismatching reflexive pronoun (*herself*) in the immediate linguistic context (Clifton & Staub, 2011; Clifton et al., 2007; Duffy & Keir, 2004), yielding longer reading times for gender-mismatching sentences.

Kreiner et al. (2008) further investigated the effects of definitional nouns (e.g., *boy* vs. *girl*, *king* vs. *queen*) and stereotypical role nouns (e.g., *nurse* or *doctor*) using sentences such as the ones in (8) :

- (8) a. Yesterday the *minister* left London after reminding himself/herself about the letter.
- b. Yesterday the *king* left London after reminding himself/herself about the letter.

The authors found effects in go-past measures at the reflexive pronoun region, as well as first-pass effects at the spillover region (... *about* ...). Mismatch costs were slightly bigger for definitional compared to stereotypical role nouns. When using cataphoric expressions (i.e., when the definitional noun/stereotypical role noun came after the reflexive pronoun) only definitional nouns elicited disruptions in reading when a mismatch was encountered. The authors concluded that the qualitative differences in processing definitional and role nouns result from the different ways in which gender is represented (in the former gender is arguably encoded in the lexical form, while in the latter gender is assigned based on probabilistic knowledge), hence the different strengths in their constraints. When syntactic constraints appear in discourse before the introduction of gender stereotypical role nouns, the integration of meaning becomes easier compared to cases in which the same role nouns come first.

Stereotypical gender is not necessarily conveyed by a single word; a more extended description of typical role nouns can also elicit gender inferences, as shown by Reali et al. (2015). The authors used sentences that described an occupation, followed by a sentence including a pronominal anaphor (see the example in (9)). Based on a prior rating study, they differentiated between high context primes (i.e., descriptions of roles which were strongly associated with a typical gender in explicit ratings, like *electrician* and *beautician*) and low context primes (i.e., descriptions of roles which were found to be

only slightly associated with a typical gender, like *lawyer* and *psychologist*), which were used in Experiment 1 and 2, respectively.

- (9) a. K. L. installs power lines and cables, checks electricity voltage.
b. In this field *he/she* has a lot of experience.

For the high priming gender contexts (Experiment 1), the first gender mismatch effects appeared at the spillover region (...*a lot of...*). Surprisingly, in the low priming gender contexts (Experiment 2), effects already appeared at the pronoun region. Correlational analyses between rating studies performed before the eye-tracking experiment and the eye movement data showed that while the explicit stereotypicality ratings did predict the eye movements in Experiment 1, no correlation was found in Experiment 2. The authors concluded that effects of gender typicality can come from two different sources: one is directly related to beliefs on the distributions of men and women in a certain occupation (as it happened with the descriptions used in Experiment 1), while the other is less explicit (as in the descriptions used in Experiment 2).

Stereotypical gender information can also have effects in cases when it is not explicitly expressed during discourse (i.e., when no anaphor confirms the gender of the referent). In the stimuli used by Garnham et al. (2002), an initial sentence introduced a stereotypical role name (e.g., *plasterer* as in example (10a) below), the second part provided further information and the third sentence introduced an item of clothing (e.g., *bikini* as in (10c)) or a biological characteristic (e.g., giving birth).

- (10) a. The plasterer, who had just finished a hard day's work,
b. went to get changed for swimming,
c. and put on a striped bikini.

After reading the last sentence, participants judged whether the final part was a sensible continuation of the other sentences by pressing a 'yes' or a 'no' button. The results showed that when the gender implied in the role name mismatched the one implied

by the clothes or the biological feature, ‘yes’ judgments were fewer and judgment times increased, compared to matching conditions (e.g., if the role name was *midwife* instead of *plasterer*). The same pattern emerged when the elements in the sentence switched positions; thus, even when a biological feature was introduced before the stereotypical noun, a mismatch in gender still caused an effect. According to the authors, these findings support the idea that certain inferences during sentence comprehension are made elaboratively i.e., upon the encounter of single lexical items and not necessarily due to discourse requirements. This is opposed to minimalist approaches of language processing, which predict that inferences are only made if necessary for local cohesion (e.g., McKoon & Ratcliff, 1992). Besides, morphological information (e.g., *he* or *she*) is not necessary for readers’ commitment to gender-stereotype interpretations.

Further evidence for the claim that comprehenders can make use of gender-stereotype knowledge immediately upon the encounter of stereotypical role nouns comes from the visual-world paradigm. Pyykkönen et al. (2010) used groups of three sentences in Finnish (see (11) translated into English except for the final, critical sentence) as auditory stimuli and four pictures presented together as visual stimuli, (a male and a female character and two objects related to the story).

- (11) a. On the screen you see Sinikka, a 35-year-old woman from Jyväskylä and Mikko, 40-year-old man from Tampere.
- b. While doing yard work Sinikka evaluated with Mikko the dangerous situations a chimney sweep gets into on slippery roofs.
- c. *Kouluttauduttuaan nuohoojaksi hän oli oppinut monia keinoja hoitaa työnsä turvallisesti.*
 ‘After having graduated to become a chimney sweep, he/she (amb) had learned many ways to work safely’.

The first sentence introduced a pair of characters (one male and one female). The authors manipulated discourse salience by changing the order of mention of the two

characters involved. In a second sentence, participants were talking about a particular stereotypical role noun (e.g., *nuohoojaksi*, ‘chimney sweep’); no reference was made to any of the participants. The third sentence started with a third person anaphoric verb (*Kouluttauduttuaan*, ‘After having graduated’) the suffix of which (*-aan*) was gender-ambiguous; then the stereotypical role noun was repeated referring to the same character as the anaphoric suffix. In half of the trials, the gender of the stereotypical role noun and the gender of the salient character matched, while there was a mismatch in the other half of the sentences. In all cases, the bridging inference⁴ that was required was the one between the anaphoric expression and one of the previously mentioned characters, while it was not necessary to establish reference when encountering the role noun. Results showed that participants looked at the more salient character upon the encounter of the anaphoric verb (i.e., *Kouluttauduttuaan*). However, after the onset of the repeated stereotypical role noun and sometimes even before the offset of it, more looks were directed towards the stereotypically consistent character, reflecting revision over the previously established relations. Similar results were obtained when the story mentioned objects stereotypically associated to characters instead of stereotypical role nouns (*motorcycle* vs. *hair clip*). Stereotypical knowledge in this case seemed to override the inferences regarding the discourse salience of characters in the story. This evidence shows that comprehenders exploit stereotypical gender information early even when it is not necessary for establishing coherence in the discourse, thus supporting the view of Garnham et al. (2002) and challenging minimalist approaches.

As evidenced by discourse processing experiments and also eye-tracking studies using the visual-world paradigm, gender stereotypes seem to have a strong influence in language comprehension, not just in the case of single lexical items such as in *surgeon* or *babysitter*, but also during the description of actions and behaviours. Some studies provide evidence indicating that knowledge about gender stereotypes can be used in an elaborative manner (i.e., without being strictly necessary during the comprehension process), arguably

⁴Bridging inferences are usually defined as the resolution of anaphoric relations by means other than explicit linguistic coreference between entities (Irmer, 2011).

because they might allow for earlier disambiguation of referents during comprehension, i.e., we try to assign a gender to the referent at the earliest opportunity. However, when wrongly used, this strategy has obvious costs in processing.

4 | Accounts and models of situated language comprehension

Although some accounts and models¹ on language processing had already put forward the interaction of several sources of information during incremental language comprehension (e.g., Jackendoff, 2002; MacDonald et al., 1994; McRae et al., 2005), not much emphasis had been placed on the non-linguistic, visual context until recently. As seen throughout chapter 2, evidence obtained from examining language in the context of a visual world reveals a need for integrating representations derived both from language and the non-linguistic visual context in accounts of language processing, as well as assessing the relative importance of one versus the other during comprehension.

One early connectionist proposal that has inspired later models for language comprehension is the simple recurrent network (SRN) by Elman (1990). This network accounts for two concepts we have been discussing, i.e., the role of representations from prior context at a certain point in time during the processing of further external input, and the ability of the comprehender to predict the input that will follow the next point in time. Input units and context units activate hidden units, and these units feed forward to output units which attempt to predict the next input. The patterns of the hidden units are saved as context for the next point in time: the context will act as memory and

¹A working definition for both terms *account* and *model* is fitting, since these are sometimes used interchangeably and confused in the literature. We will define an account as a (usually high-level) report on how language processing may develop and how its different aspects may be represented based on findings (e.g., patterns of data) from psycholinguistic research, i.e., an explanation of psycholinguistic phenomena. By contrast, we will define a model as a formalized version of an account, often computationally implemented, which can put theoretical predictions to test (e.g., see Crocker, 2010).

have recurrent connections with the hidden units, to allow for incremental processing over time. This SRN only dealt with linguistic information, but this type of network has later been implemented by Dienes et al. (1999) to account for the influence of non-linguistic domains (see Figure 4.1). Their version of the network includes the input and output units from the non-linguistic domain, as well as an encoding unit (or layer) that recodes the domain-dependent input into a common representation shared across domains.

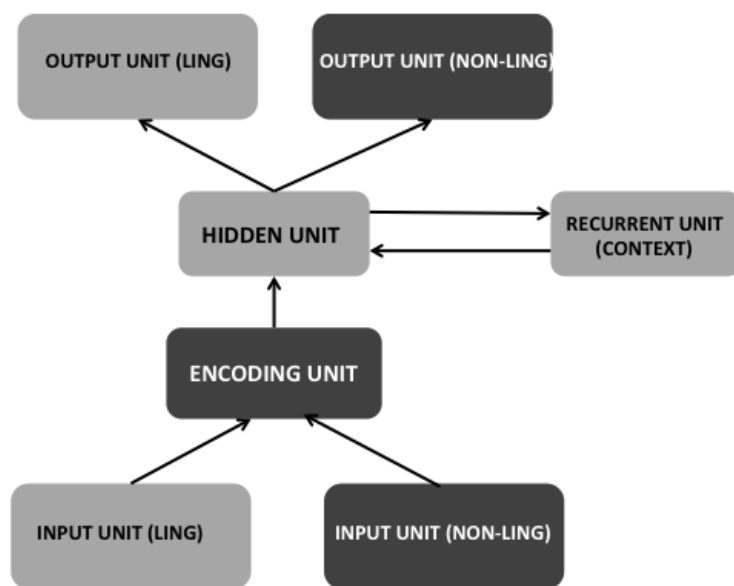


Figure 4.1.: Representation of a simple recurrent network. The light grey units are original from Elman (1990); the darker units form the implemented version by Dienes et al. (1999), p.58

A high-level (non-implemented) processing account directly addressing the interaction between a visual scene and language is the *Coordinated Interplay Account* (CIA, Knoeferle & Crocker, 2006, 2007; Knoeferle et al., 2014, see Figure 4.2). Unlike the model from Dienes et al. (1999) and other accounts on situated language comprehension (e.g., Altmann & Mirković, 2009), the CIA identifies independent representations derived from language and vision, which are nonetheless coindexed and reconciled at each point in time during comprehension. The CIA consists of three processing stages (or steps) that

although informationally dependent on each other, may "partially overlap and occur in parallel" (Knoeferle & Crocker, 2007, p. 540), namely, *sentence interpretation*, *utterance mediated attention* and *scene integration*. The first stage, *sentence interpretation* (step i), takes place incrementally. It integrates currently processed words based on prior states in order to generate new interpretations (instantiated in *int* in the account) and expectations (*ant*), which will serve the integration of following words. The second stage, *utterance mediated attention* (step i'), is a search for referents both in working memory² and in the visual scene. It is motivated by the interpretation obtained in the previous stage as well as long-term knowledge and can also reflect predictive processes (e.g., after interpreting a verb, a suitable role filler for that verb might be anticipated visually). Objects and events that are no longer present in the scene (and might be perceived as completed) may experience a decay in working memory at this point when it comes to guiding attention. A last step, *scene integration*, consists of reconciling the generated interpretation with the scene. The different stages of the CIA are enriched with a working memory (WM) component, which maintains representations of the ongoing interpretation (*int*) process, the expectations (*ant*), and the scene that is (or was recently) perceived.

The CIA can accommodate most of the phenomena encountered in situated language comprehension studies, from direct referential strategies to anticipatory eye movements, to the preference for inspecting the objects or agents from recent visual scene- or event-based representations. Initial versions of the CIA did not give specific details about the mechanisms involved during situated language comprehension and their possibly different outcomes depending on the type of information being processed. However, attempts have been made by means of ERP studies exploring the influence of the scene on syntactic disambiguation (Knoeferle, Habets, Crocker, & Münte, 2008) and, as explained in the previous section, by manipulating mismatches between the visual and the linguistic information, both in the form of thematic role relations, as well as action-verb congruency

²Working memory has been defined as a limited capacity system that "maintain[s] and manipulate[s] information over the short term" (Morrison, Conway, & Chein, 2014, p.1). WM representations are generally believed to be more accessible compared to representations from long-term memory, either because they are stored differently, or because they enjoy residual activation given their recency (Baddeley & Hitch, 1974; McElree, 2006).

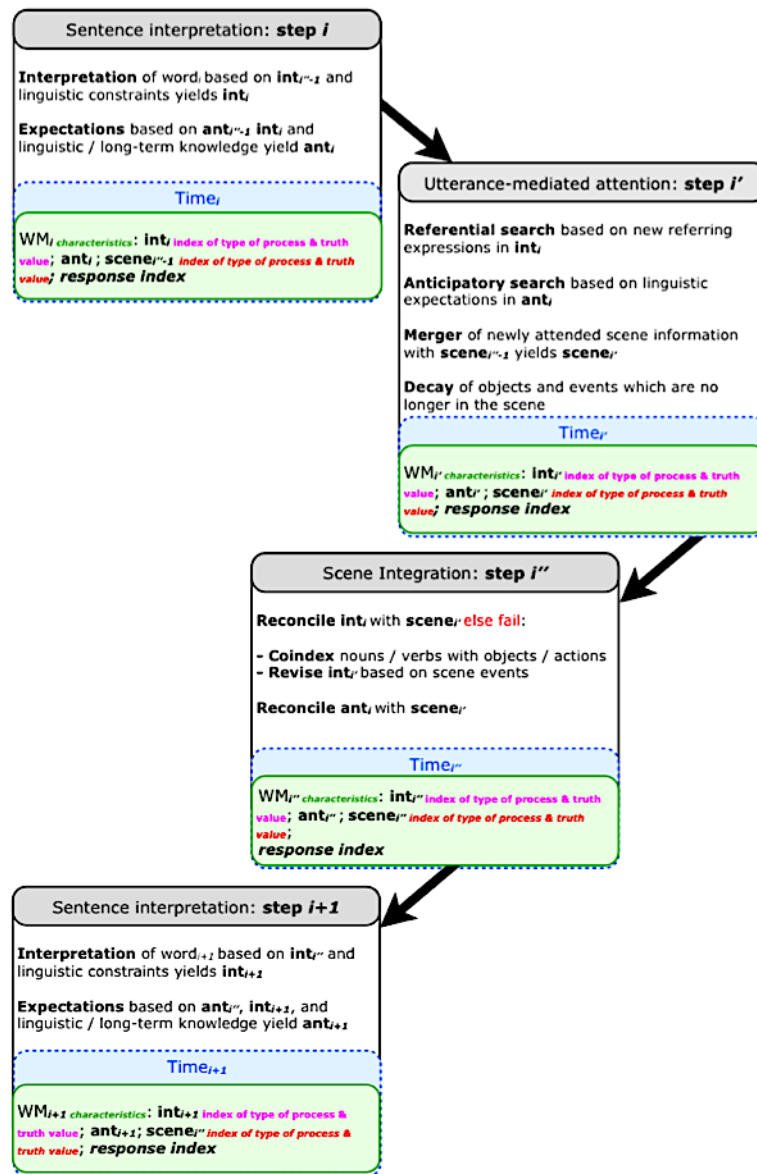


Figure 4.2.: The Revised Coordinated Interplay Account (Knoeferle et al., 2014)

(Knoeferle et al., 2014). The latest version of the CIA provides certain details on how the account can deal with situations in which there is a clash between language-based and event- or scene-based representations; it includes a system that reacts to mismatches of different kinds and establishes the *truth value* of the linguistic input (Knoeferle et al., 2014), by virtue of a *verification* parameter, something worth extending, as verification processes are "part and parcel of language comprehension" (Knoeferle et al., 2014, p.143).

As for the relative preference for recent-event representations over other cues, as Alt-

mann and Mirković (2009) argued, one important principle that underlies both the SRNs as well as accounts like the CIA is that none of the input domains (i.e., be it linguistic or non-linguistic) or the context (i.e., where the interpretations and expectations generated at one point in time are saved to be used in the processing of the upcoming word) is privileged when it comes to language processing, unless one is more predictive than the others with regards to the subsequent input. The preference for depicted events as opposed to long-term knowledge, stemming from findings like the ones summarized in section 2.3, has been contemplated in the CIA (Knoeferle & Crocker, 2007), and even implemented in computational models based on this same account (i.e., CIANet; Crocker, Knoeferle, & Mayberry, 2010; Mayberry, Crocker, & Knoeferle, 2009). Upon the encounter of a verb (e.g., ‘spy’), language-based expectations may support the anticipation of a stereotypical agent (e.g., a detective, whose stereotypical occupation is that of spying) however, a search in the scene-based representation in working memory³ will also take place, which might point towards a non-stereotypical, depicted agent (e.g., a wizard that has been seen spying upon a pilot). The competition or interaction between these two types of information is somewhat unspecified, i.e., possible modulations of the recent-event preference have not yet been fully examined. Some aspects of linguistic and long-term knowledge might have more weight than others when interfering with the representations grounded in visual events (i.e., knowledge about occupational stereotypes might be more easily discarded in the presence of prior events than, for instance, knowledge about gender stereotypes). The resulting interactions between event-based and world-knowledge representations could therefore manifest themselves in distinct ways as we incrementally understand language and anticipate possible referents.

A recent theoretical proposal has been put forward trying to extend the CIA to information about the social characteristics of the comprehender (i.e., age) and perceived

³Throughout this work we use the term ‘event-based’ instead of ‘scene-based’ representation to refer to the representation based on information provided prior to situated language comprehension, which is what the listener has to reconcile with language. Note that in our experiments, unlike previous studies, the target scene (i.e., the concurrent scene during language comprehension) differs considerably from the configuration of prior events (e.g., see Figure 5.1), and we don’t want the term ‘scene-based’ be mistakenly associated with such a scene.

social (i.e., emotional) cues (the social Coordinated Interplay Account; Münster, 2016). In this version of the CIA, the properties of the comprehender (i.e., *ProCom*, age in their example), and expectations (which as in the CIA are instantiated in *ant*) based on (long-term) social knowledge are included at every stage of processing. In the sCIA, the probabilistic weight of *ant* is instantiated via a subscript p (range 0 to 1). Several contextual factors could in principle affect the weight of this subscript, one of them being *ProCom*. For instance, in a sentence like *Den Marienkäfer kitzelt vergnügt der Kater* ('The ladybug_{obj} tickles happily the cat_{subj}'), when encountering the positive-biased adverb *vergnügt* ('happily') during sentence comprehension, older adults and children's expectations regarding the next thematic role filler - the cat - might be more strongly weighted than those of younger adults which, according to previous literature (Langeslag & van Strien, 2009; Reed & Carstensen, 2012), have a negativity bias. Although highly important for the comprehension process and its interaction with the visual world, the greater or lesser use of social information in this version of the account relies heavily on the properties of the comprehender, rather than on the properties of the entities involved in the events described (i.e., the more or less stable characteristics of the entities mentioned in a sentence and how familiar those characteristics are to the human comprehender). It would therefore be interesting to see how the idea of the probabilistically weighted anticipation parameter extends to other information sources (e.g., gender knowledge) in situated language comprehension, as well as to the reconciliation of event-based and language-based representations regarding those sources.

In the previous chapter we have seen how information about gender has a strong influence during comprehension in several discourse (and some visual) contexts. One recent model for language comprehension that puts a particular emphasis on gender features in the processing of lexical stimuli is the Cognitive-Cultural Model (or CC Model; Bojarska, 2013). In this model, the main claim is that humans are trained to pay attention to gender information in language, as such feature is virtually never absent when referring to people (the author's line of reasoning is that it is therefore fairly stable and aids comprehension processes). Direct semantic cues (e.g., linguistically specifying a nurse's gender as

male, or providing visual cues that convey gender information) may be given priority during comprehension in cases in which cognitive resources are available. However, if overt semantic information about gender is missing, or when lacking cognitive resources, the comprehender may take a less conscious, more automatic route, and they will try to infer gender via so-called *extrasemantic* cues (i.e., inferential cues based on long-term knowledge, like stereotypes). Also, even if direct semantic cues are present, some extrasemantic factors may either contribute to comprehension, slow down processing (i.e., in stereotypically incongruent situations) or even cause misunderstandings, i.e., the different sources may need to undergo *negotiation*. Although the model has a narrow focus (it focuses on certain types of lexical stimuli like *nurse* and morphological aspects like masculine plurals intended as generic) and was not conceived specifically its application in visually situated language comprehension, it backs up the idea of gender as a relevant feature to investigate during language processing at different levels, both perceptual, event-based (or semantically direct) and world-knowledge related.

We can identify some aspects in these accounts and models for situated language comprehension that could be implemented or at least addressed. These relate to the relative strength of visually grounded representations vs. long-term knowledge when generating expectations as a function of the type of information processed during comprehension, as well as the effects of incongruences between language and visually grounded information. Additionally, the mechanisms or processes that might be involved when reconciling (or failing to reconcile) representations stemming from visual and linguistic domains may need further evidence from experimental settings in order to refine existing models and accounts. Extant theories (e.g., the Monitoring Theory; Kolk et al., 2003; Vissers et al., 2008) and neurocomputational models (Crocker et al., 2010) have only partially addressed some of the phenomena that can be involved when the comprehender processes language in visual contexts (e.g., syntactic disambiguation processes, or the reconciliation of spatial, verb or thematic role information in language with a scene). These observations further motivated our research.

5 | Gendered expectations: mismatches in situated language comprehension

The main goal in our first two experiments was to check the robustness of the recent-event preference applied to gender for the first time. We wanted to see how participants established visual (anticipatory) reference with gendered agents as a function of prior gender and action cues during comprehension. Moreover, we wanted to examine how this process could be affected by different manipulations that pertain to the reconciliation between event-based and language-based representations. Participants inspected videos of a pair of hands performing an action¹ (e.g., female hands baking a cake), and then they saw a visual scene with a female and a male face. We measured participants' eye movements during the comprehension of non-canonical German object-verb-subject (OVS; *Den Kuchen backt gleich Susanna*, 'The cake bakes soon Susanna') sentences while looking at the pictures of the two agents' faces, who could be potentially mentioned at sentence-final position (see Figure 5.1).

As we explained at the beginning of the introduction, we manipulated the referential congruence between prior events (i.e., the videos) and the sentence. Sentences either matched prior events, or they could contain some sort of mismatch, i.e., mismatches between the action seen in prior events and the action described by the sentence (*action-verb match*, Experiment 1)² or mismatches between the gender of the agent in the video

¹See the Materials section for further details on the use of hands as the main visual gender cue in the current experiments.

²By "the action described by the sentence" we mean the verb phrase, that is, the object-verb word combination. The factor manipulating the match between prior visual actions and the verb-phrase will be called *action-verb match* throughout the experiments.

(conveyed by the hands) and the gender of the final subject (i.e., a proper noun) in the sentence (*hand-subject match*, Experiment 2). Furthermore, we also manipulated whether the described actions matched or mismatched stereotypically with the gender of the agent seen in prior events (conveyed by the hands; *stereotypicality match*). Participants' task was to verify via button press whether the sentence matched the video they just saw ('yes' or 'no').

5.1 | Experiments 1 and 2

In order to be able to explore gender-based agent expectations during comprehension in the current experiments, we depended on the participants' successful association between the dimorphic gender cues from the agent in prior events (i.e., the hands on the video) and the faces of the later display (i.e., the *target scene*) during comprehension (see Figure 5.1). Biological gender or sex categorization as such has been claimed to be fairly feasible in adults (Martin & Macrae, 2007; Stangor et al., 1992; Wild et al., 2000), even when it comes to subtle visual cues like the appearance of hands, i.e., their size, shape and texture (e.g., whether they are big or small, thin or thick, smooth or rough; Gaetano, van der Zwan, Blair, & Brooks, 2014). Because of the documented robustness of the recent-event preference (i.e., the preference for anticipating agents and objects that were part of prior events), we expected it to affect the visual anticipation of the agents' faces accordingly during comprehension. Such preference should be reflected in participants predominantly looking at the face whose gender features match those from the hands seen in prior events. We called this face the *target agent*, while the face from the opposite gender was labelled as the *competitor agent*.

If gender categorization took place successfully (therefore allowing for anticipatory and referential strategies) and we were to replicate the preference for event-based information (i.e., recent-event preference) over stereotype knowledge during sentence comprehension (e.g., Knoeferle & Crocker, 2007), we should see a greater proportion of inspections

to the target agent photograph (e.g., the female face when the hands in the video belonged to a woman) relative to the other character (or competitor, a man in this case) early on during the incremental comprehension of the sentence, regardless of whether the action described was stereotypically congruous or incongruous with the gender of the target agent.

We further expected that a manipulation of congruence between the prior events and the subsequent linguistic input would influence participants' attention in the concurrent, target scene during comprehension. In Experiment 1, our experimental items contained mismatches between prior events and action information from the sentence (i.e., mismatches were at the initial verb-phrase, see Table 5.1), but the gender implied by the final noun still matched the gender of the hands in prior events. Although in this experiment some of the fillers did contain complete mismatches (i.e., action and final noun), participants could opt for inspecting the target agent over the competitor regardless of the mismatching linguistic content. However, prior research and the (referential) anchoring hypothesis (Dumitru et al., 2013; Tversky & Kahneman, 1973) suggest that mismatches between prior events and language encountered at the initial part of the sentence should modulate the probability of anticipating the following gendered agent (i.e., reliability to anticipate the upcoming agent based on prior visual information should be affected). Therefore, mismatches (by virtue of describing actions different from those in prior events) should reduce the preference for the target agent, drive this preference to an at-chance level or even shift the attention from the target to the competitor agent. As for the manipulation of video-sentence matches at the final noun region (as it is the case in Experiment 2, see Table 5.2), we expected to see how incremental comprehension was affected at its final point of verification, by tapping more strictly referential processes. At this final region, we expected to find more consistent referential strategies taking place, i.e., looks at the appropriate agent, target or competitor, as implied by the final proper name Tanenhaus et al. (1995), although mismatches between gender cues and the final proper name (referring to the competitor) could lead to a reduction or a delay in attention.

Differences between the mismatching regions from Experiment 1 to Experiment 2 could as well experience time differences in their emergence relative to their onset. Both verb(-phrase) information and subjecthood are central for thematic (i.e., agent) role assignment. However, video-sentence mismatches between prior events and the described actions (Experiment 1) involve the consideration of two pieces of linguistic information (i.e., object and verb), while subject mismatches only involved one word region (i.e., the final noun). Moreover, one could say that the two types of mismatches take place at different points in the comprehension process: action-verb mismatches happen at the beginning of the sentence, which as aforementioned means that attention towards agents at this point is anticipatory in nature (i.e., before the agent's name is revealed; recall that our sentences have an OVS word order) and reactions towards mismatches may slow down. On the other hand, mismatches at the subject happen at the end of the sentence, as the agent is revealed. Because it is the point where thematic resolution takes places unequivocally, more immediate effects might be elicited relative to the onset of this region as compared to Experiment 1.

When it comes to the role of gender stereotypicality (i.e., our second manipulation), we reasoned that if gender stereotypes were used during the anticipation of agents in situated language comprehension, these could potentially modulate our visually grounded expectations (e.g., female agents cued by female hands in prior events would be preferred over male agents to a greater extent when the sentence described stereotypically female events compared to male events). Effects of stereotypicality could potentially also be seen in conditions where the reliability to inspect the target agent can be significantly reduced (i.e., in video-sentence mismatches). For instance, in Experiment 1, when participants cannot rely on the information from prior events to anticipate the agent of the sentence (action-verb mismatches), the action described might still favor the target agent whenever this is stereotypically matching, as compared to the cases when it is stereotypically mismatching. In the latter case, if stereotypicality information was strong enough to guide our expectations, participants could even resort to looking at the agent of the opposite gender (i.e., the competitor). For instance, female hands appeared in the video

performing an action different from that described in the sentence, but the sentence is about baking a cake, which is still stereotypically congruent with a female agent. In this situation participants might still prefer to look at the female face (the target agent) over the male (the competitor) to a greater extent compared to when the sentence is about building a model. If the sentence is about building a model, participants might even opt to look at the male face (the competitor agent) over the female (the target).

5.1.1 | Methods and Design

Participants 32 participants in Experiment 1 (16 females, 19-32 years, $M=26.37$) and another 32 in Experiment 2 (16 females, 19-32 years, $M=25.8$) took part, all German native speakers with normal or corrected to normal vision. They all gave informed consent before starting the experiments and received 6 Euro for participation.

Materials Using E-Prime 2.0 (Psychology Software Tools Inc.), we prepared a list of 104 verbally described actions to assess their gender stereotypicality. Actions were initially assigned an orientative label as either “female”, “male” or “neutral” by the experimenter. A group of participants that did not take part in the eye-tracking experiments ($N=20$, 10 female, mean age 26.05) evaluated these actions for gender stereotypicality prior to the eye-tracking experiment. Descriptions were presented in written form in an object-verb manner in the middle of the screen (e.g., *Den Kuchen backen*, 'Baking a cake'; *Das Modell bauen*, 'Building a model'). Participants' ratings were on a bipolar 7-point scale; they were asked to respond as fast as possible. The scale was counterbalanced across participants, e.g., 1 would stand for “very typically female” while 7 would be “very typically male” or vice versa; 4 would stand for “typical for both or neither”. After data collection, the counterbalanced scales were readjusted so that 1 would stand for “very typically female” and 7 would be “very typically male”. Based on the ratings, any action initially labelled as “female” or “male” with a score between 3 and 5 was moved into the “neutral” label. The mean scores for “female” and “male” actions were 2.26 (SE:1.28) and 5.77 (SE:1.17)

respectively. Pairwise comparisons including the “neutral” actions (M:3.86, SE:1.32) and using the False Discovery Rate (“fdr”), revealed significant differences among all three action-type groups ($ps < .01$).

Based on the rating results, we selected the top 32 stereotypically female and the top 32 stereotypically male action sentences as our experimental stimuli (see Appendix A.1), and selected 16 “neutral” actions (rated around 4) to be part of the fillers. For the experimental materials, we recorded 128 action videos. Videos were close-ups of pairs of hands (each action was videotaped once with a female and once with a male actor) acting upon objects on the surface of a table, from an external perspective and centered on screen. The use of hands as an index for gender in the same visual environment was motivated by two main reasons: a) to keep gender cues as minimal as possible (yet recognizable for gender categorization to take place, e.g., Gaetano et al., 2014) and b) to keep the visual setting where the events (actions) take place as similar across items as possible (for visual materials, see Appendix A.3).

