

Effects of Emotional Facial Expressions and Depicted
Actions on Situated Language Processing
Across the Lifespan

by

Katja Münster

A thesis submitted in partial fulfillment of the requirements for the degree of

Doktor der Philosophie

at the Faculty of Linguistics and Literature

Universität Bielefeld

Bielefeld, 2016

Evaluators:

Prof. Dr. Pia Knoeferle,
Humboldt-Universität zu Berlin
Prof. Dr. Stavros Skopeteas,
Universität Bielefeld

Contents

List of Tables	vi
List of Figures	vii
List of Graphs	viii
Abstract	xii
Acknowledgments	xv
1. Introduction	1
1.1 Thesis Motivation	2
1.2 Thesis Aims	3
1.3 Thesis Outline	4
2. Incremental Language Processing	6
2.1 Incrementality in Relation to Syntactic Structure	7
2.2 Incrementality in Relation to Thematic Role Assignment	10
2.3 Language Processing: Language Development	14
3. The Use of Visual Context for Language Processing	26
3.1 The Visual Context – Basic Assumptions	26
3.1.1 Defining the Visual Context	27
3.1.2 Present Working Definition	29
3.2 Information Extraction from the Visual Context	30
3.3 Different Kinds of Visual Contextual Cues – Direct Cues: Actions and Objects	32
3.4 Different Kinds of Visual Contextual Cues – Indirect Cues: Social Cues	34
3.5 Cumulative Use of Direct and Indirect Cues	39
3.6 Differences between Direct and Indirect Cues	40
3.7 Emotions and Emotional Facial Expressions	43
3.7.1 Identification and Processing of Emotions	43
3.7.1.1 Younger Adults	46
3.7.1.2 Older Adults	48
3.7.1.2.1 Underlying Reasons	52

3.7.1.3 Children	54
3.7.2 The Degree of Naturalness in the Depiction of Facial Expressions	60
4. Arguments for the Inclusion of Social Cues into Real-time Situated Language Processing Accounts	63
4.1 The Interplay between Language Processing and Non-linguistic Visual Context	64
4.2 Language Processing Accounts and their Limitations	66
4.3 The Importance of Including Social Cues in Real-time Language Processing Accounts	69
4.3.1 Developmental Arguments for the Inclusion of Social Cues into Real-time Language Processing Accounts	72
4.4 A Suitable Language Processing Account for the Enrichment with Social Cues – The CIA	75
5. Motivation and Research Questions	82
5.1 Design and Conditions	83
5.2 Analyses	87
5.2.1 Regions of Interest	87
5.2.2 Word Regions	88
5.2.3 Inferential Analyses	89
5.2.4 Descriptive Analyses	91
5.3 Linking Hypothesis	91
5.4 Predictions	92
5.4.1 Eye-movement Predictions	92
5.4.2 Accuracy Predictions	94
6. Experiments – Younger Adults	95
6.1 Procedure	95
6.2 Experiment 1	96
6.2.1 Participants	96
6.2.2 Materials	96
6.2.2.1 Critical Items	96

6.2.2.2 Filler Items	98
6.2.2.3 Counterbalancing	98
6.2.3 Results and Discussion	99
6.2.3.1 Eye-movement Results	99
6.2.3.1 Accuracy Results	101
6.2.4 Summary and Conclusion	103
6.3 Experiment 2	105
6.3.1 Participants	105
6.3.2 Materials and Hypotheses	105
6.3.3 Results and Discussion	106
6.3.3.1 Eye-movement Results	106
6.3.3.2 Accuracy Results	106
6.3.4 Summary and Conclusion	107
6.4 Experiment 3	108
6.4.1 Participants	108
6.4.2 Materials	108
6.4.3 Hypotheses	109
6.4.4 Results and Discussion	110
6.4.4.1 Eye-movement Results	110
6.4.4.2 Accuracy Results	112
6.4.4.3 Effect Size Differences: Schematic vs. Natural Prime	113
6.4.5 Summary and Conclusion	115
7. Experiments – Children	117
7.1 Pretest: Character, Action and Valence Recognition	117
7.1.1 Character and Action Recognition	117
7.1.2 Valence Recognition	118
7.1.3 Pretest Results	118
7.2 Experiment 4	119
7.2.1 Participants	119
7.2.2 Materials and Cognitive Tests	119
7.2.3 Hypotheses	119
7.2.4 Results and Discussion	120

7.2.4.1 Eye-movement Results	120
7.2.4.2 Accuracy Results	123
7.2.4.3 Cognitive Test Results	126
7.2.5 Summary and Conclusion	128
7.3 Experiment 5	129
7.3.1 Participants	129
7.3.2 Materials and Cognitive Tests	129
7.3.3 Hypotheses	129
7.3.4 Results and Discussion	130
7.3.4.1 Eye-movement Results	130
7.3.4.2 Accuracy Results	131
7.3.4.3 Cognitive Test Results	132
7.3.5 Summary and Conclusion	133
8. Experiment – Older Adults	134
8.1 Experiment 6	134
8.1.1 Participants	134
8.1.2 Materials and Cognitive Tests	134
8.1.3 Hypotheses	135
8.1.4 Results and Discussion	135
8.1.4.1 Eye-movement Results	135
8.1.4.2 Accuracy Results	137
8.1.4.3 Cognitive Test Results	138
8.1.5 Summary and Conclusion	141
9. Age Comparison for Experiments 3, 5 and 6	142
9.1 Results and Discussion	142
9.1.1 The Action Effect	142
9.1.2 The Positive Prime Effect	146
9.1.3 Cumulative Effects	148
9.2 Summary and Conclusion	150
10. General Discussion	151
10.1 The Action Effect	151

10.2 The Positive Prime Effect	154
10.3 Cumulative Visual Context Effects	156
10.4 The Influence of Cognitive Abilities	157
10.5 The Naturalness of the Emotional Prime Face	158
10.6 Age Differences in Emotional Prime Use	160
10.6.1 Younger Adults' Positive Prime Effect	161
10.6.2 Older Adults' Positive Prime Effect	161
10.6.3 Children's Positive Prime Effect	162
10.7 Time Course Differences	164
10.8 Towards an Extended Coordinated Interplay Account	167
10.8.1 The Social Coordinated Interplay Account	168
10.8.1.1 The sCIA – an Example	170
11. Final Conclusions and Impact	175
12. German Summary	182
Appendix	190
Appendix A: Critical Item Sentences and Word Onsets	190
Appendix B: Primes	192
Appendix C: Critical Item Images	194
Appendix D: Valence Pretest Items	211
Appendix E: Complementary Accuracy Analyses	214
References	221

List of Tables

Table 1: Analyzed German word regions, their English translation, mean word onsets and longest duration for each word region.

89

List of Figures

Figure 1: The revised Coordinated Interplay Account.	77
Figure 2: Two potential agents acting upon a patient.	78
Figure 3: Study procedure for all experiments.	85
Figure 4: Experimental conditions for experiments 1, 2 and 4.	86
Figure 5: Experimental conditions for experiments 3, 5 and 6.	87
Figure 6: The social Coordinated Interplay Account.	170

List of Graphs

Graph 1: Time course for the eye-movement results by condition (exp. 1).	99
Graph 2: Mean log ratio of looks for the main effect of action in the Verb-Adverb region (F_1 , exp. 1).	100
Graph 3: Mean log ratio of looks for the interaction between prime and action in the Adverb region (F_2 , exp. 1).	101
Graph 4: Accuracy for the main effect of prime (exp. 1).	102
Graph 5: Accuracy for the voice x prime interaction (exp. 1).	103
Graph 6: Accuracy for the voice x prime x action interaction (exp. 2).	107
Graph 7: Time course for the eye-movement results by condition (exp. 3).	110
Graph 8: Mean log ratio of looks for the main effect of prime in the Verb-Adverb region (F_1 , exp. 3).	111
Graph 9: Mean log ratio of looks for the interaction between prime and action in the Verb-Adverb region (not sign., exp. 3).	112
Graph 10: Accuracy for the main effect of action (exp. 3).	113
Graph 11: Mean generalized eta squared of the action effect for the natural face (exp. 3) compared to the schematic face (exp. 2) of the real-time data.	114

Graph 12: Mean generalized eta squared of the prime effect for the natural face (exp. 3) compared to the schematic face (exp. 2) of the real-time data.	114
Graph 13: Time course for the eye-movement results by condition for children and younger adults (exp. 1, 4).	121
Graph 13a: children's time course (exp. 4).	121
Graph 13b: younger adults' time course (exp. 1).	121
Graph 14: Mean log ratio of looks for the main effect of action in the Adverb region (F_1 , exp. 4).	121
Graph 15: Mean log ratio of looks (agent / distractor) for the main effect of prime in the NP1 region (F_2 , exp. 4).	123
Graph 16: Mean log ratio of looks for the effect of prime in the NP1 region (not sign., exp. 4).	123
Graph 17: Accuracy for the main effect of action (exp. 4).	125
Graph 18: Accuracy for the voice x action interaction (exp. 4).	125
Graph 19: Accuracy for the voice x prime interaction (exp. 4).	126
Graph 20: Accuracy for the voice x prime x action interaction (exp. 4).	126
Graph 21: Cognitive test scores for the K-ABC test (exp. 4).	127
Graph 22: Correlation between K-ABC and accuracy scores (exp. 4).	128
Graph 23: Time course for the eye-movement results by condition (exp. 5).	130
Graph 24: Mean log ratio of looks for the main effect of action in the Adverb region (F_1 , exp. 5).	130

Graph 25: Accuracy for the main effect of action (exp. 5).	132
Graph 26: Cognitive test scores for the K-ABC test (exp. 5).	132
Graph 27: Time course for the eye-movement results by condition (exp. 6).	136
Graph 28: Mean log ratio of looks for the interaction between prime and action in the Adverb region (not sign., exp. 6).	136
Graph 29: Mean log ratio of looks for the main effect of prime in the NP2 region (F_1 , exp. 6).	137
Graph 30: Accuracy for the main effect of action (exp. 6).	138
Graph 31: Cognitive test scores for the WAIS test (exp. 6).	139
Graph 32: Correlation between WAIS and accuracy scores (exp. 6).	139
Graph 33: Mean log ratio of looks for the interaction between cognitive test results and action in the NP2 region (exp. 6).	140
Graph 34: Mean log ratio of looks for the interaction between cognitive test results, action and prime in the NP2 region (exp. 6).	141
Graph 35: Time course for the eye-movement results by condition for each age group (exp. 3, 5, 6).	143
Children's time course (exp. 5): Graph 35a.	143
Younger adults' time course (exp. 3): Graph 56b.	143
Older adults' time course (exp. 6): Graph 35c.	143


Graph 36: Mean log ratio of looks for the interaction between age (older vs. younger adults, exp. 6 vs. exp. 3) and action in the Verb region (F_1).	144
Graph 37: Mean log ratio of looks for the interaction between age (children vs. younger adults, exp. 5 vs. exp. 3) and action in the Verb region (F_2).	144
Graph 38: Mean generalized eta squared for the interaction between the visual cue effect (action and prime) and age of the real-time data (exp. 3, 5, 6).	145
Graph 39: Accuracy for the action x age interaction (exp. 3, 5, 6).	145
Graph 40: Mean generalized eta squared for the visual cue effects (action and prime) of the real-time data (exp. 3, 5, 6).	146
Graph 41: Mean log ratio of looks for the main effect of prime in the Long region (F_1 , exp. 3, 6).	147
Graph 42: Mean log ratio of looks for the interaction between age and prime in the Adverb region (F_2 , exp. 3, 6).	147
Graph 43: Mean log ratio of looks for the interaction between action and prime in the NP2 region (F_2 , exp. 3, 6).	148
Graph 44: Mean log ratio of looks for the interaction between age, action and prime in the Long region (F_1 , exp. 3, 6).	149
Graph 45: Mean generalized eta squared of the prime effect of the real-time data (exp. 3, 6).	149

Abstract

Language processing does not happen in isolation, but is often embedded in a rich non-linguistic visual and social context. Yet, although many psycholinguistic studies have investigated the close interplay between language and the visual context, the role of social aspects and listener characteristics in real-time language processing remains largely elusive. The present thesis aims at closing this gap.

Taking extant literature regarding the incrementality of language processing, the close interplay between visual and linguistic context and the relevance for and effect of social aspects on language comprehension into account, we argue for the necessity to extend investigations on the influence of social information and listener characteristics on real-time language processing. Crucially, we moreover argue for the inclusion of social information and listener characteristics into real-time language processing accounts. Up-to-date, extant accounts on language comprehension remain elusive about the influence of social cues and listener characteristics on real-time language processing. Yet a more comprehensive approach that takes these aspects into account is highly desirable given that psycholinguistics aims at describing how language processing happens in real-time in the mind of the comprehender.

In 6 eye-tracking studies, this thesis hence investigated the effect of two distinct visual contextual cues on real-time language processing and thematic role assignment in emotionally valenced non-canonical German sentences. We are using emotional facial expressions of a speaker as a visual social cue and depicted actions as a visual contextual cue that is directly mediated by the linguistic input. Crucially, we are also investigating the effect of the age of the listener as one type of listener characteristics in testing children and older and younger adults.

In our studies, participants were primed with a positive emotional facial expression (vs. a non-emotional / negative expression). Following this they inspected a target scene depicting two potential agents either performing or not performing an action towards a patient. This scene was accompanied by a related positively-valenced German Object-Verb-Adverb-Subject sentence (e.g.,: *The ladybug*_(accusative object, patient) *tickles happily the cat*_(nominative object, agent)). Anticipatory eye-movements to the agent of the action, i.e., the sentential subject in sentence end position (vs. distractor agent), were measured in order to investigate if, to what extent and how rapidly  positive emotional facial expressions and depicted actions

can facilitate thematic role assignment in children and older and younger adults. Moreover, given the complex nature of emotional facial expressions, we also investigated if the naturalness of the emotional face has an influence on the integration of this social cue into real-time sentence processing. We hence used a schematic depiction of an emotional face, i.e., a happy smiley, in half of the studies and a natural human emotional face in the remaining studies.

Our results showed that all age groups could reliably use the depicted actions as a cue to facilitate sentence processing and to assign thematic roles even before the target agent had been mentioned. Crucially, only our adult listener groups could also use the emotional facial expression for real-time sentence processing. When the natural human facial expression instead of the schematic smiley was used to portray the positive emotion, the use of the social cue was even stronger. Nevertheless, our results have also suggested that the depicted action is a stronger cue than the social cue, i.e., the emotional facial expression, for both adult age groups. Children on the other hand do not yet seem to be able to also use emotional facial expressions as visual social cues for language comprehension. Interestingly, we also found time course differences regarding the integration of the two cues into real-time sentence comprehension. Compared to younger adults, both older adults and children were delayed by one word region in their visual cue effects.

Our on-line data is further supported by accuracy results. All age groups answered comprehension questions for ‘who is doing what to whom’ more accurately when an action was depicted (vs. was not depicted). However, only younger adults made use of the emotional cue for answering the comprehension questions, although to a lesser extent than they used depicted actions.

In conclusion, our findings suggest for the first time that different non-linguistic cues, i.e., more direct referential cues such as depicted actions and more indirect social cues such as emotional facial expressions, are integrated into situated language processing to different degrees. Crucially, the time course and strength of the integration of these cues varies as a function of age.

Hence our findings support our argument regarding the inclusion of social cues and listener characteristics into real-time language processing accounts. Based on our own results we have therefore outlined at the end of this thesis, how an account of real-time language comprehension that already takes the influence of visual context such as depicted actions into account (but fails to include social aspects and listener

characteristics) can be enriched to also include the effects of emotional facial expressions and listener characteristics such as age.



Acknowledgments

During the last three years I have learned and achieved more than I expected to be possible. I had to push the boundaries – sometimes to the extreme – to be able to redefine these boundaries and to go beyond of what I thought I was capable of doing. Yet, setting my own limits higher and higher strengthened my confidence and showed me how good it feels to “go beyond”. Taking on the adventure of a PhD can sometimes be frightening, full of despair and self-doubt. But without failure there is no success. Overcoming difficulties, no matter how small or big they may seem leads to highly valuable experiences and are victories of their own. Nevertheless, no one wins a battle on his or her own. Many people were involved in this project in very different ways and I am deeply grateful for the support and guidance of each of them. Without these people I would have never come this far, I wouldn’t be who I am today and this project wouldn’t have become what it is now.

First of all, my profoundest thanks go to my supervisor Prof. Dr. Pia Knoeferle. If it wasn't for her, I wouldn't have started this PhD in the first place. Her classes already sparked my interest in psycholinguistics, language processing, emotions and cognition during my master's degree. She encouraged me to take the next step and start a PhD. As my supervisor, she gave me guidance and support whenever I needed it. She prepared me for my future academic career and her door was always open. She pushed my boundaries but never forced me to cross them. Yet, if I did she always had an open ear and a pep talk to guide me across the boundaries.

Secondly I would like to thank Dr. Maria Nella Carminati. She helped me immensely during the starting phase of this project and taught me everything I need to know about empirical and academic work. Profound thanks also go to my colleagues in the Language and Cognition Group. Thank you so much for our interesting discussions but also for forcing me to take lunch and coffee breaks, providing me with cookies and taking my mind off work during our breaks.

I would also like to say thank you to the numerous people who have advised and encouraged me on my work and career during my research stay at the University of California, San Diego. Many thanks go to Prof. Dr. Marta Kutas for welcoming me in her lab and sharing her story about giving talks. My thanks go also to Dr. Tom Urbach for our great discussions and sessions on and about using EEG and measuring

ERPs and to Prof. Dr. Arielle Borovsky for our fruitful and interesting discussions whenever and wherever we have had the chance to meet.

Last but not least, my deepest thanks go to my family and friends. Without my family I wouldn't be who I am today and without my friends I wouldn't have a life outside of work. Thank you for your endless support and encouragements. Thank you for patiently listening to my complaints and sorrows. Thank you for bringing me back on track whenever I needed it. Thank you for making sure I have a work-life balance and most of all, thank you for your love and patience!

1. Introduction

Humans are social beings, and as such are guided by a strong desire to build social relations as well as to engage in social and communicative events. Yet, social communication requires the rapid real-time processing and interpretation of our interlocutors' utterances. As adults we process and interpret linguistic and non-linguistic visual input on a daily basis and with seemingly minimal effort. Often we are not even aware of the complex processes that are happening in our brain while we engage in social interactions.

Just imagine listening to your cousin who approaches you with the broadest smile on her face that you have ever seen and tells you how she met the love of her life. Even before she starts talking you have already seen and processed her facial expression. That facial expression, this huge smile on her face, tells you an essential part of the story she is going to tell you, namely that what she is about to say will most definitely be something positive. You have hence interpreted her smile and (implicitly) set up an expectation of what will follow even before your cousin has said a word. When she starts talking, your interpretation of the meaning of her smile will quickly be confirmed by the content and the matching positive valence of the linguistic input. Imagine now further, that after having told you all about how she met the love of her life, she abruptly stops, smiles even more broadly (although you thought that wasn't even possible), shows you her hand and says: "... and we just got engaged!". Now, what presumably happens within milliseconds in your brain is that you again process, identify and interpret her even broader smile and set up an (implicit) expectation for what will come next, i.e., something even more positive than what she has already told you. Her outstretched hand draws your attention, you look at it and immediately fixate the salient beautifully sparkling ring on her finger. Even before your cousin had a chance to utter the word "engaged" you have probably correctly interpreted the link between her huge smile, the ring on her finger and the beginning of the utterance ("...and we just got...") so that it only needed the word "engaged" to confirm what you had already (unconsciously) expected to hear next.

In psycholinguistic research, we try to understand how utterances such as the example from this story are processed in real-time in the mind of the comprehender. A considerable amount of studies has shown that there is a rapid interplay between the visual context, such as the engagement ring from our story, and the linguistic

input we process while looking at this visual context. Moreover, we do make extensive use of this visual context to facilitate language comprehension and thematic role assignment in real-time (e.g., Kamide, Scheepers, & Altmann, 2003b; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

Although we know how important the visual context is for language processing, researchers have mainly focused on investigating one type of information in the visual context and its effect on comprehension, namely depicted objects and events that can be directly mediated by the linguistic input (such as an engagement ring and the word “engage”, e.g., Knoeferle et al., 2005, Zhang & Knoeferle, 2012).