Fillers (N=68) included trials similar to the experimental ones but with no stereotypical valence (i.e., neutral actions), videos with two pairs of hands engaged in an action followed by sentences with dative constructions (e.g., *Susanna reaches the boy the pencil*) and pictures of objects and scenes alone (i.e., no hands) followed by a range of sentence structures (e.g., *The chair is blue*; see Appendix A.4). Like the experimental trials, half of the fillers contained video-sentence mismatches of some sort (final name, described action, color, shape, etc.). For the target scenes in trials where videos of hands were used (i.e., the visual stimuli presented during sentence presentation), we took six close-up photographs of male and female faces (two pairs were used for the experimental items while the filler trials included an additional pair).

From all these materials we created 32 experimental items consisting of video pairs (one stereotypically female and one stereotypically male action video) and their corresponding German sentence pairs with a non-canonical German object-verb-subject (OVS)

structure³. We used this structure in order to monitor participants' expectations regarding the upcoming subject (i.e. the agent). All sentences were recorded by a female German speaker with neutral intonation. We paired within an item (see Table 5.1) sentences with a similar number of syllables per word, and the onsets of the different constituents were then synchronized (see Appendix A.2).

We implemented the eye-tracking experiment using Experiment Builder (SR Research). In Experiment 1 we manipulated two factors. The first was action-verb match (the action described in the sentence as expressed by object-verb combinations i.e., the verb phrase, either matched or mismatched the action in the video); the second factor was stereotypicality match (the action described by the sentence either matched or mismatched stereotypically with the gender implied by the hands in the video). For instance, a congruous condition would feature female hands in the video and a sentence about a stereotypically female action; an incongruous example included female hands in the video and a sentence about a stereotypically male action. The sentence-final subject in the experimental items always matched the agent of the video in terms of gender in this experiment. Crossing these factors yielded 4 conditions (see Table 5.1)⁴, which were counterbalanced across experimental lists in a Latin Square manner. As for the target agent's position, there was a version of each list with the target to the right and another version with the target to the left. Word order and the use of the postverbal adverb *gleich* were constant across conditions.

For Experiment 2, the verb-(phrase) always matched the actions from the videos.

³We checked the relative frequencies per million words (pMW) of both object nouns and verbs in the present tense, third person singular form (as they were used in the sentences) in the experimental items using the COSMAS II database (web version: <http://www.ids-mannheim.de/cosmas2/web-app/>, checked during April, 2014). Pairwise t-tests were run to compare the frequency of the words between the stereotypically female and male action sentences. For nouns, there was a marginally significant difference between noun types ($M_f=6.11$, $SD=8.15$; $M_m=10.06$, $SD=12.6$), $t(31)=1.83$, $p=.08$. For verbs, there was also a marginally significant difference between types of verbs ($M_f=4.12$, $SD=9.14$; $M_m=12.75$, $SD=28.07$), $t(31)=1.91$, $p=.065$. As each experimental item contained one stereotypically female and one male action sentence, any potential confound pertaining to frequency was controlled for.

⁴Given that the actors of both genders (female and male) were recorded performing the two action types (stereotypically female vs. male), and action-verb match was manipulated (match vs. mismatch), this gave rise to eight conditions. As we only focused on action-verb match and stereotypicality match, the eight conditions were collapsed into four (i.e., across both genders) for analysis. For the sake of simplicity, the tables with the experimental items only show a female example of the conditions.

We instead manipulated the match between the gender cued by the hands in the video and the gender of the sentential subject (i.e., hand gender - subject gender match). The stereotypicality match factor from Experiment 1 was retained (see Table 5.2).

Table 5.1.: Example item for Experiment 1

Video	Sentence				Action-verb match	stereo- match
Female hands baking a cake	Den Kuchen _{NP1} <i>the cake</i> (obj)	backt _V <i>bakes</i>	gleich _{ADV} <i>soon</i>	Susanna _{NP2} <i>Susanna</i> (subj)	yes	yes
Female hands building a model	Das Modell _{NP1} <i>the model</i> (obj)	baut _V <i>builds</i>	gleich _{ADV} <i>soon</i>	Susanna _{NP2} <i>Susanna</i> (subj)	yes	no
Female hands building a model	Den Kuchen _{NP1}	backt _V	gleich _{ADV}	Susanna _{NP2}	no	yes
Female hands baking a cake	Das Modell _{NP1}	baut _V	gleich _{ADV}	Susanna _{NP2}	no	no

Table 5.2.: Example item for Experiment 2

Video	Sentence				Hand-subj. gend. match	stereo- match
Female hands baking a cake	Den Kuchen _{NP1} <i>the cake</i> (obj)	backt _V <i>bakes</i>	gleich _{ADV} <i>soon</i>	Susanna _{NP2} <i>Susanna</i> (subj)	yes	yes
Female hands building a model	Das Modell _{NP1} <i>the model</i> (obj)	baut _V <i>builds</i>	gleich _{ADV} <i>soon</i>	Susanna _{NP2} <i>Susanna</i> (subj)	yes	no
Male hands building a model	Das Modell _{NP1}	baut _V	gleich _{ADV}	Susanna _{NP2}	no	yes
Male hands baking a cake	Den Kuchen _{NP1}	backt _V	gleich _{ADV}	Susanna _{NP2}	no	no

Procedure Upon entering the lab, participants had to fill out a form (their name, age, studies, etc.) and sign the consent form. The experimenter then told participants that they were going to take part in a video-sentence verification study. They were also asked to pay attention to the videos as well as the static pictures during the experiment, which would take about an hour (100 trials). An EyeLink® 1000 Desktop Mounted Eye-Tracker (SR Research) recorded participants' eye movements with a sampling rate of 1000Hz. Viewing was binocular but only the right eye was tracked. A chinrest bar was provided for each participant to minimize head movement. In both Experiments 1 and 2 participants completed 10 practice trials including feedback before starting the experiment. Trials started with a video of the action (or a picture, as in some of the filler trials) for 3500 ms, then the video stopped and the final frame (displaying both the hands in resting position and the object) stayed for another 1500 ms. After that a cross appeared for 1000 ms and then a target screen was shown, with one picture of a female face and another of a male face along the horizontal axis. After a 1500 ms preview time, the sentence was presented and eye movements to the pictures recorded. Participants verified whether the video they just saw matched the sentence that they listened to ("yes" or "no") via button press (Cedrus RB 834). The position of the response buttons was counterbalanced across participants (see Figure 5.1). Once the experiment finished, participants were asked to respond to some additional questions on their experience, e.g., whether they found anything strange or surprising during the experiment, whether they figured out the purpose of the study, etc. At this point the experimenter could tell participants what they were tested on if participants asked for it.

5.1.2 | Analysis and Results

Analysis Reaction times were calculated from sentence onset and computed per condition (response times that were more than 2 standard deviations away from the mean were removed), and then subjected to by-subjects (F_1) and by-items ANOVAs (F_2). Accuracy was analyzed using Generalized Linear Mixed Models (suitable for binomial data)

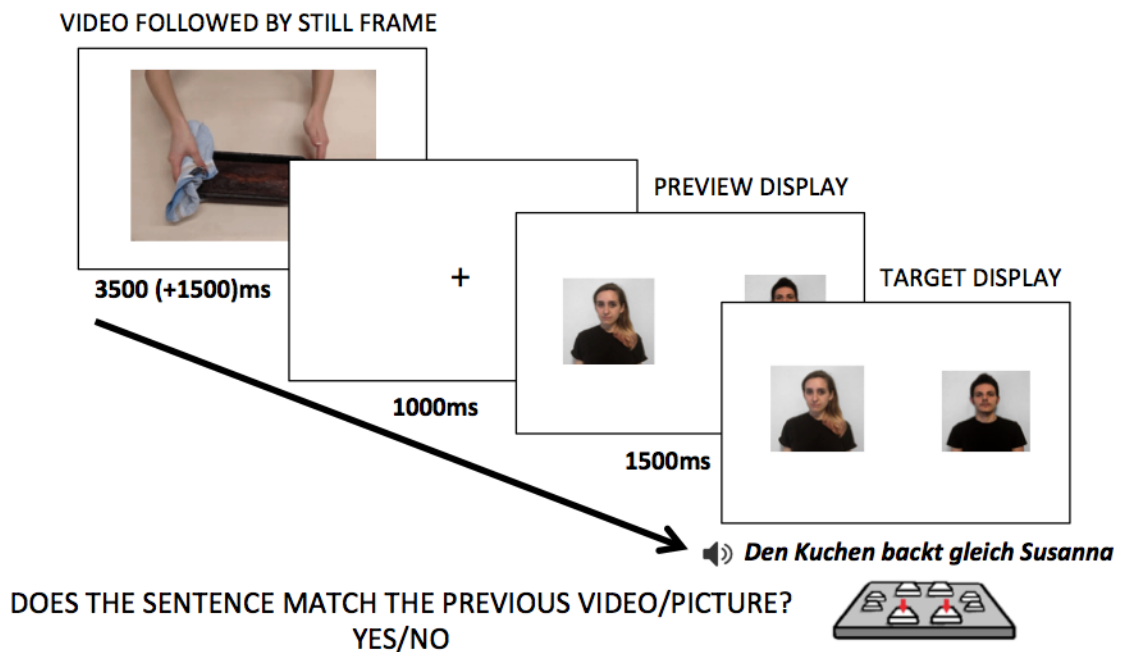


Figure 5.1.: Example of an experimental trial, Experiments 1 and 2.

using R (R Core Team, 2016, see Appendix B.1). For the eye-tracking data, we divided each experimental sentence into four time regions, (the object noun phrase: NP1, the verb: V; the adverb: ADV; and the subject noun phrase: NP2). Each region extended from its onset to the onset of the next region except for NP2, which ended at sentence offset. Fixations that started before the onset of NP1 were taken out of the analysis, as they could not be ascribed to linguistic processing⁵. Additionally, time windows were shifted forward by 200ms, to account for saccadic planning (Ferreira, Foucart, & Engelhardt, 2013; Matin, Shao, & Boff, 1993). Because looks to one of the characters implied fewer looks to the other character in the visual scene, we computed the mean log-gaze probability ratios for each separate sentence region to measure the bias of inspecting the target agent (i.e., the face which matched in gender the hands in the previous video) over the competitor agent (the other face; $\ln(P(\text{target agent})/P(\text{competitor}))$). Values above zero reflect a target agent preference, while values below zero represent a preference for the competitor. These scores are suitable for parametric tests such as ANOVAs (Arai, Van Gompel, & Scheepers, 2007; Knoeferle, Carminati, et al., 2011). We calculated mean

⁵We did not remove fixations starting before NP1 when plotting the time-course graphs.

log-probability ratios per region by subjects (F_1) and by items (F_2), which we subjected to repeated measures ANOVA analyses, with video-sentence match (action-verb match in Experiment 1 and hand-subject gender match in Experiment 2) and stereotypicality match as fixed effects. As we controlled for the gender of participants in our eye-tracking experiments (we tested the same amount of female and male participants), we included gender as a between-subjects factor for F_1 and as a within-subjects factor for F_2 (e.g. Carminati & Knoeferle, 2013; Jegerski, VanPatten, & Keating, 2016)⁶. We reported both analyses together with their effect size (partial eta squared). For the time-course graphs, we plotted the gaze probability ratios in successive 20 ms time slots from the beginning of the sentence. Missing and incorrect responses were excluded from both the eye movement and response-time analyses. Marginally significant as well as non-significant results will be reported only when relevant for the purposes of this work.

Results Experiment 1 *Accuracy and response times*: Participants responded correctly on 92% of the trials, more accurately to action-verb mismatches than matches (see Appendix B, Table B.1). Reaction times were significantly shorter for action-verb mismatches ($M=3515.58$, $SD=66.88$) than matches ($M=4640$, $SD=23.5$), $F_1(1, 30)=47.07$, $p<.001$, $\eta^2=.611$; $F_2(1,31)=370.06$, $p<.001$, $\eta^2=.923$. We also found an interaction between gender and action-verb match, $F_1(1, 30)=14.58$, $p<.001$, $\eta^2=.327$; $F_2(1,31)=237.19$, $p<.001$, $\eta^2=.738$. The interaction was driven by male participants, who responded faster than female participants to action-verb mismatches.

Eye-movement analysis: The time-course graph from Experiment 1 (see Figure 5.2) shows the attentional behaviour during sentence comprehension across participants from the beginning of the sentence per condition⁷. From the graph, we can infer that log gaze

⁶For reaction times and the eye movements, we also conducted Linear Mixed Effects analyses on the data as an alternative analysis, which is to be found in Appendixes B.2 and B.4, respectively.

⁷The time-course graphs are based on the mean onsets of word regions; mean onsets (and standard deviations) for the verb, adverb and final noun were 1401ms ($SD:155.72$), 2566 ($SD:181.09$) and 3581 ($SD:185.57$), respectively. For that reason, the time-course graph does not reflect the exact eye-movement behaviour time-locked for each item and can therefore only be taken as visual aid. In order to perform inferential statistical analyses, we conducted the time-region analyses with the log-gaze probability ratios adjusted for the word onsets of each item.

probability ratios were positive in all time regions, suggesting a target agent preference over the competitor during sentence comprehension (for statistical tests on the grand means per word region, see Appendix B.3). However, for action-verb mismatches, participants started to look away from the target agent during the verb region. Therefore, there seems to be a slight delay in the reaction towards mismatches relative to their onset; this difference between matching and mismatching conditions seems to be maintained during the final regions, even once the final subject (i.e., proper name, which in Experiment 1 refers to the target agent) is revealed.

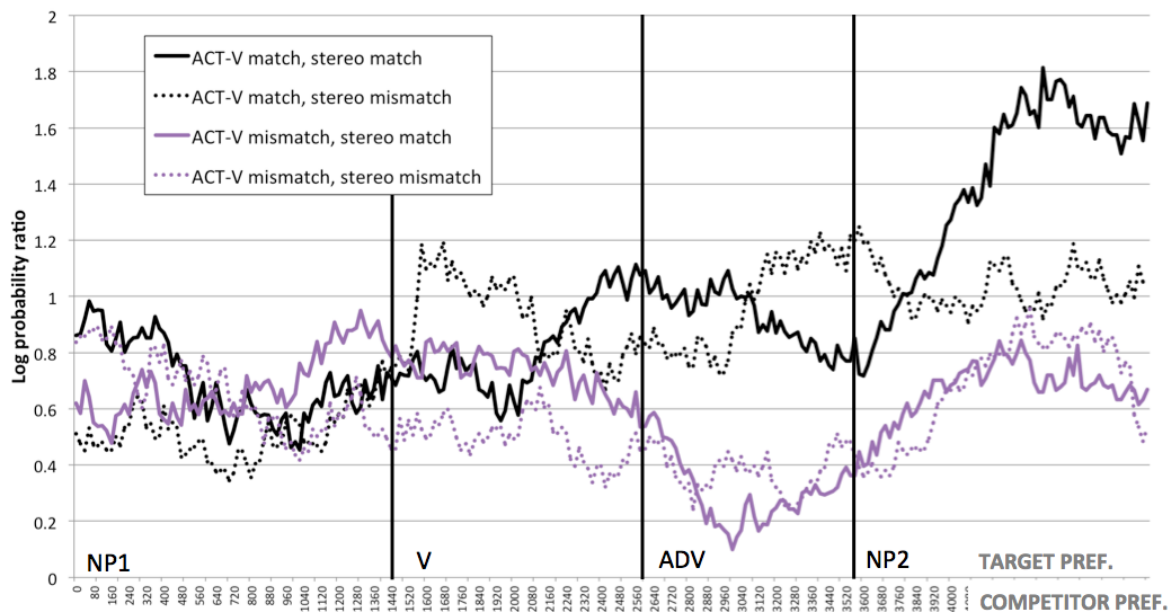


Figure 5.2.: Time-course graph for Experiment 1.

In the region analyses, no significant effects of the independent variables emerged for the NP1 region. The first mismatch effect emerged at the verb region, with a main effect of action-verb match marginal in the by subjects analysis and significant by items, $F_1(1, 30)=4.017$, $p=.054$, $\eta^2=.118$; $F_2(1,31)=4.59$, $p<.05$, $\eta^2=.129$ (see Figure 5.3). No interaction between action-verb match and stereotypicality match was found (see Figure 5.4). Participants directed more looks to the target agent for action-verb matches compared to mismatches. A marginal interaction between gender and stereotypicality did also emerge in the by items analysis, $F_1(1, 30)=2.18$, $p=.1$, $\eta^2=.068$; $F_2(1,31)=3.65$,

$p=.065$, $\eta^2=.105$ ⁸. Female participants seemed to look at the target agent to a greater extent in the stereotypically incongruent condition compared to the congruent condition, while the male participants showed the opposite pattern.

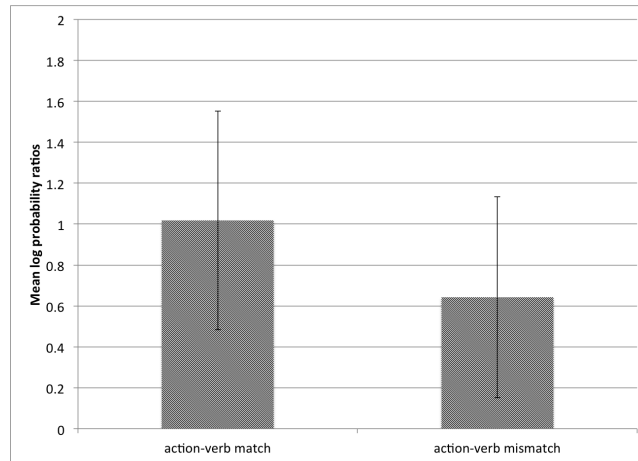


Figure 5.3.: By-subject mean log-probability ratios at the verb region, Experiment 1 (error bars indicate 95% confidence intervals).

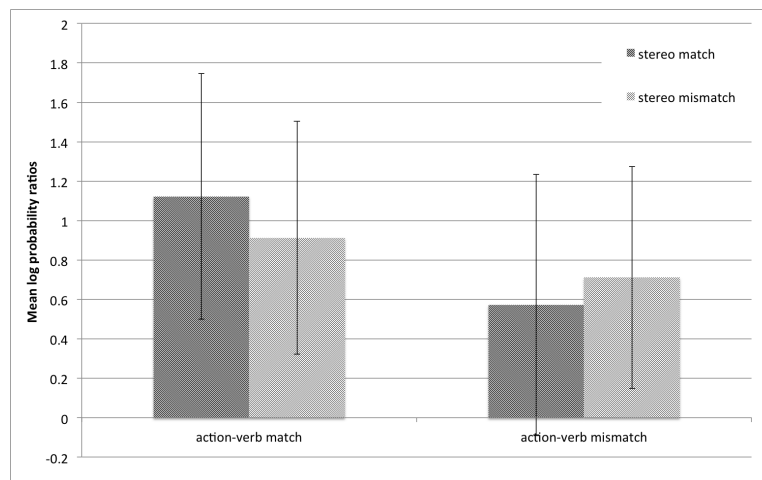


Figure 5.4.: By-subject mean log-probability ratios at the verb region per condition, Experiment 1 (error bars indicate 95% confidence intervals).

The effect of action-verb match persisted post-verbally, both at the adverb, $F_1(1, 30)=20,75$, $p<.001$, $\eta^2=.409$; $F_2(1,31)=22,81$, $p<.001$, $\eta^2=.424$ (see Figure 5.5), and NP2 regions, $F_1(1, 30)=16,59$, $p<.001$, $\eta^2=.356$; $F_2(1,31)=19,09$, $p<.001$, $\eta^2=.381$ (Figure 5.6).

⁸This interaction appeared as significant in the mixed models analysis (see Appendix B, Table B.9).

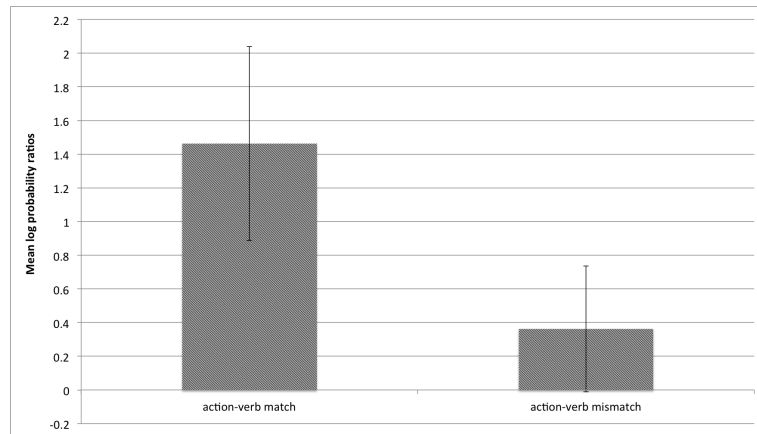


Figure 5.5.: By-subject mean log-probability ratios at the adverb region, Experiment 1 (error bars indicate 95% confidence intervals).

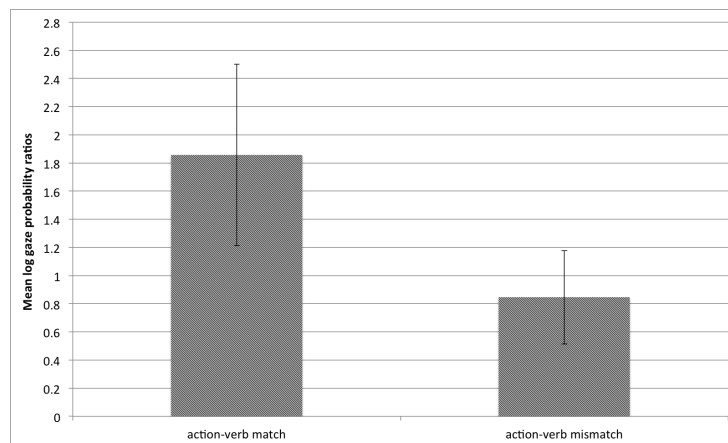


Figure 5.6.: By-subject mean log-probability ratios at the final noun (NP2) region, Experiment 1 (error bars indicate 95% confidence intervals).

Results Experiment 2 *Accuracy and response times:* Participants responded correctly on 91% of the trials. No significant effects between conditions were found for either accuracy (see Appendix B.1, Table B.2) or reaction times.

Eye-movement analysis: Like in Experiment 1, time-course graphs displayed positive values throughout the sentence (see Figure 5.7). However, the divergence between hand-subject gender matching and mismatching conditions is apparent in the final noun (NP2) region. When the final subject mismatched in gender with the hands from prior events, there was a clear decrease in preference for the target agent, yet within positive values.

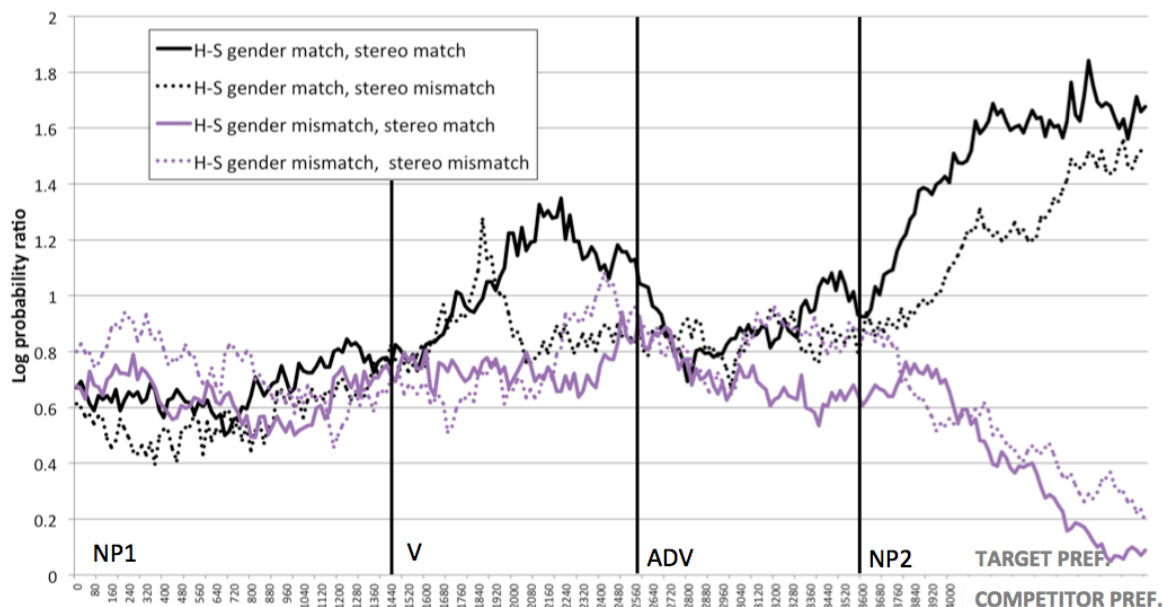


Figure 5.7.: Time-course graph for Experiment 2.

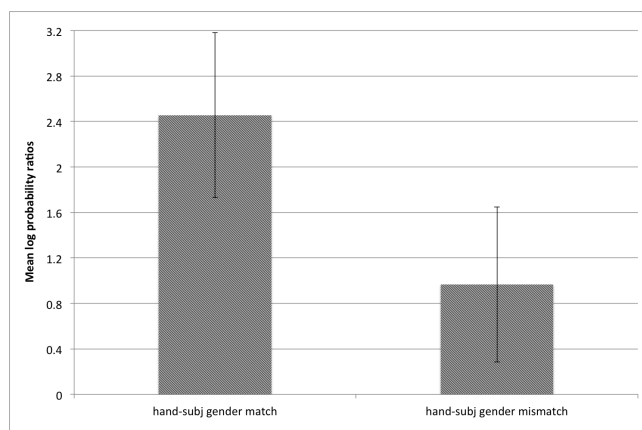


Figure 5.8.: By-subject mean log-probability ratios at the final noun (NP2) region, Experiment 2 (error bars indicate 95% confidence intervals).

We only found effects of our experimental manipulations at the final, NP2 region. We found an effect of gender in the by items analysis, $F_2(1,31)=19,99$ $p<.001$, $\eta^2=.392$, but no interaction with the other factors. Female participants inspected the target agent to a greater extent compared to male participants. There was a main effect of hand – subject gender match, $F_1(1,30)=31,56$, $p<.001$, $\eta^2=.513$; $F_2(1,31)=48,27$, $p<.001$, $\eta^2=.609$ (see Figure 5.8). Participants directed more looks to the target agent in hand-subject gender matches compared to mismatches.

5.1.3 | Discussion

In these two eye-tracking experiments, we assessed the generality of the preference for prior event-based representations in situated language comprehension by using another visual cue that had not yet been tested. We used visual gender and action cues from action videos, and pitted these against linguistic representations and gender-stereotype knowledge. We varied the video-sentence referential match from action-verb (Experiment 1) to hands-subject gender match (Experiment 2) and we furthermore assessed in both experiments the influence of gender stereotypical match between the sentence and the agent whose gender features were cued in prior events (i.e., the hands). Contrary to our predictions, participants responded faster and were more accurate for action-verb mismatches than matches (Experiment 1). This might be due to judgement facilitation for utterly mismatching verbal information compared to action-verb matches; such results have also been seen in studies using a similar paradigm (Dumitru et al., 2013; Münster et al., 2014). Unlike in Experiment 1, we did not find effects in the reaction times and accuracy results in Experiment 2. It is possible that the reliable mismatch effect in reaction times for Experiment 1 came about because in mismatching conditions, participants could detect (and thus respond to) the mismatch as early as in the first noun. For the matching conditions, participants had to wait until the end of the sentence to verify the match, as mismatches could still be found in the final subject region (although this only happened in some of the filler items for this experiment). In Experiment 2, as mismatches for the experimental items were only present in sentence-final position (i.e., hand-subject match was the main manipulation), both responses for matches and mismatches could only be given late, perhaps eliminating significant response-time differences.

Surprisingly, in Experiment 1, male participants were faster at responding to action-verb mismatches than female participants. It seems that given that participants were free to respond whenever they wanted during the sentence, participants of different genders adopted different strategies; while male participants opted for responding as soon as the

mismatch was encountered, female participants seemed to prefer to wait until the end of the sentence to make their response.

Regarding the eye movements, as the recent-event preference would predict, participants preferred to look at the target agent (i.e., the face whose gender matched that of the hands in the video) relative to the competitor throughout the sentence. This preference emerged in both experiments and regardless of the stereotypical content of the sentence, which suggests that prior visual gender cues were sufficient and strong enough to allow participants to visually anticipate the gender-matching agent during the incremental comprehension of the sentence. Importantly, video-sentence mismatches modulated this preference. Participants tended to look away from the target agent when a mismatch in language was encountered: both when the (sentence initial) verb-phrase or the final subject (i.e., the proper name) mismatched prior events, participants' preference for the target agent was affected. This would be in line with the referential anchoring hypothesis as explained by Dumitru et al. (2013): when linguistic information mismatches prior events, the reliability of the target agent as the entity to be mentioned decreases. Mismatch effects emerged with a slight delay relative to the onset of its appearance in action-verb mismatches (Experiment 1) and rapidly at the final subject region for hands-subject gender mismatches (Experiment 2). The slight delay in the emergence of action-verb mismatch effects in Experiment 1 (at the verb region rather than at the first noun where a mismatch could already be detected) could in principle indicate processes of integrating the non-canonical object with the verb (i.e., a compositional process) while reconciling both object and verb with the representation of the previous event, leading to a reconsideration of the expectations that were first generated. Note that delays in visual attention (albeit not in a mismatch design) have been reported in studies using the same OVS word order (Kamide, Scheepers, & Altmann, 2003). Perhaps the non-canonical structure was partly responsible for the delays in visual attention and the action-verb match effect in Experiment 1. Mismatch effects at the final subject region (Experiment 2), unlike those of action-verb mismatches (Experiments 1 and 3), seemed more immediate relative to the onset of the mismatching region, arguably because the former type of mismatch involves

the visual referent directly (i.e., it is more strictly referential). However, together with looking away from the agent from event-based representations (i.e., the target agent), a referential hypothesis would in addition predict a rapid shift of attention towards the competitor agent (e.g., if male hands were in prior events and the final noun is Susanna, it would be the female face) closely time-locked to its referring word (Tanenhaus et al., 1995), which was not the case. This suggests that even for mismatching conditions, discarded/residual representations of the recent event are kept in working memory, which interfere with the referential biases of the comprehender.

In the current experiments, we found little influence of gender stereotypes; only in Experiment 1 did we find an interaction between participants' gender and stereotypicality match at the verb, indicating that female participants looked at the target agent to a greater extent in stereotypically mismatching conditions compared to matching conditions, and that male participants showed the opposite pattern. This result might have come about because given the disadvantageous outcomes that gender stereotypes tend to have for women in society, they might try to confront them, even if unconsciously, to a greater extent than men, by reversing a more predictable situation in which stereotypes are preferred when emphasized (de Lemus et al., 2013).

It is important to note that from our experimental design, another outcome could have been possible if gender-stereotype knowledge had been used already when inspecting prior events. Participants could have seen our videos showing events as already portraying internally congruent or incongruent scenarios (e.g., female hands baking a cake in the video would be considered a stereotypically congruent visual event; if female hands were seen building a model, the event could be considered already internally incongruent). That could have also affected our results (e.g., people might have been prone to anticipating the female agent to a greater extent if the prior video was internally congruent compared to when it was not congruent). However, no such thing happened, as no stereotypicality match effects which could be ascribed to this influence emerged in any of the sentence regions. Participants seemed to rather rely more on the *directly verifiable* (i.e.,

referential, non-inferential) aspects (i.e., action and gender cues) between the event-based and language-based representations.

Some caveats are in order in light of the present experimental design. Time-course graphs depicting the relative preference between target and competitor agents show that log-probability ratios are not at an at-chance level at the beginning of the sentence (i.e., they are already positive). In other words, the prior events elicit anticipatory baseline effects (ABEs; Barr, Gann, & Pierce, 2011). The configuration of the target scene (i.e., where the pictures of the faces were shown, see Figure 5.1) might have been too simplistic and therefore too *unconstrained*: the presence of only the pictures of the potential agents may have favoured these ABEs, biasing participants to recruit only one type of information (i.e., the gender of the agent in the previous video) when verifying the content of the sentence. The limited set of visual stimuli in the target display (e.g., no object images that could be related to events were present) might have discouraged participants from taking a more active role in the interpretation of parts of the sentence that did not involve the agent, particularly when mismatches were encountered. Future studies on such anticipations should strive to minimize those ABEs. A richer concurrent visual context in which participants can inspect both target and competitor themes (i.e., objects) and agents (i.e., subjects) during sentence comprehension, might motivate a more active interpretation process and enrich representations during comprehension. Besides, intervening object information, which in our experiments is referred to in sentence initial position, may allow for more genuine anticipatory processes regarding the agents (objects may catch attention before agents, and that might reduce ABEs). By adding visual objects, we could moreover gain some insights into how attention towards those objects is affected by visual gender cues from prior events and stereotypical knowledge. With that in mind, we changed the configuration of the target display in the following experiment.

6 | The concurrent visual context: constraining participants' expectations

From previous experiments including our own, we can see that the recent visual context strongly influences people's attentional behaviour over co-present scenes during sentence comprehension. When the linguistic input refers back to prior events, we tend to anticipate the concurrent characters that took part in those events, or whose features match those from prior events, even when actions themselves are no longer depicted and long-term knowledge may contradict our visually grounded expectations (Knoeferle & Crocker, 2007). However, we tend to disengage our attention from referents accordingly when the linguistic input is at odds with previous visual information, while such incongruence does not eliminate the initial recent-event preference. In other words, the evidence from our experiments, similar to prior research, leads to the interpretation that prior visual events seem to leave strong episodic traces in working memory and that these interfere with later reconciliation processes between event-based and language-based information.

However, much like a recent visual context can constrain visual attention and anticipation processes, comprehension and visual attention can also be influenced by how the concurrent scene (i.e., the scene in which language unfolds) is configured (i.e., more or less constrained). We also discussed previously that the concurrent visual scene can contain relevant information to disambiguate locally ambiguous syntactic information (Knoeferle et al., 2008), but also differences in its configuration can motivate different interpretations of a sentence (e.g., whether a prepositional phrase like *on the towel* is interpreted as a

goal for an object or as its modifier depending on the amount of referents available in the scene; Tanenhaus et al., 1995). In our previous experiments (Experiments 1 and 2), the target display contained only one type of entities (i.e., the potential agents of the event). Although this could be a valid laboratory-proxy for a real-world situation like any other, it is often the case that several objects are also present in our visual environment and are as strongly related to prior events as agent information (Abashidze et al., 2014; Knoeferle & Crocker, 2007). Moreover, from an experimental point of view, if our main aim is to measure anticipation towards gendered agents during language comprehension, enriching the concurrent visual scene with more intervening visual information may allow for a reduction of anticipatory baseline effects, which tend to obscure results in visual-world studies. Recall that our sentences refer to object information first, and the pictures of such objects may make participants direct their attentional resources to these entities before they proceed to the anticipation towards the agents.

More intervening visual information during comprehension (i.e., additional target objects vs. competitors) may be more demanding for the comprehender (i.e., attention may need to be distributed across more visual entities and this might require a greater cognitive effort); however, it might also allow for a more active attempt to reconcile event-based and language-based representations. For instance, in the cases where language described prior events, the appearance of objects, together with verbally expressed actions (i.e., the verb phrase) may allow for a better reconstruction of event-based representations. Additionally, these additional objects might also help listeners with constructing new alternative representations derived merely from the linguistic input, i.e., when language is at odds with prior event information (e.g., when a cake baking action had taken place but the sentence is about a model building action, the image of the model might facilitate the mental representation of a model building event). Under these circumstances, long-term knowledge (i.e., stereotypes) might be used in order to guide expectations (e.g., if the target agent was female, constructing a representation with a model building event may reduce the preference for this agent, and even divert attention towards the competitor agent, i.e., the male face).

In sum, previous research has shown that listeners' visual anticipation of entities, and more generally their attention to entities during language comprehension can be constrained by multiple information sources. The visual context is one of them, sometimes in the form of prior visual events or visual scenes that are no longer present during the visual context concurrent with language comprehension. However, the configuration of the concurrent scene itself, i.e., which potential referents and how much contrasting information is available, can also impose contextual constraints on real time language processing, leading to different ways of using available information (event-based information and long-term knowledge derived from language).

6.1 | Experiment 3

The caveats identified in our first experiments, together with the findings from some studies on how the configuration of the visual scene affects situated language comprehension (Knoeferle et al., 2008; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995), motivated a modification in the target scenes from those employed in Experiments 1 and 2. Like in our previous experiments, participants first inspected a videotaped event that showed a pair of hands interacting with objects (e.g., baking a cake or building a toy model).

Recall that in the two previous experiments, the target scene following the video only contained photographs of the faces of a male and a female character. By contrast, in the present experiment, we added two pictures: one photograph showed the object that had been acted upon in the preceding video (i.e., the target object); the other was a photograph of an object that had not been seen before, but that could potentially be mentioned (i.e., the competitor object). This latter object was part of an action with opposite stereotype valence from that of the object in the preceding event. For example, after showing a video of a pair of female hands baking a cake, the target scene would contain a female face (the target agent), a male face (the competitor agent), the

cake (from the stereotypically female action in the preceding video) and a toy model (a competitor object, part of a model building action which would be stereotypically male, see Figure 6.1). As in the previous experiments we measured visual attention to the agent picture pairs during OVS sentence comprehension, but also to the object picture pairs. Participants answered via button press whether the sentence matched the video they had just seen (“yes” or “no”).

The factor manipulation in the present experiment was the same as in Experiment 1: action-verb (phrase) match was manipulated together with stereotypicality match. Accordingly, we predicted to replicate its findings of shorter response times to action-verb mismatches compared with matches. Additionally, we reasoned that the present constraints in the target scene (i.e., the presence of objects), together with verbally expressed actions, could help boost the representations of action events, including the stereotypical knowledge associated with them. If our new configuration motivated the use of gender-stereotype knowledge, reaction times could also be modulated by the stereotypicality of the described actions, e.g., longer reaction times for stereotypically incongruous conditions compared to congruous conditions.

For the eye movements over the agents, we initially expected that participants would prefer to inspect the target agent over the competitor, in line with the results from Experiments 1 and 2, and supporting accounts of visually mediated language comprehension (Knoeferle & Crocker, 2007). This preference should decrease when the action described by the sentence mismatches the previously depicted event. However, if the presence of the objects enriching the concurrent visual scene (together with the verbal information) maintains access to the representation of the recently inspected events (in the cases of action-verb match) and allows the elaboration of alternative representations via the competitor objects (in cases of action-verb mismatch), then mismatch effects could occur earlier compared to Experiment 1 (during or by the end of the first noun).

Additionally, if these enriched representations derived from the new visual configuration (where more intervening information is present) boosted the intervention of stereo-

typical knowledge in predicting upcoming information (i.e., the agent in final position) we could see more anticipatory looks to the target agents (over competitor agents) when the action described is stereotypically congruent compared to incongruent with the target agent (e.g., when female hands were in prior events and the described action is about baking a cake, participants may look at the female character more than when the sentence is about building a model). In the case of action-verb mismatch conditions (where the sentence would no longer support the maintenance of event-based representations and consequently, the anticipation of the target agent), and assuming a predominance of the recent-event preference, the influence of stereotypical knowledge could also result in the reduced preference for looking at the target agent in stereotypically incongruent conditions. Alternatively, if stereotypical knowledge gained more relevance under the enriched configuration of the concurrent scene, fully mismatching conditions might divert the attention of participants towards the competitor agent. For example, if participants saw female hands baking a cake (a stereotypically female action), but the following sentence described a model building action (stereotypically male), then the presence of an object such as a toy model in the concurrent visual scene might help participants with constructing an alternative representation of a toy model building event, additionally activating the gender-stereotype associated with such an event and diverting anticipatory looks from the female agent (i.e., the target) to the male one (i.e., the competitor). Although weak between-subject differences appeared in Experiment 1, these lead us to think that if any stereotypicality effects emerged, they could interact with participants' gender.

When the target scene also contained pictures of the objects involved in the events, participants could arguably spend a substantial amount of time inspecting them. However, for the relative preference in favour for target objects over competitor objects, we expected slightly different results, as object information comes first during comprehension in our experiments (due to the OVS word order) and it does not allow for linguistically driven anticipatory processes. We argued that anticipatory baseline effects might take place, i.e., participants could prefer to inspect the target object over the competitor object early on, by virtue of appearing in prior events. However, as soon as the sentence

started, we predicted that both target (in action-verb match conditions) and competitor objects (in action-verb mismatch conditions) would be looked at fast when participants listened to their referring expressions (Tanenhaus et al., 1995), perhaps with some delay delay for the competitors (in action-verb mismatches).

Also, it is possible that unlike the agents, attention towards the objects might be affected by the internal stereotypical congruence of the videos. If participants used gender cues from prior events and stereotypical gender knowledge of the objects (in relation to the agents and events they typically take place with), attention towards visual objects could be modulated. In action-verb match conditions, for example, this would be translated into more looks to the target object (over the competitor object) in stereotypically matching conditions compared to mismatching conditions. For example, if prior events featured female hands, looks to the cake when hearing ‘The cake_{OBJ} bakes_V’ (a stereotypically female action) would be more frequent compared to looks to a toy model when hearing ‘The model_{OBJ} builds_V’ (stereotypically male action). Looks to the objects in action-verb mismatches may show a similar pattern assuming that the competitor object would be preferentially inspected over the target (because in these cases it is the competitor, i.e., the one that did not appear in prior events, that is mentioned). In this case competitor objects might be looked to a greater extent when prior events featured stereotypically matching hands compared to stereotypically mismatching hands. For instance, if prior events showed female hands building a model, but the sentence is about a cake baking action, looks to the cake (which in this case would be the competitor object) when hearing ‘The cake_{OBJ} bakes_V’ would be more frequent compared to the looks towards the toy model if female hands were baking a cake and the sentence was about building a model.

6.1.1 | Methods and Design

Participants A further 32 participants took part in the experiment (16 females, 18-32 years, M=22.37). All were German native speakers and had normal or corrected to

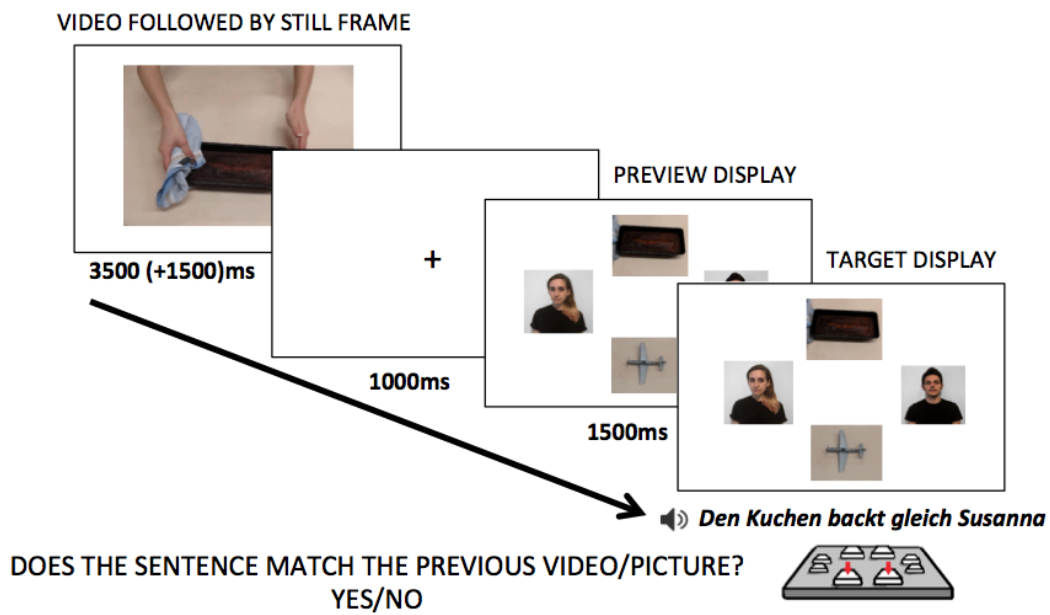


Figure 6.1.: Example of an experimental trial in Experiment 3.

normal vision. They all gave informed consent before the experiment.

Materials We used the materials from our previous studies (Experiments 1 and 2). However, for the target scene we did not use just the close-up pictures of the actors, but also the snapshots of the objects that belonged to the same item or action pair. One would be the target object (the one that appears in prior events), while the other object would be the competitor object, part of the unseen action with the opposite stereotypical gender valence. The experimental manipulation followed that of Experiment 1. We manipulated: a) action-verb match (the action described by the sentence matched or mismatched the action previously seen in the video) and b) stereotypicality match (the action described by the sentence either matched or mismatched stereotypically with the gender implied by the hands performing the action in the video, see Table 5.1 above).

Procedure The procedure was as in the previous Experiments (see Figure 6.1).

6.1.2 | Analysis and Results

Analysis We followed the same analysis based on the log gaze probability of looking at the target vs. the competitor agent as in previous experiments. Additionally, we performed analyses over the objects in which preferential looks to the target object (i.e., the object that appeared on the previous video) compared to the competitor object (the object that was only present in the target screen) were measured. For the objects, we focused on the two regions most related to object (i.e., theme) information, mainly NP1 and Verb regions. Apart from the time-course graphs using the log-probability ratios for agents and objects separately, we also created time-course graphs with the percentages of looks towards the four elements on the target scene (target and competitor agents and objects, see Appendix B.5).

Results *Accuracy and response times:* Participants responded correctly on 98% of the trials and were more accurate with action-verb mismatches than matches (see Appendix B.1, Table B.3). Reaction times were significantly shorter for action-verb mismatches ($M=3330$, $SD=69.75$), than matches ($M=4547$, $SD=22.85$); $F_1(1, 30)=30.32$, $p=.001$, $\eta^2=.503$; $F_2(1,31)=525.32$, $p<.001$, $\eta^2=.944$.

Time-course graph with percentages of looks for agents and objects: In terms of percentages of looks, participants looked at the objects (both targets and competitors) to a great extent throughout the sentence, while attention towards the agents, especially at the beginning, was small (see Appendix B.5)¹. Overall target objects and agents seem to get more attention than their counterparts, although in action-verb mismatching conditions, attention towards the competitor objects does increase at NP1 (i.e., when the object is mentioned). For the agents, it is only in the completely mismatching condition that we see an increase of attention towards the competitor object that is slightly greater

¹It is important to note that just as with time-course graphs displaying log-probability ratios, the time-course graphs with the percentages of looks are only orientative visual information, as they are based on the average duration of the sentence across all the experimental items.

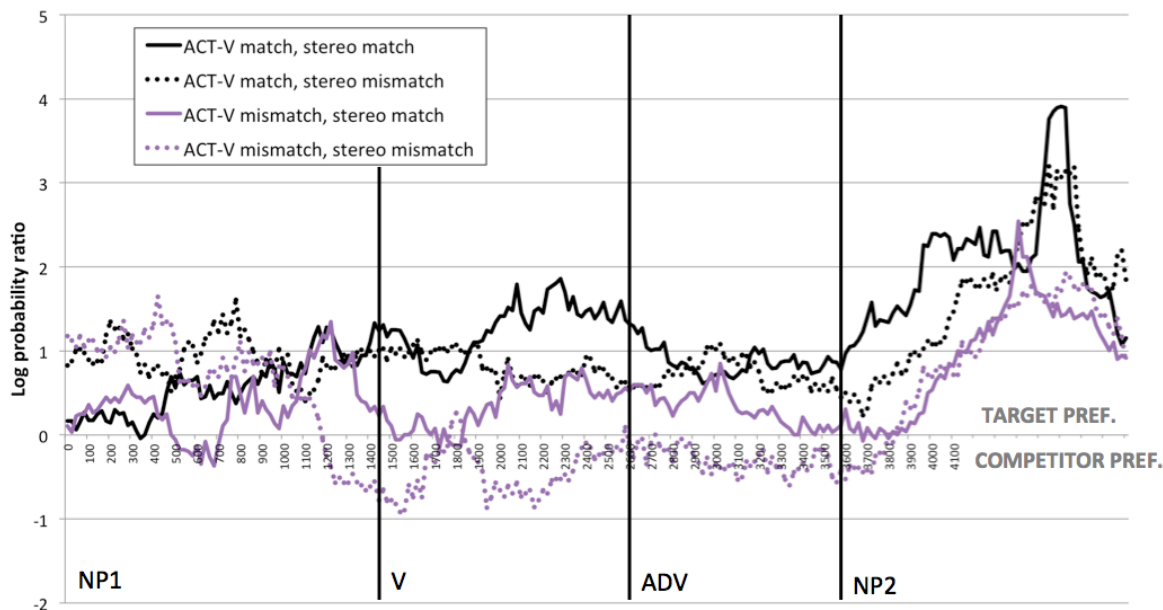


Figure 6.2.: Time-course graph for the **agents**, Experiment 3.

than the attention towards the target agent. Attention towards the target agent gets boosted at the end of the sentence by virtue of the final, proper noun (i.e., NP2).

Eye-movement data for the agents: When inspecting the time-course graph of the log-probability ratios for the agents (see Figure 6.2), we can see an overall preference for looking at the target agent compared to the competitor, similar to what we observed in Experiments 1 and 2 (see Appendix B.3). However, the graphs look substantially different. Particularly at the verb region we see a gradual modulation of log-probability ratios based on condition, with the fully matching condition (action-verb match, stereotypicality match, solid black line) showing the most positive going values (target agent preference). In the same region, the fully mismatching condition (action-verb mismatch, stereotypicality mismatch, dashed purple line) exhibits negative values of the log gaze probability ratio (which would suggest that for this condition there is a slight preference for looking at the competitor agent), while the conditions where there are mismatches of one kind (action-verb mismatch, stereotypicality match, solid purple line; action-verb match, stereotypicality mismatch, dashed black line) find themselves in between the fully matching and the fully mismatching conditions. All values turn clearly positive by the end of the sentence as the final noun (therefore, the target agent) is revealed.

As for the analyses on sentence regions, an action-verb match effect emerged as early as in the NP1 region, $F_1(1, 30)=5,41$, $p<.05$, $\eta^2=.153$; $F_2(1,31)=9,61$, $p<.01$, $\eta^2=.237$; Figure 6.3). Participants were more likely to inspect the target agent when the object mentioned in the sentence matched (vs. mismatched) prior events.

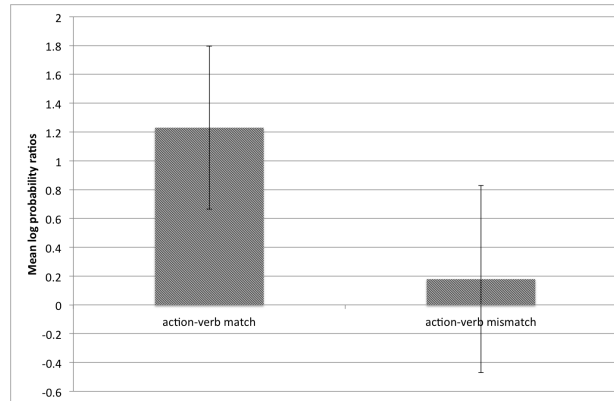


Figure 6.3.: By-subject mean log-probability ratios for the **agents** at the NP1 region, Experiment 3 (error bars indicate 95% confidence intervals).

At the verb, main effects of both action-verb, $F_1(1, 30)=28,62$, $p<.01$, $\eta^2=.206$; $F_2(1,31)=21,63$, $p<.001$, $\eta^2=.411$; see Figure 6.4a) and stereotypicality match, $F_1(1, 30)=9,78$, $p<.01$, $\eta^2=.246$; $F_2(1,31)=12,16$, $p<.01$, $\eta^2=.282$ (see Figure 6.4b) emerged. The target agent received more looks compared to the competitor agent in action-verb matches compared to mismatches; additionally, the target was looked at more when the action described by the sentence was stereotypically congruent compared to incongruent.

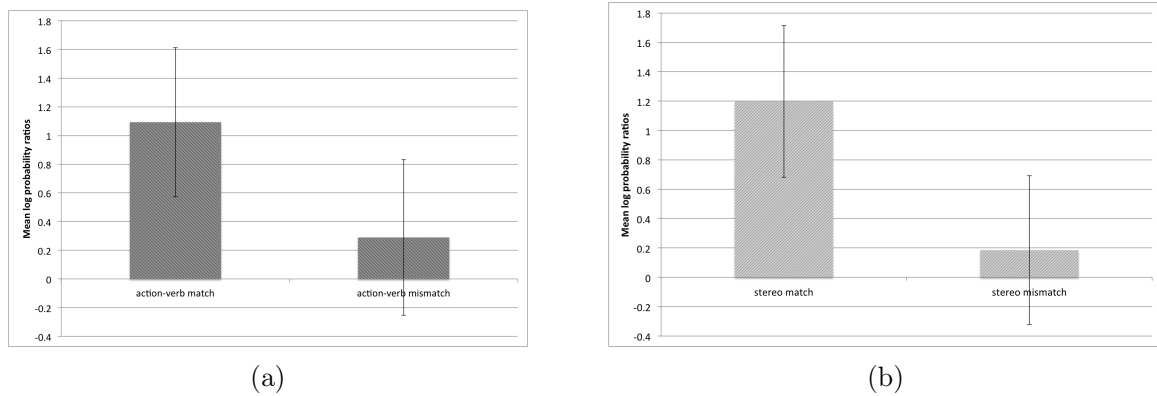


Figure 6.4.: By-subject mean log-probability ratios for the **agents** in the action-verb match condition (a) and the stereotypicality match condition (b) at the verb region, Experiment 3 (error bars indicate 95% confidence intervals).

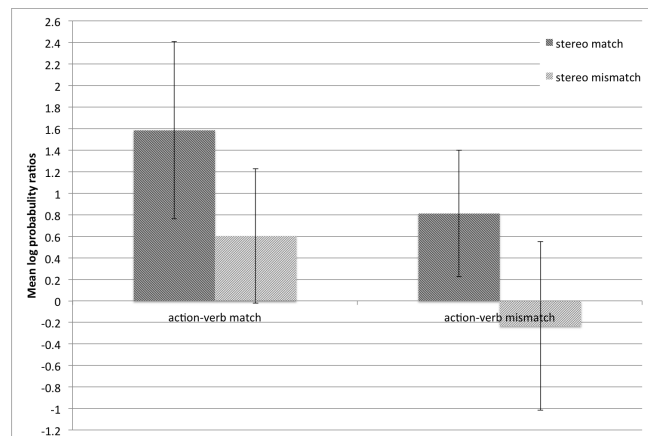


Figure 6.5.: By-subject mean log-probability ratios for the **agents** at the verb region per condition, Experiment 3 (error bars indicate 95% confidence intervals).

While in the adverb region the action-verb effect, $F_1(1, 30)=6,59$, $p<.05$, $\eta^2=.180$; $F_2(1,31)=17,49$, $p<.001$, $\eta^2=.361$; (see Figure 6.6a), and the stereotypicality effect prevailed, the latter was marginally significant by subjects, $F_1(1, 30)=3,46$, $p=.07$, $\eta^2=.104$; $F_2(1,31)=4,63$, $p<.05$, $\eta^2=.130$; (see Figure 6.6b).

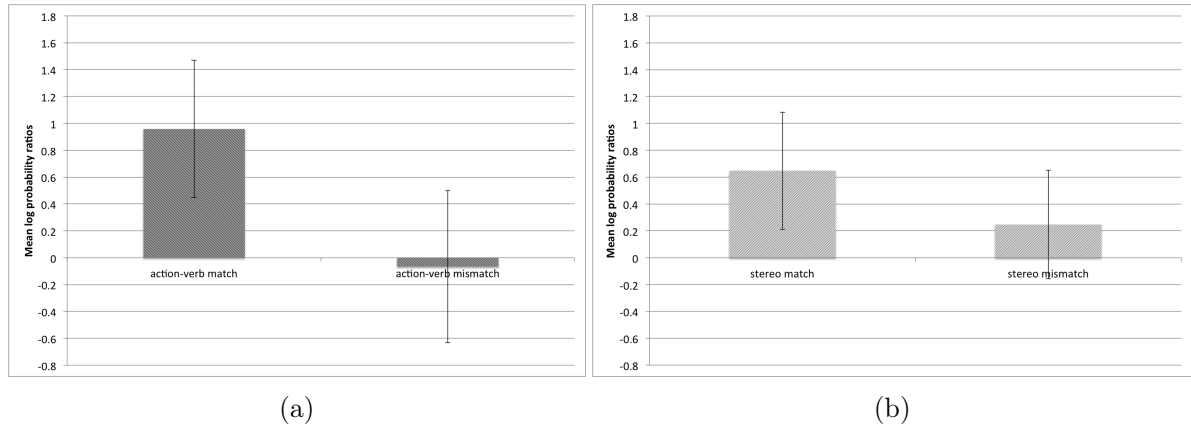


Figure 6.6.: By-subject mean log-probability ratios for the **agents** in the action-verb match condition (a) and the stereotypicality match condition (b) at the adverb region, Experiment 3 (error bars indicate 95% confidence intervals).

For the final, NP2 region, there was a main effect of action-verb match, $F_1(1, 30)=6,87$, $p<.05$, $\eta^2=.186$; $F_2(1,31)=15,08$, $p<.001$, $\eta^2=.327$ (see Figure 6.7). An interaction between action-verb match and stereotypicality match also emerged by subjects $F_1(1, 31)=4,16$, $p=.05$, $\eta^2=.122$; $F_2(1,31)=1,75$, $p=.2$, $\eta^2=.053$; (see Figure 6.8).

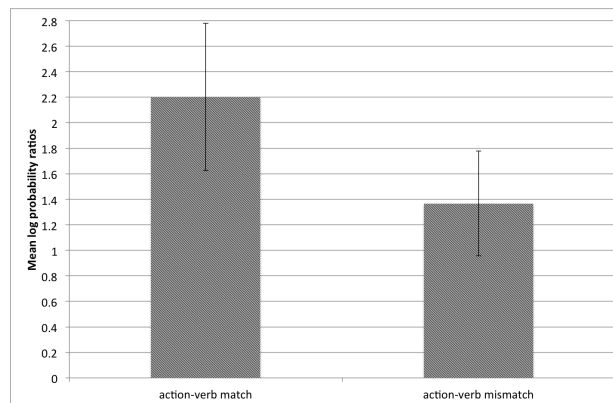


Figure 6.7.: By-subject mean log-probability ratios for the **agents** at the final noun (NP2) region, Experiment 3 (error bars indicate 95% confidence intervals).

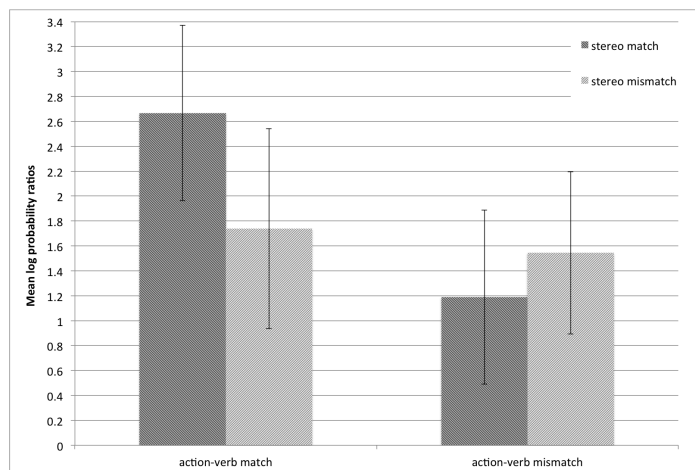


Figure 6.8.: By-subject mean log-probability ratios for the **agents** at the final noun (NP2) region per condition, Experiment 3 (error bars indicate 95% confidence intervals).

Eye-movement data for the objects: The time-course graph for the objects shows a clear split in the preference for looking at the target object (i.e., the objects that appeared in the prior events) over the competitor object. When the verb(-phrase, and therefore, the object mentioned at NP1) matches prior events (action-verb match, the two black lines), the clear preference for the target object is reflected in positive values. In the action-verb mismatching conditions (e.g., purple lines) participants shifted to the competitor object, as indexed by the negative values at NP1, see Figure 6.9). This would reflect a direct referential strategy between the nouns and their referents, although this shift of attention takes place somewhat later as a consequence of the mismatch between the sentence and prior events, i.e., the purple lines start differing from 0 half-way through NP1.

At the NP1 region, we found both main effects of action-verb match, $F_1(1,30)=142.74$, $p<.001$, $\eta^2=.826$; $F_2(1,31)=174.89$ $p<.001$, $\eta^2=.849$, and stereotypicality match, significant by subjects and marginal by items, $F_1(1,30)=6.46$, $p<.05$, $\eta^2=.177$; $F_2(1,31)=3.037$, $p=.09$, $\eta^2=.089$, as well as an interaction between the two factors significant by subjects, $F_1(1,30)=7.85$, $p<.01$, $\eta^2=.207$; $F_2(1,31)=1.069$, $p=.3$, $\eta^2=.033$. The target object preference was greater for action-verb matching conditions compared to mismatching conditions. Additionally, target objects were preferred to a greater extent in stereotipi-

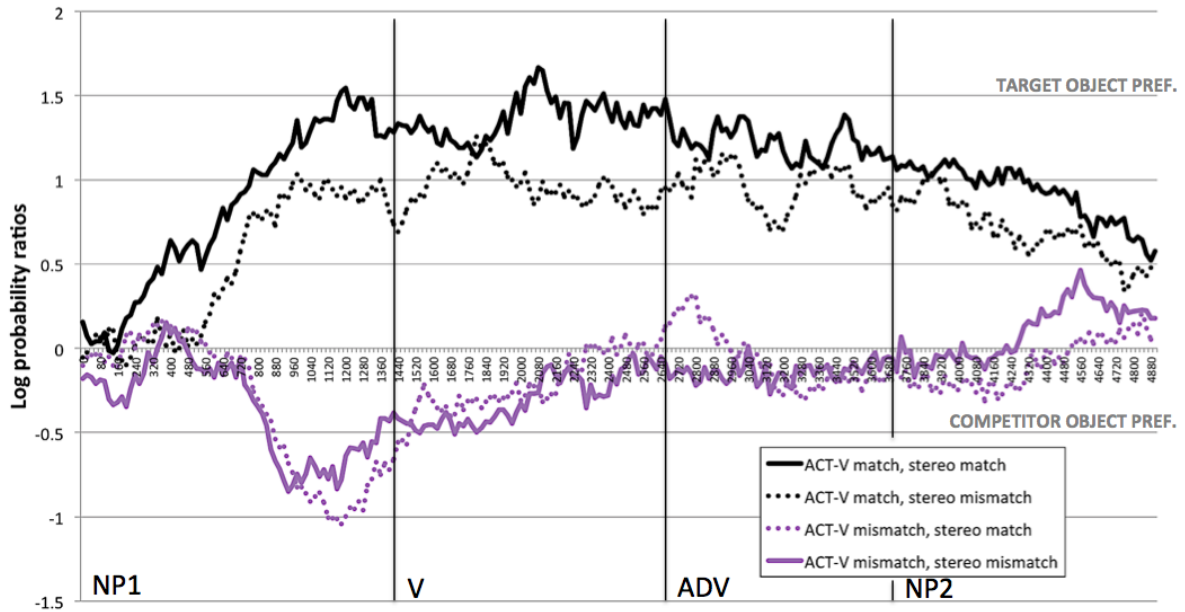


Figure 6.9.: Time-course graph for the **objects** in Experiment 3.

cally matching conditions (i.e., when the hands shown in prior events were stereotypically matching with the objects) compared to mismatching conditions (see Figure 6.10)².