Nevertheless, just imagine your cousin's facial expression was not happy, but that she approached you crying and with a sad face. Now, would you expect the same story? Would you interpret her utterance (“...and we just got engaged!”) in the same way as you would have done if she had smiled? Maybe not. Imagine further then, a slightly different situation in which you are five years old. Now your grown-up cousin approaches you with a broad smile and tells you how she met the love of her life, shows you her hand and says: “... and we just got engaged!”. Would your 5-year old self be able to interpret the facial expression and the ring on your cousin's hand as quickly and accurately as your adult self would have done? Would your 5-year old self be able to quickly draw the right conclusions and interpret the sentence correctly using the facial expression and the ring as visual cues even before the word “engaged” was heard? Maybe, but maybe not.

1.1 Thesis Motivation

Thus, even though we would intuitively think that social cues, such as an emotional facial expression and listener characteristics such as the age of the comprehender have an effect on real-time language processing, psycholinguistic research remains largely elusive regarding their roles in real-time language comprehension. Yet, determining how, when during sentence processing and to which extent social visual cues affect real-time utterance comprehension is crucial given that psycholinguistic research aims at modeling how language comprehension happens in the mind of the listener. Furthermore if and how representations of these social cues interact with representations of other visual contextual cues, such as objects or actions events, is

1. Introduction

also still unclear. We do not yet know if we can use emotional facial expressions in the same way and with the same time course as for example depicted action events to facilitate sentence processing in real-time. Crucially, we moreover do not yet know if the age of the comprehender can also affect the use of social cues for on-line language comprehension. Can children and older adults use visual social cues in the same way, to the same extent and with the same time course as younger adults when processing a sentence?

Although psycholinguistic research remains elusive regarding these questions, their answers would provide crucial information for real-time language processing accounts. A few extant accounts of real-time language processing, such as the Coordinated Interplay Account, do already include the impact non-linguistic visual information has on language comprehension, yet they do not explicitly include social context (Knoeferle & Crocker 2006, 2007). However, language processing seldom happens in isolation and is situated in a rich, social, visual context. Hence, these accounts have to be extended to also include social information and listener characteristics in order to arrive at a more comprehensive understanding about how language processing functions in the mind of the comprehender.

1.2 Thesis Aims

In six studies, this thesis will therefore investigate the effect of social (emotional facial expressions) and non-social (depicted actions) visual context effects and their possible interaction during real-time sentence comprehension. We will focus on the question if these two visual context effects can be used to facilitate sentence processing and thematic role assignment. Moreover, we will investigate if the age of the comprehender modulates the use of these two visual cue types for language processing, and if the way the social cue is portrayed affects its usage in sentence processing. The results of these studies will furthermore be used to extend an account of real-time language processing that already includes visual context (Knoeferle & Crocker, 2007) to additionally include social cues and listener characteristics.

1.3 Thesis Outline

To motivate our research further, we will first discuss the incremental nature of language processing (Section 2). Section 2 will demonstrate that we can rapidly assign syntactic structure to a sentence and determine its thematic roles. We will further discuss that language processing is influenced by the structure of the linguistic input. Moreover, we will argue that language processing does not only happen incrementally in adults, but also already in children. Yet, children are still in the process of acquiring their native language and hence face processing difficulties especially when attempting to assign thematic roles in structurally challenging sentences.

Having looked at language comprehension mainly from the linguistic side, in Section 3 we turn to the use of the visual context for language processing. We will not only discuss how children and adults use objects and actions but also how adults use visual social cues for language processing. Furthermore we will argue for a distinction between visual context that is directly mediated by the linguistic input and more indirect visual social context that cannot be directly referenced to linguistic input. Since our studies use emotional facial expressions as one kind of social cue, we will argue for the complexity of social cues exemplified as emotional facial expressions. Here we will demonstrate the importance of listener characteristics such as age since the three tested age groups, i.e., children, younger adults and older adults, portray different preferences regarding the valence of emotional information. Finally, Section 3 will discuss how the degree of naturalness of emotional facial expressions might modulate their effects on language processing.

While Section 3 looked at the use of and differences in social and non-social visual cues for language processing across the lifespan, Section 4 will argue – also from a developmental perspective – in more detail for the inclusion of social cues into real-time processing accounts. We will discuss in detail how we link the linguistic and the visual context in language comprehension and discuss extant accounts of language processing and their limitations, especially regarding the inclusion of social cues. Before moving on to the empirical part of this thesis, we will outline the importance for extending language processing accounts to also include social information and listener characteristics. Finally, we will present the Coordinated Interplay Account

1. Introduction

(Knoeferle & Crocker, 2007) as an appropriate account for the enrichment with social cues and listener characteristics.

Section 5 will then prompt the presented studies and lay out their research questions. We will explain the design of the experiments, how we analyzed the data and what our hypotheses were.

Sections 6 to 9 will present the studies grouped by the age of the comprehender. We will start with the younger adults' experiments, followed by the children's studies and present the older adults' study in Section 8. Section 9 will look more closely at the age differences between younger adults, children and older adults and their use of emotional facial expressions and depicted actions for on-line sentence processing and thematic role assignment.

Taking previous research into account, we will discuss each of our findings in Section 10. Moreover, based on the results obtained in our studies, we will introduce a possible extension for the Coordinated Interplay Account (Knoeferle & Crocker, 2007). We will demonstrate and discuss the proposed new social and listener characteristic features and changes and exemplify how the enriched account could function using our own findings. Finally, we will present our conclusions and outline the importance of the present results for (psycho)linguistic research.

2. Incremental Language Processing

The ability to communicate with each other through the use of language is one of the most fundamental phenomena that distinguish humans from other primates. Even though we make extensive use of this ability on a daily basis, we are rarely aware of the underlying complex physical and cognitive mechanisms and processes that are necessary to successfully communicate with our kin. Setting aside language production, language comprehension alone is a topic that has been studied for many decades and still we are only beginning to understand how and why we can make sense of the linguistic input.

However, what we do know is that incoming linguistic input is processed incrementally and in real time (Kutas & Hillyard, 1984; Marslen-Wilson & Tyler, 1980; Tanenhaus et al., 1995). We analyze this input as it unfolds and make rapid use of different kinds of constraints, such as the structural combination of words, how to combine them into meaningful phrases and how to reference them in a larger discourse context (Hagoort, 2003; Kamide, Altmann, & Haywood, 2003a; Kamide et al., 2003b). Thus, new linguistic input contributes and shapes the unfolding interpretation of language. Moreover, we do not only incrementally integrate semantic meaning but make rapid use of references from preceding linguistic contexts (Altmann & Steedman, 1988).

Yet, how exactly and with which time course language can be processed incrementally, depends on factors such as the structure of the sentence and the ordering of thematic roles in a sentence. As the focus of this thesis is on the question of how different age groups can use visual cues to overcome processing difficulties regarding structurally challenging sentences and the assignment of thematic roles, Section 2.1 reviews literature on syntactic processing and shows that non-canonical and syntactically ambiguous sentences are more difficult to process than canonical and unambiguous sentences. Following up on processing difficulties, Section 2.2 will demonstrate that language processing is not solely a matter of assigning syntactic structure to a sentence. Hence, focusing on the assignment of thematic roles, we will demonstrate that canonical sentences often go hand in hand with an agent before patient ordering of thematic roles. In contrast, if this ordering is reversed in a sentence – as it can be the case in non-canonical sentences – language processing is more difficult. However, whereas Section 2.2 will show that adults can still reliably and

quickly assign thematic roles in structurally challenging sentences, Section 2.3 will demonstrate that young children face difficulties in processing non-canonical sentences and in assigning thematic roles, especially when the patient role precedes the agent role. Contrasting adults' and children's language processing is furthermore relevant regarding their differing use of visual contextual cues, especially when the processing of emotions is involved (this will be discussed in more detail in Section 3.7.1).

2.1 Incrementality in Relation to Syntactic Structure

The question of how exactly and to what extent we use the linguistic input to guide and constrain upcoming input and the inferences we make about what should come next, are still a matter of debate. What we know already from numerous word and sentence processing studies is that different kinds of linguistic units¹ exhibit different constraints on the upcoming input and thus shape our expectations. Verbs for example have been found to play a major role in the organization of the underlying event structure of sentences (Gentner, 1982). They impose syntactic constraints on the arguments they take and semantic constraints regarding, among others, the thematic roles they take. Numerous studies have shown that these constraints have an impact on language comprehension (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983).

Furthermore, verbs are not the only entities that structure and shape the way we incrementally understand and interpret language. In English, the structure of a typical canonical sentence is largely fixed in that a subject is followed by a verb, which in turn is followed by an object. Thus, in a Subject-Verb-Object (SVO) sentence like (1) *The ladybug tickles the cat*, the ladybug is the subject, i.e., the initiator of the tickling action, denoted by the verb and the cat is the object, namely the receiver of the action. Due to the poor English case-marking system, changing the ordering of the words in the sentence to (2) *The cat tickles the ladybug* also reverses its meaning. Now the cat is the do-er of the tickling action, namely the subject, whereas the ladybug is the object and thus the receiver of the action. Studies have for example suggested that reversing the ordering of a sentence such as in (3) *The eraser bites the turtle* results in

1 The focus here lies on words, constituents, phrases and sentences rather than on smaller units such as phonemes, syllables or morphemes.

2. Incremental Language Processing

English participants choosing the first noun phrase as the agent of the action, regardless of its animacy or likelihood to function as an agent of the action denoted by the verb (MacWhinney, Bates, & Kliegl, 1984, but see Trueswell, Tanenhaus & Garnsey, 1994 for animacy constraints in syntactic ambiguity resolution). Moreover, these sorts of sentences elicit rapid P600 effects (which are associated with syntactic reanalysis) in EEG recordings (e.g. in *The meal was devouring*, Kim & Osterhout, 2005). Here, the SVO ordering of the words in the sentence restricts the interpretation of the sentence.

However, because other languages, such as German have a rich case marking system, the ordering of words in a sentence is much more flexible. In German, the subject of a sentence is marked with the nominative case, whereas the object takes the accusative or dative case marking. For transitive verbs, which take only one argument apart from the subject, the object has to be in the accusative case. This way, the subject and the object of a sentence are clearly marked, so that a strict ordering of subject-before-object is not necessary. Using case marking, Object-Verb-Subject (OVS) sentences (1) and SVO sentences (2) convey both that the cat is the subject:

(1) *The ladybug tickles the cat*

Transl.: ‘Den Marienkäfer kitzelt der Kater’

The (determiner, masculine, accusative case) ladybug (object) tickles (transitive verb) the (determiner, masculine, nominative case) cat (subject).

(2) *The cat tickles the ladybug*

Transl.: ‘Der Kater kitzelt den Marienkäfer’

The (determiner, masculine, nominative case) cat (subject) tickles (transitive verb) the (determiner, masculine, accusative case) ladybug (object).

However, even though in German both SVO and OVS sentences convey the same meaning, SVO ordering is the typical and canonical sentence structure for marking a simple transitive relation in both language production and comprehension. Canonical, i.e., unmarked sentence structure, is the structure of a sentence that speakers preferably use when answering questions such as: “What happened?” (Siwierska, 1988).

2. Incremental Language Processing

Adults do not have problems understanding and interpreting non-canonical sentences. Still, many studies have found that OVS sentences for which object case marking is ambiguous between nominative and accusative are initially interpreted as having an SVO structure, i.e., they are disambiguated by word order. SVO structure is thus preferred over the OVS structure (Hemforth, 1993; Hemforth & Konieczny, 2000). In German, only the masculine determiner *der* is unambiguously case-marked for the accusative case, i.e., nominative *der* turns into accusative *den*. However, the feminine and neutral determiners, *die* and *das*, respectively, do not change. Hence (4) and (5) can have different meanings, although the surface structure is identical:

(4) *Die Frau tröstet das Mädchen.*

Transl.: ‘The woman comforts the girl.’

The (determiner, feminine, nominative case) woman (subject) comforts the (determiner, neutral, accusative case) girl (object).

(5) *Die Frau tröstet das Mädchen.*

Transl.: ‘The girl comforts the woman.’

The (determiner, feminine, accusative case) woman (object) comforts the (determiner, neutral, nominative case) girl (subject).

In (4) the woman is the subject, i.e., the sentential agent, and is comforting the girl, who is the object, the sentential patient. In (5) the girl is the subject and is comforting the woman, who is the object of the sentence.

However, comprehenders not only prefer canonical over non-canonical sentence structures, the latter also need more time and cognitive resources to be processed (Scheepers & Crocker, 2004). Using non-canonical German OVS sentences with case-ambiguous feminine noun phrases (NP) in NP1 position and disambiguating masculine noun phrases in NP2 position, event-related potential (ERP) studies demonstrated that readers (Matzke, Mai, Nager, Rösseler, & Münte, 2002) and listeners (Knoeferle, Habets, Crocker, & Münte, 2008) initially interpret these locally ambiguous OVS sentences as SVO sentences. Syntactic reanalysis – in the form of a P600 effect, which was present for locally ambiguous but not unambiguous OVS sentences – only happened when encountering the unambiguously case-marked determiner of the second NP. Thus, listeners structurally analyze the incoming

linguistic input within a few hundred milliseconds on the basis of the linguistic information given at the time of encounter.

Moreover, Bornkessel, Schlesewsky, and Friederici (2002) suggested that this processing difficulty for OVS compared to SVO sentences is not due to the lower frequency of OVS (compared so SVO) sentences, but rather reflects a detailed structural analysis. Their ERP study demonstrated that in non-canonical, embedded German clauses, accusative objects and dative objects elicit different ERP responses. Despite the fact that both structures are infrequent and non-canonical, their data suggests that embedded clauses with accusative but not dative objects are more difficult to process than canonical sentences. Hence, in addition to frequency as a source of processing ease / difficulty, the higher processing costs for non-canonical sentences with an accusative object seem to reflect a fine-grained syntactic analysis.

Additionally, in a self-paced reading study, Schriefers, Friederici, and Kühn (1995) had participants read German subject and object relative clauses, which were either semantically biased towards the subject or the object or did not elicit a semantic bias. Results indicated that participants preferred the subject relative clause interpretation for the unbiased sentences and that reading times for the object relative clauses (of the auxiliary) were slower compared to the subject relative clauses. They thus demonstrated that German readers have a preference for subject relative clauses (vs. object relative clauses).

2.2 Incrementality in Relation to Thematic Role Assignment

Processing sentences is thus not just a matter of purely assigning syntactic structure to the linguistic input as it unfolds, but we also need to identify, ‘who is doing what to whom’. We do so by assigning thematic roles, such as agents and patients to the syntactic functions of the linguistic input². Thematic roles thus build a bridge between the syntactic and the semantic structure of a sentence (Schipke, 2012, see also Dürscheid, 2000; Van Valin, 1999).

In a canonical SVO sentence the grammatical functions of subject - verb - object map onto an agent - action - patient ordering of thematic roles. In case-marking languages such as German, case can be used to assign these grammatical functions

2 In the present context thematic roles are generalized into an agent and a patient group, following Dowty (1991). The argument of the verb that is most agent-like will be called the agent, the argument that is most patient-like, will be called the patient.

and their thematic roles in unambiguously case-marked sentences. Yet, if case marking is ambiguous (as it is for example, the case in German feminine noun phrases), we encounter structural and thematic role ambiguities. In these cases and in cases in which the prototypical role ordering of agent-before-patient is reversed, the language user may encounter processing difficulties.

Crucial for thematic role assignment is that the verb of a sentence constrains the arguments and thus also the thematic roles it can take. A transitive verb, such as *tickle* needs an entity that can perform the action denoted by the verb and an entity that is the receiver of the action denoted by the verb³. The former thus takes the thematic role of the agent whereas the latter is characterized as the patient.

In an SVO sentence such as (2) the sentential subject, i.e., the cat, therefore equals the agent, and the object, i.e., the ladybug, equals the patient. In a canonical sentence, such as (2) the thematic role of the agent is usually in sentence initial position, whereas the thematic role of the patient is in sentence final position. However, because of the German case-marking system, the position of the thematic roles in a sentence can be varied, leading to a non-canonical order of the words in the sentence. Thus despite varying in sentence position, in both (1) and (2) the agent of the sentence is denoted by the nominative case marking and the patient is denoted by the accusative case marking of the determiner. Since masculine noun phrases are unambiguously case-marked, their thematic role in the sentence is also unambiguous (e.g., (1) and (2)). However, because feminine and neutral case marking is ambiguous for nominative and accusative cases in German, thematic roles for (4) and (5) cannot be determined on the basis of the morphological markers of the first noun phrase (NP1) and the second noun phrase (NP2). Regarding the preferred SVO interpretation compared to OVS readings of sentences (Hemforth, 1993; Hemforth & Konieczny, 2000), in both sentences the role of the agent would be assigned to NP1, i.e., the woman, and the role of the patient to NP2, i.e., the girl, if no other (non-) linguistic context elicited a bias to the other interpretation.

So just like interpreting and anticipating the upcoming syntactic structure of a sentence (Staub & Clifton, 2006), adults quickly predict⁴ the thematic roles assigned

3 Due to the focus of transitive verbs in the presented experiments, intransitive and ditransitive verbs will not be covered here.

4 In this thesis, the term ‘prediction’ is used interchangeably with ‘anticipation’. Prediction as used here is not assumed to happen at a conscious level. The discussion about language processing as

2. Incremental Language Processing

by a verb (Ferretti, McRae, & Hatherell, 2001). Hence the expectations we derive while language unfolds are not purely linguistic but also include the conceptual relations of events described by the linguistic input (Altmann & Mircović, 2009). These thematic concepts are formed through generalizations across many experiences of an event and the agents and patients that are involved in this event (Altmann & Mircović, 2009; McRae, Hare, Elman, & Ferretti, 2005). So, similar to verbs priming agents and patients with which they typically occur, nouns can also prime event knowledge and generate expectancies about typical agents and patients (McRae et al., 2005). That nouns can also generate expectancies about thematic roles is especially important given that in some languages like German the verb can be placed at the end of the sentence in a subordinate clause. Thus, expectations about the upcoming events in a sentence are generated in real time as the sentence unfolds (Kamide et al., 2003a) and are not solely based on constituent order. They moreover involve conceptual correlates such as the thematic roles encountered in the (non-) linguistic context.

Kamide et al., (2003b) for example presented participants with visual scenes of objects, potential agents and potential patients. While participants inspected the scenes they listened to unambiguous German OVS and SVO sentences (similar to (1) and (2)) about these scenes. Their eye-movement data yielded good evidence that participants reliably assigned thematic roles to the correct entities using case marking, the constraints imposed by the verbs and their real-world knowledge. Moreover, participants did so even before the target character was mentioned, i.e., they launched anticipatory eye-movements towards the target character. In another visual-world eye-tracking study, Knoeferle et al. (2005) demonstrated that participants can use the non-linguistic visual context alone to resolve structural and semantic role ambiguities in German SVO and OVS sentences. The NP1 of the sentence was always feminine and thus case- and role- ambiguous between nominative and accusative, agent and patient respectively. The NP2 of the sentence was always masculine und thus disambiguated the SVO or OVS sentence structure. The visual scenes depicted two potential agents performing actions and a patient. Participants could use the thematic role information provided by these scenes to reliably resolve the sentence initial role-ambiguity even before they heard the disambiguating masculine determiner of the second noun phrase

incremental, partially incremental or inherently predictive is not in the focus of this work and will hence not be addressed.

2. Incremental Language Processing

(for more details on the study see Section 3.3). Moreover, when Knoeferle and Crocker (2006) additionally introduced stereotypical role-knowledge of agents and patients, as for example a detective spying versus a wizard spying, participants preferred to rely on the visual information⁵ rather than on the stereotypical world-knowledge introduced by the sentence. These findings suggest that people make use of all available information in order to quickly assign thematic roles, although some cues might be more important than others or might seem more reliable respectively (for more details see Section 3.2).