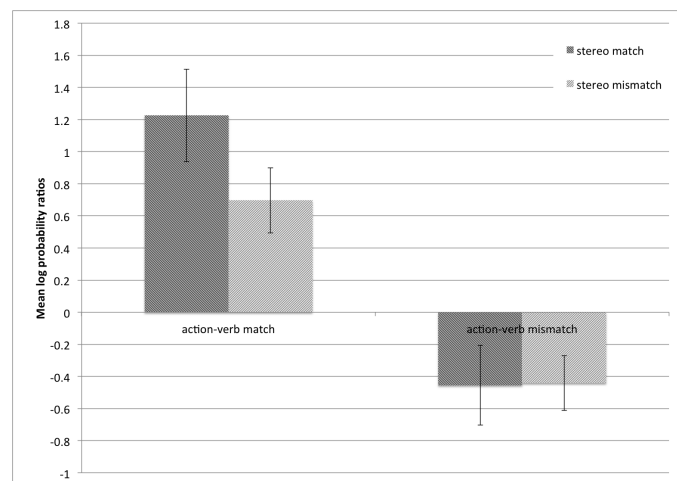


Figure 6.10.: By-subject mean log-probability ratios for the **objects** at the NP1 region per condition, Experiment 3 (error bars indicate 95% confidence intervals).

At the verb the effect of action-verb match persisted, $F_1(1, 30)=79.27$, $p<.001$, $\eta^2=.725$; $F_2(1,31)=107.57$ $p<.001$, $\eta^2=.776$ (see Figure 6.11a), as did the stereotypicality match effect, significant by subjects and marginal by items $F_1(1, 30)=5.22$, $p=.05$,

²The effects of stereotypicality were not reliable in the mixed models effects analysis for the objects, see Appendix B.4, Table B.17.

$\eta^2=.119$; $F_2(1,31)=10,50$, $p=.08$, $\eta^2=.094$ (see Figure 6.11b).

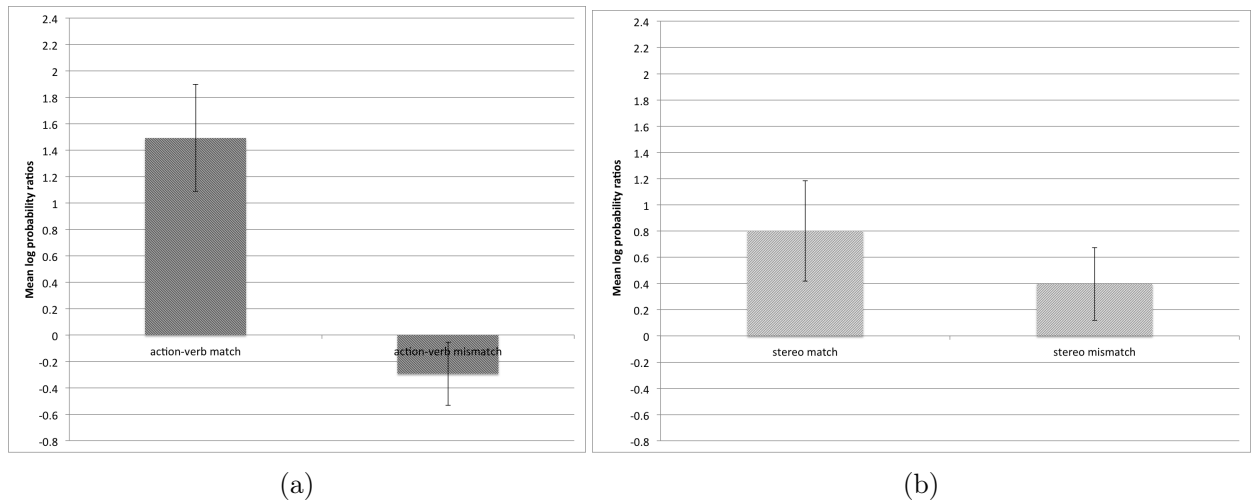


Figure 6.11.: By-subject mean log-probability ratios for the **objects** in the action-verb match condition (a) and the stereotypicality match condition (b) at the verb region, Experiment 3 (error bars indicate 95% confidence intervals).

6.1.3 | Discussion

Experiments 1 and 2 had contrasted the effects of gender and action cues from recently inspected events with those of gender-stereotype knowledge in situated language comprehension. Analyses of the data from these studies have revealed participants' reliance on representations derived from the preceding events to anticipate potential agents during comprehension, affected by referential mismatches between prior events and language but with weak influence of stereotypicality. In the present experiment, we changed the configuration of our visual scene concurrent with language to reduce anticipatory baseline effects and to study the extent to which constraints posed by the concurrent scene might affect the use of event-based representations as well as knowledge on gender stereotypes.

Based on the design from Experiment 1, we manipulated the match between the previous event and the sentence (i.e., action-verb match), as well as between the stereotypicality of the action described and the gender features of the agent cued in prior events (i.e., stereotypicality match). Unlike in Experiments 1 and 2, participants could not only inspect the photographs of the faces of two potential agents in the target scene, but also

a photograph of an object from the previously seen event and a so-called ‘competitor’ object (from an action with opposite gender stereotypicality valence from that of the object in prior events). We reasoned that the additional object presentation together with the sentential verb phrase would a) allow keeping the representation of the recently perceived event, and b) motivate building alternative representations when language mismatched prior events (i.e., action-verb mismatches). The enriched visual scene could additionally motivate the use of (long-term) stereotypical knowledge, resulting in more anticipatory looks to the target in stereotypically congruous compared to incongruous conditions. Additionally, how the inspections towards the objects developed would also provide insight into the effects of action-verb match and gender stereotypicality in establishing noun-object visual reference.

The added contextual constraints gave rise to different results in the anticipatory eye-movement behaviour of participants compared with the results reported in Experiment 1. The accuracy and reaction-time results were similar for both Experiments 1 and 3, and in the latter, we did not find between-subject (gender) differences. Moreover, as it can be seen in the eye-movement results, we also obtained an overall preference for looking at the target agent (vs. the competitor agent). Crucially, however, we did observe earlier effects of action-verb match manipulations in the inspection of agents in the present study (already in the NP1 region). In addition, the present analyses confirmed effects of stereotypicality match, which had been absent from the previous studies; these emerged as early as in the verb region. Participants' preference for inspecting the target agent (vs. the competitor) was more pronounced when the verbally expressed action matched in terms of gender stereotypicality (e.g., when female hands had performed an action, participants preferred to look at the female agent more when the following sentence mentioned a cake baking action compared to a model building action). The inferential statistics at this region, together with the plotted data per condition (Figure 6.5), suggests that the effects of both action-verb match and stereotypicality match may be (super-)additive rather than interactive in the context of recent events: Both types of congruence can work either independently or in tandem to increase, maintain or decrease

participants' visual attention towards a target agent (e.g., Casasanto, Hofmeister, & Sag, 2010; Chow et al., 2014; Yap & Pexman, 2016). When both cues (a congruent linguistic and stereotypical meaning) favour the anticipation of the agent of a particular gender, log-probability ratios are high, while when both cues are incongruent, this preference is hindered the most (see Figure 6.5). Although the completely mismatching condition did exhibit a negative mean (suggesting a tendency to look at the competitor agent), this was numerically close to 0, suggesting that in this condition, participants looked at the two potential agents at an at-chance level. Not even when participants cannot rely on prior visual events and have to interpret the linguistic content alone, and the content of language is stereotypically biased towards the opposite agent from that recently seen, does the presence of visual objects that could help participants in constructing an alternative mental representation prompt looks to the competitor agent. What we can say is that at least under these circumstances, the preference for event-based visual gender cues seems to disappear.

The objects presented on the target scene did gain substantial visual attention during comprehension, although differently from the agents. Attention towards the objects was ruled by more consistent referential strategies (i.e., both target and competitor objects were looked at as they were mentioned), as referential accounts would predict (Tanenhaus et al., 1995). Interestingly, however, for the objects we also found effects of stereotypicality, both at the NP1 region as well as at the verb. Participants preferred to inspect the target object to a greater extent when the gender features of the agent that appeared in prior events (i.e., the hands) was stereotypically matching (vs. mismatching). For the objects, we did see signs of an interaction between action-verb match and stereotypicality. By looking at the graphs (see Figure 6.10) we could infer that stereotypical match was more effective in action-verb matching conditions compared to mismatching conditions, suggesting that stereotypical knowledge may help maintain event and object representations during comprehension inasmuch as language is consistent with prior events.

To date, prior visual events have shown a virtually invariant influence on visual at-

tention during situated language comprehension. These events are strong cues to guide attention, as they provide detailed information about actions and their associated target objects, as well as individuating information (e.g., about gender of an agent's hands) that can serve to identify associated agents. Upon the encounter of linguistic information, this recent information, rather than world-knowledge derived from language, predominantly guides people's attention over entities in an anticipatory and referential manner (e.g., a verb phrase can identify the agent that was involved in prior events, and a referent that matches those features will be fixated during the unfolding of the sentence). However, referential mismatches are detected quite fast and they do affect this preference. Moreover, changes in the constraints of the visual context concurrent with language may change the speed with which we react to the detection of those mismatches. Not only that, differences in configuration do also seem to determine whether world-knowledge plays a role, i.e., whether such knowledge may modulate the inspection of anticipated agents.

The results obtained in this experiment support the idea that manipulating visual constraints in situated language comprehension may not only make a difference in terms of how comprehenders resolve structural ambiguity (Spivey et al., 2002; Tanenhaus et al., 1995). Constraining the visual environment via additional (object) referents (as implemented in the present study compared with Experiment 1) can boost concurrent visiolinguistic interactions and help maintain more vivid event-based representations. This can modulate the time-course with which mismatch effects between event-based representations and language emerge. The presence of additional entities in the concurrent scene during language comprehension seems to also motivate a greater use of inferences based on world-knowledge about gender stereotypes as compared to less constrained visual environments, which can modulate the extent to which event-based information is trusted.

7 | The electrophysiological correlates of visual gender verification in language comprehension

So far we have seen evidence on how comprehension of sentences conveying action and gender information is affected as a function of prior visual events by means of tracking participants' attention in the visual world. However, another very fruitful methodology that can further complement our findings, as mentioned in section 2.4, is the measurement of event-related brain potentials. In this way, we could further explore the mechanisms involved during sentence processing when the comprehender tries to reconcile the gender features of an agent with the gender implied in language, and further inform accounts on situated language comprehension.

The role of gender information (grammatical and semantic) in language processing has already been looked at using ERPs, with most of the research focusing on the study of anaphoric resolution (Hammer, Jansma, Lamers, & Münte, 2008; Kreiner, Mohr, Kessler, & Garrod, 2009; Lamers, Jansma, Hammer, & Münte, 2006; Schmitt, Lamers, & Münte, 2002; Streb, Hennighausen, & Rösler, 2004; Xu, Jiang, & Zhou, 2013). Studies in this area have encountered mixed results, arguably due to the difficulties in disentangling grammatical and semantic gender information. In a study in German carried out by Schmitt et al. (2002), they tested gender agreement between a person referent in either a non-diminutive form (e.g., *Der Bub*, 'The boy') or in a diminutive form (which is

usually expressed in the grammatically neuter form e.g., *Das Bübchen*). A following pronoun was presented in either of the three German grammatical gender forms (*er/es/sie*). Mismatches between the non-diminutive form and the pronoun with both semantic and syntactic constraints (e.g., *Der Bub-sie*, ‘The boy - she’) compared to matching counterparts elicited negativities peaking around 400ms and positivities peaking around 600ms (a biphasic N400/P600 response). The other types of mismatches (e.g., between the neuter form and a pronoun) only elicited P600 effects. The authors concluded that in the processing of gender agreement, both syntactic (as reflected in P600 effects) and semantic aspects (as reflected in N400 effects) are involved, the latter in cases where biological gender information is more salient.

Another study (Hammer et al., 2008) using persons versus things as antecedents also found N400 effects for the former in cases of a mismatch with the pronoun (*Der Häuptling_{mas} ist kriegerisch, weil sie_{fem} gewinnen will*, ‘The chief_{mas} is martial, because she to win wants’), suggesting that agreement processing between person (i.e., biological gender) antecedents and pronouns, unlike things, required semantic integration processes. However, not all studies using this type of antecedents have obtained the same results. For example, Xu and colleagues (2013) used sentence pairs where the first sentence introduced a protagonist (either a neutral noun marked for gender, or nouns with definitional gender like *mother* or *uncle*) and a second sentence contained a pronoun that referred back to it (example translated from Chinese: ‘*This woman patient was in low spirits, doctors encouraged him/her to cheer up*’). At the 550-650 time window relative to pronoun onset (*him*), the authors found more positive going ERPs for gender mismatches between the pronoun and its antecedent compared to matches. However, in their second experiment, where the mismatch occurred with the plural form of the pronoun (‘*These woman patients - them_{masc}*’) the authors did find negative going amplitudes for mismatching conditions in the 250-400 time window in addition to the P600 effects.

Gender information, much like an antecedent noun, can be alternatively conveyed through a visual context. In a way, visual information preceding sentence comprehension

is arguably analogous to a discourse processing environment when it comes to the flow of the information; i.e. two entities presented one after the other, albeit from different modalities, need to be coindexed in order to succeed in interpretation. Recent ERP studies have also explored the different processes involved in the verification of gender from two different sources, visual and auditory, during sentence comprehension. Picture-sentence verification studies have manipulated the different points at which mismatches can occur, i.e., whether they affect verb information, thematic roles, spatial relations and so forth (Coco, Araujo, & Petersson, 2016; Knoeferle, Urbach, & Kutas, 2011; Knoeferle et al., 2014; Vissers et al., 2008; Willems, Özyürek, & Hagoort, 2008). These studies, just like in discourse processing research, have also found mixed results: while some studies encountered N400 effects for incongruencies between visual and linguistic stimuli (e.g. Coco et al., 2016; Knoeferle et al., 2014), other studies have also found P600 effects for mismatches compared to matches (e.g., Vissers et al., 2008; Wassenaar & Hagoort, 2007). Moreover, Knoeferle et al. (2014) had already compared different types of mismatches between pictorial stimuli and language within their experiments (verb-action relations vs. thematic role relations), and concluded, given the somewhat different topographies of their N400 effects, that distinct cognitive mechanisms might be involved in the different picture-sentence relations (see section 2.4).

If we observe how both studies on anaphoric resolution and picture-sentence verification are temporally organized, we can see that the flow of information between the prior context and target sentences (i.e., where a word referring back to an antecedent appears) might take place in an analogous manner. Based on both lines of research (i.e., studies on gender information in anaphoric resolution and picture-sentence verification), this experiment aimed at exploring gender processing in sentence comprehension as a function of prior visual gender cues from an electrophysiological perspective. More specifically, we aimed at identifying the ERP components that might be related to this particular verification process and how effects of these visuosyntactic mismatches can compare to other types of picture-sentence verification processes and to the findings obtained in the literature on gender processing through discourse.

7.1 | Experiment 4

In order to explore the electrophysiological responses of gender verification during sentence comprehension, we adopted the same manipulation used in Experiment 2 over the manipulation in Experiments 1 and 3 (where the match between prior visual actions and the verb-phrase was manipulated). Recall that in Experiment 2, the main manipulation was the match between the gender of the final noun (i.e., the NP2; e.g., *Susanna* vs. *Tobias*) and the hands seen in the previous video (*hand-subject gender match*). Therefore, in this experiment, we also focused our attention on the resolution of the agent. Unlike in Experiment 2, in the current experiment we did not include the gender stereotypicality factor. Participants were first presented with videos in which a pair of hands performed an action and then they had to listen to OVS sentences describing those action events. Also unlike in our eye-tracking experiments, this time we did not have a target scene presented together with the sentence. As the presence of visual objects in the target scene during sentence comprehension could elicit non-neural artifacts (e.g., blinks and eye movements) that can distort the data, we opted for a cross in the middle of the screen that participants had to fixate during comprehension (see Figure 7.1). As in previous experiments, at the end of the sentence participants verified whether the sentence matched the previous video.

As we argued above, the flow of information processing between the prior visual context and language might take place in an analogous manner to discourse comprehension. However, in our experiment, rather than linguistic antecedents we have a visually-grounded context (i.e., prior events) that needs to be verified with a referring expression (i.e., the final proper name). Visual gender cues convey a meaningful content about a person's identity, especially when being verified with lexical-semantic content in a sentence. It would therefore follow that mismatches between the final noun (e.g., *Susanna/Tobias*) and the gender of the hands from the previous video should give rise to a greater negativity in the ERP response compared to matching conditions, resembling an N400 effect.

This would reflect the involvement of semantic integration processes (Hammer et al., 2008; Kutas & Hillyard, 1980; Schmitt et al., 2002; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Some of the studies mentioned above report P600 effects, which have mainly been ascribed to structural processing (e.g., morphosyntactic violations) of language. We should in principle not expect effects related to this process in the current experiment, as no revision of the linguistic structure should be required for the verification of visual gender cues and language.

7.1.1 | Methods and Design

Participants 16 participants (13 female, 19-32 years, $M=24.62$), all German native speakers, right-handed, with normal or corrected to normal vision and no known neurological disease received 15 Euro each for participation. They all gave informed consent before starting the experiment^{1,2}.

Materials The 32 experimental items from previous experiments were turned into 64 (each action video, regardless of stereotypicality, performed by both a female and a male actor)³. The main manipulation was congruence between the gender implied by the hands and that implied by the name in sentence-final position, the first factor in Experiment 2 (see Table 7.1). Like in the previous experiments, different fillers were used ($N=120$), containing videos of actions similar to the experimental ones with the same sentence structure, videos with two pairs of hands engaged in an action with dative constructions, and pictures of objects and scenes paired with a range of sentence structures. Just like the experimental items, half of the fillers contained video-sentence mismatches of some sort (final name, described action, color, shape, etc.).

¹Consent forms were partially based on the guidelines from the *Ethikkommission der Deutschen Gesellschaft für Psychologie für die Teilnehmerinformation für EEG-Studien*.

²The study followed the provisions of the Declaration of Helsinki (October 2013, 64th Meeting, Fortaleza, Brasil).

³As in the current experiment we only manipulated the match between the final noun and the hands of the video regardless of stereotypicality, we could make use of both of the sentences that formed an item in our previous experiments (e.g., see Table 5.2), permitting us to increase the number of experimental items.

Table 7.1.: Example item for Experiment 4

Video	Sentence				Hand-subj. gend. match
Female hands baking a cake	Den Kuchen _{NP1} <i>the cake</i> (obj)	bakt _V <i>bakes</i>	gleich _{ADV} <i>soon</i>	Susanna _{NP2} <i>Susanna</i> (subj)	yes
Female hands baking a cake	Den Kuchen _{NP1} <i>the cake</i> (obj)	bakt _V <i>bakes</i>	gleich _{ADV} <i>soon</i>	Tobias _{NP2} <i>Tobias</i> (subj)	no

Procedure After participants' preparation for EEG, they were seated in front of the computer in a sound attenuated room while the experimenter sat behind a panel for data recording. to pay attention to the videos, and then fixate the cross in the middle of the screen during sentence comprehension. They were also asked to avoid blinking and facial movements during sentence comprehension, as well as any other abrupt body movement. After the instructions, participants had 10 practice trials where they received feedback. Trials would start with a video of the action (3500 ms), then the video would stop and the final frame (displaying both the hands in resting position and the objects) stayed for another 1500ms. Next a cross appeared on screen. 1500 ms later the sentence was presented auditorily while the cross remained on screen. The ERP responses were time-locked to the final name region (i.e., NP2), also considered the *critical word* (CW). When the cross turned green, participants could respond whether the sentence matched the events via button press (see Figure 7.1). Response buttons were counterbalanced across participants⁴. Given that the duration of the session was long (2 hours with preparation), participants had a pause three times during the experiment, and could ask for further pauses whenever needed.

7.1.2 | Recording, Analysis and Results

Recording and Analysis The experiment was implemented and run using E-prime (Psychology Software Tools, Inc.). The EEG data was recorded from 26 active electrodes (together with 4 eye-electrodes and 2 electrodes for the mastoids) embedded in an elastic

⁴By mistake one participant was assigned the wrong button-press configuration. As we did not find differences in the results by removing the data from this participants, we kept them in the analysis.

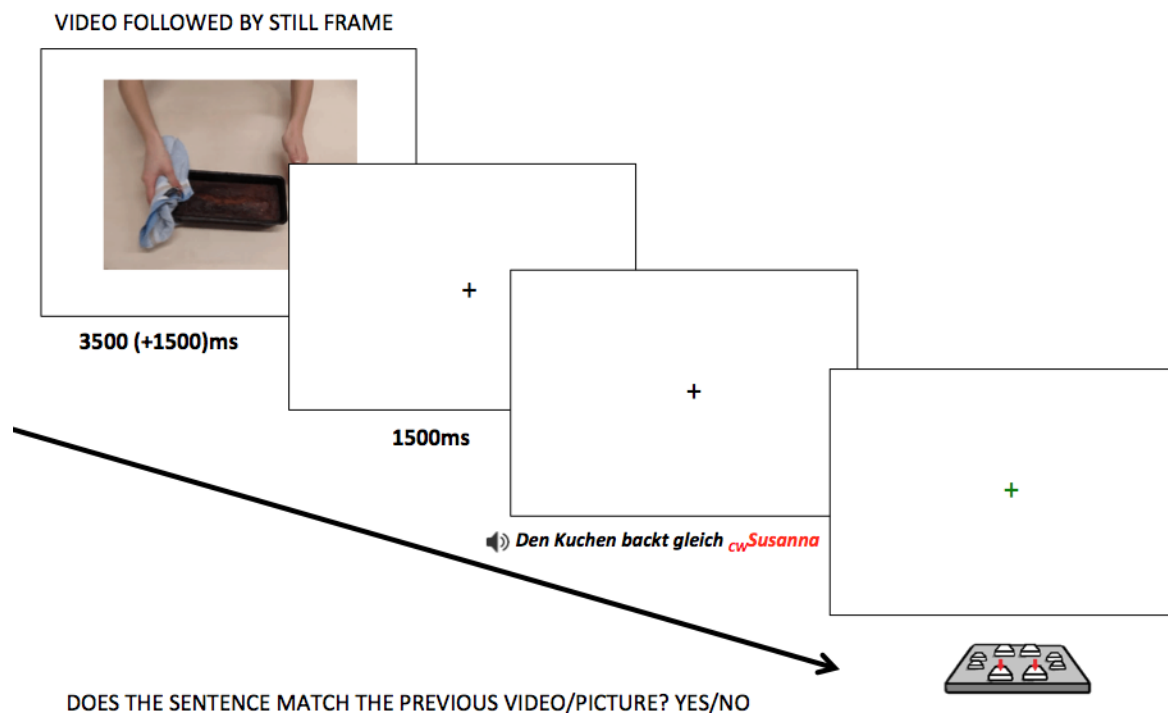


Figure 7.1.: Example of an experimental trial in Experiment 4.

cap (Acticap, Brain Products, see electrode distribution in Figure 7.2) and placed over the scalp, using BrainVision Recorder (Brain Products) at a sampling rate was of 500 Hz. The signal was amplified by a BrainAmps DC amplifier. Horizontal eye movements (HEOGs) and blinks (VEOGs) were monitored by electrodes placed on the outer canthi on both eyes and above and below the left eye. The impedance for all electrodes was kept below 5Ω . The on-line reference electrode was the left mastoid (TP9). Brain Vision Analyzer was used to perform off-line preprocessing. EEG data were off-line re-referenced to the average activity of both mastoid electrodes (TP9-TP10). Signal was bandpass filtered between 0.1 and 30 Hz. Epochs of interest lasted from -200 ms before the onset of the target word to 1000 ms post-stimulus onset (-200 to 0 ms baseline corrected). Artifact activity was excluded based on visual inspection of each trial. No participant was kept if the total percentage of discarded trials or the discarded trials in any of the conditions was above 25 (out of 32 trials per condition). Based on visual inspection and the research hypotheses, we conducted omnibus ANOVAs for the 300-500 and the 500-900 time windows. Repeated measures-ANOVA were first performed taking Condition (hand-subject

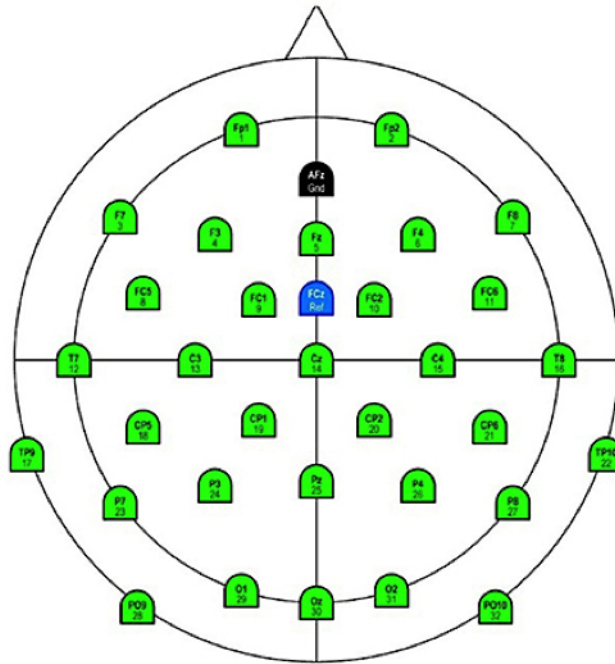


Figure 7.2.: Electrode configuration, using Acticap 32-channel active electrode system (Brain Products). Two electrodes were moved to the outer canthi (T7 and T8), two to the left eye (PO9 and PO10) and two to the left and right mastoids (TP9 and TP10).

gender match vs. mismatch) and Electrode (26 electrodes: Fp1, Fp2, F7, F8, F3, Fz, F4, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, O2) as factors. We also conducted repeated measures ANOVAs with Condition (2 levels: match vs. mismatch), Hemisphere (2 levels: left vs. right) and Anteriority (3 levels: frontal, central and posterior) as factors. Greenhouse-Geisser adjustments to degrees of freedom were applied to correct for violations of the assumption of sphericity (Greenhouse & Geisser, 1959). Interactions were followed up with separate pairwise comparisons for the factor Condition within four quadrants: left-frontal (F7, F3, FC5, FC1), right-frontal (F4, F8, FC2, FC6), left-posterior (CP5, CP1, P7, P3) and right-posterior (CP2, CP6, P4, P8).

Results Accuracy: There were no significant differences in accuracy between the two conditions (see Appendix B.1, Table B.4).

EEG data: Visual inspection of the ERP waveforms time-locked to the final noun

(see Figure 7.3) revealed more negative going amplitudes for hand-subject gender mismatches compared to matches starting between around 300 and 500 ms. Between 500 and 900 ms, hand-subject gender mismatches also showed more positive going amplitudes compared to matching conditions. Topographical distributions of hand-subject gender match effects for both time windows can be seen in Figure 7.4. At first glance, effects in both time windows seem to be quite broadly distributed and more pronounced in posterior areas.

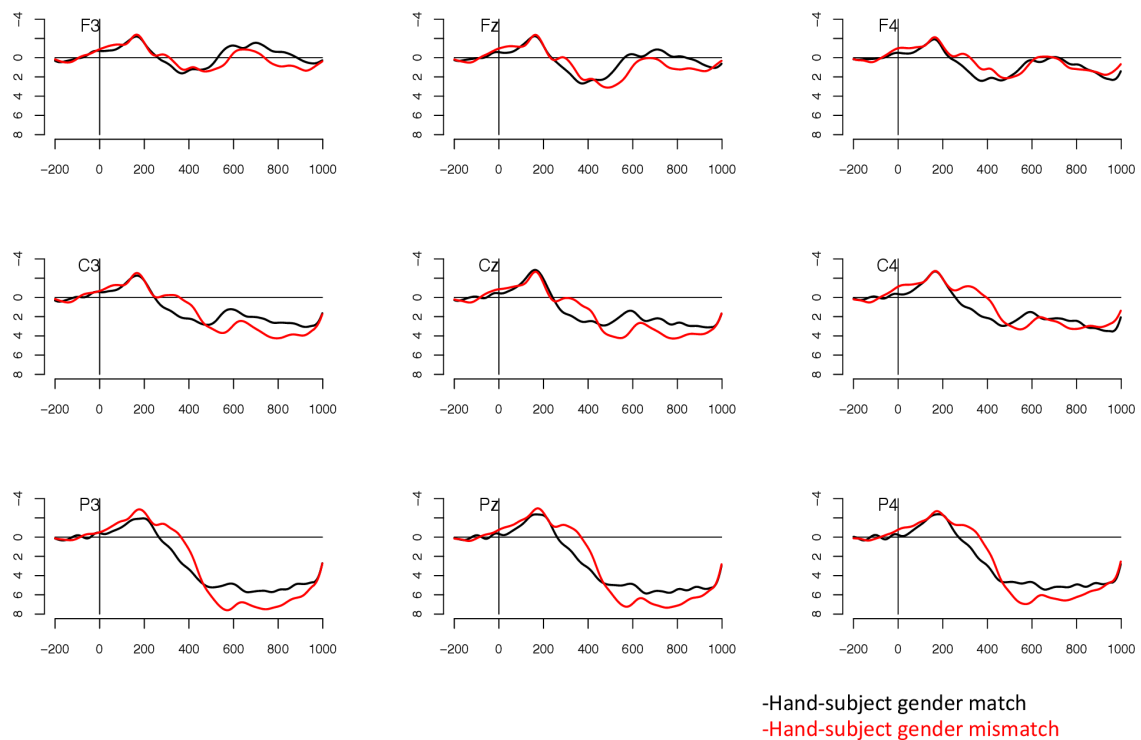


Figure 7.3.: Grand average ERPs (mean amplitude) for 9 electrodes (3 frontal, 3 middle and 3 posterior) time-locked to the final noun (NP2).

300- 500 time window. Analyses in this time window revealed a significant main effect of Condition, $F(1,15)=7.40$, $p<.05$, $\eta^2=.330$, but no interactions.

500- 900 time window. Analyses in this time window revealed a marginally significant effect of Condition, $F(1,15)=3.49$, $p=.08$, $\eta^2=.189$, and an interaction between

Condition and Hemisphere, $F(1,15)=8.88$, $p<.01$, $\eta^2=.372$. There was also a marginal interaction between Condition and Anteriority, $F(1.67, 25.14)=2.83$, $p=.085$, $\eta^2=.159$. Pairwise comparisons on the quadrants revealed a significant effect of Condition at the left-posterior area; $t(15)=-3.30$, $p<.01$, and a marginal effect at the right posterior area; $t(15)=-1.95$, $p=.07$.

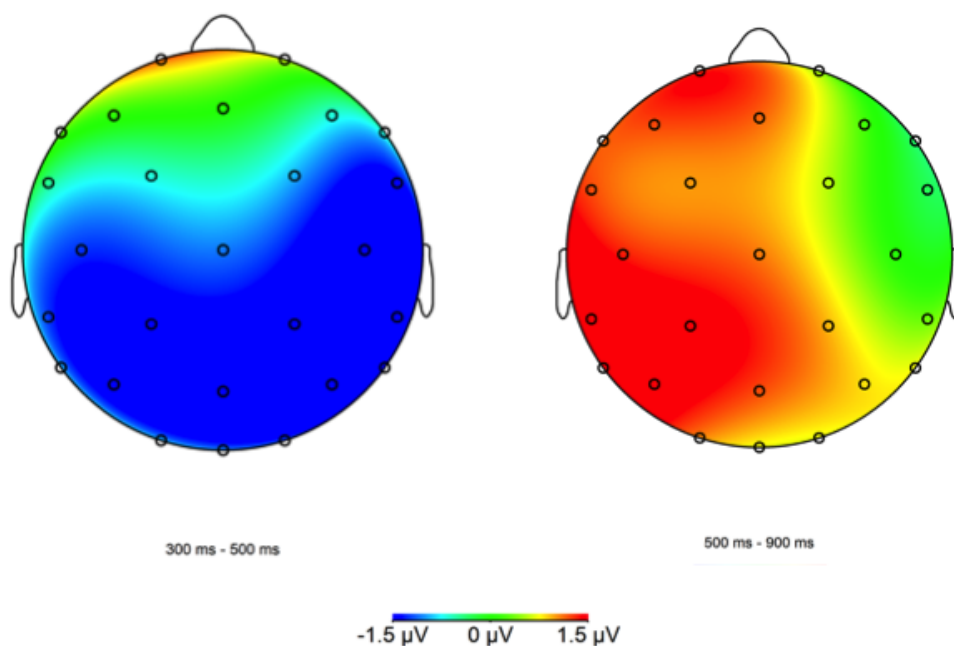


Figure 7.4.: Grand average ERPs (mean amplitude) across the scalp at the final noun region (300-500 and 500-900 time windows), obtained by subtracting the matching condition from the mismatching condition.

7.1.3 | Discussion

Mismatches in the verification of visual gender cues with subsequent referring expressions during comprehension elicited clear effects at the 300-400 and 500-900 time windows. Latencies of those effects had a somewhat earlier onset than those typically found for N400 and P600 effects. However, this might be due to the fact that auditory presentation

of the stimuli sometimes induces faster brain responses in comparison to written stimuli (e.g., Holcomb & Neville, 1990, 1991; Osterhout & Holcomb, 1993). Although the effects seem to be broadly distributed, a closer visual inspection of the scalp distributions as well as the interactions we found between hand-subject gender match and some electrode sites, suggests that amplitude differences were more pronounced at posterior compared to anterior sites, similar to the ones usually found in linguistic contexts (Hammer et al., 2008; Xu et al., 2013, but see Schmitt et al. 2002). Although both N400 and P600 components have been a focus of interest in countless studies on sentence processing, not all studies have shown a biphasic response like the one in the current study. In principle we could, as predicted, ascribe the N400 effects to the violation of semantic expectations (i.e., about biological gender) at the final noun (i.e., NP2) based on prior visual gender cues. However, the appearance of the P600 effect in the current study, and most particularly its appearance after an N400 effect, makes the interpretation of the results more difficult. Unlike for studies making use of anaphoric expressions, we cannot relate the P600 effects obtained in this study to morphosyntactic violations, as such a manipulation was absent in this experiment. Studies like the one from Wassenaar and Hagoort (2007), and even Experiment 1 from Knoeferle et al. (2014) did find what could be considered P600 effects in their picture-sentence verification studies, which they related to thematic role assignment processes. Still no such interpretation could be drawn from our data, as we did not manipulate role relations (the final subject, albeit from a different gender, was still the agent of the events).

Vissers et al. (2008) did, like in the current study, find a biphasic response after picture-sentence mismatches about locative relations between objects (e.g., *The triangle stands in front of the square*). While the N400 effect was not predicted in their study, they explained it in terms of a violation of high semantic expectations. The (predicted) P600 effects, on the other hand, were explained in terms of the Monitoring Theory (Kolk et al., 2003; Van De Meerendonk, Kolk, Vissers, & Chwilla, 2010; Vissers et al., 2008). They argued that the P600 effect does reflect reanalysis processes, but not necessarily related to syntactic properties of the language. Rather, under this account the P600

would represent a more general *monitoring mechanism* that has to deal with two different information sources, one coming from the representation of the picture, and the other from the representation of the sentence. This monitoring mechanism would arguably take care of checking whether the initial interpretation of the sentence has been correct. Applying this interpretation to our results, we could say that visually grounded gender information establishes a communicative context that generates high expectations for sentence verification. During comprehension, facing a noun (i.e., NP2) that defies such semantic expectations makes the processor launch a first "warning flag", reflected in the N400. A second stage in processing would be devoted to the confrontation of both the gender representation that created the semantic expectation (prior visual events) and the gender representation based on the gender implied by NP2 in the sentence. Failing to reconcile both representations would give rise to the P600 effect.

The interpretation for the biphasic response given by Vissers et al. (2008), despite being somewhat ad hoc, might be a plausible explanation for our results. However, as Van Petten and Luka (2012) pointed out, rather than costs from an erroneous expectation, N400 effects could in turn be measured in terms of how easily a word is incrementally integrated based on prior context. A recent proposal to describe the functional properties of the N400 and the P600, provides yet another sensible explanation, namely, the Retrieval-Integration account (Brouwer, Fitz, & Hoeks, 2012; Brouwer & Hoeks, 2013). This account interprets N400 effects as an index of memory retrieval at a certain word as a function of semantic memory and prior context (i.e., to what extent the retrieval of the semantic features of a word are facilitated), while semantic integration would be reflected in P600 effects. Brouwer et al. (2012) argued that P600 effects can be explained in terms of the construction, revision or updating of mental representations from language, which uses retrieval information as an input. In this way, for our current study we could argue that the semantic features of the gender implied by NP2 were facilitated in cases where prior events had matching visual gender cues. In the case of a mismatch, difficulties to retrieve the semantic features from NP2 would be indexed by the N400 effect. As for the P600 effect, this would reflect additional processing costs associated with an increasing

difficulty in the construction of a mental representation based on the sentence, or an update of the existing representation based on the new linguistic input, i.e., in establishing the gender of the agent of the event described in language.

In essence, semantic gender representations from visual cues seem to be similarly generated and maintained as those from prior linguistic context during comprehension. Although not so common, biphasic responses to mismatches of biological gender have sometimes been seen in the literature of anaphoric resolution (Lamers et al., 2006; Schmitt et al., 2002). However, given the manipulation used for our experiment (i.e., the match between visual and linguistic cues), we cannot relate our P600 effects to syntactic reanalysis, which challenges traditional views of the P600 as exclusively indicating structural processing (e.g., Hagoort et al., 1993; Neville et al., 1991; Osterhout & Holcomb, 1992; Osterhout et al., 1994). The above interpretations of the P600 need to be taken carefully, as P600 effects have also been related to sentence-end wrap up effects, which might obscure the local effects of our manipulation (e.g., Hagoort, 2003; Osterhout & Nicol, 1999). This observation notwithstanding, our results suggest that the conciliation of visual (event-based) and language-based gender information during sentence comprehension may require two semantically-induced processing stages⁵.

In the current experiment we have not directly addressed the role of stereotypical knowledge and how this could modulate the results. Although the eye-tracking data from our previous experiments did not always provide evidence in favour of the use of stereotypical knowledge (but see Experiment 3), ERPs might reveal covert processes and changes of cognitive states not visible in the eye movements. An increasing number of studies is beginning to study the role of stereotypical gender knowledge in sentence comprehension using ERPs (e.g., Kreiner et al., 2009; Molinaro et al., 2016; Osterhout et al., 1997, White, Crites, Taylor, & Corral, 2009). Kreiner et al. (2008), for example, argued given their results that the processing of definitional and stereotypical nouns was qualitatively similar, but might have graded differences at a representational level. Manipulating

⁵The use of the word *stage* in this context does not necessarily relate to the steps outlined in the CIA. Such a comparison remains to be performed.

determiner-noun gender match, Molinaro et al. (2016) found N400 effects for grammatical, yet stereotypically incongruent combinations (e.g., *las mineras*, ‘the miners_{fem}’) with a different topography from the classic N400 distributions (anterior instead of posterior), which led them to the conclusion that there might be a difference between how we process (gender) stereotype knowledge and other types of semantic information (but see Kutas & Federmeier, 2011, who claim that this account might not be viable). Further research on this direction is therefore necessary to see whether gender information has distinctive prints in our brain compared to other relations between visual cues and events and language, and whether it is the case that stereotypical and other aspects of semantic knowledge affect comprehension processes differently.

8 | General discussion

In this thesis, we examined the influence of prior event-based representations and language-based long-term knowledge, by means of visual gender cues embedded in gender-stereotypical actions and (OVS) sentences. In order to do so, we measured participants' visual attention over gendered agents (i.e., one female and one male) during the verification of sentences more or less related to prior events (Experiments 1 to 3), as well as the ERP responses of participants while these verified the gender cued in prior events with the gender implied by the proper noun from the sentence (Experiment 4). Previous research exploring the influence of prior events on situated language processing (Abashidze et al., 2014; Knoeferle, Carminati, et al., 2011; Knoeferle & Crocker, 2007) has found that during comprehension, event-based representations (i.e., representations stemming from prior/recent events) have a strong influence on how participants predict and establish reference with entities in a scene, with little to no influence coming from long-term knowledge associated with language alone (i.e., the semantic or world-knowledge that comprehenders would use if prior events were not present).

Although the preference for event-based representations, also called *recent-event preference* (Knoeferle, Carminati, et al., 2011), is robust, the exact nature of the predominance of event-based information over long-term knowledge still remains unresolved. For instance, questions regarding the relative strength of both sources of information do still need to be addressed (i.e., does this preference for prior/recent events always apply or are there situations in which long-term knowledge takes precedence even if both information sources are present?). In fact, it has also been found that certain manipulations, like

verb tense and frequency of post-sentential events modulate the extent to which entities from recent events are preferred over other potentially plausible candidates (Abashidze et al., 2014; Knoeferle & Crocker, 2007; Knoeferle, Urbach, & Kutas, 2011), although the preference for recently-seen entities replicated. Given the extensive literature on the effects of gender information in language processing, we reasoned that information on such a biologically and socially relevant aspect, which might be considered *inherent* to the comprehender and their understanding of the world, might have an important role during situated language comprehension.

Our experiment differed from prior research in essential points (i.e., we used OVS instead of SVO sentences, we measured anticipation towards potential agents instead of objects in the events, and we had a video-sentence verification task, which other studies did not use). These differences notwithstanding, our results support the idea that the preference for event-based representations generalizes to another type of information, i.e., gender features from the hands of an agent. During comprehension, participants generally preferred to look at the target agent (i.e., the agent whose gender was cued by the hands in prior events) compared to the competitor. Results further show that this recent-event preference does not just rely on representations of full objects, agents and events, but also subtler (gender) features that serve to identify feature-matching targets (i.e., faces of agents are inspected based on the gender features from hands seen in prior events). Importantly, we found some additional results stemming from two different types of manipulations, a) mismatches between (gender and action) representations from prior events and language and, under certain circumstances, b) mismatches between the stereotypical knowledge from language and event-based gender representations (i.e., the gender of the agent that was part of prior events). In what follows, we will address these effects more closely and we will describe their potential implications.

8.1 | Preference for prior visual cues

In line with prior research, our results support the idea that when recent-event information serves as a cue during situated language comprehension, representations from this visually grounded information gain greater importance over long-term, experiential knowledge when it comes to guiding our attention in a scene (Knoeferle, Carminati, et al., 2011; Knoeferle & Crocker, 2006, 2007). In determining who does what, we found that perceptual gender cues (i.e., the hands of an agent) were used early on during sentence comprehension to preferentially inspect one potential agent’s face over another from the opposite gender (Experiments 1, 2 and 3). As indicated by the grand means (i.e., the intercepts) from the log-probability ratios throughout the sentence and at different sentence regions, this preference stayed constant and significantly different from an at-chance level, (i.e., zero; $ps \leq .05$).

The preference for event-based representations as opposed to world-knowledge from the linguistic input in inspecting/predicting upcoming entities (i.e., the *recent-event preference*) has been taken as an *epistemic or cognitive bias* of the comprehender. As past events are verifiable during comprehension via visually grounded information, representations from those events are favoured over other plausible, yet less tangible representations of yet-to-happen events during spoken language comprehension (Abashidze et al., 2014; Staub & Clifton, 2011). In our experiments, given that participants had the explicit task of verifying the content of the sentence with prior events, it is not surprising that such a bias for retaining event-based representations was enhanced, as said bias would help in the verification task. This bias would be further supported by working memory (as accommodated in accounts like the CIA; Knoeferle & Crocker, 2007), as this component, given the sometimes limited capacity of the comprehender to concurrently process several types of information, will temporarily maintain the representations from recent (visual) events (McElree, 2006), leading to the anticipation of the most likely referents during utterance-mediated attention.

It is important to point out that a slight difference in the present experiments as compared to previous situated language comprehension studies is that prior events did not actually contain the same visual information as the visual scene during comprehension, nor did the experiment use blank-screen presentation (Altmann, 2004). While prior events featured the hands of an agent (i.e., subtler cues than whole body images), in the visual scene concurrent with language the faces of potential agents were shown (and no hands). Apparently, after gender categorization from hand information takes place, the gender-matching features are kept active and can be reconciled with the agent's face as the sentence unfolds. In this sense, working memory would not only contain representations of scene objects and events, but also subtler features of those elements, which may allow for a *spread of activation* towards further feature matching target elements that have not been seen before but have a relation with elements in prior visual events. Similar to cases where semantically related elements like a trumpet are fixated when a target word of the same category like *piano* is heard (e.g., Huettig & Altmann, 2005; Yee & Sedivy, 2006), or to how a pointer towards a location in a blank scene is kept in memory for later use during comprehension (e.g., Altmann & Kamide, 2004; Spivey & Geng, 2001), linking the gender features from one visual property to the other via anticipatory eye movements may assist language processing, arguably by *projecting* the recent action events onto the available visual entity, i.e., the agent's face (e.g., Spivey & Geng, 2001).

One caveat that might be considered in our eye-tracking experiments, as we did at the end of Chapter 5, is that of anticipatory baseline effects (ABEs). We tried to minimize these effects by removing fixations that started prior to sentence presentation from our analyses, and had more intervening material (i.e., objects) for participants to inspect in the target scene of our third experiment (something that is believed to reduce ABEs). Nevertheless, time-course graphs indicate that log-probability ratios departed from an at-chance level, favouring the target agent, already prior to sentence presentation (i.e., the onset of NP1), regardless of the experimental condition. The nature of the task (i.e., paying attention to details in prior events to verify them with language) may accentuate this tendency. This is a common issue in visual-world studies in which atten-

tional processes during comprehension are measured as a function of prior context (see Barr et al., 2011, for a discussion on ABEs). However, as it has also been argued, these ABEs, although not always desired, can still be interpretable and informative. First, they serve as an indicator of participants paying attention to prior context, which is usually indispensable in these type of designs. Second, they can still reflect participants' ability to link feature-matching cues, as well as their expectations during comprehension, which however early, are in line with our experimental predictions. The fact that our experimental manipulations (i.e., video-sentence match and stereotypicality match) did have an effect on participants' eye-movement behaviour, further suggests that participants did not ignore the linguistic input (i.e., mismatches between prior events and language reduced the initial target agent preference). Not only that, ABEs did not prevent participants from recruiting information that was neither explicitly required for the task, nor strictly necessary for integrating the visual and the language-based event representations, as it happened in Experiment 3. In this experiment, stereotypical gender knowledge modulated the target agent preference, but crucially not at NP1, where such effects could have also been confused with ABEs. The results provide strong support for the active role of language during its reconciliation with event-based representations.

8.2 | Mismatch effects

However robust the preference for event-based representations is during comprehension, we further saw that this preference is not invariant. The preference for looking at the agent whose gender features matched recent visual cues was modulated by mismatches in language, i.e., whenever the actions described or the gender implied by the final noun were at odds with prior events. Recall that we used OVS sentences to explore participants' anticipatory processes towards the gendered agents concurrent with language. While in Experiments 1 and 3 mismatches were encountered prior to the resolution of the final agent's name (i.e., the initial verb-phrase), in Experiment 2 (and Experiment 4) mismatches took place at the final noun (i.e., the proper name).

At least for the former two cases, where eye-movement behaviour towards the agent can be considered anticipatory, following the cognitive bias for event-based representations at the time of establishing visual reference, an additional heuristic strategy might apply, namely, *anchoring* (Tversky & Kahneman, 1973). The anchoring hypothesis (Tversky & Kahneman, 1973) broadly describes a rule of thumb by which at the time of making an estimation of some sort, individuals will rely heavily on a previously considered estimate (i.e., an ‘anchor’). In the context of visual attention during comprehension, a referential anchoring hypothesis (Dumitru et al., 2013) would predict that early mismatches between the visual and the linguistic domains will affect the extent to which upcoming entities are expected, consequently influencing how visual reference towards these entities is established. Applying the referential anchoring hypothesis to Experiments 1 and 3, one might argue that the level of attention towards a gendered (target) agent in an anticipatory manner would be affected by the match between the prior action events that agent was involved in and the verbal information processed before subject resolution in the sentence.

Indeed, results suggest that the inspection of a gendered agent, albeit still maintained to a certain degree, significantly decreases. This mismatch effect persists even during the final noun, i.e., when the subject is revealed (in Experiments 1 and 3, the gender implied by the proper name in final position of the experimental sentences did match the gender cues from prior events). As for Experiments 2 or 4, mismatch effects could in principle be more closely related to strictly referential processes. However, mismatches at the final region in Experiment 2 did not result in a shift of attention towards the competitor agent (i.e., in hand-subject gender mismatching conditions, the agent whose gender was implied by the final proper name), as a referential account would predict. Just like in Experiment 1, the preference for inspecting the agent from recent events, even if significantly reduced upon the encounter of a mismatch, was maintained. This suggests that in both cases, it might be the case that a discarded/residual representation of prior events is kept in memory, interfering with the anticipatory/referential processes taking place during situated language comprehension.

In Experiments 1 and 2, we saw that in a concurrent context where only the faces of the two potential agents are present, mismatches taking place early on in the sentence (i.e., at the verb-phrase region, Experiment 1) elicited effects with a slight delay relative to the onset of the mismatching region (i.e., the initial object noun) as compared to the effects elicited by mismatches at the final region (i.e., the final, proper name, Experiment 2), which were immediate. One could argue that it is simply the non-canonical word order (i.e., OVS) that causes the delayed mismatch effects at the beginning of the sentence in Experiment 1 (Kamide, Altmann, & Haywood, 2003) and that the comprehender might have tried to integrate both object and verb information before reconsidering the weight of their initial expectations regarding the agent.

However, as suggested by the results obtained in Experiment 3, how the visual scene concurrent with language is configured does seem to also influence how rapidly mismatch mechanisms are put to work when these are anticipatory in nature. In Experiment 3, together with the faces of the potential agents of the events, pictures of objects were included, one of which appeared in the prior event (i.e., the target object) and the picture of another object which in cases of video-sentence mismatches could be referred to by language, and was part of an action with the opposite stereotypical valence (i.e., the competitor object). Constraining the visual scene in this way, anticipatory processes may become more narrowly focused, helping comprehenders to retain a more vivid representation of the prior event. Also, by virtue of having additional contrasting information in the visual scene (target objects and agents vs. competitors), interpretation and conciliation of the language-based and visual representations might take place more actively. Indeed, attention towards the target agent's face was significantly reduced as soon as the object name (or theme) at the initial position of the sentence mismatched prior events, and this effect once again persisted throughout the sentence.

Worth mentioning is how participants inspected the objects in Experiment 3 as compared to how the agents were inspected in Experiment 2 (i.e., the one with the hand-subject gender manipulation). Objects captured a considerable amount of attention from

the beginning of the sentence, and we also found action-verb match effects at NP1, just like we did for the agents at NP2 in Experiment 2. These effects on the objects were however more consistent with the referential account than were the effects involving the agents in Experiment 2. While in Experiment 2 the target agent was still preferred even if the gender implied by the final noun was mismatching, in Experiment 3 competitor objects were inspected as they were mentioned (note that the objects in the mismatching conditions were fixated with a slight delay). The negative going log ratios (i.e., reflecting a competitor object preference) at NP1 for the action-verb mismatching conditions support this interpretation (see Figure 6.10). An *ad hoc* explanation behind this difference between agents and objects might be that as the latter are more concrete, and only two object images are in the scene, it is easier to establish reference to them. Looks towards the agents (i.e., to their faces) based on prior gender feature information (i.e., the hands) may require further inferential processes. That together with the possible interference from residual representations of the prior event could make it less likely for participants to inspect the competitor agent in hand-subject gender mismatch conditions.

In an attempt to further study the potential mismatch mechanisms underlying the effects found in eye-tracking, we implemented the design from Experiment 2 (i.e., where mismatches were found between the gender cued in prior events and the final proper name of the sentence) in an ERP study (Experiment 4). In broad terms, in Experiment 4 we found a semantically induced biphasic response (i.e., N400 and P600 effects) to mismatches compared to matches. Knoeferle et al. (2014) had already investigated different types of mismatches between visual events and language (i.e., thematic role and verb) in order to provide further evidence that could enrich situated language processing accounts like the CIA (Knoeferle & Crocker, 2006, 2007). The authors discussed their ERP results in terms of a comparison between the different distributions in the N400 effects found for visiolinguistic mismatches in thematic role vs. verb relations (more central vs. more posterior), which were ascribed to distinct mismatch mechanisms involved during situated language comprehension. Interestingly, however, when taking a closer look at their results, in their first Experiment (with a 500ms word onset asynchrony and a word

duration of 200ms), role mismatches compared to matches also seemed to elicit a posterior positivity following an anterior negativity from the onset of the initial noun. This positivity started at around 400ms after the onset of the region (and extended to the beginning of the verb).

Unlike in their study, the negativities observed in our Experiment for gender mismatches had a posterior distribution; however, the biphasic response obtained in our study does resemble the results elicited by role mismatches. In Knoeferle et al. (2014)'s study, P600 effects were associated with a potential structural revision elicited by this type of mismatch (similar to the thematic role assignment processes suggested in Wassenaar & Hagoort, 2007); by contrast, no such interpretation can be drawn from our ERPs in Experiment 4, as no structural manipulation was made. If we were to relate our gender verification study to thematic role relations, we might need to abandon the idea of P600 effects as implying structural/syntactic processes. In Chapter 7 we discussed the results of Experiment 4 on the basis of two different, yet not entirely exclusive models for language comprehension, namely, the Monitoring Theory (Kolk et al., 2003; Vissers et al., 2008) and the Retrieval-Integration account (Brouwer et al., 2012; Brouwer & Hoeks, 2013). The main difference between these two theories is that the latter can better accommodate biphasic responses (the Monitoring Theory would have predicted P600 effects in our experiments as reflecting conflicting visiolinguistic representations, but it is not so clear about the N400 effects). By contrast, Brouwer et al. (2012)'s account might explain the data, as a function of retrieval-integration (i.e., N400/P600) cycles. Our eye-tracking data (e.g., Experiment 2) already suggested that the gender cues from prior events do create a disposition (i.e., anticipation) for retrieving and integrating the agent of a certain gender during language processing. If indeed a retrieval of the gender features of the agent was to take place when processing the final proper noun, a clash between *preactivated* gender features from prior events (as suggested by the anticipatory eye movements towards the target agent early in the sentence) and those retrieved at the final noun position would be indexed by the N400 effects in the ERP data. As for the P600 effects, after the initial difficulties indexed by the N400, the conflicting representations from prior events would

need to give way to an attempt for integrating the newly acquired information, which would result in an integration effort, indexed by greater positive-going amplitudes for mismatches compared to matches.

Although the idea that language comprehension takes place in retrieval-integration cycles is theoretically sound and compatible with our results, not enough testing of the Retrieval-Integration account incorporates the visual domain (i.e., picture-sentence verification studies or situated language comprehension studies). Furthermore, although the Retrieval-Integration account makes predictions regarding the magnitude of effects depending on the type of word being processed (e.g., the account predicts more pronounced responses for linguistic phenomena involving open vs. close class words; Brouwer & Hoeks, 2013), it would need to further explain the different scene-based mismatch effects for thematic role vs. action information, which seemed to elicit biphasic vs. single ERP responses, respectively (Knoeferle et al., 2014).

8.3 | Contribution of stereotypical gender knowledge

Echoing the preference for event-based representations or the mismatch effects between such events and language, changes from Experiment 1 to Experiment 3 in the configuration of the visual scene during comprehension (both experiments manipulated action-verb match) did give rise to additional findings. The new configuration, where not just the images of potential agents, but also objects were available in the visual scene concurrent with language, led to a pronounced contribution of stereotypical gender knowledge in modulating the comprehender's visual attention towards a target agent.

At the verb, in addition to action-verb match effects already present in the previous region, we also observed stereotypicality match effects¹, i.e., participants looked at the

¹Although fully significant in the ANOVAs, some of the stereotypicality match effects were rather marginal in the Mixed Models analyses (see Appendix B.4.).

target agent to a greater extent when the action described by the sentence stereotypically matched the gender cue (i.e., the hands) from prior events (i.e., if female hands were shown and the sentence was about a stereotypically female action, participants tended to attend to the female face over the male face to a greater extent than when the sentence was about a stereotypically male action). In the absence of interaction effects, we could say that we observed (super-)additive, rather than interactive effects of the two manipulated factors: action-verb match and stereotypicality match. When both the linguistic input matched prior events and the stereotypical content of the action described matched the agent favoured by event-based representations (i.e., the target agent), this target character was anticipated to a greater extent compared to the other conditions during comprehension, at least numerically (see Figure 6.5). Also when both cues were incongruent (i.e., there was a mismatch between the action described and prior events, and the action described was stereotypically incongruent with the gender cues from prior events), the preference for the target agent did seem to be cancelled out (i.e., the mean log-probability ratio for the fully mismatching condition in this experiment seems to hit negative values, suggesting a competitor agent preference, however, these values are close to an at-chance level, i.e., zero).

Why is it that under the latest target scene configuration, unlike in our previous experiments (i.e., Experiments 1 and 2), stereotypicality effects emerged? Although we cannot completely discard the idea that gender stereotypes might have been activated all along in our experiments, the fact that their effects were only evident in our third experiment suggests that at least in situated language comprehension, gender stereotypes may not always be automatically used, as suggested by other psycholinguistic studies (Banaji et al., 1993; Bargh et al., 1996; Duffy & Keir, 2004). The use of gender stereotypes might rather be context-dependent; i.e., motivated by the constraints present in the visual scene.

At least two (not necessarily orthogonal) reasons could be behind the effects obtained in Experiment 3. On the one hand, it might be that the richer a visual context

is, and the more contrasting entities (i.e., competitor or distractor agents and objects) are present at the time of comprehension, the more engaged a comprehender will be during situated language processing, to the point of recruiting additional, world-knowledge information in order to reconcile representations from prior events and language. On the other hand, it could also be argued that rich visual contexts pose greater cognitive demands than those that have more simplistic setups (as it might have been the case in Experiments 1 and 2). In order to ease the task, stereotypical knowledge might have come to the surface. Some studies in the social psychology field have claimed that information processing tends to be easier in those cases where stereotypical information is present, and stereotypical knowledge can be used to increase efficiency in certain cognitive activities (Andersen, Klatzky, & Murray, 1990; Macrae, Milne, & Bodenhausen, 1994; Sherman, Lee, Bessenoff, & Frost, 1998; Sherman, Macrae, & Bodenhausen, 2000). For instance, it has been reported that participants' reading or reaction times are shorter when processing stereotype-consistent compared to inconsistent information with regards to an individual (e.g., a 'skinhead' or a 'priest'), particularly in cases of cognitive load, i.e., when participants performed an additional, unrelated task while reading/forming impressions (Sherman et al., 2000).

It might actually be the case that participants used gender-stereotype knowledge in our experiment to execute the verification task efficiently; however, this knowledge did not elicit reaction-time differences (i.e., stereotypically congruent vs. incongruent conditions had similar response times). Nor did gender-stereotype knowledge suffice to "turn the tables" in the eye movements, i.e., the competitor agent did not attract more visual attention than the target agent as a function of stereotypical knowledge. This type of stereotypical knowledge might not be strong enough to override representations from events that listeners' have recently witnessed. At best, we can say that they might use gender-stereotype information in certain contexts to enrich event-based representations, in so far as the stereotype conveyed by language is congruent with the cues from such an event. If incongruent, gender-stereotype knowledge might disfavour event-based representations to some degree, even to the point of hindering them during situated language

processing. All in all, it might be the case that in situated language comprehension, certain contexts activate inferences of gender stereotypes, which may be elaborative (i.e., not strictly necessary, but still used for the task in hand; Garnham et al., 2002; Pyykkönen et al., 2010).

In addition to the effects encountered in the eye movements to the agents, in Experiment 3 gender stereotypicality knowledge did also seem to have an influence on establishing visual reference with objects as these were mentioned (i.e., at the beginning of the sentence, NP1)². The interaction between verb-action match and stereotypicality (see Figure 6.10) likely came about because the effects of stereotypicality were visible inasmuch as the linguistic input matched prior events (i.e., in the action-verb matching conditions). This might mean that instead of gender-stereotype knowledge stemming from language, gender-stereotype knowledge from prior events themselves might have guided attention, (i.e., the gender cues from prior events could have *primed* or at least facilitated establishing reference with objects as a function of stereotypical knowledge). Although this finding was not central for the aims of the current thesis, and prior research has already studied the influence of stereotypical knowledge in relating words for objects and gendered characters or names (e.g., Leinbach, Hort, & Fagot, 1997; Most, Sorber, & Cunningham, 2007), this is, to our knowledge, the first study showing differences in establishing visual reference with objects as a function of prior events containing gender and action cues and the stereotypical knowledge stemming from them.