However, even though we can quickly determine the ‘who does what to whom’ in both OVS and SVO sentences, the fact that OVS sentences are non-canonical makes these sentence structures inherently more difficult to process than canonical SVO sentences. Additionally, apart from unambiguous German object-first sentences, there are numerous other constructions that are grammatically challenging but unambiguous concerning the assignment of thematic roles. In a series of reaction time studies, Ferreira (2003) tested how quickly and accurately participants can assign thematic roles in unambiguous English active, passive, subject-cleft and object-cleft sentences. The main finding was that the difficulty (based on higher reaction times) to assign thematic roles is not just due the frequency of a construction (cf., Bornkessel et al., 2002) but that constructions in which the patient role has to be assigned before the agent role are more difficult to understand than sentences with agent - patient ordering. Thus, identifying thematic roles in passive and object-cleft sentences led to more errors and also took more time compared to active and subject-cleft sentences. Moreover, whenever participants made errors in assigning agent and patient roles, they erroneously assigned the first mentioned noun phrase the role of the agent. Thus, the processing difficulties of German OVS sentences might not just be due to the non-canonical construction, but more so to the fact that the prototypical order of agent-before-patient is violated in these constructions.

We have thus far seen that adults – who are highly proficient in their mother tongue – still face difficulties when processing grammatically challenging sentences and when assigning thematic roles in non-canonical structures. Yet, this thesis does

5 Note that the terms ‘information’ and ‘cue’ are used interchangeably. As used in the present thesis, unless otherwise stated they refer to linguistic and / or visual and / or social context that is present during language processing, was present prior to language processing or that is stored in working memory or as general world-knowledge. We are refraining from a more specific definition at this point.

not just focus on the language processing difficulties that adults face during real-time language processing. In order to investigate how language processing in non-canonical sentences and the assignment of thematic roles can be facilitated across the lifespan, we also need to focus on how children deal with structurally challenging sentences. The next Section (2.3) will hence focus on children's language processing difficulties and will demonstrate that children do not yet seem to be able to use case marking for thematic role assignment in purely linguistic contexts. Yet, we will also suggest that children are very well able to assign agent and patient roles if they can draw on supporting non-linguistic visual context.

2.3 Language Processing: Language Development

If even adult native speakers struggle with these non-canonical constructions, a central and still open question is how exactly children cope with this kind of linguistic input. We know that already by the age of four children have a robust, basic understanding of their native language. When communicating with them, they seem to comprehend most of the linguistic input and have thus already acquired and learned to use an impressive amount of vocabulary (Snedeker, 2013). Moreover, just like adults (Kamide et al., 2003a), children (Mani & Huettig, 2012; Nation, Marshall, & Altmann, 2003) can very quickly anticipate upcoming linguistic information given a supportive linguistic and visual context. When 10-11-year old children listened to sentences like *Jane watched her mother eat the cake* and the visual context showed only one edible object among distractors, children launched eye-movements to the only edible object in the scene before having heard the object's name, i.e., already at the verb region (Nation et al., 2003). Moreover, 2-year old children's anticipation skills showed a significant correlation with their productive vocabulary size (Mani & Huettig, 2012). Although less skilled comprehenders between 10 and 11 years of age did not differ in how rapidly they anticipated the target, their gaze pattern differed as a function of their comprehension scores (compared to more skilled children, Nation et al., 2003). Thus it might be that individual differences in language comprehension and production only play a role early on in language acquisition, at least for simple grammatical structures such as those used in Mani and Huettig (2012) and Nation et al. (2003).

2. Incremental Language Processing

Yet, because children are still in the process of acquiring their native language, handling difficult sentence structures presents a challenge for them. Savage, Lieven, Theakston, and Tomasello (2003) for example primed⁶ 3-, 4- and 6-year old children with a picture and listened to a related sentence in either the active or the passive voice. Then they were asked to describe a target picture. Prime and target sentences could either overlap both lexically and structurally or only structurally. Using adult participants, classic structural priming studies have shown that sentences of the same syntactic structure prime one another even when prime and target sentence have no lexical items in common and share only the structural properties of the sentence (Arai, van Gompel, & Scheepers, 2007; Bock, 1986). Savage et al. (2003) however found that only 6-year olds, like adults, showed both lexical and structural priming, whereas 3- and 4-year olds showed lexical priming only. Hence, the authors conclude that by the age of 6 children already have fairly abstract structural representations of active and passive transitive sentences, whereas 3- and 4-year olds still rely on specific lexical items to form representations of these structures.

Another study that tested the abstractness of 5- and 7-year old children's and adults' representations of novel structures found similar results. Here, Boyd and Goldberg (2011) employed a novel construction learning paradigm, allowing constant input across different learner groups. Children and adults watched short movie clips and listened to descriptions of the performed novel actions in a novel NP1-NP2-VERB construction. After these exposure trials, in the test trials participants performed a forced-choice comprehension task. This task required them to listen to a voiceover sentence and to pick which of the two simultaneously played movies corresponded to the meaning of the heard sentence. Each test trial was repeated until a response was made. Additionally, the test trials were either completely novel, only featured a novel verb or were the same as the exposure trials. Boyd and Goldberg (2011) demonstrated that 5-year olds were less likely to pick the correct movie than 7-year olds and adults, especially in the novel verb and completely novel test trials. They concluded that 5-year olds are hence less likely to generalize a given construction than 7-year olds and adults. Like Savage et al. (2003), they suggest that

6 The term 'priming' (as a method or phenomenon) refers to the effect that the exposure to a stimulus (the prime) beneficially influences the perception of / response to a related subsequent stimulus (the target). For definitions see e.g., Lashley, 1960; McNamara and Holbrook, 2003; Rickheit, Sichelshmidt and Strohner, 2002.

2. Incremental Language Processing

younger children still portray item-based behavior and cannot yet identify more abstract generalizations.

Another phenomenon with which children around the age of 5 still have problems is the so-called garden-path sentence. Garden-path sentences contain a structural ambiguity so that the initial interpretation might have to be revised and their meaning reconstructed to match the intended meaning, depending on the wider (linguistic) context (Trueswell & Gleitman, 2004). An example of a garden-path sentence from Trueswell, Sekerina, Hill, and Logrip (1999) is

(6) *Put the frog on the napkin in the box*

Here, *on the napkin* can either be interpreted as the modifier of the frog, i.e., a frog sitting on a napkin, or as a future destination of the frog. That is, in the destination interpretation the frog should be put on the napkin before it is put in the box. Once the adult listener hears *in the box*, s / he should interpret *on the napkin* as specifying the location. Given a visual display of an apple, an apple on a towel, a box and an empty towel, when adults listened to a garden-path sentence like (7)

(7) *Put the apple on the towel in the box*

while looking at the visual display, their eye-movement behavior indicated that they preferred to look at the apple on the towel (vs. the single apple). This suggests that adults could use the visual context to quickly resolve the syntactic ambiguity in favor of the modifier interpretation, which, given no supportive visual context, is typically dispreferred compared to the destination interpretation. Hence, the authors conclude that a suitable visual context can modulate listeners' preference for the destination interpretation (Tanenhaus et al., 1995). In a similar eye-tracking study, which employed an act-out task using sentences like (6), 5-year old children on the other hand failed to choose the modifier interpretation. Their eye-movements as well as their actions demonstrated that they instead put the frog first on the empty napkin and then into the box, despite the visual act-out scene favoring the modifier interpretation of the sentence (Trueswell et al., 1999). Hence, 5-year old children still seem to pursue a different kind of referential strategy in which they directly link the current

2. Incremental Language Processing

auditory linguistic input to the best matching related visual context (Zhang & Knoeferle, 2012) in their use of the (non-) linguistic context for syntactic ambiguity resolution than adults.

Trueswell and Gleitman (2004) conclude from this study that children still rely heavily on lexical cues (instead of using the referential context provided by the visual scene), especially on the grammatical preferences imposed by the verb *put*, which usually denotes a goal as a prepositional phrase. Snedeker, Thorpe, and Trueswell (2001) and Snedeker and Trueswell (2004) investigated children's and adults' verb-driven syntactic parsing preferences further in that they manipulated verb bias and referential scene information in an eye-tracking act-out study. They used instrument bias (8), equi bias (9) and modifier bias sentences (10)

(8) *Tickle the pig with the fan*

(9) *Feel the frog with the feather*

(10) *Choose the cow with the stick*

and crossed them with referential scenes either depicting two potential referents (e.g., a frog, a frog with a feather, a feather and a distractor) or only depicting one potential referent (e.g., a frog with a feather, a feather, and two distractors). Snedeker et al. (2001) and Snedeker and Trueswell (2004) provided good evidence that unlike adults, children heavily relied on the verb bias and not on the referential scene manipulation. In contrast to children's eye-movements, adults initial eye-movements and actions showed sensitivity to both the verb bias and the referential scene information.

However, regarding the fact that the sentences used by Snedeker et al. (2001) and by Snedeker and Trueswell (2004) were all English imperative sentences and thus verb-initial sentences, it could not be ruled out that children rely heavily on verb information. This would result in two competing cues, in which the sentence-initially encountered verb provides the first and stronger cue compared to the visual scene. Due to their strong verb-specific bias children might not have been able to revise their initial syntactic parsing and thus might have stuck with the verb-biased interpretation of the sentence rather than taking the referential context into account. On the other hand, children's parsing preference for the destination ((6), Trueswell et al., 1999) or instrument interpretation (see (8) and (9), Snedeker & Trueswell, 2004) could also be

2. Incremental Language Processing

du to a lack of inhibitory control, i.e., the inability to override a certain (dominant) behavioral response in children at that age, regardless of a verb-specific bias.

To test this assumption Choi and Trueswell (2010) conducted the Trueswell et al. (1999) study in Korean, a head-final language. Readers and listeners of head-final languages process sentences incrementally (just like comprehenders of head-initial languages) and hence start predicting the syntactic structure of the whole sentence without having encountered the verb (Choi & Trueswell, 2010). Thus, conducting the study in Korean should shed light on the question of why children are led down the garden-path, i.e., due to a verb-specific bias or due to a lack of inhibitory control. Korean children showed difficulties in recovering from garden-path sentences even though the disambiguating verb was at the end of the sentence. Thus, their interpretation difficulties cannot be due to a verb-specific bias. Their data suggested that the preference to select the destination rather than the modifier interpretation of the verb's argument does indeed seem to support the assumption that children still show a lack of inhibitory control. It seems that they cannot revise their initial destination interpretation in favor of the modifier interpretation provided by the context (Choi & Trueswell, 2010). We will discuss this matter further in Section 10.7, taking the results of the studies presented in Sections 6 and 7 into account.

Another structure that is especially difficult for children acquiring a case-marking language is the OVS sentence structure and hence assigning thematic roles based on case marking. In Section 2.1 we have seen that OVS sentences are already more difficult to process compared with SVO sentences for adult native speakers. However, using case marking and determining who the agent and who the patient of a sentence is, poses even more problems for children. Assigning thematic roles with the help of case marking takes up to the age of 7 to be fully developed in a rich case-marking language such as German (Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008a).

Thus the acquisition of these sentence structures (SVO and OVS) has been a matter of debate for more than 40 years. It has been argued in the beginning that children rely first on the ordering the words in a sentence and only later develop the competence to correctly interpret sentences on the basis of their case-marking system alone (Bruner, 1975). However, the picture does not seem to be that clear-cut. Weist (1983) for example argued against the claim that the ordering of the words precedes the use of inflections for the comprehension of sentences in child language. Using

2. Incremental Language Processing

Polish sentences he demonstrated that even 2-3-year old children are already able to use inflections as a cue to determine thematic roles in a sentence.

An influential model of language acquisition that attempts to deal with the variation of natural languages and that seems to be well-suited to address the word-order discussion is the competition model by Bates and MacWhinney (1987). Their model of linguistic performance attempts to describe the mapping between form and function, and production and acquisition. They developed a number of claims on the basis of cross-linguistic Italian - English - French investigations on the acquisition of SVO and OVS sentences. The most important of these claims are:

1. Direct Mapping
2. Cue Validity
3. Cue Strength

The first of these (1.) specifies a direct mapping between form and function. These two processing levels do not need to stand in a one-to-one relation; meaning that e.g., a single form, such as the German feminine determiner *die* can have multiple functions (indicating accusative or nominative case). 2. Cue Validity is composed of cue availability and cue reliability and is supposed to be a property of the cue itself. Cue Validity thus refers to questions such as: “How often does this cue appear in this particular context?” and “How accurate are the conclusions drawn when this cue is used?”. 3. Cue Strength refers to the weight of a specific cue relative to a goal, i.e., nominative case marking would be an example of a strong cue towards agenthood (see (1) and (2)). The model moreover assumes that different cues compete with each other on the basis of these claims and that conflicts are resolved based on the strength of the competing cues. This competition happens in real time as the sentence is processed incrementally such that, for example the assignment of nouns to case roles is continuously updated as the sentence unfolds.

Regarding language acquisition, the most important claim here is number 2, Cue Validity, as it will determine the order of acquisition of the different grammatical cues available in a language. The authors assume that children are sensitive to linguistic patterns in their native language and to the information value of these patterns from very early on. Regarding the word-order debate, this does not mean that word order is always acquired before inflectional morphology for the assignment of thematic roles. It rather depends on the grammatical and semantic features of a given language and

2. Incremental Language Processing

on the Cue Validity of these features to determine the order of acquisition. Bates and MacWhinney (1987) discuss their competition model based on Italian and French. These languages are rather flexible and make extensive use of morphological cues instead of relying on word order. English on the other hand is clearly dependent on the ordering of the words in a sentence to correctly interpret sentence meaning due to its rather poor case-marking system. Thus in Italian, children as young as 2 years of age already make extensive use of word order variation using case marking instead of relying on a strict ordering of the words in the sentence (Bates & MacWhinney, 1987).

The question of how children acquire different languages with regard to word order, grammatical cues and the assignment of thematic roles has been studied widely since Bates and MacWhinney (1987). Gertner, Fisher, and Eisengart (2006) for example investigated if 21- and 25-month old toddlers can use their knowledge of English word order to generalize their syntactic knowledge to new verbs. In a preferential looking study they presented the toddlers with 2 simultaneously playing videos. Each video featured 2 characters engaged in a novel action, while the toddlers listened to a transitive sentence with a new verb referring to one of the novel actions in a 'who does what to whom' fashion (e.g., *The duck is gorping the bunny* for agent=subject matches, *He is gorping the bunny* for patient=object matches). Gertner et al. (2006) found longer looking times to the correct (agent and patient) videos than to the incorrect videos in which the displayed thematic roles did not match the thematic roles of the sentence for both 21- and 25-month old toddlers. They suggest that toddlers can use their abstract knowledge of the fixed English word order to interpret new action verbs and to assign agent and patient roles correctly.

Moreover, they argue against a lexical account. In their view children's linguistic knowledge must be abstract enough in order to generalize syntactic patterns to unknown events and in order to assign agent and patient roles in these events even in the case of unknown verbs. A lexical account would assume that toddlers this young would not be able to correctly assign thematic roles in their study as they do not know the meaning of the new verbs, and thus do not know how to use them for syntactic structuring.

Furthermore, the authors point out that the vocabulary size of 2-year-olds is too small to start generalizing syntactic rules based on their lexicalized inventory (Gertner

2. Incremental Language Processing

et al., 2006). Dittmar, Abbot-Smith, Lieven, and Tomasello (2008b) furthermore conducted the Gertner et al. (2006) study with German children and transitive familiar verbs. In Gertner et al.'s (2006) study, the English children were prepared for the actual test phase using a training phase with familiar verbs in the training sentences. However, the nouns, i.e., agents and patients, of the test phase were the same as in the training phase. Dittmar et al. (2008b) thus suggested that the results that children at that age can build abstract syntactic rules from novel verbs might be due to the children's prior experience from the training phase of who is the agent and who is the patient. Hence, using the same general setup and age group, they divided the children into two groups. One group received a training phase as in Gertner et al. (2006), and the other group only received a familiarization phase (with the same videos) for the action verbs without naming agents or patients, e.g., *This is called 'washing'*. The test phase was the same for both groups and similar to the test phase in Gertner et al. (2006), except that all sentences were transitive German sentences with masculine NPs (cf., Dittmar et al., 2008a, see (1)). The results suggested that indeed only the group which received training on the full sentences, i.e., including naming agents and patients, looked longer at the test video (which used novel verbs) matching the transitive test sentence and thus could determine who the correct agent / patient was in the video. The group that only received the general familiarization with the action verbs did not look longer at the action-matching compared to the action-mismatching video in the test phase. Dittmar et al.'s (2008b) study first of all suggests that like English toddlers, German toddlers are already sensitive to thematic roles in transitive, unambiguous SVO sentences. Dittmar et al. (2008b) hence replicated the findings by Gertner et al. (2006) for the 2-year-olds. Moreover, this study indicated that some kind of prior linguistic experience / training is necessary for small children in order to build more abstract syntactic rules and to generalize these rules to new exemplars (Dittmar et al., 2008b).

However, Gertner et al.'s (2006) study only used simple English transitive sentences in which word order was the only cue available to use for the children. Using German, Dittmar et al. (2008a) tested thematic role assignment with both word-order and case-marking cues of 2-, 4- and 7- year olds. Creating German SVO and OVS sentences using novel verbs, both of these cues could either support each other or word order or case marking could be the only available cue. Hence, in the two-cue

2. Incremental Language Processing

condition canonical SVO sentences with masculine NPs were used, see (2). When word order was the only available cue, the NPs in the sentences were either both feminine or neuter in gender, see (4) and (5). When case marking was the only available cue, non-canonical OVS sentences with masculine NPs were used, see (1). In their study children saw two simultaneously played videos with a potential agent and a potential patient performing a novel action denoted by a novel verb. In both videos the same actions were performed but agent and patient roles were reversed in the two videos. At the same time children heard a sentence in one of the cue conditions describing the scene in a ‘who does what to whom’ fashion. Their task was to point towards the video that matched the spoken sentence. The results suggested that 2-year old German children only fully comprehended the prototypical SVO sentences when both word order and case marking supported each other, i.e., when only masculine NPs were used. They did not comprehend the single-cue sentences even when word order was the only cue that could be used, i.e., when both NPs were either feminine or neuter, or when case marking was the only cue, i.e., in OVS sentences with masculine NPs. By contrast, 4-year old children mainly relied on word order information for the comprehension of the novel-verb sentences. Their accuracy was at ceiling for both the two-cue condition and the word-order-only condition. However, they only scored at chance level in the condition in which only case marking could be used to determine agent and patient roles. Only the 7-year olds showed adult-like comprehension of all sentence types and were thus able to also use case marking on its own to assign thematic roles.

These results are in line with the competition model by Bates and MacWhinney (1987). Cue Validity is higher in German for word order compared with case marking. The latter has many different and ambiguous forms for the same grammatical function, whereas the word order of sentences in German is canonically SVO. Moreover, German children seem to rely first on Cue Availability (frequency) and only later in the development learn to also rely on Cue Reliability (case marking). This interpretation is also in line with Ferreira's (2003) suggestion about the prototypical ordering of agent before patient roles, since SVO sentence are more frequent in German than OVS sentences.

Electrophysiological evidence also supports the findings that German children struggle with non-canonical OVS sentences much longer than they do with canonical

2. Incremental Language Processing

SVO sentences. Measuring ERPs during a sentence listening study, Schipke, Friederici, and Oberecker (2011) presented 3-, 4.6- and 6-year old German children with transitive SVO and OVS sentences as well as with their grammatically incorrect double-nominative and double-accusative counterpart sentences. The results underlined the developmental process of acquiring the use of the case-marking system for thematic role assignment. Unlike the double-nominative condition, which showed an adult-like pattern for these violations in all age groups, the double-accusative condition showed different ERP responses compared to its grammatically correct OVS counterpart for each age group. These responses can be interpreted as shifting from error detection in the youngest to thematic repair strategies in the 4.6-year-olds until they are adult-like in the 6-year olds and resemble ERP responses associated with syntactic / thematic repair strategies (Schipke et al., 2011). From these studies it becomes clear that the comprehension of more difficult sentence structures, such as garden-path or non-canonical object-front sentences posit a challenge not only for adult native speakers who have already mastered their mother tongue, but also especially for small children who are still in the process of learning to make sense of their own language.

However, children do not learn a language isolated from their surroundings and thus might make use of their visual environment in order to align their linguistic input with information they get from the outside world. As became clear in Section 2.2 from the results by Kamide et al. (2003b) and Knoeferle et al. (2005), adults can very rapidly use the visual context to anticipate upcoming linguistic input.

One study that demonstrated the use of contextual, non-linguistic information for sentence processing in children moreover looked at the facilitatory effect that visual context can have on on-line sentence processing and thematic role assignment. Zhang and Knoeferle (2012) presented 4-5-year old children and adults with two potential agents and a potential patient on a visual display. The potential agents either performed an action or did not perform an action. While watching the scene children and adults listened to unambiguous German OVS and SVO sentences similar to (1) and (2) describing the scene in a ‘who does what to whom’ fashion. In half of the trials the actions denoted by the verbs of the sentences were depicted in the scenes, whereas the characters did not perform any actions in the other half of the trials. Participants’ eye-movements were measured while they watched the scenes and

2. Incremental Language Processing

listened to the sentences. Moreover, they answered a comprehension question about ‘who is doing what to whom’ after each trial. Zhang and Knoeferle's (2012) results suggest that children can indeed use the visual context of the depicted actions to overcome their processing difficulties for German OVS sentences (cf., Dittmar et al., 2008a). Adults, as well as children anticipated the correct patient denoted by NP2 of the sentence earlier when an action was vs. was not depicted in the visual scene for both canonical SVO and non-canonical OVS sentences. Moreover, accuracy scores for the children were higher when events were, compared to when they were not depicted in OVS sentences. Irrespective of the actions being depicted or not, children did not have problems answering the comprehension questions in the SVO sentence condition. Adults scored at ceiling for both sentence types and regardless of action depiction. Hence, case marking alone does not enable 4-5-year old children to assign thematic roles, as has already been shown by Dittmar et al. (2008a). However, visual contextual information such as depicted events can help children to overcome processing difficulties for challenging syntactic structures (cf., Trueswell et al., 1999 for different results and see Section 10.1 for a related discussion).

In summary, this chapter has provided insights into the notion of incrementality and language processing. We have seen that language processing is highly dependent on the linguistic input and that the way sentences are structured can determine how easily we process them. Especially in a language such as German determining thematic roles can pose a problem as different cues can be used to assign agent and patient roles. Moreover, we have seen that this task is even more challenging for children who are still in the process of mastering their native language. Throughout the chapter we have discussed several studies that used the visual context to explore how we process language as it unfolds in real time. We have seen that we can use visual cues to quickly anticipate upcoming linguistic information. We have seen that adults and more importantly even children can make rapid use of visual information in order to overcome language processing difficulties more easily than when no visual cues are present.

In order to arrive at a more detailed understanding of how language users integrate visual and linguistic information, the next Section (3) will take a closer look at the use of the visual context for language processing. As the present thesis aims to come up with a conceptual model of how comprehenders integrate and use (non-) linguistic

2. Incremental Language Processing

visual cues during real-time language comprehension (see Section 4 and 10.8), we will first determine what can be considered a visual context, and therefore will establish a working definition for the present purpose. Next, we will have a look at which information we extract from the visual context and how we use this information for language processing. Additionally, Section 3 will cover different kinds of visual cues. To date, most studies have focused on the use of directly depicted scene information for language processing. More social visual cues, such as emotional facial expressions of our interlocutors have widely been ignored in psycholinguistic research and even more so in real-time language processing accounts. Yet, a conceptual model of real-time language processing also needs to take social cues into account. Hence in the next Section (3) we will discuss the differences between different types of social and non-social visual cues and to what extent and in which way we exploit them for language processing.

Following Section 3, in Section 4 we will take a detailed look at how we actually link the visual context with our linguistic input. We will discuss different theories, accounts, frameworks and mechanisms of language processing. We will see which claims they make, in how far the different approaches differ and crucially in which way they could still be improved, especially regarding the integration of different information types into language processing. The discussions and findings in the upcoming Section 3 will provide further support for our argument to extend the CIA (Knoeferle & Crocker, 2007) – an existing account of real-time language processing (Section 4). Considering also our own findings that both social and non-social visual cues facilitate language processing depending on the age of the comprehender (Sections 6-9), we will finally exemplify in detail how such an account could be adapted in order to arrive at a more detailed understanding of the integration of visual cues into real-time language comprehension (Section 10.8).

3. The Use of Visual Context for Language Processing

When communicating with other people, most of the time we are aware of the visual world that surrounds us. We know where in the world we are situated, we know the communicative setting we are in and we perceive – be it consciously or unconsciously – our immediate visual context even as we are speaking or listening. Moreover, often when we are talking or listening, we make use of this visual information directly for example by referring to something we see, by using visual information as a replacement for a linguistic utterance, or by taking advantage of the visual information to facilitate language comprehension and linguistic ambiguity resolution.

3.1 The Visual Context – Basic Assumptions

Despite this intuitively strong association between a visual context and our understanding and processing of language, as a discipline, psycholinguistics in the past has widely neglected that language comprehension is situated in a rich non-linguistic context and was mostly purely language-driven (Knoeferle, 2015a). However, during the last 30 years studies have corroborated the view that the linguistic and visual information with which our senses provide us can overlap partially and cannot always be strictly separated. Processing sensory information does not happen discretely but rather continuously. Moreover, our saccades are tightly coupled to our linguistic processing when visual context is present. The signal streams of vision and language are roughly equally important and thus also constrain each other continuously and in real-time (Anderson, Chiu, Huette, & Spivey, 2011).

Using a paradigm dubbed the Visual World Paradigm (Tanenhaus et al., 1995, see also the review of visual-world studies in Sections 2 and in this Section) we can investigate the real-time influence language has on the visual context and vice versa. Eye fixations are recorded while participants inspect a visual scene and listen to words, sentences or instructions. While doing so, participants tend to rapidly establish a link between the visual and the linguistic input. Moreover, they even do so given complex scenes, high referential density and a fast speech rate. This rapid interplay between language and vision suggests that the linking between the visual and the linguistic input is not stimuli-dependent but a property that is fundamental for the language system (Andersson et al., 2011). This fundamental feature of language understanding makes the use of the Visual World Paradigm for research on on-line

3. The Use of Visual Context for Language Processing

language comprehension especially useful. Here, the visual context is readily available for the language user while spoken language is processed in real time. Furthermore, the linguistic input presented to the subject directs him / her to interact with the visual context, which in turn becomes highly relevant for the comprehension process. As the linguistic message unfolds moment by moment, the listener makes use of this information to reduce referential ambiguity until the linguistic input is specific enough to determine the intended referent (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). While doing so, the eye-movements the participants launch can be time-locked to the auditory linguistic input, as they are preceded by a shift in attention towards the assumed visual referent. Thus, we can determine when the auditory input triggered a saccade and which part of this input is responsible for the shift in attention (e.g., Eberhard et al., 1995; Tanenhaus & Spivey-Knowlton 1996; Tanenhaus et al., 1995).

Even though we know that our eye-movements are tightly coupled to, and provide an index of cognitive, linguistic and visual processing in real time (see e.g., Henderson, 2003, Rayner 1998, Rayner 2009), few psycholinguistic accounts attempt to explain language processing in real time and additionally integrate the visual context. One of these few accounts is the Coordinated Interplay Account (CIA, Knoeferle & Crocker, 2006, 2007). However, before we engage in a more in-depth discussion about different language processing accounts and their limitations regarding the integration of the visual context, we must first define what we mean by visual context.

3.1.1 Defining the Visual Context

In order to be used in language comprehension, the visual context has to be available to the comprehender while the linguistic input is processed on-line. This linguistic input then allows the listener to act on, react to, or interact with the visual information. Thus the context is relevant to the language comprehension process in that linguistic information can, for example, be used to constrain which referent (of a set of potential referents) is used. Hence, unambiguous reference is established as soon as the linguistic input has provided enough information to correctly select the intended referent (Eberhard et al., 1995).

3. The Use of Visual Context for Language Processing

However, because our visual working memory capacity is limited to about 4 items (Luck & Vogel, 1997), we might not always be able to take all the visual information that is available into account. These items do not have to be individual features, such as colors or shapes but can also consist of integrated objects (Luck & Vogel, 1997). Depending on the kinds and complexity of the given visual information, we might assume that our brains prioritize some cues over others (as e.g., results by Knoeferle & Crocker, 2006 suggest that visually depicted cues are preferred over stereotypical knowledge of role relation, see also Section 3.2 for more details) in order to choose which visual cues to integrate into the linguistic input.

Moreover, a set of objects in our visual field is not what we usually would call a visual context in the real world. We rather see different kinds of scenes in front of us. Henderson and Hollingworth (1999) described a scene as semantically coherent. According to these authors it is a human-scaled view of the real world and consists of background elements and objects that are spatially arranged with these background elements. Additionally, a well-formed scene has to conform to basic physical and semantic constraints that structure the environment (Henderson & Ferreira, 2004). Moreover, real scenes often consist of many different kinds of objects. These objects are sometimes very likely to be found in this particular setting (a kettle is, for example, more likely to be found in a kitchen than in a bathroom), but they can also be unpredictable (Andersson, Ferreira, & Henderson, 2011). Just imagine finding your lost car keys in the fridge after you have already given up looking for them in their usual place (e.g., the key holder).

Furthermore, natural real-world scenes are not stable, but are dynamically changing. Objects might be present at one point in time and gone the next moment. Moreover, when language comes into play and objects in the scene are referenced in speech, language will select only some objects. These objects continue to be relevant while others lose importance. Hence, the comprehender must filter out the irrelevant information, decide what can be used most efficiently from the scene and integrate this information into the process of understanding the linguistic input (Andersson et al., 2011).

3.1.2 Present Working Definition

In order to arrive at a more detailed understanding of how language users integrate visual and linguistic information, most psycholinguistic experiments investigating the interface between language and vision take place in a highly controllable laboratory setting. The scenes that the participants of these experiments usually look at while listening to language thus often differ in complexity and naturalness from real-world scenes. In this regard, Henderson and Ferreira (2004) distinguish between a ‘true scene’ and an ‘ersatz scene’. Whereas a true scene is a complete scene (either in the form of a real environment or in the form of a depiction of a scene), an ersatz scene is an incomplete (depicted) scene, which lacks some of the critical characteristics of a normal environment in the real world. These differing and / or missing characteristics in an ersatz scene could for example be a missing background or unusually large depictions of actors in the ersatz scene.

However, even these ersatz scenes, which are often used in psycholinguistic studies, can convey a sense of coherence and do not necessarily have to be composed of discrete and individual objects. Hence, most of the visual scenes used in experiments fall somewhere on a continuum between arbitrarily arranged objects in a visual display and true (depicted) scenes, such as when people are depicted as agents and patients which are actively engaged in an action (Henderson & Ferreira, 2004).

The underlying assumption of using simplified scene depictions is that they are good substitutes of natural real-world scenes. These depictions can capture important properties of real-world objects and scene perceptions like the complexity of the scene, semantic coherence and constraints, and scene meanings, while at the same time allowing the researcher to control for factors that cannot easily be controlled in the natural world and that could lead to confounds due to many possible and uncontrollable influences (Henderson & Ferreira, 2004).

One set of examples of such ersatz scene depictions with semantic coherence and scene meanings are agent-action-patient depictions. Participants can reliably and very quickly extract the necessary information from a given depicted scene that allows them to identify thematic roles if the depicted characters possess typical agent- (Hafri, Papafragou, & Trueswell, 2013), patient-, and action- characteristics. They can moreover use this scene information to correctly interpret a spoken sentence in resolving sentence initial role ambiguities by quickly integrating scene and linguistic

3. The Use of Visual Context for Language Processing

information on-line (Knoeferle et al., 2005). The present studies (see Sections 6-9) will make use of these depicted ersatz scenes as a visual context similarly to Knoeferle et al. (2005) in order to investigate the assignment of thematic roles during language comprehension.

However, (depicted) true and (depicted) ersatz scenes are not the only kind of visual contextual information that can be used in order to gain insights about the use of visual information for language processing. More often than not we also use visual information that is not necessarily present in the related scene to interpret linguistic utterances, such as the facial expression (Carminati & Knoeferle, 2013) or the gestures (Wu & Coulson, 2005) of our interlocutor who describes the scene. We know that we extensively use all kinds of available visual information to shape and facilitate the interpretation of the linguistic input (regardless of whether that information is part of a scene or not). Nevertheless, the questions of why we use certain types of information from the visual context as opposed to others or why we prioritize some visual cues whereas others in the same visual field would also have been helpful is not as straightforward as it may seem. These questions are thus addressed further in the next Section (3.2).

3.2 Information Extraction from the Visual Context

Our visual-cognitive system controls our eye-movements and directs our gaze to informative regions in our visual field. One factor that is crucial for visual and cognitive processing is attention, as we fixate regions in a scene that draw our attention. Hence eye-movements provide a behavioral measure of the ongoing processes in our attentional system (Henderson, 2003).

However, sometimes we may miss visual changes in a scene, even if we inspected the scene attentively. In a flicker paradigm Rensink, O'Regan, and Clark (1997) asked participants to look at scenes while these scenes were repeatedly alternated with a modified version of the same scene. The modified scene contained a visual change and participants were asked to name the type of change and the part of the scene that was changing. Even though the changes in the scene were substantial, participants often failed to detect them or needed considerable time to spot the changes. However, performance on change detection increased significantly for salient scene information when the change in the scene happened while the changing object was the current

3. The Use of Visual Context for Language Processing

center of attention. Performance did not increase when the object was attended at an earlier point in time (Rensink et al., 1997) supporting the tight coupling between attention and vision.

Furthermore, attention and vision are highly task-driven processes in that task knowledge can be used to control and guide the eye-movements (Rensink et al., 1997). The cognitive goals of the observer depend on the task s / he is performing. The task in turn influences the observer's eye-movement behavior, in that people for example fixate an object immediately before reaching for it (Salverda, Brown, & Tanenhaus, 2011).

Nevertheless, it is not only task-related knowledge that guides our visual attention. Our ability to control our gaze does not only rely on the present visual input. We can also access information from our short and long term memory of a scene. Additionally, we also use our world knowledge to determine which objects and things in our visual field are salient and relevant for us to look at. We can for example detect objects more quickly if they are consistent with the scene, i.e., we identify a kettle quicker in a kitchen than in a bathroom scene, because our stereotypical knowledge tells us that a kettle is more likely to be found in a kitchen than in a bathroom (Henderson, 2003; Palmer, 1975).

However, even though we draw on our world knowledge to control and guide our eye-movements, it seems that we still rely more on information that is actually depicted in a scene than on what we believe is stereotypical. In a visual-world eye-tracking study, Knoeferle and Crocker (2006) presented participants with agent-action clip-art scenes. Participants listened to sentences containing a verb that identified two agents: one was a stereotypical agent of the verb (e.g., a detective for *spying*). The other agent was not a stereotypical agent of the verb (e.g., a wizard) but was depicted as performing the action mentioned by the verb (*spying*). Thus, participants had to chose between two conflicting thematic role cues. The eye-movement data suggested that participants were able to use both cues but preferred only one of the cues when the two cues conflicted. In the conflicting condition they fixated the non-stereotypical agent performing the action mentioned in the sentence more than the stereotypical agent who performed a non-mentioned unrelated action. Participants hence relied on the depicted event information rather than on their stereotypical world knowledge in

3. The Use of Visual Context for Language Processing

assigning thematic roles. Thus, participants prioritized the depicted spying action over their world knowledge (Knoeferle & Crocker, 2006).

It seems thus that our notion of the visual context must be relatively broad. We use not just the information available from the immediate visual context, but also from past visual experiences. We draw on our world- and stereotypical knowledge and make extensive use of cognitive goals and task demands in directing our attention to objects and events in the environment (see also Knoeferle & Guerra, 2012 for more details on the notion of non-linguistic context in language comprehension).

Nevertheless, there seem to be differences between the various types of information in the visual context and in how and to what extent we exploit them during language comprehension. If we want to account for these different kinds of visual cues in language processing accounts, we need to be aware of their differences. The next Sections will hence discuss the use of two groups of visual contextual cues, namely direct visual cues (Section 3.3, see Section 3.3 for our definition of ‘direct cues’) and more indirect visual social cues (3.4, see Section 3.4 for our definition of ‘indirect cues’) and discuss how we can use them in language comprehension. Additionally, to date we do not yet know if these different visual cue types can be used to the same extent and with the same time course, or if they can rapidly interact with each other. Section 3.5 will hence discuss the cumulative use of different direct and indirect cues. As these Sections will show, not all visual cues are processed in the same way and can be processed by children and adults alike. Hence, following up on this, we will more explicitly argue that direct and indirect visual cues are processed differently (Section 3.6) and will then review and discuss this issue further in Section 3.7. Using emotions as an indirect social cue we will demonstrate that children and older adults but not younger adults prefer positive over negative emotional material. Thus Section 3.7 will focus on age differences in emotion recognition and interpretation.

3.3 Different Kinds of Visual Contextual Cues – Direct Cues: Actions and Objects

Depicted actions and events in the visual world that can be directly linked to the linguistic input can be used efficiently and very quickly to facilitate the real-time processing of this linguistic input (Knoeferle & Guerra, 2012). Adults can for

3. The Use of Visual Context for Language Processing

example, rapidly use visual information about objects and depicted action events for disambiguating structurally ambiguous sentences when these cues are directly mediated by (words in) the utterance (henceforth ‘direct cues’).

Recall that in the real-world study by Tanenhaus et al. (1995, see Section 2.1) the need to distinguish between the two apples modulated participants’ interpretation preference towards resolving *on the towel* as a modifier of *the apple*, interpreting it as a location. This was evidenced by the lack of looks to the empty towel (i.e., the destination) and looks to the apple located on the towel combined with the looks to the correct target destination (i.e., the box, Tanenhaus et al., 1995).

In addition to noun-object relations, adults can use other direct cues such as verb-mediated depicted actions to facilitate role assignments and thus the processing of German SVO and non-canonical OVS sentences (Knoeferle et al., 2005, see also Section 2.2). In a visual-world eye-tracking study, participants inspected clip-art scenes depicting for example a princess as washing a pirate while a fencer painted her. The spoken sentence played during scene inspection was initially ambiguous and either described the princess-washes-pirate event (in SVO order) or the princess-is-painted-by-fencer event (in OVS word order). Shortly after the verb had mediated one of the two depicted actions, participants either visually anticipated the pirate (if they had heard *washes*) or the fencer (if they had heard *paints*). From the anticipation of a patient (the pirate in SVO sentences) or an agent (the fencer in OVS sentences), the authors deduced that listeners had assigned a thematic role to the initially role-ambiguous noun phrase *the princess*.

Moreover, also world-knowledge and long-term linguistic knowledge can quickly guide people’s eye-movements and attention to yet unmentioned objects (Altmann & Kamide, 1999; Kamide et al., 2003a). This holds not only true for adults, who are already proficient in their native language, but also for children (Mani & Huettig, 2012; Nation et al., 2003). Furthermore, as already mentioned in Section 2.3, adults as well as children can use direct visual cues, i.e., in this case depicted actions which are mediated directly by the verb (and the associated agents), to facilitate thematic role assignment in difficult OVS sentences (Zhang & Knoeferle, 2012).

Directly depicted visual objects and action events are thus characterized by a strong and mostly direct referential link to an entity from the linguistic input. Hence, these direct cues can very rapidly influence and guide language interpretation. Their

3. The Use of Visual Context for Language Processing

interpretation and anticipation can in turn also be quickly influenced and guided by the linguistic input. Nevertheless, integrating the visual world around us with our linguistic input is not always as straightforward as in these cases.

Often, we cannot make use of or do not have access to these direct cues and instead have to rely on other and potentially subtler visual cues to help us interpret an utterance. In this way, gestures can influence and facilitate language comprehension. Measuring ERPs, Kelly, Kravitz, and Hopkins (2004) have suggested that the processing of iconic hand gestures and speech in naturalistic context are closely connected and that a match between hand gestures and linguistic content has a facilitatory effect on language processing. Additionally, Wu and Coulson (2005) provide strong EEG evidence for an integration of iconic gestures and speech. Moreover, Wu and Coulson (2007) extend their previous findings from the word level to discourse comprehension, providing evidence that speech-congruent iconic gestures elicited smaller N300 and N400 ERPs than incongruent or no gestures. These effects suggest that comprehenders make use of the speaker's gestures (congruent with the linguistic input) in order to arrive at a better understanding of his / her utterance than when no or incongruent gestures were seen.