8.4 | Implications for accounts of situated language comprehension

The findings reported in this thesis serve to inform extant models and accounts of situated language comprehension such as the ones introduced in Chapter 4 (e.g., Altmann & Mirković, 2009; Dienes et al., 1999; Knoeferle & Crocker, 2006, 2007; Knoeferle et al.,

²Stereotypicality match effects were not reliable in the Mixed Models analyses.

2014; Münster, 2016). As previously discussed, the CIA (see Figure 4.2) can account for phenomena such as the different temporal dynamics of visual and linguistic domains (e.g., whether scene and language are presented simultaneously and/or whether prior event information has been presented) and mismatches between the representations derived from both (Knoeferle et al., 2014), something we also examined in this thesis. What the CIA has yet to explain is the relative strength of the different sources of information (e.g., information coming from recent perceptual experience vs. long-term knowledge) when listeners generate expectations and anticipate entities in a scene, something we will discuss next.

To start with, based on the mismatch effects obtained both in the eye-tracking studies as well as in the ERP experiment, we certainly believe that, as suggested by Knoeferle et al. (2014), situated language comprehension in the context of prior visual events may need a verification mechanism, a mechanism that *flags* different sentence regions when conflicts occur during comprehension, and that tries to reconcile representations from visual and linguistic sources. The overall preference for visually grounded information found in prior research and replicated in our eye-tracking studies (Experiments 1 to 3), further supports the idea that, even in the case of a mismatch, a discarded visual representation of prior events is still operative in working memory, which would also serve to index the *truth value* of the interpretation derived from language, and to support the overt responses from the comprehender if needed (e.g., button-press after listening to the sentence). Whether there should also be a parameter specifically designed, as suggested in the last version of the CIA, to signal different subprocesses during situated language comprehension (e.g., verification of thematic roles, actions or gender features), depends on our interpretation of the distinct ERP mismatch responses encountered in Experiment 4 as well as in Knoeferle et al. (2014). Although Knoeferle et al. (2014) argued that the differential effects obtained for thematic role relations and action mismatches between the scene and language may indicate different processing mechanisms in situated language comprehension, we cannot entirely exclude the idea that a single mismatch mechanism exists. The different stages of such a mechanism might manifest themselves differently

depending on the type of information being processed, perhaps in a fashion on the lines of the Retrieval-Integration account (Brouwer et al., 2012). However, in order to figure this out, further research exploring different visiolinguistic relations might be needed, to see which type of information elicits which sort of response and whether all responses can be accommodated within a single-mechanism account.

Our findings moreover have implications for how expectations are generated based on prior representations and world-knowledge during comprehension, which in the CIA is instantiated in the *ant* parameter. Given our findings, we have reasons to believe that although some predictive cues like event-based representations have more weight than others when generating expectations, more fine grained influences are also involved. Although refraining ourselves from relabelling the *ant* parameter from prior versions of the CIA, we think it is appropriate to adopt the modification suggested by Münster (2016) in the sCIA. Recall that in this version of the account, the social characteristics of the comprehender (i.e., age) as well as perceived social information from depicted events (i.e., emotion) further contribute to situated language comprehension. The *ant* parameter (named *ant_s*) was implemented by means of a probabilistic weight indicating how strong a particular expectation is. The probabilistic weight of *ant* was instantiated via a subscript *p* (range 0 to 1). Several factors can determine the weight of *p*, and many of them will likely comprise information coming from social (and also biological) aspects³, which may apply to different parts of the communicative context (from the comprehender, to the speaker, as well as the informational content), e.g., age, gender or race (Münster & Knoeferle, 2018).

Although we did not observe a sufficient amount of gender differences in our participants to draw firm conclusions, we can agree that, as proposed by Münster (2016), *the properties of the comprehender* (which in the sCIA was instantiated in *ProCom*) can be a

³Münster (2016) and Münster and Knoeferle (2018) relabelled the *ant* parameter as *ant_s* from ‘social knowledge’. However, given that this implementation in the CIA does implicitly accept other contextual factors affecting the weighting of *ant*, like incongruencies between event-based and language-based representations in our case, we don’t see it necessary to adopt this new label, but acknowledge the contribution of social information.

relevant factor at the time of generating expectations during situated language processing. The listener's identity, their more or less stable features and cognitive abilities may lead to differential comprehension and attention patterns. Age differences, for instance, seem to have an influence on how visual and linguistic stimuli comprising emotional valence are processed, i.e., age differences result in distinct positivity/negativity biases (Langeslag & van Strien, 2009; Reed & Carstensen, 2012), which seem to affect the readiness for anticipation of thematic role fillers depending on the emotional valence of the adverb used.

Another factor that might influence expectations during situated language comprehension is *the characteristics of the speaker* (who utters the linguistic input). Indeed, it is generally assumed that the comprehender often considers the speaker's perspective and adopts their point of view to arrive at a successful communication (Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Ryskin, Wang, & Brown-Schmidt, 2016). More specifically, evidence from studies like the one from Van Berkum et al. (2008) and Hanulíková and Carreiras (2015) suggest that mismatches between the linguistic input and speaker's identity lead to disruptions in processing (e.g., listening to a sentence like 'Every evening I drink some wine before I go to sleep' in a young child's voice will elicit similar ERP responses at the word *wine* to those usually found for semantic anomalies). The identity of the speaker may trigger expectations in the comprehender in certain situations and affect their attentional processes accordingly (e.g., children may drink juice or milk, but not typically wine, therefore, in a visual setting it would be more likely that a person would visually anticipate any of the former drinks, but not the latter, if a child was talking). If the characteristics of the comprehender need to be specified in accounts of situated language comprehension, so should the characteristics of the speaker.

Third, a central factor for which we gained additional evidence in the current work is that of *the content of the information being processed* (i.e., the type of visual and linguistic content being processed in context) and its relevance for the comprehender. The content might include information about events, in which objects, actions and individuals

may be involved. The weight with which a particular expectation is generated during language comprehension may depend on how relevant or familiar that information is for the comprehender, as well as how the environment is configured. Biological and social features inherent to individuals like gender, race or age may once again be of importance to generate distinctive expectations and stereotypical biases⁴. However, the activation of certain types of long-term knowledge, such as stereotypical gender knowledge, may be context-dependent, i.e., the sensitivity of *ant* towards (gender) stereotypes, for instance, may only be activated under certain (visual) contexts.

8.4.1 | Example: Gender information

We have tried to make the point that gender is part of the informational content that the comprehender is particularly prone to extracting in visual events during language processing in the visual world. Both explicit (as when a gender pronoun is uttered or dimorphic visual information is provided) and implicit cues (inferences drawn from gender-stereotype knowledge) may be used, even when these are not necessary for interpretation, e.g., to establish discourse coherence (e.g., Bojarska, 2013; Pyykkönen et al., 2010). This does not need to be different for accounts on situated language comprehension, like the CIA. Based on the evidence found in previous studies and our own, we could say that explicit cues like the ones provided in prior visual events, if present, may be predominant when determining expectations towards agents during processing. In other words, if we were to determine the influence of such a cue on an expectation parameter like *ant*, the gender features extracted from dimorphic cues in prior events, which together with the objects and actions form the event-based representation, will assign a weight, favouring the feature-matching agent.

As we just mentioned, this is not all there is to it, at least not for gender. Based on our results, we can say that the weight of *ant* towards a particular agent can be

⁴Individual differences among comprehenders pertaining to the small cultures they might belong to may also be of relevance here (i.e., the "small social groupings or activities wherever there is cohesive behaviour"; Holliday, 1999)

further affected by two other manipulations. One manipulation is the match between event-based and language-based representations at the time of verifying language with prior visual events. The other manipulation, although its use is more context-dependent, pertains to the match between the stereotypical gender knowledge derived from language and the agent whose gender features have been cued in prior events. Although those manipulations did not change the course of the expectations during comprehension (i.e., expectations seem to mainly stay oriented towards the entities from the event-based representation), it was still significantly modulated. When any of the two manipulations does not support (i.e., mismatches) the event-based representation, the weight of *ant* might decrease, weakening the expectations that a particular agent would be mentioned during comprehension. This in turn will decrease the amount of fixations directed towards that agent. While effects derived from video-sentence mismatches seem to be quite persistent across our Experiments, it is not until Experiment 3, with an enriched concurrent visual setting, that we see effects of both experimental manipulations. What difference does this make for *ant*? If again we were to take the CIA as a template to explain our findings, the answer may lie both in the *working memory* component, as well as the concept of *decay* of the representations from prior events, affected by the configuration of the visual scene concurrent with language.

When a visual scene concurrent with language is simple enough (i.e., when visual constraints are few and cognitive demands low, as is likely the case when only the agents' faces are present, as in Experiments 1 and 2) prior events might suffer from a slight decay in working memory, but they can still be easily projected onto the available entities, and expectations are kept high throughout processing. When mismatches occur between the language-based representation and that of prior events, there will be a more pronounced decay, further decreasing the weight given to the *ant* parameter (as suggested by the referential anchoring hypothesis and evidenced in our findings). However, a residual weight driven by event-based representations might persist as a result of the cognitive bias of the comprehender to give preference to that information, and also as a basis for verification judgments. If the visual context concurrent with language provides a more

complex and constrained configuration, including additional contrasting entities (as in the case of Experiment 3; i.e., competitor objects not part of the prior event-based representations) the cognitive demands during processing may increase. This could arguably render event-based representations more sensitive to decay than in simpler settings, and particularly in cases where representations from language are at odds with event-based representations. Because the comprehender still has a cognitive bias to anticipate entities based on event-based representations and language, the system may be motivated to try and recruit additional resources to support anticipation based on such representations. Putting gender-stereotype knowledge from the linguistic input to work may provide such a resource, and sensitivity towards this type of long-term knowledge information may be activated in order to contribute in the weighting of *ant*⁵.

We will utilize the findings from Experiment 3, in which, unlike Experiments 1 and 2, we could argue that the context-dependent sensitivity of *ant* towards gender stereotypes played a role together with event-based and linguistic congruence. We will exemplify what may happen to *ant* upon the encounter of the initial noun and the verb in particular, for sentences like *Den Kuchen backt gleich Susanna* ('The cake bakes soon Susanna') or *Das Model baut gleich Susanna* ('The model builds soon Susanna'). Let's say we have a scenario in which, for instance, female hands were seen performing a stereotypically female action, e.g., baking a cake, and that the sentence will describe this same event (i.e., a fully matching scenario). The visual events will be tracked in the scene representation (scene i"-1). Right after, the visual context changes. Susanna and Tobias, who could be the potential agents of the event that just took place, are now present, as well as a cake and a toy model. Given the prior events, the system already has some gender features, an action, and an object activated in working memory (from scene i"-1) and will likely bias the *ant* parameter early on. However, given the current visual setting (i.e., where not just the target elements may attract attention), the event-based representation might be subject to decay as soon as sentence comprehension starts, and this might happen to

⁵Although many different expectations regarding the upcoming words (or visual entities) could be generated during situated language processing, the present examples will only focus on how the *ant* parameter may vary with regards to the agent as indicated in final position of the sentence (i.e., NP2).

a greater extent as compared to a simpler visual setting (e.g., where only Susanna and Tobias were present).

When the sentence starts with *Den Kuchen* ('The cake') (word_i), however, this is not a dramatic decay. The word will be interpreted (yielding int_i at step i) and, as the visual scene contains the cake, attention will be guided to that object in a referential manner after interpretation (step i)⁶, although some anticipatory attention might also start spreading towards Susanna (whose gender features match the gender cues from prior events), already reflecting anticipatory processes motivated by the event-based representation. By coindexing the initial object noun with the picture of the cake (step i), the event-based representation will be reinforced, and the *ant* parameter that would index anticipation of the upcoming agent (i.e., Susanna) will be weighted high (indexed by p). At the verb *bakes*, (i.e., after both the object, word_i, and verb, word_{i+1}, have been interpreted, yielding int_{i+1} at step $i+1$) the weighting of *ant* towards the agent, now ant_{i+1} , will be maintained or even increase (i.e., the value of p may be kept or be higher than before), arguably thanks to two sources, namely a) the match between the cake baking action stored in memory and the representation from language, and b) the congruency between the stereotypicality of the cake baking action and the expected agent (i.e., the one whose gender-features matched prior visual events, i.e., Susanna). Although the cake predominantly gets attention, high values at ant_{i+1} should be reflected in an increase of anticipatory eye movements towards the agent at step $i+1$. The verb representation will be coindexed with the object and, by virtue of the congruent visiolinguistic context and stereotypical information, the agent representations at step $i+1$.

Imagine, by contrast, that we were talking about a situation where prior events featured female hands building a model and the sentence described the model building action (i.e., stereotypically mismatching condition). *ant* will also favour the upcoming agent whose features match the gender cues from event-based representations as a basis of the match between these representations and those from language, and its weight will

⁶For the sake of simplicity and in lack of further evidence, we will refrain ourselves from discussing the stereotypicality effects found for the objects in Experiment 3.

also be high, always experiencing a slight decay due to changes in the visual context (from prior visual events to the concurrent visual scene). However, one of the two sources that may contribute to p is now contradictory, namely, the stereotypicality of the action (i.e., building a model) in relation to the expected agent. As one of the two sources is incongruent with the established expectation, by the time the verb is processed (i.e., *builds* at step $i+1$) ant_{i+1} won't be given as much weight as it would in a fully matching scenario, i.e., it might be reset to a lower value for p . Anticipatory eye movements to the agent will still be possible at step $i'+1$, as well as reconciliation of the verb with objects and agents at step $i''+1$, but less effectively than the previous case.

Now, let's say that the comprehender saw female hands baking a cake, but the sentence described a *toy model building event* (i.e., fully mismatching scenario). After the event has occurred, the visual context will again have Susanna and Tobias, as well as the cake and the toy model. This time again, the event-based representation (scene $i''-1$) will likely bias the *ant* parameter. But again, the current visual setting might make these representations subject to decay. When the sentence starts with *The toy model* ($word_i$), the word will be interpreted (yielding int_i at step i) and, as the visual scene contains the toy model, attention will be guided to that object in a referential manner after interpretation (step i'). By the time the word is coindexed with the image of the toy model in the scene (step i''), the value of the sentence to describe prior events will be flagged as false. Although there might still be a residual value for p towards the expected agent, the event-based representation it is based on (i.e., the cake baking action event) should have experienced a significant decay, triggered by a lack of reliance (i.e., anchoring) on this expectation. At the verb *builds*, (i.e., after both the object and verb information have been interpreted at step $i+1$, yielding int_{i+1}) the weighting for ant_{i+1} will likely decrease further. In this example, none of the sources (i.e., neither congruence between event-based and language-based representations, nor stereotypicality of the action in relation to the gender features of the agent from prior events) favour the anticipation of the event-based gendered agent. If anything, int_{i+1} might bias the

expectations somewhat towards the competitor agent in certain cases ⁷. The weight of ant_{i+1} towards the initially expected agent might be kept as minimal or non-existent. Anticipatory eye movements towards the agent, which might have originated prior to sentence comprehension, should decrease if already present, or they might not even take place.

If, however, we had the reverse situation in which female hands were seen as building a model and the sentence described a cake baking action (i.e., stereotypicality matching scenario), the ant parameter might be weighted differently at step $i+1$ (when the object and the verb have been interpreted, even if not fully reconciled with the scene). Say that the comprehender can no longer rely on a match between event-based and language-based representations in order to give a weight to ant . Unlike in the fully mismatching scenario, the residual representation of the gender features of the event-based representation might still be favoured by virtue of the stereotypicality congruency between the cake baking action described in language and the gender features from the hands in prior events (female in this example). The system might then opt to grant ant_{i+1} a greater weight as compared to the former, fully mismatching scenario.

8.5 | Conclusions

The findings from this thesis suggest that the preference for event-based representations in guiding anticipatory eye movements during situated language processing (a preference that has been replicated robustly in previous research) generalizes to yet another visual cue, i.e., gender cues from prior action events. However, our contribution pertains to two further aspects of the relation (or lack thereof) between representations from prior events and language, namely, the match at different points between event-based and language-based representations as well as the level of congruence between agents favoured by prior events and stereotypical knowledge from language. We reported evidence showing that

⁷Given that the negative log-probability ratios for this particular condition were not substantially different from zero, we cannot guarantee that this actually happens (see Figure 6.5).

when it comes to processing gender information, these two aspects can modulate the attentional behaviour of the comprehender during language processing, even when this behaviour is strongly biased by the event-based representations from working memory. Moreover, gender mismatches between prior events and language generate similar, yet slightly different neurophysiological responses to those elicited by thematic role or action relations, which invites for further exploration of the mechanism(s) involved in the process of conciliation between visually-derived and language-based representations. All in all, we have reasons to believe that not all kinds of semantic/world-knowledge-related information have the same impact in situated language comprehension, i.e., certain social aspects such as gender seem to have their own hallmark, even if their visibility may be context-dependent.

As mentioned during the discussion and exemplified above, studies like the ones reported in this thesis may significantly impact the current accounts of language processing which try to shed light on the interplay between language processing and visual perception. The evidence can give us fine grained insights into how the configurations of prior visual events and the concurrent scene, as well as the different aspects of semantics or world-knowledge are resorted to and how they are weighted when generating expectations. Biological and social aspects like gender may be the kind of information to target: as inherent properties to the human comprehender, they might be of particular relevance when processing information, and more resources may be devoted to exploiting them. Besides, because some of these properties (e.g., gender, age, race or even class) can arguably be instantiated at least at three different levels of the communicational context (comprehender, speaker and content), it seems pertinent to explore them, either in isolation or in interaction, and understand them and how they influence comprehension in relation to the visual world. After all, no proper model of language comprehension should be built in the absence of situational contexts conveying social information. Language itself is a social phenomenon, arguably evolutionary implemented for communication (e.g., Reboul, 2015; Scott-Phillips, 2014). Moreover, social knowledge is embedded in our everyday conversations, and it likely impacts comprehension in many different ways (by eliciting

mental representation of events, inferences and expectations, as argued throughout this work).

On another note, despite the fact that some researchers in the field of social cognition have long attempted to highlight the importance of language in their research area (e.g., Krauss & Chiu, 1998; Semin & Fiedler, 1988), little common ground has been given to psycholinguistics and social psychology. This is somewhat surprising, given that language is an important medium by which responses are elicited and recorded in most social psychology studies. Studies like the ones reported here could contribute to and benefit from the social sciences, by helping to uncover cognitive and perceptual biases when listeners understand and verify aspects from the world around them. Additionally, not only might this information be useful at the level of the typical comprehender, but also in atypical populations: apparently, children suffering from autism do make use of some aspects of social cognition, like social stereotypes which, though pernicious in some contexts, may be useful for these groups in trying to understand human behaviour (Hirschfeld, Bartmess, White, & Frith, 2007, White, Hill, Winston, & Frith, 2006). Moreover, there has been some evidence showing that in language comprehension, people with high functioning autism, who also tend to struggle with certain aspects of pragmatic language, do for instance show efforts of integrating linguistic content with speaker identity (e.g., gender, age, class) during sentence processing, even if they do so differently from their typical peers (Tesink et al., 2009). Therefore, studies like the ones reported in this work may have developmental, even educational implications. The path is full of possibilities at theoretical and applied levels.

For now, going back to that little boy we talked about in the introduction, we could say that watching him trying out just the pink bike will likely make us anticipate this specific bike while he is making his birthday-wish statement. Even if he ended up not choosing a bike but rather something else, like a ball, we might still entertain the idea of the pink bike being chosen by the boy for a while, as that image might have lingered in our minds, given its recency. However, depending on how busy the supermarket is

on that day, and probably even our own mental state, that idea may lose strength, as in these circumstances the idea of the pink bike may clash more easily with some of our firm conventions on what that kid might rather choose. Whether that blue bike next to the pink one remains at the back of our minds is rather unsure, but depending on how language develops in context it may still be there, as if waiting to be mentioned.

9 | German summary

Kommunikation im Alltag findet oftmals in reichhaltigen Kontexten statt, für die wir als Sprachbenutzer linguistische sowie nicht-linguistische Quellen nutzen. Vorangegangene Studien zur Satzverarbeitung in visuellen Umgebungen (d.h. *situiertes Sprachverstehen*) haben gezeigt, dass unser semantisches Wissen und unser Weltwissen (d.h. Langzeitgedächtnis) in einer Sprache, Augenbewegungen zu bestimmten Objekten und Personen leiten, auch wenn diese Aspekte noch nicht genannt worden sind. Zum Beispiel, löste ein visueller Kontext, der ein kleines Mädchen, einen Mann, ein Motorrad und ein Karussell zeigte, bei Versuchspersonen vorausschauende Blicke zu dem Karussell aus, wenn sie den Satz *The girl will ride...* ('Das Mädchen wird fahren...') hörten, im Gegensatz zu *The man will ride...* ('Der Mann wird fahren...'; Altmann, 2004; Kamide, Altmann, & Haywood, 2003). Allerdings ist es offensichtlich, dass wenn vorherige Ereignisse zur Verfügung stehen, z.B. wenn visuelle Darstellungen von Handlungen zwischen Agens und Thema/Patiens (d.h. thematische Rollen) kurz bevor der Satzverarbeitung angeschaut wurden, diese Art von Informationen für die Generierung von Erwartungen höher bewertet wird als die des Langzeitwissens. Zum Beispiel, wenn nach der Präsentation von abgebildeten Ereignissen, ein Verb (*bespitzt*) zwei Personen aus der Szene identifiziert, d.h. einen Agens einer vorher dargestellten Handlung (einen Zauberer) und einen stereotypischen Agens (einen Detektiv), betrachten die Probanden häufiger den Agens, der die Verb-bezogene Aktion ausführt anstatt den stereotypischen Agens (Knoeferle & Crocker, 2006).

Auch Studien in denen das Tempus des Verbs manipuliert wurden, haben ähnliche Ergebnisse gezeigt (Knoeferle & Crocker, 2007, Experiment 3). In diesem Experi-

ment, haben die Versuchspersonen eine Szene gesehen, in der ein Agens mit einem von zwei möglichen Objekten interagiert (z.B. ein Kellner poliert Kerzenleuchter). Zunächst hörten sie einen Satz, in dem das Verb entweder in der Vergangenheitsform (*Der Kellner polierte...*), oder im futurischen Präsens (*Der Kellner poliert demnächst...*) verwendet wurde. Die Vergangenheitsform Form bezieht sich auf das Objekt des vorangegangenen Ereignisses (z.B. die Kerzenleuchter), während das futurische Präsens sich auf das Objekt für potenziell zukünftige Ereignisse bezieht (z.B. die Gläser). Unabhängig von der Zeitform, betrachten die Versuchspersonen häufiger das Objekt des vorangegangenen Ereignisses (die Kerzenleuchter). Der bisher allgegenwärtige Einfluss eines vorangegangenen Ereignisses kann durch verschiedene experimentelle Manipulationen (z.B. durch die Häufigkeit der vergangenen/zukünftigen Verbformen, sowie durch die Präsentation von "zukünftigen" Ereignissen, die nach der Satzverarbeitung erscheinen; Abashidze et al., 2014) moduliert oder reduziert werden. Aber die Präferenz für das aktuellste Ereignis (*recent-event preference*) bleibt bestehen.

Existierende Sprachverehensmodelle, wie der *Coordinated Interplay Account* (CIA; Knoeferle et al., 2014), haben sich mit der Interaktion zwischen sprachlichen und nicht sprachlichen Hinweisen beschäftigt. Der CIA kann die Mehrheit der Sprachverstehensphänomene während situierten Sprachverstehensstudien beschreiben, z.B. die rasche Interaktion von bildhaften Informationen mit Sprachverarbeitungsprozessen, die Präferenz für visuell verankerte Ereignisse und die Verarbeitung verschiedener Inkongruenzen zwischen visuellen und sprachlichen Darstellungen. Allerdings bleiben noch offene Fragen, deren Antworten bestimmte Aspekte des Sprachverehensmodells weiter spezifizieren könnten, z.B. weitere Inkongruenzen zwischen visuellen/sprachlichen Darstellungen sowie der Einfluss des unterschiedlichen Informationsgehalts verschiedener Quellen. Zum Beispiel haben aktuelle Arbeiten zur Sprachverarbeitung die Relevanz von sozialen Aspekten (z.B. das Alters des Rezipienten und der emotionale Inhalt des Satzes) in der Interaktion zwischen visuellen und linguistischen Darstellungen herausgestellt (Münster & Knoeferle, 2018).

Informationen bezüglich des Geschlechts bieten eine weitere Untersuchungsmöglichkeit für die Ergänzung von Sprachverehensmodellen. Einige Studien in der Psycholinguistik haben die Verwendung geschlechtsrelevanter Informationen untersucht (z.B., Garnham et al., 2002; Gygax et al., 2008; Kreiner et al., 2008; Pyykkönen et al., 2010; Siyanova-Chanturia et al., 2012). Die Studien beschäftigen sich mit *expliziten* Hinweisen (z.B. Pronomen oder visuelle Hilfsmittel) bis hin zu *indirekten* Hinweisen (z.B. Wissen über Geschlechtstereotypen). Auch wenn explizite Hinweise stärker als indirekte Hinweise sind, können in bestimmten Situationen Geschlechterstereotype auch andauernde Wirkungen (z.B. Priming-Effekte) generieren (Bojarska, 2013; Cacciari & Padovani, 2007; Garnham et al., 1992; Pyykkönen et al., 2010). Da Geschlecht ein inhärentes (biologisches und soziales) Merkmal der Sprachverstehers ist, haben wir untersucht, ob diese Informationstypen besondere Effekte während des Sprachverstehens haben könnten.

Wir haben in unseren Experimenten (drei Eye-tracking *visual-world* Studien sowie ein EEG Experiment) geschlechtsrelevante Informationen eingebaut, um die *recent-event preference* zu überprüfen. Versuchspersonen haben zunächst reale Videoaufnahmen gesehen, in denen ein Paar Hände (männlich oder weiblich) mit verschiedenen Objekten interagiert. In Experiment 1-2 wurde nach diesem Video eine statische Szene mit zwei Gesichtern (ein männlicher und ein weiblicher Agens) gezeigt. Während eines Objekt-Verb-Subjekt (OVS) Satzes (z.B. *Den Kuchen backt gleich Susanna/Tobias*) wurden die Augenbewegungen der Probanden in Richtung der beiden Gesichter aufgenommen. Das Gesicht des Agens, dessen Geschlechtsmerkmale mit den Händen im Video übereinstimmen, bezeichnen wir als *Ziel-Agens*, während wir das andere Gesicht als *Competitor-Agens* bezeichnen. In Experiment 3 gab es neben den beiden Gesichtern zusätzlich Bilder von Objekten (d.h. ein *Ziel-Objekt* und ein *Competitor-Objekt*). Zwei Faktoren wurden manipuliert: a) die Übereinstimmung der visuellen Ereignisse im Video (Geschlecht und Aktion) mit den darauf folgenden OVS Sätzen, die die Ereignisse beschreiben und b) die stereotypische Übereinstimmung zwischen der beschriebenen Aktion und dem Ziel-Agens in der Szene (durch die Geschlechtshinweise der vorangegangenen Ereignisse begünstigt: die Hände im Video). In den Experimenten 1 und 3 gab es Diskrepanzen zwischen den

visuellen und den im Satz beschriebenen Aktionen (Objekt + Verb; siehe Tabelle 5.1), während sich die Diskrepanzen in den Experimenten 2 und 4 am Ende des Satzes (das Subjekt, oder Eigennamen, die den Agens ergeben) befanden (z.B. Susanna; siehe Tabelle 5.2).

Unsere Studien haben gezeigt, dass die Darstellungen von vorangegangenen Ereignissen Priorität genießen, da sie die Aufmerksamkeit der Probanden auf den passenden Agens in Bezug auf die Geschlechterkriterien lenken (d.h. Geschlechts-Hinweise aus vorangegangenen Ereignissen leiten die Augenbewegungen auf den Agens, der passende Geschlechtseigenschaften hat; Experimente 1 bis 3). Zusätzlich konnten unsere zwei experimentellen Manipulationen diese Präferenz unter bestimmten Umständen (d.h. zusätzliche Bilder von Objekten während des Sprachverstehens) (super-)additiv modulieren (Experiment 3). Wenn die visuellen und sprachlichen Darstellungen übereinstimmen und die stereotypische Valenz der beschriebenen Aktion zu dem Ziel-Agens (d.h. zu den Geschlechtshinweisen abgeleitet aus dem Ergebnis) passt, wird dieser Agens (d.h. Gesicht) mehr inspiziert, als wenn einer der beiden Hinweise fehlt. Wenn der Satz nicht mit den vorangegangenen Ereignissen im Video übereinstimmt oder wenn die im Satz beschriebene Aktion nicht stereotypisch für den Ziel-Agens ist, ist diese Präferenz erheblich reduziert (sogar getilgt; siehe Grafik 6.5). Allerdings hat Experiment 4 gezeigt, dass Diskrepanzen zwischen visuellen Geschlechtsreizen und Eigennamen, die eine semantisch-begründete biphasische (N400/P600) elektrophysiologische Reaktion hervorrufen, Gemeinsamkeiten mit der Verarbeitung von thematischen Rollen und Verb-Aktionen haben (Knoeferle et al., 2014).

Insgesamt lässt sich sagen, dass durch einen biologischen/sozialen Aspekt der visuellen und sprachlichen Domäne - geschlechtsrelevanter Informationen - Sprachverstehensberichte wie der CIA weiter ergänzt werden können. Zuerst stimmen wir zu, dass ein Überprüfungsmechanismus (Knoeferle et al., 2014), der die Übereinstimmung verschiedener Aspekte der visuellen und linguistischen Darstellung verifiziert (z.B. thematische Rolle, Verb-Aktion, oder Geschlechterbeziehungen zwischen visuellen Ereignissen

und Sprache), nötig ist. Im Bezug auf Erwartungen, die während der Satzverarbeitung generiert werden und die Augenbewegungen lenken, bedarf es einer Modellierung eines gewichteten Systems (Münster, 2016), das nicht nur von expliziten visuellen Hilfsmitteln aus vorangegangenen Ereignissen beeinflusst wird, sondern auch von verschiedenen visuellen/sprachlichen Diskrepanzen und, je nach kognitivem Anspruch und der Art des Informationsinhalts, auch von bestimmten Aspekten des Langzeitgedächtnisses (z.B. Geschlechtsstereotype). Die Ergebnisse unserer Studien legen eingehendere Untersuchungen zu den verschiedenen Informationsarten im (situierten) Sprachverstehen, wie *biosoziale* Faktoren, die auf verschiedenen Ebenen und Domänen der Kommunikation gefunden werden können (z.B. Sprachverstehender, Sprecher, Sprachlicher Inhalt, etc.), nahe.