Yet, even though these authors do not appear to examine the function of gestures as visual social cues (see Section 3.4 for our definition of indirect social cues), gestures also do not fit perfectly into our definition of direct cues, i.e., visual cues that can be directly mediated by the linguistic input. We hence see that strictly separating direct from indirect (social) cues might not be possible. Rather the way in which these different types of visual cues relate to language seems to be more fluid. Nevertheless, trying to categorize different kinds of visual cues into one or the other category represents a first step towards integrating them into real-time language processing accounts.

3.4 Different Types of Visual Contextual Cues – Indirect Cues: Social Cues

Hence, in addition to direct visual cues – such as depicted objects or depicted action events when they are overtly referred to by the utterance – we are also able to use more indirect social visual cues (henceforth ‘indirect cues’) to guide our interpretation of an utterance. By indirect cues we mean that these cues are not directly referential

3. The Use of Visual Context for Language Processing

(in contrast to, e.g., a depicted object and its linguistic referent). Specifically, we are focusing on visual social cues, such as emotional facial expressions as indirect cues here (but see also Section 11 for a discussion on the influence of indirect social but non-visual cues on language processing in adults and children). Without explicitly stating an emotion through speech, the link between a speaker's facial expression and her / his linguistic utterance is arguably less direct (in contrast to noun-object or verb-action relations)⁷.

That we can also use indirect cues to guide language interpretation becomes obvious when looking at language as social behavior (Tomasello, 1992). In the social-pragmatics, usage-based view, language is not an isolated system that refers to things in the world. We as human beings make use of language in order to communicate with each other. Hence it seems essential to consider language use in relation to human beliefs, mental states, characteristics and actions. However, how, when and to what extent we make use of more indirect social information in language processing is unfortunately still under-examined in (real-time) psycholinguistic research (Knoeferle & Guerra, 2012; Van Berkum, Brink, Tesink, Kos, & Hagoort, 2008).

Yet, there are a few studies that have investigated the influence of visual social information on language use. Findings by Staum Casasanto (2008) for example suggest that visual social information can affect the resolution of referential ambiguity. Wolfram (1991) demonstrated that the reduction of a consonant cluster interacts with social factors such as ethnicity. Staum Casasanto (2008) made use of this, showing participants in a first experiment a sentence with either a deleted consonant (*The mis' predicted by the weatherman surprised me*) or the same sentence without a deletion (*mist* instead of *mis'*). She simultaneously presented a photo of a Caucasian person and a photo of a person of color above the written sentences and asked subjects to circle the picture of the person they imagined uttering the sentence. This first experiment suggested that participants associate the ethnicity of the speaker with *t/d* deletion: The sentences containing a consonant cluster reduction were significantly more likely to be attributed to the photo of a person of color than the standard sentences.

⁷ Note that we refrain from defining direct and indirect cues further at this point. Additionally, we are not arguing that direct cues are necessarily non-social or that indirect cues must be social. The focus here lies on the nature of the link between linguistic input and a visual context (see Section 3).

3. The Use of Visual Context for Language Processing

In her second experiment, participants listened to the ambiguous part of one sentence in a pair (e.g., *The mast probably lasted though the storm.* / *The mass probably lasted an hour on Sunday.*). The spoken sentence contained no final stop (e.g., *The [mas] probably lasted...*). Simultaneously participants saw a picture, which represented the speaker of that sentence on the screen. After the sound clip, one of the two potential sentence endings (e.g., *...through the storm.* / *...an hour on Sunday.*) appeared on the screen and subjects indicated via a button press if it made sense in relation to the sentence beginning. Subjects responded faster to the end of the sentence, supporting the *mast* interpretation when they saw the face of a person of color (rather than a Caucasian face) and faster to the sentence ending supporting the *mass* interpretation of the sentence when they saw a Caucasian face (rather than the face of a person of color). In sum, the time it takes to arrive at a decision about the interpretation varies depending on who people perceive as the speaker (Staub Casasanto, 2008).

Furthermore, using two non-standard subject-verb agreement variables, singular NP + *don't* and *there's* + plural NP (both attributed to working class speakers and an informal register vs. singular NP + *doesn't* and *there are* + plural NP as standard register), Squires (2013) conducted experiments on how visually depicted social status affects sentence processing and whether speaker perception has an influence on linguistic perception. In her first experiment the prime trials consisted of a photograph of a high-status or low-status speaker and an unambiguous (standard or non-standard) audio clip of a sentence in which a portion of the sentence containing a noun (but not the subject noun phrase) was masked by white noise. The photo of the high-status speaker was always presented together with the standard sentence, whereas the low-status speaker picture was presented together with the non-standard sentence. After the sound file had ended, two pictures of objects appeared on the screen. These pictures were possible completions for the missing part of the sound file. Participants had to perform a picture-sentence verification task. The speaker photograph and the subject-verb agreement in the sentence should prime participants' sociolinguistic knowledge. The subsequent target trials consisted of an image of a high or low status speaker and an orally presented sentence (e.g., *In the yard, the {white noise} don't sit on the feeder.*). The subject noun phrase, which was ambiguous regarding agreement with the verb and masked by white noise, was followed by either *don't* or preceded

3. The Use of Visual Context for Language Processing

directly by *there's*. The subsequent response screen showed two pictures, one depicting the masked noun in the sentence in its singular version (e.g., *a bird*) and one depicting it in its plural version (e.g., *birds*). The task was to choose between the two pictures and decide which one corresponded better to what the speaker in the photo said. Results suggested that for *there's* (but not for *don't*) trials participants were more likely to respond with the non-standard version when the prime was also non-standard / low-status. However, participants target response was unaffected by the target speaker photograph.

Experiment 2 used the same design and included a speaker choice task. In a prime trial, participants first listened to a sentence with standard or non-standard agreement. Then two pictures of either two low-status people (for non-standard agreement) or two high-status people (for standard agreement) were displayed and participants had to choose which one of them uttered the previously heard sentence. The pictures of high- and low status people were used in order to activate subjects' implicit sociolinguistic knowledge. In the prime trials there was hence no correct or incorrect answer. In the target trials, the target sentence was displayed on-screen to avoid the influence of vocal cues about speaker identity. After participants had read the sentence, a picture of a low-status and a picture of a high-status person appeared and participants chose which speaker best represented the previous sentence. Results indicated that for *don't* trials participants chose low-status speakers more often when preceded by a non-standard / low-status prime and a non-standard target sentence (compared to standard / high-status prime and a non-standard target sentence). Thus, people were influenced by the low-status / non-standard prime in their speaker choice for a non-standard agreement sentence. For *there's* (but not for *don't*) trials, participants were more likely to choose low-status speakers following non-standard target trials than following standard target trials, regardless of high- or low-status condition in the prime trials.

In summary, *there's* trials but not *don't* trials elicited a linguistic (but no social) priming effect. However, for *don't* trials, low-status speaker choice and the non-standard agreement seemed strongly linked (Squires, 2013). Squires suggested that participants may have had stronger sociolinguistic knowledge for *there's* than *don't* trials, since the sociolinguistic prime trials were not affective in the *there's* trials (Squires, 2013).

3. The Use of Visual Context for Language Processing

Using emotional priming in a Visual World Paradigm, Carminati and Knoeferle (2013) suggested that videos of dynamically unfolding emotional facial expressions of real faces (presented as primes) facilitated the on-line processing of emotionally valenced canonical SVO sentences when the emotional valence between prime face and target sentence matched. Their results demonstrated that participants fixated an emotional picture more often when the presented prime was congruent in emotional valence with the target sentence and the related picture (see Section 3.7.1 for more details on this study). Thus, emotional facial expressions as another kind of indirect visual social cue can be integrated incrementally with the linguistic input and modulate its processing on-line. These results are also supported by a recent fMRI study indicating that implied emotions in sentences led not only to activation in emotion related areas but also to activation in language related areas. This activation pattern was however only observed for sentences that implied an emotional content, such as *The boy fell asleep and never woke up again* whereas there was no activation in emotion related areas for neutral sentences like *The boy stood up and grabbed his bag*. These results suggest a strong connection between emotions and language in that although the emotional sentences did not contain a single emotionally valenced word, the meaning of the sentence as a whole has a strong emotional valence which cannot be ignored (Lai, Willems, & Hagoort, 2015).

These studies have so far suggested that both direct referential cues such as depicted objects (adults and children, Zhang & Knoeferle, 2012) and more complex, indirect visual social cues, such as social status (adults, Squires, 2013), ethnicity (Staub Casanto, 2008) and emotional facial expression (adults, Carminati & Knoeferle, 2013) have an effect on how language is comprehended and processed. However, only very little research has so far investigated the combination of direct and indirect cues and their effect on sentence processing. To date we do not yet know if more indirect visual social cues can be integrated into language processing to the same extent and with the same time course as direct visual cues. Moreover, we do not yet know if these two cue types interact with each other, i.e., if an accumulation of different kinds of cues further facilitates and enhances sentence processing or if language comprehension is in fact inhibited the more (distinct) cues are available for use. Additionally, almost nothing is yet known about the developmental differences in using these visual cue types for language processing.

3.5 Cumulative Use of Direct and Indirect Cues

One of the few studies that already looked at the influence of both direct (object depiction) and indirect (speaker gaze shift) cues on sentence processing is by Kreysa, Knoeferle, and Nunnemann (2014) and Kreysa, Knoeferle and Nunnemann (in prep.) They demonstrated that both direct cues (object depictions) and indirect cues (speaker gaze shift) individually have an effect on sentence processing. In a visual-world eye-tracking study, they presented participants with videos of a speaker uttering sentences about two depicted virtual characters (e.g., *The waiter congratulates the millionaire in the afternoon*). They manipulated whether actions related to the sentential verb were (vs. were not) depicted and whether the speaker's gaze fixated the characters or was obscured. Thus, the baseline condition contained neither the gaze nor the action cue, one condition only contained the gaze cue, one condition only contained the action cue and a fourth condition combined both cues. The cues were always presented at the same time just after the onset of the verb, making it possible for participants to use the available cue to anticipate the upcoming patient (*the millionaire*) even before it was mentioned in the sentence. Results indicated that both cues were affective on their own in that listeners used them to anticipate the upcoming patient of the sentence before it was actually mentioned. However, presenting both cues at the same time was not more helpful in anticipating the upcoming character than the gaze cue alone.

Another study that investigated the integration of two different (social) cue types into language comprehension (but not in combination / interaction with direct referential cues, i.e., depicted objects or action events) was carried out by Holler, Schubotz, Kelly, Hagoort, Schuetze, and Özyürek (2014). They crossed social speaker gaze with co-speech gestures. Participants had to decide which object out of a pair was mentioned in the cue-accompanying sentence. The results showed that reaction times were faster in the condition in which speaker gaze was averted but the sentence was accompanied by a gesture than when no gesture was present. However, in the conditions in which the gaze was not averted, i.e., the listener felt addressed by the speaker, reaction times did not differ between a condition in which the gesture was also present and one in which no speech-accompanying gesture was present.

Thus, although the direct cue (object depiction) and the indirect cues (gaze shift, co-speech gesture) were exploited and used for sentence processing, the specific way in which they are used seems to differ. Moreover, in both studies (Kreysa et al., 2014)

and (Holler et al., 2014), two cues do not seem to be more helpful than one. However, the nature of these two cues differs greatly. Due to the diverse nature of different kind of cues we do not yet know if it is always the case that direct and indirect cues cannot be used cumulatively to ease utterance interpretation. Moreover, the age at which we begin to see the first evidence for the comprehenders ability to process and use them seems to differ depending on the type of visual cue. These processing differences become especially evident when looking at the ability to process faces and emotional facial expressions.

3.6 Differences between Direct and Indirect Cues

These two kinds of cues, i.e., direct referential cues such as depicted objects and more indirect social cues, are not just processed differently with regard to language comprehension, but also seem to be processed in partially distinct brain regions. Looking at the face processing literature, this becomes evident. The ability to recognize faces compared to other objects is widely seen as an exceptional process. Whereas common objects, such as cars, toasters and computers are visually very distinct, human faces, although differing in detailed features, all share uniform and general features (Grelotti, Gauthier, & Schultz, 2002). Even though we are able to discriminate faces on the basis of their detailed features, most people are not able to do the same with individual objects belonging to a certain category, such as discriminating different types of racing cars or songbirds (Archambault, O'Donnell, & Schyns, 1999; Diamond & Carey, 1986). However, we interact with people on a daily basis and can hence be called experts in face processing and recognition. This is on the other hand not the case for non-face objects, as we rather discriminate them into separate general categories like bird or dog as opposed to Robin or Great Dane (Grelotti et al., 2002). Yet, if the motivation is strong enough, we seem to be able to become experts in virtually any field.

The motivation for becoming an expert in the recognition and processing of faces is likely socially motivated, allowing us to interact and communicate with each other (Grelotti et al., 2002). Nevertheless, arguably because not everyone is an expert on everyday common objects, indeed, distinct brain regions have been found to be active during the perception of objects in contrast to faces. Pitcher, Charles, Devlin, Walsh, and Duchaine (2009) delivered transcranial magnetic stimulation (TMS) over

3. The Use of Visual Context for Language Processing

different brain regions of participants, while they performed face, body and object discrimination tasks. They found that face discrimination was impaired when the right occipital face area was stimulated but not when the right lateral occipital area was stimulated. By contrast, object discrimination was impaired when the right lateral occipital area was stimulated but discrimination was not impaired when TMS was delivered over the right occipital face area. These results led to the conclusion that faces, objects and bodies are processed in functionally different brain areas (Pitcher et al., 2009).

A single case study of a patient with developmental deficits in object but not face recognition supports this assumption (Germine, Cashdollar, Düzel, & Duchaine, 2011). The patient did not suffer from brain damage or memory impairment and underwent an otherwise normal development. Still, in a series of face and object recognition tests, her performance in the object recognition tests across different object categories was impaired. Her face recognition scores by contrast did not show any deficits (Germine et al., 2011).

Scherf, Behrmann, Humphreys, and Luna (2007) looked at the organization of category-selective regions in the brain from a developmental perspective. In an fMRI study, they showed children (5-8 years), adolescents (11-14 years) and adults (20-23 years) videos of natural faces, objects and places and measured their brain activity in the ventral visual cortex. Adults and adolescents both showed activation in classical face-related areas, i.e., the fusiform face area, the occipital face area and the superior temporal sulcus. Children by contrast, although showing adult-like activation for places and objects, did not show activation in any of the classical face-related areas. Hence, these results suggest that category-selectivity develops differently depending on the class of visual objects. Moreover, Scherf et al., (2007) see their results as compatible with the notion of expertise and the acquisition of expertise for face recognition.

In light of these studies, it becomes clear that not all visual cues are processed in the same way and that they also differ in terms of when we develop the ability to process them. With regard to language processing, it is then not surprising that direct visual cues such as objects and indirect visual cues such as faces might be used to different extents in order to comprehend language.

3. The Use of Visual Context for Language Processing

However, being experts in the recognition of faces due to social motivation (Grelotti et al., 2002) and extensive practice throughout our lives does not tell the whole story when it comes to integrating social cues such as faces into language processing because faces as social cues seem to be particularly complex. When talking, we do not just “have a face” so that our conversation partner can recognize it, but we arguably use our face conversationally to (consciously or unconsciously) convey a non-verbal message alongside our verbal message. In turn, the comprehender interprets the facial expression of the conversation partner (mostly unconsciously) and tries to integrate it into the linguistic input in order to correctly understand and interpret it or even to facilitate sentence processing (Carminati & Knoeferle, 2013).

This facial expression is often not neutral but portrays some kind of emotion depending on the emotional valence of the utterance. However, the use of emotional facial expressions as a cue is not an easy task, as the processing, recognition and interpretation of emotions is highly complex in and of itself.

Recall that the studies we will present in this thesis (Sections 6-9) investigated the use of emotional facial expressions as an indirect visual social cue on real-time language comprehension and thematic role assignment across the lifespan. Hence, the following Section (3.7) will focus on emotion processing in more detail. We will demonstrate that emotions are processed differently depending on the age of the comprehender and will discuss potential underlying reasons. While younger adults seem to prefer negative over positive emotional stimuli, older adults and children appear to show an opposite emotional bias⁸, i.e., they favor positive over negative emotional material. This is especially relevant in the light of our results (see Sections 6-9) and will be discussed further in Section 10.6 taking our results into account. Moreover, these age dependent processing differences of emotional facial expressions as an indirect social cue have clear implication for the development of a more complete real-time language processing account and hence need to be taken into account. This will also be discussed further in Section 10.8.

8 Throughout this thesis we will use the terms (emotional) bias and preference interchangeably and define them as synonyms. For a more detailed discussion on this matter see Murphy & Isaakowitz (2008).

3.7 Emotions and Emotional Facial Expressions

Emotions are complex psychological and physiological states that can vary not only from person to person but also in their degree of positivity or negativity according to our current emotional state. Emotions cannot be classified into discrete entities or feelings but are rather situated on a continuum, hence, the border between positive and negative emotions is fluid (Dolan, 2002). Moreover, identifying emotions from facial expressions is not always easy. Most situations elicit more than one emotion and multiple emotions can be more or less prominently portrayed through a single facial expression (Plutchik, 1980). This is even more the case in natural social interactions, in which emotions can be expressed very subtly and the facial expressions of our interlocutors are highly dynamic and fast-changing (Hareli & Hess, 2012). In this line, (Carroll & Russell, 1996) suggest that even though we can easily determine the ‘quasi-physical’ features of an emotional facial expression, such as a smile or a frown, inferring the underlying emotion of a person portraying these quasi-physical features, as for example, joy, anger, surprise or rage is arguably difficult.

Nevertheless, recognizing the facial emotion our interlocutors display during an interaction is vital for building and maintaining social relationships. They are crucial for nonverbal communication and moreover essential for interpreting other people’s feelings (Lamy, Amunts, & Bar-Haim, 2008). Thus, identifying and interpreting emotional facial expressions is highly important (Kissler & Keil, 2008).

3.7.1 Identification and Processing of Emotions

Hence, it is not surprising that emotional material is also treated differently (behaviorally, but also neurally) from neutral facial expressions, pictures or words. An example illustrating the relevance of emotional cues can be found in research on emotionally valenced pictures: When subjects inspected a display of an emotional and a neutral picture, their gazes were directed first to the emotional picture. Even when people were instructed to explicitly look at the neutral pictures first, looks were directed first and also longer to the emotional pictures (Nummenmaa, Hyönä, & Calvo, 2006). Furthermore, pupil size increases for emotionally arousing pictures (Bradley, Miccoli, Escrig, & Lang, 2008) and sounds (Partala & Surakka, 2003) compared to neutral stimuli. These findings thus demonstrate that we unintentionally react to emotional stimuli more than we react to neutral ones. Additionally, we also

3. The Use of Visual Context for Language Processing

integrate word pairs better into a single mental representation when one of the words has an emotional valence than when both words of a pair are neutral (Murray & Kensinger, 2012).

Furthermore, a great number of emotional priming studies have shown that emotionally valenced positive and negative information (primes) can facilitate and / or speed up the processing and recognition of emotionally congruent subsequent targets (see e.g., Hermans, Houwer, & Eelen, 1994; Lamy et al., 2008). Emotional priming is even possible across modalities and across emotional stimuli from different categories (see Kim & Sumner, 2015 for effects of semantically neutral words spoken with emotional prosody on semantically related but emotionally and prosodically neutral target words; Hermans, Spruyt, & Eelen, 2003 for face-word priming; Carroll & Young, 2005 for face-word, face-picture, non-verbal emotional sounds and words, non-verbal emotional sounds and face / picture priming; and Carminati & Knoeferle, 2013 and Münster, Carminati, & Knoeferle, 2015 for (dynamic) face - sentence priming).

In an emotional priming paradigm, Aguado, Garcia-Gutierrez, Castañeda, and Saugar (2007) for instance used faces as primes and words as targets. Primes and targets were presented once with a stimulus onset asynchrony (SOA) of 300 ms and once with an SOA of 1000 ms. Participants saw a positive or a negative face followed by a positive or negative word or a question mark. If the word appeared on the screen, subjects had to judge the valence of the word. If the question mark appeared, subjects had to make a decision about the gender of the previously seen emotionally valenced face. As participants did not know in advance whether they had to detect the gender of the face or judge the valence of the word, the task was unpredictable. For both SOAs the measured reaction times were shorter for the valence congruent face-word trials i.e., a priming effect was observed. In the experiment using an SOA of 300 ms, priming was moreover observed even in the non-evaluative task, when detecting gender. Hence, in early stages of valence processing, the evaluation of emotional faces appears to be automatic and unintentional. In a second experiment, when SOA was increased from 300 ms to 1000 ms, a prime effect was found when the valence of the face had to be judged, while there was no prime effect when participants had to detect the gender of the face. Aguado et al. (2007) thus suggested that the affective priming effect can be modulated by the prime task. Crucially however, they also

3. The Use of Visual Context for Language Processing

suggested that emotional priming in the early stages of processing happens unintentionally and only persist for longer SOAs if the emotional valence is made conscious (see also Hermans et al., 2003; Li, Zinbarg, Boehm, & Paller, 2008; Wong & Root, 2003 for similar results).

Furthermore, emotional information conveyed through faces, pictures or words also elicits different ERP responses compared to neutral information. Kissler, Herbert, Peyk, & Junghofer (2007) found enhanced early cortical responses during uninstructed reading of positive and negative words compared to neutral words. Furthermore, Zhang, Guo, Lawson and Jiang (2006) employed an emotional priming paradigm and presented positive, negative or neutral words or pictures as primes and positive, negative or neutral words as targets. Participants' brain waves and reaction times were measured. Participants indicated via button press if the emotional valence of the target word was happy or sad. In line with the emotional priming studies cited above, reaction times were faster for valence congruent than valence incongruent trials. Moreover, participant's ERP responses were different for emotional compared to neutral trials. Furthermore, the N400 component, which is sensitive to linguistic semantic integration, was delayed for prime-target incongruent compared to congruent trials. This delay suggests that the N400 is also sensitive to affective mismatches and not only to semantic incongruence (Zhang et al., 2006). Treese, Johansson, & Lindgren (2010) also found differences in ERPs for emotionally valenced faces compared to neutral faces. In a continuous recognition study, they presented participants with repeated runs of positive, negative and neutral facial expressions to investigate memory distortions induced by emotion. They found that emotionally valenced faces were better and faster remembered than neutral faces. Additionally, negative and positive faces elicited greater positive going ERPs between 300 and 700 ms compared to neutral faces (Treese et al., 2010).

Hence, emotional information conveyed through faces, pictures, words or sounds is (un)consciously treated differently from neutral information. This is not only the case for reaction time, recognition and memory effects, but crucially also for underlying electrophysiological responses and can hence have implications – especially regarding the time course – for the integration of emotional information into real-time language processing.

However, emotional information is not just different from neutral information, there are furthermore differences in how we react to and perceive different emotions and emotional valences. Moreover, the age of the interpreter also seems to have a considerable impact on how emotionally valenced material is processed.

3.7.1.1 Younger Adults

The literature reviewed above showed that emotional information in words, pictures, faces and even sounds is more salient than neutral information. There is, however, also a difference as to which emotions are perceived and recognized better and crucially also preferred as a function of the perceiver's age. Younger adults for example, in contrast to older adults and children, seem to prefer negative over positive emotional material.

Grühn and Scheibe (2008) asked younger and older adults to rate positive and negative emotional pictures (IAPS, Lang, Bradley, & Cuthbert, 1998) and to perform a picture recognition task. As each participant saw each picture equally often as a target and as a distractor item in the recognition task, they computed an index showing how easily a picture was discriminated as a target or as a distractor. Their results showed a correlation between these memorability scores and the rating scores for younger adults, indicating that they recognized negative pictures better than positive pictures. Testing young adults, Finn, Roediger and Rosenzweig (2012) moreover provide evidence that negative pictures can enhance picture recall when they are shown right after a retrieval attempt. Hence, negative pictures can modify memory after retrieval. However, this was only the case after negative, but not after positive pictures.

Additionally, Kim and Sumner (2015) presented young adults with semantically neutral words spoken in neutral, angry or happy prosody followed by semantically-related (but also neutral) target words without emotional prosody. They demonstrated that recognition of target words was facilitated only when primes were spoken with a neutral or angry prosody. There was however no facilitation for primes spoken with happy prosody. Kim and Sumner (2015) do not explicitly state which age group their participants belonged to, but mention that they were recruited from the Stanford University Community and participated for money or course credit. Hence, participants' ages were likely between 18 and 30 years.

3. The Use of Visual Context for Language Processing

Bach, Schmidt-Daffy, and Dolan (2014) found enhanced memory access for angry compared to happy facial expressions and enhanced identity recognition of neutral faces when they were followed by angry faces. Another study showing the preference for negative emotion in younger adults is Lamy et al. (2008). In this intertrial priming study participants either had to detect an emotionally valenced face among an array of neutral faces or a neutral face among an array of emotionally valenced faces. The reaction times were shorter when an emotional expression was followed by a similar emotional expression than when followed by a neutral expression on subsequent trials. Moreover, this effect was significantly bigger for angry faces than for happy faces.

Hence, younger adults seem to be biased more towards negative compared to positive information, as the former tend to elicit faster and stronger behavioral responses. One explanation for these results is that negative stimuli seem to need more processing resources, have a privileged access to attentional resources and are cognitively more complex than positive stimuli, thus leading to a negativity bias for younger adults (Finn et al., 2012; Holt, Lynn, & Kuperberg, 2008).

Furthermore, this negativity bias has not only been found using off-line measures, but moreover, many ERP studies have noted that the Late Positivity is bigger for negatively than for positively valenced stimuli (e.g., Bernat, Bunce, & Shevrin 2001; Kanske & Kotz, 2007). A bigger Late Positivity for negative material (vs. positive) was also found even when people did not have to pay explicit attention to the affective meaning of the stimuli, or when the degree of arousal was balanced over negative and positive stimuli (Bernat et al., 2001; Ito, Larsen, Smith, & Cacioppo, 1998). In another ERP study, Holt et al. (2008) embedded neutral, positive or negative words in a preceding neutral sentential context. Participants (younger adults) had to judge the valence of the sentence (experiment 1) and passively read the sentences (experiment 2). In both experiments Holt et al. (2008) found a Late Positivity with a larger amplitude for negative compared to positive words. However, the additional N400 effect they observed was only modulated by the emotionality of the words and not by their specific valence. They thus suggest that although the N400 is sensitive to emotional information (vs. neutral valence), only the Late Positivity is sensitive to specific (positive vs. negative) valence information.

In line with these results, Martín-Loeches et al. (2012) found differences in the

amplitude of the left anterior negativity (LAN) for emotionally valenced words with morphosyntactic violations embedded in sentences compared to neutral words. A larger LAN amplitude was found for negative words compared to neutral words and a smaller LAN amplitude was found for positive words compared to neutral words. However, when the emotional words did not violate the sentence syntactically but semantically, the N400 was not affected by the specific valence of the emotional words. Hence, the authors suggested that the increase in LAN amplitude for negative words reflected greater difficulties in syntactic processing and greater attentional resources needed for the integration of negative compared to neutral and positive words into sentences. These findings thus demonstrate that younger adults are influenced more strongly in their behavior by negative emotional information compared to neutral and positive information, but that this negativity bias also has an impact on the underlying information integration processes in the brain.

3.7.1.2 Older Adults

However, the older we get, the less good we seem to be able to recognize emotions and the more we seem to shift our preferences for emotionally valenced information from a negativity towards a positivity bias. Unlike many other studies concerned with aging effects, Paulmann, Pell and Kotz (2008) did not test younger and older adults, but used younger (18 to 28 years) and middle aged (38 to 50 years) participants to show the effects of aging on emotional speech recognition. In their study, participants listened to sentences and identified the emotional prosodic category (anger / disgust / sadness / fear / happiness / pleasant surprise / neutral) as quickly and as accurately as possible. Younger adults performed more accurately and had fewer time-outs than middle-aged adults, thus showing a decline in the recognition of emotional speech with increasing age (Paulmann et al., 2008; see also Kiss & Ennis, 2001). Moreover, older adults also rated positive pictures as more positive and less arousing and negative emotional pictures as more negative and arousing than younger adults (Grühn & Scheibe, 2008; see also Di Domenico, Palumbo, Mammarella, & Fairfield, 2015).

In another emotion recognition study, Mill, Allik, Realo and Valk (2009) investigated in more detail at which age the decline in emotion recognition abilities begins. They asked participants between 18 and 84 to recognize facially or vocally

3. The Use of Visual Context for Language Processing

expressed emotions and found that sadness, anger, fear and disgust recognition was significantly worse in participants over 60 years of age compared to the younger age groups. At what age range participants started performing worse, however, depended on the specific negative emotion. Recognition of facial expressions worsened from 30 years onwards. On the other hand, the recognition of positive emotions from facial and vocal cues only slightly decreased at age 61 compared to the other age groups.

In line with this are findings from many other studies, showing that older people have trouble recognizing certain, especially negative, emotions (for recent evidence, see Di Domenico et al., 2015). Many studies on the other hand suggest that positively valenced facial emotional expressions are memorized better and more often than negatively valenced ones in older age (Mather & Carstensen, 2003). It is, however, still debated whether the positivity effect found in older adults is in fact due to a shift away from negative information, i.e., a reduced negativity bias, or if it rather represents a bias towards positive information (Reed, Chan, & Mikels, 2014).

In a meta-analysis Ruffman, Henry, Livingstone, and Phillips (2008) found a trend for older adults having difficulties in identifying anger, sadness and fear. However, the differences between the age groups cannot be reduced to age-related problems in face processing (Calder et al., 2003) or to the explanation that these emotions are inherently difficult to identify, because younger adults did not show any problems in recognizing these emotions (Ruffman et al., 2008).

Hence these conclusions rather suggest a positivity bias in the sense that negative emotions are avoided (yet we will discuss this issue further in the remainder of this Section and in Section 3.7.1.2.1). Isaacowitz et al. (2007) reinforce these conclusions in their own meta-analysis. They also ran experiments using facial expressions and sentences with an emotionally valenced word. In addition to using a facial and a lexical recognition task, they also included a group manipulation in testing younger, middle aged and older adults. The results indicated that the younger group performed better in all tasks than the middle-aged and the old group. Concerning emotion recognition, all groups were better at recognizing happiness compared with the other emotions, followed by sadness.

Furthermore, Isaacowitz, Allard, Murphy, and Schlangel (2009) showed older and younger adults neutral - emotional synthetic face pairs and measured their eye-movement patterns. Their results indicated that older but not younger adults looked

3. The Use of Visual Context for Language Processing

preferably towards positive and away from negative faces. However, for older adults, the preference to look at positive faces only emerged 500 ms after face onset and increased over time. The tendency to look more at the negative facial expression compared to the neutral one needed even more time, i.e., emerged approximately 3 seconds after stimulus onset. The authors concluded that the tendency to avoid negative emotions is costly in terms of cognitive control resources (see also Isaacowitz, Wadlinger, Goren, & Wilson, 2006).

Another recent meta-analysis of 100 empirical studies on the other hand concluded that the positivity effect is reliable, i.e., not a reduced negativity bias for older adults. Interestingly, the authors demonstrated that the positivity effect is strongly influenced by task constraints. They found that the positivity effect was often completely absent in studies with cognitively costly tasks, whereas it was strongest when participants' emotion processing was unconstrained (Reed et al., 2014).

Still, the question about the positivity preference for older adults has not yet reached a consensus. While measuring their pupillary responses, Ziaei, Hippel, Henry, and Becker (2015), for example, asked older and younger adults to encode blocks of either positive or negative pictures and to ignore emotionally valenced or neutral distractor pictures. After the encoding phase, participants performed a memory recognition test. Even though their results suggested a memory advantage for positive versus negative pictures for the older but not the younger adults – thus speaking in favor of a bias towards positive information – pupil dilation for older adults was only significantly increased in response to negative compared to positive pictures. Moreover, the positive distractor images did not alter older participants' pupil size, suggesting that they do not automatically attract attention.

The assumption that this study points towards avoidance for negative information, rather than a true preference for positive information is furthermore supported by the fact that the positivity effect found for the memory task and its relation to the age groups is mediated by pupillary response. Ziaei et al., (2015) reason that the pupillary responses suggest that older adults exerted more cognitive effort in encoding negative compared to positive information, maybe in order to suppress negative information (Ziaei et al., 2015).

The suppression of negative emotion and the resulting bias towards positive information has also been demonstrated by studies on first impressions of faces and

3. The Use of Visual Context for Language Processing

on mood in older compared to younger adults. Older adults' face ratings for first impressions were more positive and they rated even the most negative faces as more healthy, trustworthy and less hostile than younger adults (Zebrowitz, Franklin, Hillman, & Boc, 2013). Regarding mood, whereas younger adults preferably look at emotional faces that are congruent with their current mood, older adults display a bias towards positive facial expressions and away from negative ones when they are in a bad mood. It seems thus that older adults can also use the bias towards positive material to regulate their current mood (Isaacowitz, Toner, Goren, & Wilson, 2008).

Moreover, it has been suggested that the preference to attend to positive information more than to negative also has an effect on subsequent language processing in older age (Carminati & Knoeferle, 2013; Münster, Carminati, & Knoeferle, 2014). Carminati and Knoeferle (2013) and Münster et al. (2014) presented younger and older adults with positive or negative natural facial expressions as primes. Following this prime face, two pictures of opposite emotional valence were depicted on the screen. Additionally, participants heard an either positively or negatively valenced sentence describing one of the pictures while they inspected the images. Hence, the emotional prime face could be either positively or negatively congruent or incongruent with the emotional picture-sentence target. Participants' eye-movements were measured while they listened to the sentence and inspected the images. They also had to perform a sentence-picture matching task after every trial. Eye-movement results for the younger adults were in line with findings presented in Section 3.7.1.1, in that younger adults fixated the negatively valenced picture more during the processing of a negatively valenced sentence only when they were also primed with a negative facial expression. The eye-movement data for the older adults suggested in contrast that they were facilitated only in processing the positive sentence when preceded by a positive emotional prime face. There was however no facilitation for the negatively congruent prime-target combination. Hence, it seems that older adults cannot only use the emotional facial expression to regulate their mood but they make also use of their positivity bias for processing emotionally valenced linguistic input.

3.7.1.2.1 Underlying Reasons

The reviewed studies have thus far clearly demonstrated that older adults differ in their processing of emotional information from younger adults. Although the question of why this age-related change occurs is not the focus of this work, we will briefly address this issue. Just like the debate about the nature of the positivity bias – as either a bias towards positive or away from negative information – the question of the nature of this shift in bias is still under discussion.

One theory is the socioemotional selectivity theory (SST) by Carstensen, Fung, and Charles (2003). They suggest that with advancing age, people have gained substantial experience, especially emotional experience and thus pay more attention to emotional goals. Concerns for the future on the other hand become less relevant due to the limited time that is left in their lives. Crucially, the SST claims that this motivational shift affects cognitive processing. That is, a regulation of emotional arousal leads to a shift in focus towards positive events and present experiences. Negative experiences and feelings become less relevant (Carstensen et al., 2003).

Another explanation is based mainly on the cognitive decline and assumes that the positivity bias occurs because older adults are worse at recognizing and processing negatively valenced information due to amygdala dysfunction (Cacioppo, Berntson, Bechara, Tranel, & Hawkley, 2011) – which is suggested to play an important role for biased attention (see e.g., Todd, Cunningham, Anderson, & Thompson, 2012). However, if the positivity bias were only due to cognitive decline in the elderly, this effect should be automatic and not be observed 500 ms after stimulus onset, as in Isaacowitz et al. (2009).

Moreover, older adults should not be able to regulate their mood by avoiding negative information (Isaacowitz et al., 2008), since avoiding something presupposes that this information has been processed and interpreted in the first place. However, it is indeed the case that several cognitive abilities such as fluid intelligence, memory and reaction times are also said to start worsening already at the age of 30 (Mill et al., 2009) and that older adults are less accurate at recognizing negative emotions than younger adults. However, a number of studies have reported no correlation between emotion recognition deficits and crystallized and / or fluid intelligence, even though fluid intelligence has been found to be affected by age (see e.g., Calder et al., 2003; Isaacowitz et al., 2006; Salthouse, 2000; Sullivan & Ruffman, 2004; Zebrowitz et al.,

3. The Use of Visual Context for Language Processing

2013). This change in cognitive processing resources goes hand in hand with a neurological decline in aging (Calder et al., 2003). Especially the frontal and temporal brain regions have to deal with considerable age-related changes. Thus it is possible that these changes explain at least partly the difficulties older adults have in recognizing and identifying certain emotions (Di Domenico et al., 2015; Kiss & Ennis, 2001; Paulmann et al., 2008; Ruffman et al., 2008).

Moreover, it has been suggested that older and younger adults differ in their amygdala activation when processing positive and negative stimuli (Leclerc & Kensinger, 2011; Mather et al., 2004). While older adults show a higher amygdala activity in attending to positive (vs. negative) pictures (Mather et al., 2004), younger adults did not show a difference in amygdala activation or showed higher activation for negative compared to positive images (Leclerc & Kensinger, 2011).

Hence, even though theories based on motivational shifts and theories based on neural and cognitive decline make distinct assumptions about the reasons underlying the age-related shift in emotion processing, the empirical evidence suggests that neither one can fully explain the positivity bias without taking the other into account. In fact, it is much more likely that both explanations complement each other (Beer & Ochsner, 2006; Mather, 2012).

This becomes even more evident when looking at the development of emotion recognition and face processing in children. Unlike adults, children are not just still in the process of mastering their mother tongue, but they moreover also have to learn how to use and interpret their social environment. In addition, this all happens during a period in which the child's cognitive and neural functions are still in the process of maturation (see e.g., Adolphs, 2006; Grossmann & Johnson, 2007 for a review and Beer & Ochsner, 2006; Tzourio-Mazoyer et al., 2002). Hence, it is highly likely that processes such as face and emotion recognition and communication are tightly coupled and closely interact with each other from the very beginning of our lives until the end.

Before we discuss the interaction and effect of emotions on language processing further in our own studies (Section 6-9), we will look at children's emotion processing skills and preference in the next Section (3.7.1.3) so that we can discuss our own findings in light of the reviewed literature. Even though children can from a very early age onwards distinguish positive from negative emotions, they seem to find

positive emotions easier to process than negative ones. This might be one of the reasons why they show, similar to older adults, a preference towards positive emotions.

3.7.1.3 Children

It is crucial for the child's social and linguistic development to be able to recognize her mother's facial emotional expression and to know that when she is smiling (vs. looking angry) while saying "Come here!", something good (vs. something bad) is going to happen. This social and cognitive development and learning has been shown to begin at the earliest moments of a child's life. Even newborns, only minutes after birth already attend to faces more than to non face-like stimuli (Johnson, Dziurawiec, Ellis, & Morton, 1991; Mondloch et al., 1999). By the age of 5, children can already correctly interpret non-verbal cues as complex as those indicating that someone is in a more powerful position, i.e., "in charge" than someone else. Even 3-year-olds can already determine other people's social roles when non-verbal cues associated with "being in charge" such as an expansive body posture or the head tilted back are supported by linguistic cues (Brey & Shutts, 2014). Furthermore, 5-year old children also already have and make use of their social preferences when building social networks. Kinzler, Shutts, DeJesus, and Spelke (2009) have indicated that 5-year olds prefer to be friends with children of their own native language compared to another language or their own native but accented language, regardless of comprehensibility. Additionally, Caucasian children preferred to be friends with children of color when these spoke with a native accent over Caucasian children who spoke with a foreign accent (Kinzler et al., 2009). Hence, at the age of 5 children already seem to have a good understanding of some social cues.