Appendices

A | Experimental materials (Experiments 1 to 4)

A.1 | Experimental sentences

Table A.1.: Sentences for the experimental items

Item	Stereotype	Sentence	
1	female	Die Mütze strickt gleich <i>The cap knits soon</i>	Susanna/Tobias
	male	Die Kette ölt gleich <i>The chain oils soon</i>	Susanna/Tobias
2	female	Den Brotteig knetet gleich <i>The bread dough kneads soon</i>	Susanna/Tobias
	male	Den Schaltkreis verlötet gleich <i>The circuit solders soon</i>	Susanna/Tobias
3	female	Den Schmuck bewundert gleich <i>The jewelry admires soon</i>	Susanna/Tobias
	male	Das Metall bearbeitet gleich <i>The metal handles soon</i>	Susanna/Tobias
4	female	Die Plätzchen verziert gleich <i>The biscuits adorns soon</i>	Susanna/Tobias
	male	Das Radio repariert gleich <i>The radio repairs soon</i>	Susanna/Tobias
5	female	Das Baby füttert gleich <i>The baby feeds soon</i>	Susanna/Tobias
	male	Die Stange verbiegt gleich <i>The bar bends soon</i>	Susanna/Tobias

Item	Stereotype	Sentence	
6	female	Die Hose bügelt gleich <i>The trousers irons soon</i>	Susanna/Tobias
	male	Die Kiste lackiert gleich <i>The box varnishes soon</i>	Susanna/Tobias
7	female	Die Schokolade raspelt gleich <i>The chocolate rasps soon</i>	Susanna/Tobias
	male	Den Durchmesser bestimmt gleich <i>The diamiter calculates soon</i>	Susanna/Tobias
8	female	Das Schminketui öffnet gleich <i>The jewelry case opens soon</i>	Susanna/Tobias
	male	Den Fahrradschlauch flickt gleich <i>The bicycle tube fixes soon</i>	Susanna/Tobias
9	female	Das Ei schlägt gleich <i>The egg whisks soon</i>	Susanna/Tobias
	male	Die Latte bohrt gleich <i>The batten drills soon</i>	Susanna/Tobias
10	female	Die Bluse faltet gleich <i>The blouse folds soon</i>	Susanna/Tobias
	male	Den Hammer verwendet gleich <i>The hammer uses soon</i>	Susanna/Tobias
11	female	Die Kekse formt gleich <i>The cookies forms soon</i>	Susanna/Tobias
	male	Die Batterie lädt gleich <i>The battery charges soon</i>	Susanna/Tobias
12	female	Das Törtchen dekoriert gleich <i>The tartlet decorates soon</i>	Susanna/Tobias
	male	Die Glühbirne installiert gleich <i>The light-bulb installs soon</i>	Susanna/Tobias

Item	Stereotype	Sentence	
13	female	Die Möhre schält gleich <i>The carrot peels soon</i>	Susanna/Tobias
	male	Die Schraube schraubt gleich <i>The screw tightens soon</i>	Susanna/Tobias
14	female	Die Socken stopft gleich <i>The socks mends soon</i>	Susanna/Tobias
	male	Das Schild befestigt gleich <i>The plaque attaches soon</i>	Susanna/Tobias
15	female	Den Kuchen backt gleich <i>The cake bakes soon</i>	Susanna/Tobias
	male	Das Modell baut gleich <i>The model builds soon</i>	Susanna/Tobias
16	female	Die Erdbeeren zuckert gleich <i>The strawberries sugars soon</i>	Susanna/Tobias
	male	Das Komikheft liest gleich <i>The comic reads soon</i>	Susanna/Tobias
17	female	Das Mehl siebt gleich <i>The flour sieves soon</i>	Katharina/Sebastian
	male	Das Holz sägt gleich <i>The wood saws soon</i>	Katharina/Sebastian
18	female	Den Hemdknopf schneidert gleich <i>The shirt button tailors soon</i>	Katharina/Sebastian
	male	Die Kante schleift gleich <i>The edge sands soon</i>	Katharina/Sebastian
19	female	Den Nagellack verdünnt gleich <i>The nail polish blends soon</i>	Katharina/Sebastian
	male	Die Krawatte bindet gleich <i>The tie binds soon</i>	Katharina/Sebastian

Item	Stereotype	Sentence	
20	female	Die Duftkerze löscht gleich <i>The perfumed candle extincts soon</i>	Katharina/Sebastian
	male	Den Rasierer säubert gleich <i>The razor cleans soon</i>	Katharina/Sebastian
21	female	Die Puppe kleidet gleich <i>The doll dresses soon</i>	Katharina/Sebastian
	male	Den Nagel hämmert gleich <i>The nail hammers soon</i>	Katharina/Sebastian
22	female	Das Räucherstäbchen verbrennt gleich <i>The incense stick burns soon</i>	Katharina/Sebastian
	male	Die Werkzeugkiste ordnet gleich <i>The toolbox organizes soon</i>	Katharina/Sebastian
23	female	Den Lippenstift testet gleich <i>The lipstick tries soon</i>	Katharina/Sebastian
	male	Die Holzfigur schnitzt gleich <i>The wooden figure carves soon</i>	Katharina/Sebastian
24	female	Den Kaschmirschal befühlt gleich <i>The cashmere scarf palpates soon</i>	Katharina/Sebastian
	male	Die Alarmanlage montiert gleich <i>The alarm mounts soon</i>	Katharina/Sebastian
25	female	Das Potpourri kreierte gleich <i>The potpourri creates soon</i>	Katharina/Sebastian
	male	Die Klinge ersetzt gleich <i>The blade replaces soon</i>	Katharina/Sebastian
26	female	Das Kleid kurzt gleich <i>The dress shortens soon</i>	Katharina/Sebastian
	male	Den Draht windet gleich <i>The wire twists soon</i>	Katharina/Sebastian

Item	Stereotype	Sentence	
27	female	Die Handtasche schliesst gleich <i>The handbag closes soon</i>	Katharina/Sebastian
	male	Das Videospiele spielt gleich <i>The videogame plays soon</i>	Katharina/Sebastian
28	female	Das Halstuch näht gleich <i>The scarf sews soon</i>	Katharina/Sebastian
	male	Das Gewicht hebt gleich <i>The weight lifts soon</i>	Katharina/Sebastian
29	female	Die Rose beschnuppert gleich <i>The rose sniffs soon</i>	Katharina/Sebastian
	male	Die Kabel verbindet gleich <i>The cables connects soon</i>	Katharina/Sebastian
30	female	Die Blume gießt gleich <i>The flower waters soon</i>	Katharina/Sebastian
	male	Das Messer wetzt gleich <i>The knife sharpens soon</i>	Katharina/Sebastian
31	female	Das Geschenk verpackt gleich <i>The present wraps soon</i>	Katharina/Sebastian
	male	Das Stativ demontiert gleich <i>The tripod dismounts soon</i>	Katharina/Sebastian
32	female	Den Zucker wiegt gleich <i>The sugar weights soon</i>	Katharina/Sebastian
	male	Die Pfeife stopft gleich <i>The pipe packs soon</i>	Katharina/Sebastian

A.2 | Onsets and offsets of experimental sentence regions

Table A.2.: Onsets and offsets of sentence regions (in msec)

Item	NP1on	NP1off	Von	Voff	Advon	Advoff	NP2on	NP2off
1a	0	797	1237	1934	2357	2954	3370	4090
1b	0	797	1237	1823	2357	2954	3370	4090
2a	0	1227	1660	2481	2907	3447	3932	4690
2b	0	1219	1660	2435	2907	3482	3932	4690
3a	0	707	1205	1999	2490	3013	3477	4160
3b	0	770	1205	2005	2490	3013	3477	4160
4a	0	966	1394	2218	2686	3220	3713	4455
4b	0	944	1394	2192	2686	3220	3713	4455
5a	0	823	1249	1910	2411	2929	3441	4176
5b	0	752	1249	1993	2411	2966	3441	4176
6a	0	809	1294	1944	2464	3033	3510	4246
6b	0	794	1294	2044	2464	3033	3510	4246
7a	0	1083	1544	2306	2799	3346	3852	4569
7b	0	1111	1544	2322	2799	3346	3852	4569
8a	0	1140	1609	2281	2744	3242	3712	4438
8b	0	1134	1609	2281	2744	3242	3712	4438
9a	0	775	1271	1941	2379	2923	3413	4173
9b	0	780	1271	1883	2379	2923	3413	4173

Table A.3.: Onsets and offsets of sentence regions (in msec)

Item	NP1on	NP1off	Von	Voff	Advon	Advoff	NP2on	NP2off
10a	0	765	1190	1883	2356	2884	3354	4063
10b	0	730	1190	1920	2356	2884	3354	4063
11a	0	890	1358	1996	2407	2979	3427	4176
11b	0	890	1358	1934	2407	2979	3427	4176
12a	0	975	1446	2327	2812	3355	3840	4581
12b	0	975	1446	2346	2812	3355	3840	4581
13a	0	857	1326	2039	2519	3082	3512	4291
13b	0	892	1326	2039	2519	3082	3512	4291
14a	0	758	1266	2083	2572	3154	3624	4384
14b	0	761	1266	2113	2572	3127	3624	4384
15a	0	897	1363	1856	2339	2896	3368	4133
15b	0	897	1363	1927	2339	2896	3368	4133
16a	0	1074	1548	2204	2629	3174	3642	4335
16b	0	1120	1548	2147	2629	3174	3642	4335
17a	0	1035	1393	2164	2631	3218	3674	4606
17b	0	915	1393	2164	2631	3218	3674	4606
18a	0	860	1291	1986	2434	3015	3467	4396
18b	0	918	1291	1957	2434	3015	3467	4396
19a	0	900	1375	1970	2448	2944	3415	4341
19b	0	900	1375	1970	2448	2944	3415	4341
20a	0	1192	1619	2255	2728	3257	3738	4630
20b	0	1173	1619	2361	2728	3257	3738	4630
21a	0	820	1277	1957	2368	2928	3389	4307
21b	0	859	1277	1897	2368	2928	3389	4307
22a	0	1403	1872	2586	3011	3597	4026	4884
22b	0	1439	1872	2546	3011	3597	4026	4884

Item	NP1on	NP1off	Von	Voff	Advon	Advoff	NP2on	NP2off
23a	0	1079	1534	2185	2647	3194	3648	4533
23b	0	1079	1534	2230	2647	3171	3648	4533
24a	0	1220	1636	2362	2864	3427	3852	4736
24b	0	1193	1636	2424	2864	3373	3852	4736
25a	0	887	1306	2073	2513	3029	3498	4432
25b	0	836	1306	2044	2513	3029	3498	4432
26a	0	932	1396	2048	2509	2996	3478	4376
26b	0	900	1396	2025	2509	2996	3478	4376
27a	0	1032	1536	2271	2706	3261	3756	4637
27b	0	1117	1536	2231	2706	3261	3756	4637
28a	0	976	1393	1944	2418	2892	3367	4269
28b	0	909	1393	1914	2418	2892	3367	4269
29a	0	875	1312	2157	2625	3180	3670	4595
29b	0	875	1312	2157	2625	3180	3670	4595
30a	0	845	1334	1912	2400	2893	3383	4241
30b	0	865	1334	1918	2400	2893	3383	4241
31a	0	864	1337	2097	2618	3157	3658	4525
31b	0	816	1337	2156	2618	3172	3658	4525
32a	0	837	1264	1841	2341	2900	3405	4341
32b	0	837	1264	1914	2341	2900	3405	4341

A.3 | Visual materials

A.3.1 | Snapshots of the agents' faces and hands with Consent to Use of Image forms

Figure A.1.: Snapshot of the agents' faces



(a) Katharina



(b) Sebastian



(c) Susanna



(d) Tobias

Figure A.3.: Snapshot of the agents' hands from the experimental videos



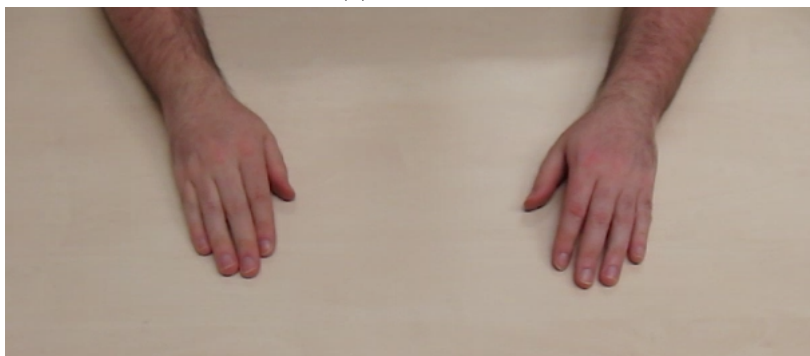
(a) Katharina



(b) Sebastian



(c) Susanna



(d) Tobias

Universität Bielefeld

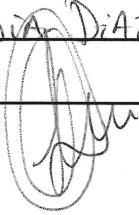
P. Knoeferle (PhD), Assistant Professor
 Cognitive Interaction Technology, Excellence Cluster, Bielefeld University, Zehlendorfer Damm 201
 33615 Bielefeld, Deutschland
 Tel: +49521106-12250,
 Email: knoeferl@cit-ec.uni-bielefeld.de
 Lab Homepage: <http://www.homes.uni-bielefeld.de/pknoeferle/Homepage/Home.html>

Consent to Use of Image

I hereby give the Language and Cognition group at Bielefeld University permission to use my image for the following research purposes:

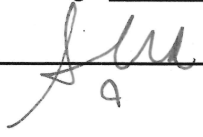
- To be displayed in experimental settings
- To serve as an illustrative example in conference posters/ presentations
- To be published as part of a research paper

Name: SONIA DIAZ DIAZ

Signature:  Date: 16.05.2014

The image will only be used as specified above and under a fictitious name. Should you no longer wish your image to be used for any of these purposes, it is possible to modify or cancel this consent form without any disadvantage. The cancellation of this consent form does not, however, guarantee retroactive effects (e.g. cases in which an experiment has been already run, published papers, etc.).

Researcher in charge: ALBA RODRIGUEZ

Signature:  Date: 16/5/2014



Universität Bielefeld

P. Knoeferle (PhD), Assistant Professor
Cognitive Interaction Technology, Excellence Cluster, Bielefeld University, Zehlendorfer Damm 201
33615 Bielefeld, Deutschland
Tel: +49521106-12250,
Email: knoeferl@cit-ec.uni-bielefeld.de
Lab Homepage: <http://www.homes.uni-bielefeld.de/pknoeferle/Homepage/Home.html>


Consent to Use of Image

I hereby give the Language and Cognition group at Bielefeld University permission to use my image for the following research purposes:

- To be displayed in experimental settings
- To serve as an illustrative example in conference posters/ presentations
- To be published as part of a research paper

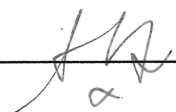
Name: Julio Revillano

Signature: _____ Date: 17/05/2014



The image will only be used as specified above and under a fictitious name. Should you no longer wish your image to be used for any of these purposes, it is possible to modify or cancel this consent form without any disadvantage. The cancellation of this consent form does not, however, guarantee retroactive effects (e.g. cases in which an experiment has been already run, published papers, etc.).

Researcher in charge: ALBA RODRIGUEZ

Signature:  _____ Date: 17/5/2014

Universität Bielefeld


P. Knoeferle (PhD), Assistant Professor
 Cognitive Interaction Technology, Excellence Cluster, Bielefeld University, Zehlendorfer Damm 201
 33615 Bielefeld, Deutschland
 Tel: +49521106-12250,
 Email: knoeferl@cit-ec.uni-bielefeld.de
 Lab Homepage: <http://www.homes.uni-bielefeld.de/pknoeferle/Homepage/Home.html>

Consent to Use of Image

I hereby give the Language and Cognition group at Bielefeld University permission to use my image for the following research purposes:


- To be displayed in experimental settings
- To serve as an illustrative example in conference posters/ presentations
- To be published as part of a research paper

Name: Marta Liceras Cervera

Signature:  Date: 16/05/2014

The image will only be used as specified above and under a fictitious name. Should you no longer wish your image to be used for any of these purposes, it is possible to modify or cancel this consent form without any disadvantage. The cancellation of this consent form does not, however, guarantee retroactive effects (e.g. cases in which an experiment has been already run, published papers, etc.).

Researcher in charge: Alba Rodríguez Llamazares

Signature:  Date: 16/V/2014



Universität Bielefeld

P. Knoeferle (PhD), Assistant Professor
Cognitive Interaction Technology, Excellence Cluster, Bielefeld University, Zehlendorfer Damm 201
33615 Bielefeld, Deutschland
Tel: +49521106-12250,
Email: knoeferl@cit-ec.uni-bielefeld.de
Lab Homepage: <http://www.homes.uni-bielefeld.de/pknoeferle/Homepage/Home.html>

Consent to Use of Image

I hereby give the Language and Cognition group at Bielefeld University permission to use my image for the following research purposes:

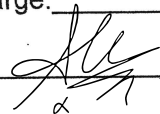
- To be displayed in experimental settings
- To serve as an illustrative example in conference posters/ presentations
- To be published as part of a research paper

Name: Antonio Longo

Signature: Antonio Longo Date: 16/05/2014

The image will only be used as specified above and under a fictitious name. Should you no longer wish your image to be used for any of these purposes, it is possible to modify or cancel this consent form without any disadvantage. The cancellation of this consent form does not, however, guarantee retroactive effects (e.g. cases in which an experiment has been already run, published papers, etc.).

Researcher in charge: Alba Rodríguez Llamazares

Signature:  Date: 16/V/2014

A.3.2 | Snapshots of the objects from the experimental videos

Figure A.5.: Snapshots of the objects from the experimental videos



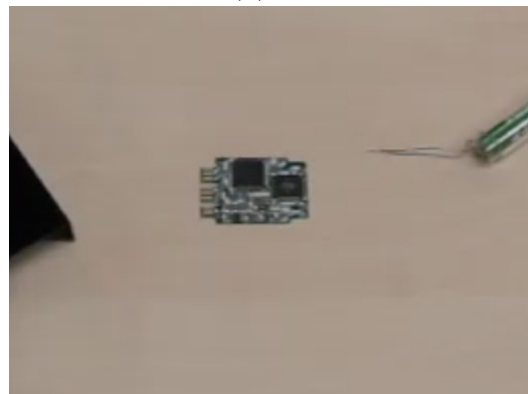
(a) 1a



(b) 1b



(c) 2a



(d) 2b



(e) 3a



(f) 3b



(a) 4a



(b) 4b



(c) 5a



(d) 5b



(e) 6a



(f) 6b



(g) 7a



(h) 7b



(a) 8a



(b) 8b



(c) 9a



(d) 9b



(e) 10a



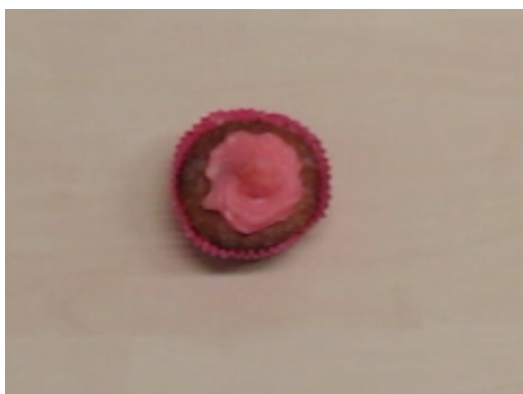
(f) 10b



(g) 11a



(h) 11b



(a) 12a



(b) 12b



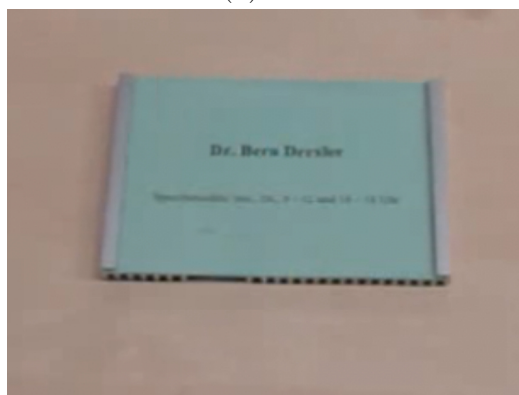
(c) 13a



(d) 13b



(e) 14a



(f) 14b



(g) 15a



(h) 15b



(a) 16a



(b) 16b



(c) 17a



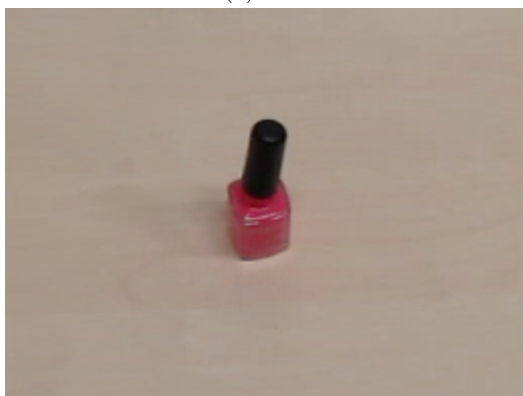
(d) 17b



(e) 18a



(f) 18b



(g) 19a



(h) 19b



(a) 20a



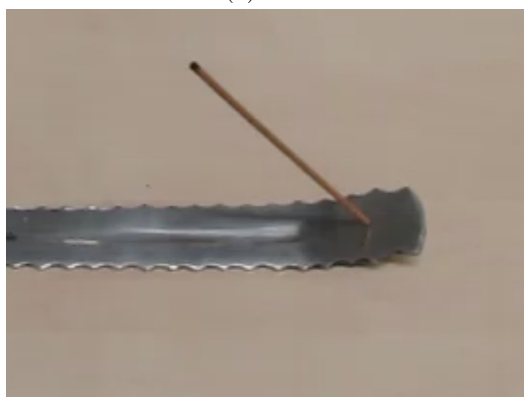
(b) 20b



(c) 21a



(d) 21b



(e) 22a



(f) 22b



(g) 23a



(h) 23b



(a) 24a



(b) 24b



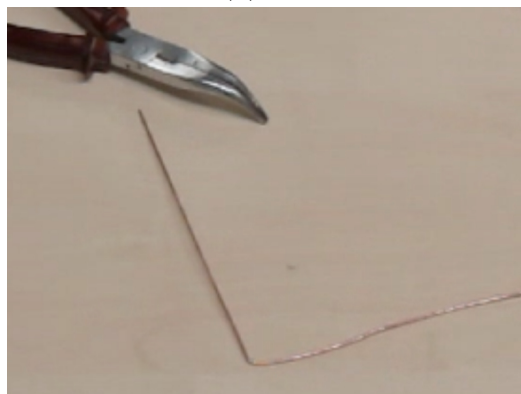
(c) 25a



(d) 25b



(e) 26a



(f) 26b



(a) 27a



(b) 27b



(c) 28a



(d) 28b



(e) 29a



(f) 29b



(a) 30a



(b) 30b



(c) 31a



(d) 31b



(e) 32a



(f) 32b

A.4 | Example of two filler trials

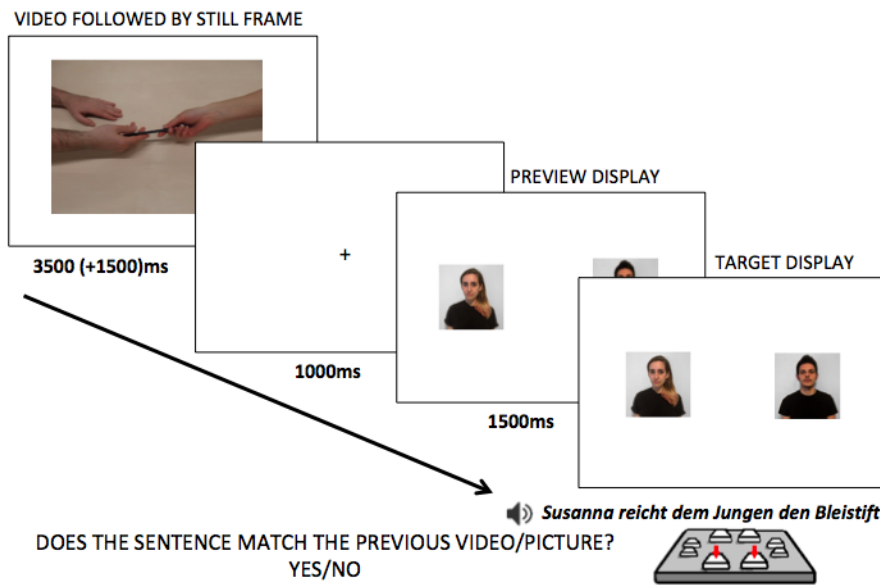


Figure A.15.: Filler trial with two pairs of hands

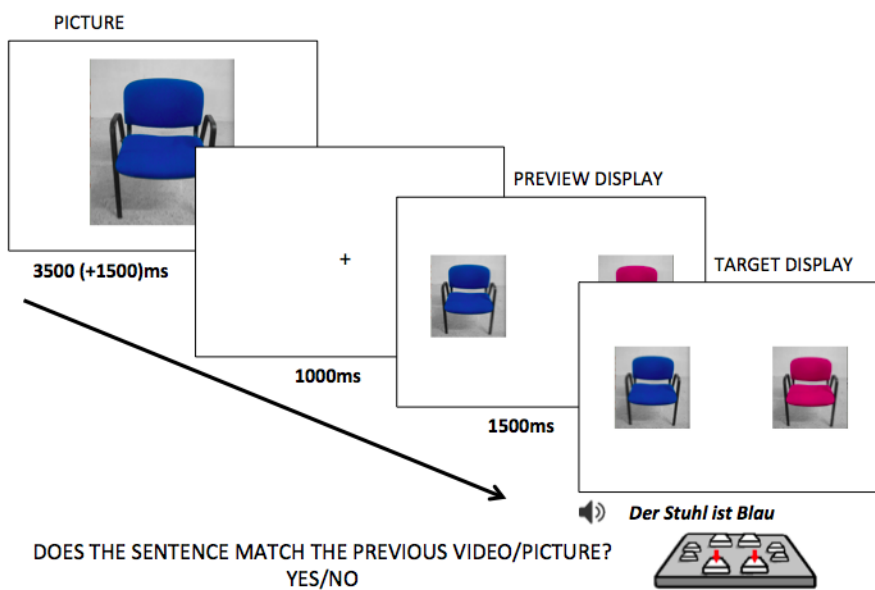


Figure A.16.: Filler trial with an object picture

B | Additional statistical analyses (Experiments 1 to 4)

B.1 | Accuracy analyses using GLME (Experiments 1 to 4)

We ran Generalized Linear Mixed Effects analyses (suitable for binomial data) on accuracy using R (R Core Team, 2016) by means of the "lme4" package (Bates, Mächler, Bolker, & Walker, 2015). Following Barr, Levy, Scheepers, and Tily (2013), we first computed the maximal converging model, in which we included video-sentence match (expressed as *v_a_match* in Experiments 1 and 3; *hand_subj_match*, Experiments 2 and 4) and stereotypicality match (*stereomatch*) as within-subjects factors, and gender as a between-subjects factor¹. Models including random slopes for participants and items were also included when converging. The first converging model was defined as the “maximal model,” against which simpler models were compared by residual maximum likelihood tests (REML), following a backward selection procedure. This procedure continued until either the removal of an element led to a significant decrease in model fit or until the model contained only fixed effects.

¹In Experiment 4 we only included *hand_subj_match* as fixed factor.

Table B.1.: Accuracy analysis, Experiment 1

Maximal model: $\text{acc} \sim (\text{v_a_match} * \text{stereomatch}) + (\text{v_a_match} * \text{gender}) + (1 | \text{participant}) + (1 | \text{item})$

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.1723	1.0594	4.882	1.05e-06 *
v_a_match	-2.4745	1.0902	-2.270	0.0232 *
stereomatch	1.1057	1.1578	0.955	0.3396
gender	-1.1134	1.1628	-0.957	0.3383
v_a_match*stereomatch	-0.6963	1.2258	-0.568	0.5700
stereomatch*gender	1.1785	1.2239	0.963	0.535

Table B.2.: Accuracy analysis, Experiment 2

Maximal model: $\text{acc} \sim (\text{hand_subj_match} * \text{stereomatch}) + (\text{hand_subj_match} * \text{gender}) + (\text{stereomatch} * \text{gender}) + (1 | \text{participant}) + (1 | \text{item})$

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.17119	0.45496	6.970	3.16e-12 *
hand_subj_match	-0.27981	0.51814	-0.540	0.589
stereomatch	0.04011	0.54565	0.074	0.941
gender	0.54042	0.61519	0.878	0.380
hand_subj_match*stereomatch	0.41248	0.64554	0.639	0.523
hand_subj_match*gender	-0.24814	0.64861	-0.383	0.702
stereomatch*gender	-0.37611	0.64553	-0.583	0.560

Table B.3.: Accuracy analysis, Experiment 3

Maximal model: $\text{acc} \sim (\text{v_a_match} * \text{stereomatch}) + (\text{v_a_match} * \text{gender}) + (\text{stereomatch} * \text{gender}) + (1 | \text{participant}) + (1 | \text{item})$

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.30832	0.77053	5.591	2.25e-08 *
v_a_match	-0.06243	0.85797	-0.073	0.942
stereomatch	1.41477	1.19553	1.183	0.237
gender	0.69647	0.96890	0.719	0.472
hand_subj_match*stereomatch	-1.31108	1.27658	-1.027	0.304
hand_subj_match*gender	-1.21178	1.12149	-1.081	0.280
stereomatch*gender	0.04745	1.01542	0.047	0.963

Table B.4.: Accuracy analysis, Experiment 4

Maximal model: $\text{acc} \sim \text{hand_subj_match} + (1 + \text{hand_subj_match} | \text{participant}) + (1 | \text{item})$

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.5857	0.4972	7.211	<5.54e-13 *
hand_subj_match	-0.7506	0.5829	-1.288	0.198

B.2 | Alternative reaction-time analyses using LME (Experiments 1 to 3)

We also ran Linear Mixed Effects analyses on the log-transformed reaction-time data using the "lme4" package (Bates et al., 2015). Potential differences between ANOVA and LME analyses may appear, as LME analyses allow for the inclusion of participants and items within the same model (instead of separate F_1 and F_2 analyses). In these analyses, we included video-sentence mismatch (*v_a_match* in Experiments 1 and 3; *hand_subj_match*, Experiment 2) and stereotypicality match (*stereomatch*) as within-subjects factors, and gender as between-subjects factor. Random slopes for participants and items were also included when converging. We applied the same model reduction process as with the accuracy data; models were compared using maximum likelihood tests (ML). The LMER output provided us with the estimates, standard errors and t-values for the fixed effects; an absolute t value equal or superior to 2 was taken as statistically significant.