However, not every social cue seems to be readily available for correct use at this young age. As became clear from Section 3.7.1.1 and 3.7.1.2, interpreting emotional information from sounds, pictures or faces is not always easy and even changes with the age of the interpreter. Thus, learning how to recognize and categorize different emotions cannot be achieved within the course of a day, but takes years to be mastered completely. Still, acquiring emotional competence is as essential for a child's development as is the acquisition of language and other social competences,

3. The Use of Visual Context for Language Processing

since these three are tightly intertwined and moreover crucial for successful human communication (Denham et al., 2011).

Children's understanding of emotions emerges quickly but has been shown to improve gradually over time. Happy facial expressions are produced and identified the earliest and most accurate followed by sadness, and then by other negative emotions such as fear and anger (Bullock & Russell, 1984; Widen & Russell, 2003). This development takes place between a few months of age until early adolescence (see Herba & Phillips, 2004 for a review). Children begin to use emotion descriptive language and talk about their emotions with around 16-20 months of age (Fivush, Brotman, Buckner, & Goodman, 2000; Ridgeway, Waters, & Kuczaj, 1985). By the age of 2, basic emotional expressions such as happiness, sadness, anger and fear are then already commonly used to describe related states such as crying and hurting (Wellman, Harris, Banerjee, & Sinclair, 1995). By the age of 4-5, children can reliably label the most basic emotional facial expressions such as happiness, sadness and anger. Moreover, at that age they also understand causal reasons for these emotions. However, the ability to comprehend the influence of false beliefs and conflicting desires on emotions, as well as the possibility to hide emotions only seems to emerge between 5 and 6 years of age (Pons & Harris, 2005).

The younger the children are, the more they seem to rely on facial expressions as opposed to situational information when emotional cues are conflicting. Gnepp (1983) showed 3-, 6- and 11- year old children a series of pictures portraying children with emotional expressions in situational setting. The emotional facial expressions of the children in the pictures could either match or mismatch with the situational setting, e.g., it could show a smiling boy with his broken bicycle or a crying boy with his broken bicycle. The children were then asked to indicate how the child in the picture felt and how they could tell that the child in the picture felt that way. The results indicated that the youngest children made judgments on the basis of the character's facial expression alone, whereas the older children made significantly more affective judgments consistent with the situation. In addition, the older the children, the better they were at creating a story that reconciled the conflicting facial and situational cues (Gnepp, 1983).

However, Morton and Trehub (2001) suggested that when emotional linguistic content and emotional prosody match or mismatch, children younger than 9 years of

3. The Use of Visual Context for Language Processing

age relied mostly on content whereas adults primarily relied on paralinguistic cues in judging the happiness or sadness of a speaker. It was only at the age of 10 that half of the children evaluated the speaker's feelings on the basis of the paralinguistic cues. Nevertheless, when the content could not be judged because it was presented in a foreign language even the youngest children of 4 years of age could accurately judge the emotional feeling of the speaker based on the emotional prosody alone (Morton & Trehub, 2001).

Gnepp (1983) and Morton and Trehub (2001) thus demonstrate that the ability to correctly use, process and identify emotional information does not only develop with age. Crucially, it also varies as a function of how the emotional information is presented and with which other information it co-occurs. Hence, the modality and way in which emotional content is presented effects the way children (and adults) process this information and the ease with which they can do so.

Apart from the fact that the processing of emotional material seems to depend on the way in which it is presented, children also already react differently to opposing emotionally valenced information. In addition to understanding and producing positive emotional expressions earlier than negative ones, the former also seem to be categorized more accurately than the latter (Durand, Gallay, Seigneure, Robichon, & Baudouin, 2007).

Durand et al. (2007) asked 5-6-, 7-8-, 9-10-, 11-12-year old children and a group of adults to recognize happy, sad, disgusted, angry, fearful and neutral facial expressions and found that the youngest group could only accurately recognize happy and sad facial expressions (see also Gross & Ballif, 1991). However, the recognition of fear, anger and disgust needed an additional 2-4 years to reach adult-level competence. Interestingly, neutral facial expressions also seem to be difficult for 5-6-year old children to recognize, since children under the age of 9 showed a tendency to assign an emotional valence even to the neutral facial expression. Hence, the ability to correctly process, identify and categorize emotional information does not seem to be an easy task and needs years of development and learning until adult competence is reached (Durand et al., 2007).

That emotional information is special compared to non-emotional information becomes even clearer when looking at children's ERP responses. In contrast to neutral stimuli, emotionally positive and negative stimuli elicited an increased positivity

3. The Use of Visual Context for Language Processing

between 500 and 1500 ms after stimulus presentation. This late positive potential (LPP) in response to emotional compared with neutral material has also been observed in adults. Thus, 5-8-year old children already show adult-like brain responses regarding the on-line processing of emotional pictures (Hajcak & Dennis, 2009) even though they might not yet be able to fully identify all emotional categories. Although children know implicitly and explicitly from a very early age onwards that there is a difference between positive and negative emotional material, to come to a full adult-like understanding of the different specific emotions takes time to develop.

However, research regarding a bias towards positive or negative emotional material – as it is the case for older and younger adults – has not come to a clear consensus yet. Thomas et al. (2001) for example investigated amygdala activation of children and adults in response to fearful and neutral facial expressions and found an increased activity to neutral expressions compared to fearful expressions in children and the opposite effect in adults. However, they did not test amygdala activation in response to positive emotional facial expressions. Moreover, children often assign an emotional valence to neutral expressions instead of categorizing them as non-emotional (Durand et al., 2007; Gross & Ballif, 1991; Hortaçsu & Ekinci, 1992; Tottenham, Phuong, Flannery, Gabard-Durnam, & Goff, 2013). Pagliaccio et al. (2013) did use positive emotional facial expressions in addition to negative and neutral expressions. They found no differences in the amygdala activation of 7-12-year olds in response to emotional faces of different valences, but an increased activation in response to emotional compared to neutral facial expressions.

In contrast to that, Vaish, Grossmann, & Woodward (2008) argue that the negativity bias in younger adults is already present in children. However, they only review previous studies which did not directly address the question of an emotional bias but claim that the negativity bias for children has been shown as a byproduct in these studies. Moreover, most studies in their review suggest that extreme negative emotions / events are avoided more than positive emotions / events are pursued. That negative events are avoided more than positive events are pursued is for example the case when infants are asked if they want to cross a visual cliff based on their mother's fearful, angry or happy facial expressions who stands on the other side of the visual cliff. Children crossed the cliff fewer times when the facial expression of the mother

3. The Use of Visual Context for Language Processing

was fearful compared to when it was happy. This effect was driven by the fearful expression, i.e., none of the infants crossed in the fearful condition but only 14 of the 19 infants crossed in the happy condition (Sorce, Emde, Campos, & Klinnert, 1985). For this reason among others, Vaish et al. (2008) argue for a negativity bias.

However, the task of crossing a cliff is a negative (and potentially dangerous) event and should be avoided regardless of the facial expression of the person standing on the other side. Hence, the claim that children show a negativity bias like adults based on studies such as Sorce et al. (1985) seems questionable. Vaish et al. (2008) also mention themselves that most of their reviewed studies do not provide intensity ratings. Moreover, the fact that children often interpret neutral emotions as negative does not generally favor the assumption of a negativity bias in children, but rather suggests that neutral emotional facial expressions as well as negative expressions are overall more complex and diverse than positive facial expressions.

This is further supported by Berman, Chambers, and Graham (2016) who measured 3- and 5-year olds' fixations towards happy, sad and neutral facial expressions while listening to happy, sad or neutral sounding but semantically neutral sentences. They found that children could not match neutral expressions with the neutral vocal emotion. However, although they did not find a behavioral difference between positive and negative face-prosody matching, children needed more time to fixate the positive facial expression when hearing a positive voice than fixating a negative facial expression when hearing a negative voice. They already fixated the matching negative face 200 ms after the onset of the sad-sounding vocal affect, but needed 600 ms longer to fixate the happy face in the positive valence condition.

This finding could be interpreted as indicating a negativity bias in children similar to that found in younger adults. Despite this, there is, however, more evidence speaking in favor of a positivity bias in children similar to that observed in older adults. First of all, children understand and use positive emotional information earlier and more accurately than negatively valenced emotions (for a review see Herba and Phillips, 2004). Moreover, when asked to generate as many words as possible for an imagined positive or negative emotional state, children produced fewer words related to negative feelings than to positive ones. Additionally, the older the children were (primary vs. secondary school) the more negative feelings they described (Doost, Moradi, Taghavi, Yule, & Dalgleish, 1999). When 7-9-year old children and adults

3. The Use of Visual Context for Language Processing

rated pictures for valence, arousal and complexity, children rated positive and neutral pictures as more positive than adults. There was however no significant difference in ratings of aversive images between age groups (Cordon, Melinder, Goodman, & Edelstein, 2013). Furthermore, children are not only more accurate, but also faster in identifying positive emotional facial expressions than negative emotions (De Sonneville et al., 2002; Richards, French, Nash, Hadwin, & Donnelly, 2007). Moreover, in a gaze-cueing task, Niedźwiecka and Tomalski (2015) suggested that 12-months old infants oriented their gaze more rapidly to targets cued by happy compared to angry and fearful faces.

In addition, research reported by Todd, Evans, Morris, Lewis, and Taylor (2010) further underlines that children, just like older adults, portray a positivity rather than a negativity bias (for a review see also Todd et al., 2012). In a functional magnetic resonance imaging (fMRI) study they analyzed 3-8-year olds' and young adults' brain responses to emotional facial expressions and found that children but not adults showed greater amygdala responses for happy than angry facial expressions. Adults were between 18 and 33 years old and showed – in line with their negativity bias – greater amygdala activation for angry compared to happy expressions. Children's activation towards angry facial expressions increased with age (Todd et al., 2010).

In conclusion, even though emotion processing, just like face processing is present from a very early age onwards, becoming an expert in a specific domain takes time and practice (for a review on expertise and face processing see Tarr & Cheng, 2003). It seems that the processing, identification and interpretation of emotions is biased with regard to its valence depending on the age of the perceiver. Children presumably start out with a preference towards positive emotional material, maybe because there is less variation and complexity within the range of positively valenced information. The older they get, there more this preference shifts towards the arguably more complex negative emotional information. However, this negativity bias for younger adults seems to change again into a positivity bias and the avoidance of negative emotions in older age.

Regardless of the perceiver's age, emotions are social cues to the understanding of our interlocutor and are hence a means for establishing successful communication. Still, not many studies have addressed the question of how emotions influence the way we understand language in real time and even fewer have taken the emotional

3. The Use of Visual Context for Language Processing

biases of different ages into account. More so, to our knowledge, there are to date no studies focusing on the real-time effect of visual emotional cues on language processing in children.

However, in two off-line studies Ruffman, Slade, Rowlandson, Rumsey, and Garnham (2003) investigated in how far language in 3-5-year old children relates to emotion understanding and specifically focused on the correlation between emotion understanding and syntax and semantics. They tested children's understanding of syntax and semantics using the comprehension of word order and embedded clauses. The sentence stimuli were carefully designed to permit a clear distinction between syntax and semantics. Moreover, Ruffman et al. (2003) asked the same children to identify emotions (happiness, fear, sadness, anger and surprise) from facial expressions. Additionally, children had to perform false belief and desire tasks indicating what they thought the protagonist of a story thought or wanted. Their results suggest a tight coupling between syntax and semantics, as well as between language use and the ability to attribute and infer mental states and desires from oneself to others, i.e., theory of mind (TOM, Premack & Woodruff, 1978). Moreover, they also found a high correlation between emotion recognition and syntax but not semantics. Correlations between syntax and TOM were expected because syntax (e.g., an embedded sentence) is used to describe and understand another character's situational representation. On the other hand, Ruffman et al. (2003) did not expect a correlation between emotion recognition and syntax as according to them emotion understanding is not needed for insights into another character's situational representations.

3.7.2 The Degree of Naturalness in the Depiction of Facial Expressions

Apart from the complexity of different emotional valences and their biases regarding the age of the comprehender, a further aspect that we should consider when examining emotional facial expressions and language is the way the former are used and depicted. Especially in experimental studies, stimuli (e.g., emotional facial expressions) cannot always be natural since they have to be strictly controlled. We must strive to eliminate as many confounds as possible so as to be able to draw clear conclusions based on the obtained results. On the other hand, the less natural our

3. The Use of Visual Context for Language Processing

stimuli are, the less we can claim that our results are also valid in the real world, i.e., outside of the laboratory. Whether or not natural versions of our stimuli would yield the same results as degraded or simplified stimuli is an empirical question (see also Adolphs, 2006 on this matter).

However, we do not know whether, for instance, schematic emotional facial expressions can be used in the same way, to the same extent and with the same time course as natural emotional facial expressions. Yet, see Section 6.4.4.3 for results of the studies we will present and Section 10.5 for further discussion on this matter (taking our results from Section 6.4.4.3 into account).

Rhodes, Brennan, and Carey (1987), as well as Benson and Perrett (1991) for example investigated the effect of caricaturing on the recognition of line drawings and natural face depictions. Both studies found a recognition advantage (in both recognition speed based on reaction time measures and likeness based on participants' subjective ratings) for slightly exaggerated faces compared to the original face. Caricatures accentuate particular details of a face (Brennan, 1985) and thus highlight the most prominent features of an expression, making it easier for participants to recognize the person the face belongs to. Moreover, in studies on human - robot interaction the appearance of the robot plays an important role for self-identification with a robot's face. It is assumed that the more schematic a robot's face is, the more humans can identify with it, as the distinction between one's own representation and that of another person becomes less and less pronounced (the more schematic a face is, the more characters it can represent, Blow, Dautenhahn, Appleby, Nehaniv, & Lee, 2006).

On the other hand, most of the time we interact with other human beings and easily attribute mental states, beliefs and feelings to our interaction partners, based on our own mental states (Premack & Woodruff, 1978). That is, we are much more experienced in our interaction with real and natural emotional faces than with schematic facial representations. However, research has also suggested that schematic (computer-generated) faces are recognized as well as natural facial expressions (e.g., Chang, 2006; Öhman, Lundqvist, & Esteves, 2001; Ruffman, Ng, & Jenkin, 2009).

Nevertheless, when it comes to the difference between static and dynamic natural faces, findings indicate that dynamic natural emotions are recognized faster and more accurately and elicit enhanced and prolonged cortical responses compared with their

3. The Use of Visual Context for Language Processing

natural static counterparts (see e.g., Harwood, Hall, & Shinkfield, 1999 for identification of emotions from moving and static videotaped and photographic displays and Recio, Sommer, & Schacht, 2011 for ERP evidence). Additionally, Sato, Yoshikawa, Kochiyama, and Matsumura (2004) conclude that dynamic in contrast to static emotional facial expressions leads to a facilitation in emotion recognition. They hypothesize that the better recognition of dynamic (vs. static) facial expressions is due to the enhanced activation of brain regions implicated in face perception. Brain regions responsible for face perception were more active when dynamic compared to static faces were presented. Bassili's (1979) study also already suggested that dynamic emotional images were better recognized than their static counterparts, meaning that all emotions displayed indicated more accurate recognition results in the dynamic face condition as opposed to the static face condition. Moreover, their study investigating the role of facial movement demonstrated that even the surface movement of a face serves as information that helps to correctly identify specific emotions (see also Schultz & Pilz, 2009).

However, children face difficulties in identifying emotions and are still in the process of development regarding both emotion and language processing. Hence, displaying stereotypical emotions, e.g., in the form of a smiley, might help them to link emotion to language. Yet, it would also be possible that this stereotypicality confuses or hinders people in making use of a specific cue, as these non-natural cues present only a limited amount of (emotional) information (Gross & Ballif, 1991). Del Giudice and Colle (2007) have for example indicated that both adults and 8-10-year old children found smiles with an open mouth (bared teeth) much more salient than closed smiles. Children judged these strong smiles as more sincere, whereas adults found it easier to judge them as fake smiles.

Hence, even though it seems clear that dynamic facial expressions have a recognition advantage over static faces, there does not seem to be a clear consensus regarding the use of stereotypical schematic versus natural facial expressions in the study of emotion and face processing. In how far and to what extent schematic and natural dynamic facial expressions have an influence on the way emotional language is processed is even less clear.

4. Arguments for the Inclusion of Social Cues into Real-time Situated Language Processing Accounts

In the previous Section (3) it became clear how diverse visual contextual information is and in how far these different kinds of direct and indirect cues are interpreted and used for language processing. Visual context ranges from direct visual cues such as a depicted or real object (e.g., a ball) that can be referenced in speech (e.g., *He kicked the ball!*) to much more indirect social cues such as a facial expression (e.g., a smile). For more indirect social cues, the interpretation in relation to the linguistic input (e.g., *He kicked the ball!*) has to be inferred by the comprehender and likely depends on his or her personal characteristics and world knowledge. In the former case, the linguistic input (*Look, a ball!*) leads to a mental representation of a ball in the mind of the comprehender and can thus elicit a visual search in the real world for the referenced object that matches the mental representation. In the latter case however, the comprehender has to interpret the smile first not just as a sum of specific physical features and movements but s / he may have to infer an underlying meaning from it (but note that this might not always be the case, such as in instances of mimicry where inferences might arguably not be necessary). Depending on whether the comprehender judges the speaker's smile as signaling happiness, annoyance or politeness, the interpretation of the linguistic input likely changes.

In this Section (4) we will now take a detailed look at how we actually link the visual context with our linguistic input. Answers to these questions provide the baseline for the claims a real-time account of language processing has to make to be enriched with richer (social) contextual information. We will further discuss different theories, accounts, frameworks and mechanisms of language processing, so that we will see which claims they make, in how far these approaches differ and crucially which limitations they have regarding the integration of different information types into language processing. The discussions and findings in this Section will hence provide further support to enrich and extend the CIA (Knoeferle & Crocker, 2007) – a suitable existing account of real-time language processing. Before considering our own findings, viz. that both direct and indirect visual cues facilitate language processing depending on the age of the comprehender (Sections 6-9), we will first exemplify in detail how the CIA functions. Having reported our own results in Sections 6-9 we will demonstrate how the CIA could be adapted in order to arrive at a

more detailed understanding of the integration of direct and indirect visual cues into real-time language comprehension (Section 10.8).

4.1 The Interplay between Language Processing and Non-linguistic Visual Context

As Section 2 has shown, the incoming linguistic input is integrated and processed incrementally and in real-time. It can influence the processing of, and be influenced by, these direct and indirect visual cues. A reason for this is that in situated language processing our eye-movements do systematically yet not directly reflect language processing (Huettig, Romners, & Meyer, 2011). In interpreting language-mediated eye-movements, we assume that they are guided by (active) mental representations (Anderson et al., 2011). Hence, eye-movements indicate a shift in attention towards the to be fixated object (Rayner, McConkie, & Ehrlich, 1978; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Moreover, as language comprehension is, at least for adults, a highly routine skill, the eye-movements mediated by language arguably reflect different degrees of automaticity, depending on task demands, the situational context and the incoming linguistic information (Mishra, Olivers, & Huettig, 2013). For adults, even in complex visual scenes it only takes a few hundred milliseconds to launch an eye movement to a depicted referent after the linguistic input has been processed (Allopenna, Magnuson, & Tanenhaus, 1998; Andersson et al., 2011). In turn, the visual context also rapidly affects language processing and comprehension (e.g., Knoeferle et al., 2005; Knoeferle et al., 2008).

Recall that Knoeferle et al. (2005) for example showed that visually depicted event relations realized through depicted actions between agent and patient characters have an influence on how linguistic input is processed and interpreted on-line. In this case eye-movements towards objects occur before the linguistic input has referred to these objects, i.e., anticipatory eye-movements can be observed. Recall also that Knoeferle and Crocker (2006) demonstrated that not only visual and linguistic information is used to form an event interpretation but that moreover also real-world knowledge plays an important role. We hence generate expectations given our world knowledge on the basis of visual and linguistic contextual information and shift our attention to (visual) objects that are in line with these expectancies (Chambers, Tanenhaus, & Magnuson, 2004; McRae et al., 2005).