Table B.5.: Reaction-time analysis, Experiment 1

Maximal model: $RT \sim v_a_match * stereomatch * gender$

$+(1 + v_a_match + stereomatch | participant)$

$+(1 + v_a_match + stereomatch | item)$

Reduced model: $RT \sim hand_subj_match * stereomatch * gender$

$+(1 + v_a_match | participant) + (1 + v_a_match | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	8.2517	0.1008	81.81*
v_a_match	0.2039	0.0922	2.21 *
stereomatch	0.0297	0.0275	1.08
gender	-0.3991	0.1382	-2.89 *
v_a_match*stereomatch	-0.0409	0.0393	-1.04
v_a_match*gender	0.3652	0.1246	2.93 *
stereomatch*gender	-0.0321	0.0379	-0.85
v_a_match*stereomatch*gender	0.0494	0.0542	0.91

Table B.6.: Reaction-time analysis, Experiment 2

Maximal model: $RT \sim \text{hand_subj_match} * \text{stereomatch} * \text{gender}$

$+(1 + \text{hand_subj_match} * \text{stereomatch} | \text{participant})$

$+(1 + \text{hand_subj_match} * \text{stereomatch} | \text{item})$

Reduced model: $RT \sim \text{hand_subj_match} * \text{stereomatch} * \text{gender}$

$+(1 + \text{hand_subj_match} | \text{participant}) + (1 | \text{item})$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	8.426110	0.014970	562.979 *
hand_subj_match	-0.001662	0.008448	-0.16
stereomatch	-0.002578	0.006911	-0.34
gender	0.010386	0.018377	0.58
hand_subj_match*stereomatch	0.004542	0.009793	0.44
hand_subj_match*gender	0.006237	0.011927	0.50
stereomatch*gender	0.005535	0.009757	0.54
hand_subj_match*stereomatch*gender	-0.012058	0.013784	-0.85

Table B.7.: Reaction-time analysis, Experiment 3

Maximal model: $RT \sim v_a_match * stereomatch * gender$

$+(1 + v_a_match * stereomatch | participant)$

$+(1 + v_a_match * stereomatch | item)$

Reduced model: $RT \sim hand_subj_match * stereomatch * gender$

$+(1 + v_a_match | participant) + (1 | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	7.9410	0.1435	55.33*
v_a_match	0.4601	0.1335	3.45 *
stereomatch	-0.0327	0.0302	-1.08
gender	0.0397	0.1907	0.21
v_a_match*stereomatch	0.038	0.0427	0.89
v_a_match*gender	-0.0104	0.1777	-0.06
stereomatch*gender	0.0587	0.0403	1.46
v_a_match*stereomatch*gender	-0.0674	0.0570	-1.18

B.3 | Statistical tests for the intercept per sentence region (Experiments 1 to 3)

Table B.8.: Statistical tests for the intercept (grand average per subjects) in the log-probability ratios per sentence region (Experiments 1 to 3)

	Sentence region	$F_{\text{inter}} (1,30)$	p	η^2
exp1	np1	12,10	<.01	.920
	verb	13,002	<.01	.302
	adv	20,42	<.001	.405
exp2	np2	38,02	<.001	.559
	np1	5,11	<.05	.920
	verb	30,77	<.001	1
exp3	adv	25,15	<.001	.998
	np2	29,23	<.001	.494
	np1	13,45	=.01	.310
	verb	12,64	=.01	.297
	adv	6,21	<.05	.172
	np2	92,25	<.001	.755

B.4 | Alternative eye-movement analyses using LME (Experiments 1 to 3)

We also ran alternative LME analyses on our eye-tracking data (log-probability ratios). Once again, potential differences between ANOVA and LME analyses may appear, as the fixation sums which are then transformed into log-probability ratios are averaged differently as compared to separate per-subject and per-item analyses. However, as we can see in the results, results from mixed models analyses follow the same pattern as in the ANOVAs. As with the accuracy and reaction-time analyses, we first computed the maximal converging model, which was compared to simpler models by maximum likelihood (ML) tests. The procedure continued until either the removal of an element led to a significant decrease in model fit or until the model contained only fixed effects. Absolute t values around or superior to 2 were taken as statistically significant.

Table B.9.: Eye-movement analysis, Experiment 1, verb region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match+stereomatch|participant)+(1+v_a_match+stereomatch|item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1|participant) + (1|item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.9448	0.5829	3.336*
v_a_match	1.1645	0.5982	1.947#
$stereomatch$	-0.2751	0.5919	0.465
$gender$	-1.1425	0.8252	1.405
$v_a_match * stereomatch$	-0.4974	0.8458	0.588
$v_a_match * gender$	-0.3515	0.8518	0.413
$stereomatch * gender$	0.9898	0.8388	1.18
$v_a_match * stereomatch * gender$	0.5139	1.201	0.428

Table B.10.: Eye-movement analysis, Experiment 1, adverb region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match * stereomatch | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1 | participant) + (1 | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.1736	0.6119	1.918#
v_a_match	2.0189	0.597	3.381*
stereomatch	0.2144	0.5895	0.364
gender	-0.5385	0.8654	0.622
v_a_match*stereomatch	-1.004	0.8433	1.191
v_a_match*gender	-0.4875	0.8474	0.575
stereomatch*gender	0.2435	0.8329	0.292
v_a_match*stereomatch*gender	0.6929	1.1926	0.581

Table B.11.: Eye-movement analysis, Experiment 1, NP2 region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match+stereomatch|participant)$

$+(1+v_a_match+stereomatch+gender|item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1|participant) + (1|item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.58995	0.58085	2.737*
v_a_match	1.16743	0.57776	2.021*
stereomatch	0.17084	0.57165	0.299
gender	0.04499	0.82239	0.055
v_a_match*stereomatch	0.42852	0.81688	0.525
v_a_match*gender	-0.08631	0.82261	0.575
stereomatch*gender	-0.3411	0.80932	0.421
v_a_match*stereomatch*gender	-0.19364	1.15756	0.581

Table B.12.: Eye-movement analysis, Experiment 2, NP2 region

Maximal model: $\log \sim \text{hand_subj_match} * \text{stereomatch} * \text{gender}$
 $+(1 + \text{hand_subj_match} + \text{stereomatch} | \text{participant})$
 $+(1 + \text{hand_subj_match} * \text{stereomatch} | \text{item}) + (1 + \text{hand_subj_match} * \text{gender} | \text{item})$
 Reduced model: $\log \sim \text{hand_subj_match} * \text{stereomatch} * \text{gender}$
 $+(1 + \text{hand_subj_match} | \text{participant}) + (1 | \text{item})$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.0986	0.7341	2.858*
hand_subj_match	2.3882	0.6493	3.678*
stereomatch	0.1404	0.5667	0.248
gender	-0.8012	1.0382	-0.772
hand_subj_match*stereomatch	-0.4603	0.8001	-0.575
hand_subj_match*gender	-0.4223	0.9183	-0.460
stereomatch*gender	-0.8797	0.7992	-1.101
hand_subj_match*stereomatch*gender	1.6246	1.1300	1.438

Table B.13.: Eye-movement analysis, Experiment 3 (agents), NP1 region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match + stereomatch + gender | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1+v_a_match | participant)$

$+(1+v_a_match | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	0.3025	0.4385	0.690
v_a_match	1.3426	0.5796	2.316*
stereomatch	0.3137	0.5157	0.608
gender	-0.4929	0.5879	-0.838
v_a_match*stereomatch	-1.1641	0.7212	-1.614
v_a_match*gender	-0.0239	0.7903	-0.030
stereomatch*gender	-0.1916	0.7160	-0.268
v_a_match*stereomatch*gender	1.0697	1.0068	1.062

Table B.14.: Eye-movement analysis, Experiment 3 (agents), verb region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1+v_a_match | participant) + (1 | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-0.42111	0.39000	-1.080
v_a_match	1.61728	0.56966	2.839*
stereomatch	0.96854	0.50650	1.912#
gender	-0.24091	0.54513	-0.442
v_a_match*stereomatch	-0.19567	0.71043	-0.275
v_a_match*gender	0.13209	0.79579	0.166
stereomatch*gender	0.05798	0.70479	0.082
v_a_match*stereomatch*gender	0.06828	0.99214	0.069

Table B.15.: Eye-movement analysis, Experiment 3 (agents), adverb region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match * stereomatch | item)$

$+(1+v_a_match * gender | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1 | participant) + (1 | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-0.39812	0.39679 9	1.003
v_a_match	1.58866	0.53347 7	2.978*
stereomatch	0.41383	0.53989	0.767
gender	0.19727	0.55125	0.358
v_a_match*stereomatch	-0.07042	0.75612	0.093
v_a_match*gender	-0.68612	0.74297	0.923
stereomatch*gender	-0.09083	0.74989	0.121
v_a_match*stereomatch*gender	0.53681	1.05549	0.509

Table B.16.: Eye-movement analysis, Experiment 3 (agents), NP2 region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match * stereomatch | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1 | participant) + (1 | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.52936	0.40805	3.748*
v_a_match	0.48019	0.46207	1.039*
stereomatch	0.16135	0.46809	0.345
gender	-0.45845	0.56984	0.805
v_a_match*stereomatch	-0.19567	0.71043	-0.275
v_a_match*gender	-0.45101	0.64346	0.701
stereomatch*gender	-0.07876	0.64983	0.121
v_a_match*stereomatch*gender	1.05159	0.91436	1.15

Table B.17.: Eye-movement analysis, Experiment 3 (objects), NP1 region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match * stereomatch + gender | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1 | participant) + (1 | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-1.09379	0.38011	-2.878*
v_a_match	2.89220	0.51636	5.601*
stereomatch	0.25367	0.51636	0.491
gender	-0.04423	0.53755	-0.082
v_a_match*stereomatch	0.67341	0.73024	0.922
v_a_match*gender	-0.00183	0.73025	-0.003
stereomatch*gender	-0.27189	0.73025	-0.372
v_a_match*stereomatch*gender	-0.51143	1.03222	-0.495

Table B.18.: Eye-movement analysis, Experiment 3 (objects), verb region

Maximal model: $\log \sim v_a_match * stereomatch * gender$

$+(1+v_a_match * stereomatch | participant)$

$+(1+v_a_match + stereomatch | item)$

Reduced model: $\log \sim v_a_match * stereomatch * gender + (1+v_a_match | participant)$

$+(1+v_a_match | item)$

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-0.3864	0.4978	-0.776
v_a_match	2.8685	0.2924	4.609*
stereomatch	-0.2759	0.5830	-0.473
gender	-0.9906	0.7039	-1.407
v_a_match*stereomatch	0.9710	0.8245	1.178
v_a_match*gender	0.8354	0.8802	0.949
stereomatch*gender	1.0774	0.8246	1.307
v_a_match*stereomatch*gender	-1.0330	1.1655	-0.886

B.5 | Time-course graphs: percentage of looks, Experiment 3

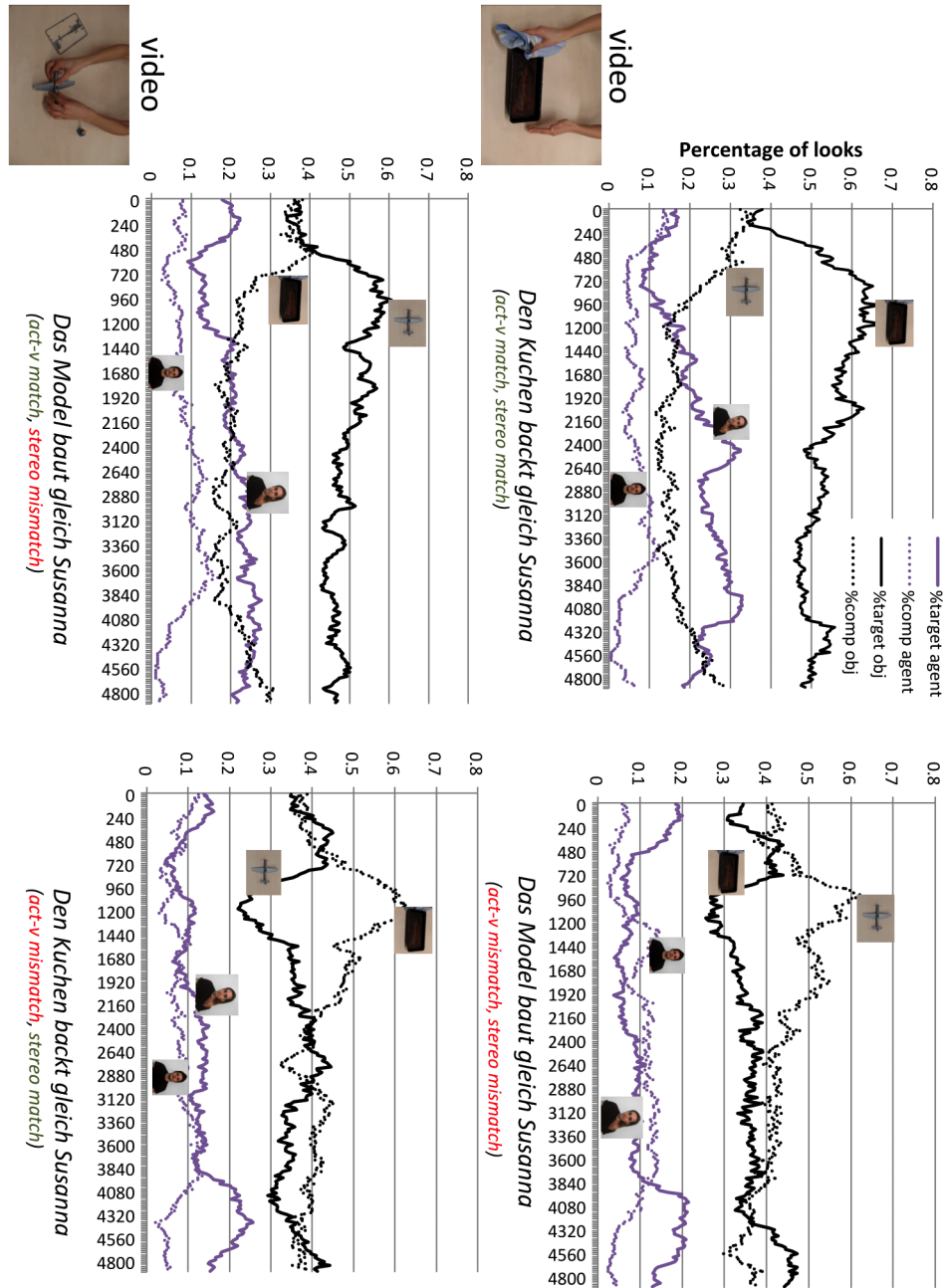


Figure B.1.: Time-course graphs per condition with percentages of looks, Experiment 3

References

- Abashidze, D., Carminati, M. N., & Knoeferle, P. (2014). How robust is the recent event preference? In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 92–97). Austin, TX: Cognitive Science Society.
- Abashidze, D., & Knoeferle, P. (2015, September). Do people prefer to inspect the target of a recent action?: The case of verb-action mismatches.. Poster presented at the Architectures and Mechanisms for Language Processing Conference (AMLaP 2015), Valetta, Malta.
- Ainsworth, C. (2015). Sex redefined. *Nature*, *518*(7539), 288.
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.
- Altmann, G. T. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41*(1), 124–145.
- Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, *93*(2), B79–B87.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.
- Altmann, G. T., & Kamide, Y. (2004). Now you see it, now you don’t: Mediating the mapping between language and the visual world. *The interface of Language, Vision, and Action: Eye Movements and the Visual World*, 347–386.

- Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.
- Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71.
- Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609.
- Andersen, S. M., Klatzky, R. L., & Murray, J. (1990). Traits and social stereotypes: Efficiency differences in social information processing. *Journal of Personality and Social Psychology*, *59*(2), 192.
- Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, *137*(2), 181–189.
- Arai, M., Van Gompel, R. P., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250.
- Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527–565.
- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.
- Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: a dynamic account of the n400. *Language and Cognitive Processes*, *26*(9), 1338–1367.
- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*(2), 272.

- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–153.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230.
- Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, *137*(2), 201–207.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.
- Bassetti, B. A. (2014). Is grammatical gender considered arbitrary or semantically motivated? Evidence from young adult monolinguals, second language learners, and early bilinguals. *British Journal of Psychology*, *105*(2), 273–294.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. *The Crosslinguistic Study of Sentence Processing*, *3*, 73–112.
- Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*(6), 1142–1163.
- Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. *The Sage Handbook of Social Cognition*, 311–329.
- Bojarska, K. (2013). Responding to lexical stimuli with gender associations: a cognitive-cultural model. *Journal of Language and Social Psychology*, *32*(1), 46–61.
- Boland, J., Acker, M., & Wagner, L. (1998). The use of gender features in the resolution of pronominal anaphora. *Cognitive Science Technical Report*, *17*.
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: re-

- thinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143.
- Brouwer, H., & Hoeks, J. C. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*(758), 1-12.
- Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676–713.
- Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, *28*(02), 277–293.
- Carminati, M. N., & Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PloS One*, *8*(9), e72559.
- 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: Section A*, *49*(3), 639–663.
- Casasanto, L. S., Hofmeister, P., & Sag, I. (2010). Understanding acceptability judgments: Additivity and working memory effects. In *Proceedings of the Cognitive Science Society* (Vol. 32).
- Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687.
- Chow, W.-Y., Lago, S., Barrios, S., Parker, D., Morini, G., & Lau, E. (2014). Additive effects of repetition and predictability during comprehension: evidence from event-related potentials. *PloS One*, *9*(6), e99199.
- Clifton, C., & Staub, A. (2011). Syntactic influences on eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of*

- Eye Movements* (p. 895—909). Oxford: Oxford University Press.
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye movements: A Window on Mind and Brain*, 341–372.
- Coco, M. I., Araujo, S., & Petersson, K. M. (2016). Disentangling stimulus plausibility and contextual congruency: Electro-physiological evidence for differential cognitive dynamics. *Neuropsychologia*, 96, 150–163.
- Cole, C. M., Hill, F. A., & Dayley, L. J. (1983). Do masculine pronouns used generically lead to thoughts of men? *Sex Roles*, 9(6), 737–750.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6(1), 84–107.
- Corbett, G. G. (1991). Cambridge University Press.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19(4), 245–264.
- Crocker, M. W. (2010). Computational psycholinguistics. In A. Clark, C. Fox, & S. Lapin (Eds.), *The Handbook of Computational Linguistics and Natural Language Processing Handbook*. London, UK: Blackwell.
- Crocker, M. W., Knoeferle, P., & Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112(3), 189 - 201. doi: <http://doi.org/10.1016/j.bandl.2009.03.004>
- de Lemus, S., Spears, R., Bukowski, M., Moya, M., & Lupiáñez, J. (2013). Reversing implicit gender stereotype activation as a function of exposure to traditional gender roles. *Social Psychology*, 109-116.
- Dienes, Z., Altmann, G., & Gao, S.-J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science*, 23(1), 53–82.
- 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.
- Dumitru, M. L., Joergensen, G. H., Cruickshank, A. G., & Altmann, G. T. (2013). Language-guided visual processing affects reasoning: The role of referential and spatial anchoring. *Consciousness and Cognition*, 22(2), 562–571.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of psycholinguistic research*, 24(6), 409–436.
- Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology*, 32(2), 247–255.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, 69(3), 165–182.
- Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44(4), 516–547.
- Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology*, 32(1), 18–31.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. 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(2), 178–210.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84.
- Gaetano, J., van der Zwan, R., Blair, D., & Brooks, A. (2014). Hands as sex cues: Sensitivity measures, male bias measures, and implications for sex perception mechanisms. *PloS One*, 9(3), e91032.
- Garnham, A. (1981). Mental models as representations of text. *Memory & Cognition*, 9(6), 560–565.

- Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes*, 7(3-4), 231–255.
- Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, 30(3), 439–446.
- Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, 4(2), 161–178.
- Gernsbacher, M. A. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of Learning and Motivation*, 27, 217–263.
- Gough, P. B. (1965). Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, 4(2), 107–111.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95–112.
- Gygax, P. M., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, 23(3), 464–485.
- Gygax, P. M., Garnham, A., & Doehren, S. (2016). What do true gender ratios and stereotype norms really tell us? *Frontiers in Psychology*, 7(1036).
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: Erp effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883–899.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483.
- 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.

- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.
- Hanulíková, A., & Carreiras, M. (2015). Electrophysiology of subject-verb agreement mediated by speakers' gender. *Frontiers in Psychology*, *6*, 1396.
- Harper, M., & Schoeman, W. J. (2003). Influences of gender as a basic-level category in person perception on the gender belief system. *Sex roles*, *49*(9-10), 517–526.
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836.
- Hirschfeld, L., Bartmess, E., White, S., & Frith, U. (2007). Can autistic children predict behavior by social stereotypes? *Current Biology*, *17*(12), R451 - R452.
- Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes*, *5*(4), 281–312.
- Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*(4), 286–300.
- Holliday, A. (1999). Small cultures. *Applied Linguistics*, *20*(2), 237–264.
- Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.
- Huettig, F., Mishra, R. K., & Olivers, C. N. (2011). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology*, *2*.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.
- Irmer, M. (2011). *Bridging inferences: Constraining and Resolving Underspecification in Discourse Interpretation* (Vol. 11). Walter de Gruyter.
- Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. NY: Oxford University Press.

- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research, 1146*, 2–22.
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*(11), 1161–1166.
- Jegerski, J., VanPatten, B., & Keating, G. (2016). Relative clause attachment preferences in early and late Spanish-English bilinguals. *Advances in Spanish as a Heritage Language, 49*, 81–99.
- Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior, 10*(3), 244–253.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2*(4), 647–670.
- Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133–156.
- Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37–55.
- Knoeferle, P., Carminati, M. N., Abashidze, D., & Essig, K. (2011). Preferential inspection of recent real-world events over future events: evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology, 2*, 376.
- Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*(3), 481–529.
- Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*(4), 519–543.
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*(1), 95–127.

- Knoeferle, P., & Guerra, E. (2016). Visually situated language comprehension. *Language and Linguistics Compass*, 10(2), 66–82.
- Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: evidence from erps during situated spoken comprehension. *Cerebral Cortex*, 18(4), 789–795.
- Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from erps and picture-sentence verification. *Psychophysiology*, 48(4), 495–506.
- Knoeferle, P., Urbach, T. P., & Kutas, M. (2014). Different mechanisms for role relations versus verb–action congruence effects: Evidence from erps in picture–sentence verification. *Acta Psychologica*, 152, 133–148.
- Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and language*, 85(1), 1–36.
- Köpcke, K.-M., Panther, K.-U., & Zubin, D. A. (2010). Motivating grammatical and conceptual gender agreement in German. *Cognitive Foundations of Linguistic Usage Patterns*, 171–194.
- Krauss, R. M., & Chiu, C.-Y. (1998). Language and social behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (Vol. 1-2). McGraw-Hill.
- Kreiner, H., Mohr, S., Kessler, K., & Garrod, S. (2009). Can context affect gender processing? erp differences between definitional and stereotypical gender. *Brain Talk: Discourse with and in the Brain*, 107–119.
- Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, 58(2), 239–261.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, 4(12), 463–470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in

- the n400 component of the event-related brain potential (erp). *Annual Review of Psychology*, *62*, 621–647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–3.
- Lamers, M. J., Jansma, B. M., Hammer, A., & Münte, T. F. (2006). Neural correlates of semantic and syntactic processes in the comprehension of case marked pronouns: evidence from German and Dutch. *BMC Neuroscience*, *7*(1), 23.
- Langeslag, S. J., & van Strien, J. W. (2009). Aging and emotional memory: The co-occurrence of neurophysiological and behavioral positivity effects. *Emotion*, *9*(3), 369.
- Leinbach, M. D., Hort, B. E., & Fagot, B. I. (1997). Bears are for boys: Metaphorical associations in young children's gender stereotypes. *Cognitive Development*, *12*(1), 107 – 130. doi: [https://doi.org/10.1016/S0885-2014\(97\)90032-0](https://doi.org/10.1016/S0885-2014(97)90032-0)
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676.
- Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*(1), 37.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102.
- Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*(5), 806–816.
- Matin, E., Shao, K., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380.
- Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, *33*(3), 449–496.

- McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, 46, 155–200.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440.
- McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7), 1174–1184.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283–312.
- Molinaro, N., Su, J.-J., & Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language*, 155-156.
- Morrison, A. B., Conway, A. R., & Chein, J. M. (2014). Primacy and recency effects as indices of the focus of attention. *Frontiers in Human Neuroscience*, 8, 6.
- Most, S. B., Sorber, A. V., & Cunningham, J. G. (2007). Auditory stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology*, 43(2), 287 - 294.
- Münster, K., & Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: towards socially situated interpretation. *Frontiers in Psychology*, 8, 2267.
- Münster, K. (2016). *Effects of Emotional Facial Expressions and Depicted Actions on Situated Language Processing Across the Lifespan* (Unpublished doctoral dissertation). Bielefeld University.
- Münster, K., Carminati, M. N., & Knoeferle, P. (2014). How do static and dynamic emotional faces prime incremental semantic interpretation?: Comparing older and younger adults. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2675–2680). Austin, TX: Cognitive Science Society.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials.

- Journal of Cognitive Neuroscience*, 3(2), 151–165.
- Oakes, P. (1996). Social groups and identities. In W. P. Robinson & H. Tajfel (Eds.), *Social groups and identities: Developing the legacy of Henri Tajfel* (pp. 95–120). Butterworth Heinemann.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25(3), 273–285.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785–806.
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, 8(4), 413–437.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786.
- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14(3), 283–317.
- Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General*, 115(3), 281.
- Pykkönen, P., Hyönä, J., & van Gompel, R. P. (2010). Activating gender stereotypes during online spoken language processing. *Experimental Psychology*, 57(2), 126–133.
- Pykkönen-Klauck, P., & Crocker, M. W. (2016). Attention and eye movement metrics in visual world eye tracking. In P. Knoeferle, P. Pykkönen-Klauck, & M. W. Crocker (Eds.), *Visually Situated Language Comprehension*. John Benjamins.
- R Core Team. (2016). *R: A Language and Environment for Statistical computing*. Vienna, Austria. Retrieved from <https://www.R-project.org/>

- Reali, C., Esaulova, Y., Öttl, A., & von Stockhausen, L. (2015). Role descriptions induce gender mismatch effects in eye movements during reading. *Frontiers in Psychology*, *6*(1607).
- Reboul, A. C. (2015). Why language really is not a communication system: a cognitive view of language evolution. *Frontiers in Psychology*, *6*(1434).
- Reed, A. E., & Carstensen, L. L. (2012). The theory behind the age-related positivity effect. *Frontiers in psychology*, *3*, 339.
- Rodríguez, A., Burigo, M., & Knoeferle, P. (2015). Visual gender cues elicit agent expectations: different mismatches in situated language comprehension. In G. Airenti, B. Bara, & G. Sandini (Eds.), *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science (EAPcogsci 2015)* (p. 234—239). Aachen: CEUR-WS.org.
- Rodríguez, A., Burigo, M., & Knoeferle, P. (2016). Visual constraints modulate stereotypical predictability of agents during situated language comprehension. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 580–585). Austin, TX: Cognitive Science Society.
- Ryskin, R. A., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition*, *147*, 75–84.
- Sanford, A. J., & Garrod, S. C. (1981). *Understanding Written Language: Explorations of Comprehension Beyond the Sentence*. John Wiley & Sons.
- Schmitt, B. M., Lamers, M., & Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research*, *14*(3), 333–346.
- Schneider, J. W., & Hacker, S. L. (1973). Sex role imagery and use of the generic "man" in introductory texts: A case in the sociology of sociology. *The American Sociologist*, 12–18.
- Scott-Phillips, T. (2014). *Speaking Our Minds: Why Human Communication Is Different, and How Language Evolved to Make It Special*. Palgrave MacMillan.

- 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*(4), 558.
- Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development, 25*(1), 7-15. doi: 10.1080/01650250042000078
- Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75*(3), 589.
- Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2000). Attention and stereotyping: Cognitive constraints on the construction of meaningful social impressions. *European Review of Social Psychology, 11*(1), 145–175.
- Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS One, 7*(12), e48712.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological research, 65*(4), 235–241.
- Spivey, M. J., & Huettenlocher, S. a. (2016). Toward a situated view of language. In P. Knoeferle, P. Pyykkönen-Klauck, & M. Crocker (Eds.), *Visually Situated Language Comprehension*. John Benjamins Publishing Amsterdam.
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*(4), 447–481.
- Squires, L. (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics, 17*(2), 200–237.
- Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62*(2), 207.

- Staub, A., & Clifton, C. (2011). Processing effects of an indeterminate future: Evidence from self-paced reading. *University of Massachusetts Occasional Papers in Linguistics*, 38, 131–140.
- Steele, S. (1978). Word order variation: A typological study. *Universals of Human Language*, 4, 585–623.
- Streb, J., Hennighausen, E., & Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, 33(3), 175–201.
- Tanenhaus, M. K., Boland, J., Garnsey, S. M., & Carlson, G. N. (1989). Lexical structure in parsing long-distance dependencies. *Journal of Psycholinguistic Research*, 18(1), 37–50.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632.
- Taylor, S. E., & Hamilton, D. L. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive Processes in Stereotyping and Intergroup Behavior* (pp. 83–114). Taylor & Francis.
- Tesink, C. M., Buitelaar, J., Petersson, K. M., Van der Gaag, R., Kan, C., Tendolkar, I., & Hagoort, P. (2009). Neural correlates of pragmatic language comprehension in autism spectrum disorders. *Brain*, 132(7), 1941–1952.
- 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(3), 285.
- Tversky, A., & Kahneman, D. (1973). Judgment under uncertainty: Heuristics and biases. In D. Wendt & C. Vlek (Eds.), *Utility, Probability, and Human Decision Making* (pp. 141–162). Boston: Dordrecht-Holland.
- Van Berkum, J. J. (1996). *The psycholinguistics of grammatical gender: Studies in language comprehension and production* (Unpublished doctoral dissertation). University of Nijmegen.

- Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.
- Van De Meerendonk, N., Kolk, H. H., Vissers, C. T. W., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different erp patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *International Journal of Psychophysiology*, *83*(2), 176–190.
- Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*(4), 455–478.
- Vissers, C. T. W., Kolk, H. H., Van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: evidence from erps in a picture–sentence matching task. *Neuropsychologia*, *46*(4), 967–982.
- Wannemacher, J. T. (1974). Processing strategies in picture-sentence verification tasks. *Memory & Cognition*, *2*(3), 554–560.
- Wassenaar, M., & Hagoort, P. (2007). Thematic role assignment in patients with broca’s aphasia: Sentence–picture matching electrified. *Neuropsychologia*, *45*(4), 716–740.
- Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63–B72.
- Wendler, K., Burigo, M., Schack, T., & Knoeferle, P. (2016, June). Role versus action mismatches in situated language comprehension: A blank screen study.. Presented at the MODELACT conference on Action, Language and Cognition, Rome, Italy.
- White, K. R., Crites, S. L., Taylor, J. H., & Corral, G. (2009). Wait, what? assess-

- ing stereotype incongruities using the n400 erp component. *Social Cognitive and Affective Neuroscience*, nsp004.
- White, S., Hill, E., Winston, J., & Frith, U. (2006). An islet of social ability in asperger syndrome: judging social attributes from faces. *Brain and Cognition*, *61*(1), 69–77.
- Whorf, B. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. (J. B. Carroll, Ed.). MIT Press.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, *39*(3), 483–508.
- Wild, H. A., Barrett, S., Spence, M., O’Toole, A., Cheng, Y., & Brooke, J. (2000). Recognition and sex categorization of adults’ and children’s faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology*, *77*(4), 269–291.
- Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235–1249.
- Xu, X., Jiang, X., & Zhou, X. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain and Cognition*, *81*(2), 223–236.
- Yap, M. J., & Pexman, P. M. (2016). Semantic richness effects in syntactic classification: The role of feedback. *Frontiers in Psychology*, *7*(1394).
- Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.