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

Moreover, not only adults as language experts, but also children and even very young infants direct their eye-movements in response to linguistic and visual input (e.g., Bergelson & Swingley, 2012). Children's eye movement behavior resembles that of adults' in that they launch an eye-movement to the corresponding visual target immediately (about 150-200 ms, Allopenna et al., 1998; Tanenhaus et al., 1995) after hearing the visual referent's name (Snedeker & Trueswell, 2004; Trueswell et al., 1999). Recall also that they can moreover already predict upcoming linguistic input given a constraining visual display, e.g., looking at a visual display of one edible object among distractor objects upon hearing *The boy will eat the...* elicits anticipatory eye-movements at the verb region towards the edible object before this object is mentioned (e.g., Mani & Huettig, 2012; Nation et al., 2003).

That the processing of language and visual context information is tightly coupled in both adults and children is also supported by ERP evidence. Showing adults and 19-months old infants pictures of objects that were either named congruently or incongruently, Friedrich and Friederici's (2004) results indicated a similar N400 effect for incongruous (vs. congruous) object names in infants compared to adults (however, topography and latency differed). Hence, already 19-months old infants show the beginnings of adult-like mechanisms regarding semantic interpretation.

However, simply measuring people's eye-movements while they inspect a given scene and listen to language does not allow us to draw conclusions about the underlying nature of language processing. In order to do so, we need to specify a so-called linking hypothesis that describes the relationship between eye-movements and linguistic processes in more detail, i.e., we set up an assumption about the relationship between our eye-tracking data and the underlying cognitive processes (Knoeferle, 2015a) depending on the type of hypothesis we want to investigate. Our interrogation of the visual environment is highly sensitive to pragmatic, semantic, and referential processes, and even to the interpretation of abstract language. As a result, eye gaze can provide a very useful tool for investigating visually situated language comprehension (Knoeferle, 2015a). Making use of such a linking hypothesis should therefore answer questions about when, how and why a shift in visual attention towards visual contextual information occurs and what the nature of its relationship to the unfolding linguistic input is (Salverda et al., 2011).

Nevertheless, investigating how and when language processing (of different linguistic structures), and the visual context (i.e., the perception of different visual

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

contextual cues) interact and influence each other in real time only provides the basis for developing a real-time language processing account that will eventually model how language comprehension functions.

Even though many psycholinguistic processing accounts, models and frameworks have been developed over the past three decades, not all of them consider language processing in real time. Moreover, even those that look at on-line language processing are mostly language-centric (see Knoeferle, 2015a for a related argument) or only include the influence and impact of direct referential non-linguistic visual cues.

However, as Section 3 has argued, visual contextual cues are highly diverse and do not just comprise visual contextual information such as depicted characters and actions. Even though we are aware of the impact that our social environment and the indirect social visual cues (such as emotional facial expressions) have on the way we process linguistic input, to date, real-time language processing accounts remain largely elusive regarding the integration of these cues.

This is however essential when the goal is to develop an account that can explain how language processing with all its associated real-world influences happens in real time in the mind of the comprehender. One way this could be achieved will be exemplified in Section 10.8.1. Using our own findings, we will demonstrate how the CIA could be extended to also include more indirect visual social cues, as well as listener characteristics such as age (Knoeferle & Crocker, 2006; 2007). The following Section (4.2) will therefore present the most prominent language processing accounts, frameworks and models and point out their limitations. We will further argue for an adaptation of real-time language processing accounts regarding the integration of social (visual) information (4.3) based not only on the findings discussed in Section 3, but also from a developmental angle (4.3.1). Following our argument that language processing and social information are tightly linked from the beginning of our lives, we will discuss the CIA as a suitable candidate to be enriched with the influence of social cues and listener characteristics (4.4).

4.2 Language Processing Accounts and their Limitations

Up to this point no real-time language processing account has taken into account the impact that social factors (such as discussed in Section 3) or listener characteristics such as age might have on our language processing mechanisms. This is true for

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

constraint-based models of language processing (cf., MacDonald, 1994), for situation models (cf., Zwaan, 2014) for expectancy generation models (cf., McRae et al., 2005), for situated theories of cognition (cf., Myachykov, Scheepers, Fischer, & Kessler, 2013), and for processing accounts of situated language comprehension (CIA, Knoeferle & Crocker, 2006, 2007). However, if language is a means to maintain social behavior (Semin, 1995), and the knowledge we have about social interaction is inseparably linked to language (Semin & Fiedler, 1992), what we need is an interface between, or an integration of, accounts on language processing and accounts on processing social information.

Theories of language comprehension and communication have contributed insights into how context affects language processing. The Communication Adaptation Theory (CAT) by Giles, Coupland, and Coupland (1991) focuses on the adaptation of interlocutors, e.g., through speech, vocal patterns, gestures, accents, but also through social norms and social situations. According to CAT, people converge and diverge with their communicative partners. The more the partners' behavior and speech patterns resemble each other, the more they converge and sympathize.

Likewise, the linguistic style matching (LSM) that Niederhoffer and Pennebaker (2002) propose, argues that the use of words of a speaker unconsciously primes the response, and thus the use of words, of the listener. This unconscious priming implies, according to Niederhoffer and Pennebaker (2002), that conversation partners who match their linguistic styles organize their minds psychologically in a similar manner.

Even though CAT relies heavily on social factors and we unintentionally seem to match our linguistic styles when we interact with an interlocutor, little attention has gone to the effects of the visual context (e.g., the visual features hinting at social status, gender, race, emotional expression etc.) on language processing. Moreover, CAT and LMS rather focus on how communication works as a social phenomenon but much less on the underlying language processing mechanisms. The indexical hypothesis by Glenberg and Robertson (1999) already highlights the importance of experiential components (which imply social constructs, beliefs and experiences) and their indexical nature for language comprehension but do not make any claims regarding the time course of the integration of this kind of information into language processing.

Situation models (Zwaan, 2014) and expectancy generation models (e.g., McRae et al., 2005) and models of interactive alignment in dialog (Garrod & Pickering, 2009;

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

Pickering & Garrod, 2009) take the world knowledge, the situational (visual and non-visual, linguistic and non-linguistic) context, the expectancies we generate, our experiences, and the behavior of our interaction partner into account, when incrementally integrating and processing information. Also, subsequent input can change how we interpret previous input. Information processing in these latter accounts occurs on-line and can be updated at the same time (see also Anderson et al., 2011).

Although these approaches make use of rich social and situational cues, they miss out on assumptions about how language processing happens in detail, i.e., which information is processed when during language comprehension and how exactly these different cues affect the comprehension and interpretation of an utterance in real time. However, a situation model as discussed by Zwaan (2014) might be a suitable higher-level environment to accommodate a more specific account on language processing like the CIA. This issue will be addressed further in Section 10.8.1.1.

Vice versa, most serial language comprehension accounts (e.g., Frazier & Fodor, 1978; Friederici, 2002) include little contextual information and delay context effects to relatively late processing ‘stages’. Other theories foreground a rapid interaction between syntactic and non-syntactic information sources (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). In these probabilistic theories the input is analyzed and ranked according to probabilistic constraints, which are competing with each other. However, constraint-based accounts only hold when a linguistic competitor is present. Extralinguistic information is not directly linked to the linguistic input and the interpretation of a sentence cannot be updated. Rather, these accounts model the competition between two linguistic interpretations and not the way interpretations are built incrementally. Yet, one could argue that the interpretation of a sentence can be updated through the competing constraints. How much support a specific interpretation receives during sentence processing can be updated as a function of the constraints that become available at each word. Nevertheless, they do not explicitly include a rich non-linguistic context that could provide social information. Furthermore, even though they might be extended to include social and visual information, they do not directly model language comprehension in real-time. Thus a real-time language processing account arguably provides a more suitable basis for the enrichment with direct and indirect visual cues.

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

Recall that Tanenhaus et al. (1995) investigated the rapid interaction of available structural and visual constraints on language processing. Using eye tracking, they showed that people shift their gaze towards an object immediately after its mention, suggesting incrementality of comprehension. Additionally, the visual context disambiguated local structural ambiguity in the linguistic input. These results support a theory of language comprehension that combines extralinguistic and linguistic information in real-time comprehension (see also Novick, Thompson-Schill, & Trueswell, 2008).

However, as mentioned, existing approaches to real-time language processing have not yet considered social cues. There are, however, accounts (Knoeferle & Crocker, 2006, 2007) and computational models (Crocker, Knoeferle, & Mayberry, 2010) that include at least actions and events in addition to objects as visual contexts (but no indirect social cues of the kinds we discussed in Section 3.4). Visual cues can help listeners to rapidly disambiguate sentences and assign thematic roles (see also Altmann & Mirković, 2009). Still, all these approaches fail to include *social* aspects into the processing cycle.

4.3 The Importance of Including Social Cues in Real-time Language Processing Accounts

Nevertheless, language is always situated in the social and socio-cultural context. Already in 1972, Průcha, just like Labov (1972), argued that studying language must also involve studying the language user and the social context. However, unlike Labov who focused on society as a whole, Průcha postulates that among other factors the IQ, the temperament, the character, the emotional state and the verbal memory of the encoder shape the use of language unconsciously. Social factors and listener characteristics such as age, sex, education, status and ethnicity arguably also contribute to how exactly the producer forms his utterance. Moreover, he assumes that the producer's attitude towards his interlocutor, the social situation they are in, and the intended meaning, will shape the message. Although Průcha (1972) mainly focused on the speaker, the above-mentioned factors are also crucial in the process of comprehending a message. How an utterance is processed is not just a matter of decoding the linguistic input, but it might depend also on the above-mentioned social factors and listener characteristics. That said, we should not only consider modeling

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

language comprehension per se, but also take social factors and listener characteristics such as social status, race, age or emotional facial expression for language processing into account. Note that this thesis focuses on the influence of visual contextual cues on language processing. We will outline how an account of real-time language processing can be adapted to include social information and listener characteristics focusing on visual contextual information in Section 10.8.1.1. Nevertheless, we acknowledge that much social information is non-visual and hence argue for the enrichment of real-time language processing accounts with more indirect social information, regardless of the information being visual or not in Section 4.3 and 4.4. See Section 11 for a discussion on the integration of non-visual social information into our proposed adaptation (Section 10.8.1.1).

Apart from the discussed indirect visual information (Section 3.4 and 3.5), some acoustic features of the speaker can affect speech perception. A study by Martin, Mullennix, Pisoni, and Summers (1989) demonstrated that listeners remembered fewer words in a recall task when there were multiple speakers compared to single-talker lists of words. Additionally, speech sounds were slower and less accurately identified when talkers varied randomly compared to a single talker (Green, Tomiak, & Kuhl, 1997). Moreover, already in 1977 Geiselman and Bellezza (1977) suggested that sentences read out by male (vs. female) speakers were judged to be more “potent”. Walker and Hay (2011) furthermore demonstrated that lexical access happens faster and more accurate when ‘word age’ and voice age matched (vs. mismatched). In an auditory lexical decision task, they presented participants with words predominantly used by older and younger people respectively, spoken by either an older or younger voice. ‘Word age’ and voice age could thus either match or mismatch. The voice of the speaker acted as a social cue for the participant and facilitated lexical access, i.e., shorter reaction times of the particular exemplar due to the match with the word’s distribution. Thus social knowledge together with speech has an impact on the interpretation of an utterance.

A further claim comes from the sociophonetic side. The claim is that linguistic representations are closely linked to / contain social information. Since linguistic representations are connected to social knowledge, our representations of linguistic ‘variants’ should also contain social knowledge. If our linguistic representations also contain social knowledge it in turn means, that our language processing mechanism could be influenced by the social information linked to / contained in our linguistic

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

representations (Foulkes & Docherty, 2006; Hay, Warren & Drager, 2006; Johnson, 2006). Recall that Squires (2013) used this approach and tested sociolinguistic links between subject-verb agreement and the inferred visual social status of the speaker of an utterance. Both her experiments combined “insights from sociolinguistic research on speech perception with psycholinguistic research on sentence processing” (Squires, 2013, p. 204) using a method she labeled ‘sociolinguistic priming’ (see Section 3.4).

According to De Bot (2000) understanding how exactly social factors impact different levels of language processing is a first step towards a universal theory of language use. While social cues play a role in language use, not many studies have focused on how they interact with incremental language processing and production (including the implicated mechanisms, see Section 3.4).

Two approaches in the psychology of language (Clark, 1996; Pickering & Garrod, 2004) have undertaken a more fine-grained embedding of language within the social context. Although these approaches focus on grounding communication within the shared knowledge of the interlocutors and on alignment during conversation (e.g., through actions), they fail to consider social aspects. By contrast, Mehl and Pennebaker (2003) focus on the effect of social cues on language, such as how the gender of interlocutors and their social context (student life) respectively affect language use.

That said, none of the above approaches considers language processing across time, and thus inferences about how and when social aspects influence our use and comprehension of language remain elusive. Keysar, Barr, Balin, and Brauner (2000) and Keysar, Lin, and Barr (2003) for example measured participants’ eye-movements in a dialogue setting; they did not, however, focus on social cues of the interlocutors’ interpretation but instead focused on common ground (much like Clark, 1996 or Pickering and Garrod, 2004).

To investigate the real-time interplay of social and linguistic processes we must thus enrich research on real-time language processing with social information. First support for an account that combines aspects that could be argued to be social and the processing of language in real time comes from recent ERP studies about the integration of (iconic) gestures into language processing. Recall that Kelly et al. (2004) have suggested that the processing of iconic hand gestures and speech in naturalistic context are closely connected even at early processing stages, and that a

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

match between hand gestures and linguistic content has a facilitatory effect on language processing. Additionally, recall also that Wu and Coulson (2005) provide strong EEG evidence for an integration of iconic gestures and speech. Moreover, Wu and Coulson (2007) extend their previous findings from the word level to discourse comprehension, showing that congruent iconic gestures elicited smaller N300 and N400 ERPs than incongruent or no gestures. These results suggest that comprehenders make use of the speakers' gestures (congruent with the linguistic input) in order to arrive at a better understanding of his utterance than when no or incongruent gestures were seen (see the end of Section 3.3 also for a note on gestures as indirect (social) cues). Indeed, continuous measures such as eye tracking and the recording of event-related brain potentials during the comprehension of spoken sentences in rich visual contexts, lay the foundation for examining social effects on real-time language processing.

4.3.1 Developmental Arguments for the Inclusion of Social Cues into Real-time Language Processing Accounts

More support for the idea of the integration of social aspects into a language processing account comes from children studies suggesting that social conventions affect vocal tract properties. Before children reach puberty, females have higher formant frequencies than males even though there is not yet a difference in f_0 (Perry, Ohde, & Ashmead, 2001). These higher formant frequencies may be the consequence for learned, i.e., conventionalized, gender-typical dialects that generate different speech patterns between males and females even before puberty affects f_0 (Creel & Bregman, 2011).

Another argument for integrating social aspects into accounts of real-time language processing comes from developmental studies of learning. Already neonates react differently to the voice of a stranger than to their own mother's voice (DeCasper & Fifer, 1980). Furthermore, in a preferential looking study, Kinzler, Dupoux, and Spelke (2007) demonstrated that infants preferred looking at faces associated with their native language compared to faces associated with accented (but comprehensive) speech or a foreign language. Their results hence indicate that familiar acoustic speech patterns are used for the social evaluation of the speaker in language processing. Furthermore, the developmental aspect strengthens the argument that

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

language processing and the processing of social information should be integrated together into an account of language processing. That is, social information seems to have an impact on the evaluation of the linguistic input, which in turn influences the way we build our social networks. Hence, social cues and language processing are inseparably linked and should not be considered separately without taking their mutual influence into account. Although these studies mainly focus on the integration of talker identification through speech sounds into language processing, this could be extended to other listener characteristics not just of the speaker but also of the comprehender, such as age and other non-visual social cues such as emotional prosody.

Yurovsky, Wu, Yu, Kirkham, and Smith (2011) for example suggested that infants as young as eight months of age learned faster from a visual social (facial) cue than from a non-social cue (square) in an attentional cueing task on audio-visual learning. Moreover, Yu and Ballard (2007) argue for an integration of indirect (social) cues, such as gaze direction, gesture and facial expression into statistical learning models⁹. They showed that in an analysis of CHILDES video clips of mother-infant interactions, a model including social cues outperformed the results of the purely language-driven statistical learning method when word-meaning associations were computed.

More support for a processing account that includes both linguistic and social information comes from emergentist and usage-based perspectives (see also Section 2.3). Recall that these perspectives assume that linguistic representations are formed using an individual's experience of each representation together with the linguistic mechanisms that constrain its usage. Moreover, language can be learned from language use which is itself dependent on social skills like joint attention and the ability to understand others as intentional agents (Behrens, 2009; Tomasello & Rakoczy, 2003). Additionally, emergentist and usage-based perspectives assume powerful generalization mechanisms, meaning that children quickly generalize new structures using previous experiences (O'Grady, 2005). Hence, emergentism is seen as essential for language acquisition. It is the interaction with language and the communicative intention behind language that eventually emerges into competent linguistic knowledge (Behrens, 2009). This view is similar to exemplar-based

⁹ Statistical learning in language acquisition refers to the assumption that children can quickly discover and learn from patterns in the underlying structure of the language system (Saffran, 2003).

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

perspectives in that formed exemplars then contribute to the linguistic competence and can change slightly with each new experience gained regarding the specific exemplar (Abbot-Smith & Tomasello, 2006). Writing about the acquisition and use of social knowledge, Wyer (1980) states that social knowledge is inseparable from world knowledge and experience. Incoming input, no matter from which modality, needs some kind of context or previously acquired knowledge to be interpreted. Learning in isolation is extremely difficult as we use contextual information and previously acquired knowledge to interpret the world meaningfully (Wyer, 1980). The importance of context, experience and knowledge thus suggests that language cannot be separated from other cognitive abilities, which we, just like language, use to understand language and the world around us (Bybee, 2006).

In line with this is the view of Abbot-Smith and Tomasello (2006) who propose a hybrid categorization model to account for syntactic acquisition. Their model combines dynamic exemplar learning with a permanent and static schema abstraction, which can be frequently accessed to build new exemplar categories. Learning linguistic representations is thus highly dependent on the (non-) linguistic, social, situational and contextual environment in which the linguistic representation is encountered, as it is the communicative function that is underlying language acquisition and use (Tomasello, 2000). This view is supported by empirical evidence from a number of studies showing that children of different ages are able to produce SVO sentences using a novel verb which they only heard previously with a different syntactic structure (e.g., Akhtar & Tomasello, 1997; Olguin & Tomasello, 1993; Tomasello & Brooks, 1998, see also Section 2.3).

Additionally, Lavie (2005) proposed the ‘Analogical Speaker’ which applies the exemplarist view to syntax, stating that children do not learn a specific language, but rather learn how to use speech. In his model, he stresses the dynamics of linguistics and the progressive generalization a child constantly has to perform in order to achieve adult speaker competence. However, this cannot be done in a single instant but new syntactic structures emerge slowly with intermediate steps before the structure can be produced correctly (Lavie, 2005).

As is evident from Section 2, children only gradually learn to comprehend challenging syntactic structures. However, recall that Zhang and Knoeferle (2012) showed that children can integrate visual contextual information in the form of depicted actions into their sentence processing to boost comprehension of these

challenging structures. Weighall and Altmann (2011) additionally asked 6-8-year old children to look at a display and listen to sentences with an embedded structure of varying complexity. The scenes either showed the depicted characters engaged in an action or not, i.e., they were passive. Just as in Zhang and Knoeferle (2012) their offline comprehension task showed that children performed better in the task when the characters were performing an action and hence can make use of extralinguistic cues to form event representations.

Language processing accounts should therefore integrate social cues and listener characteristics as we are using both direct and indirect cues to incrementally update our interpretation of utterances. Moreover, we seem to do so from a very early age onwards. Next, we will introduce the Coordinated Interplay Account, which posits a suitable basis as it already includes the visual context as a means for updating utterance interpretation. We will outline how the CIA could be adapted to accommodate visual social aspects and listener characteristics used in the present studies, i.e., emotional facial expressions and listener age, in Section 10.8.1.

4.4 A Suitable Language Processing Account for the Enrichment with Social Cues – The CIA

Psycholinguists and Sociolinguists (cf., Gaskell, 2008; Labov, 1972; Van Berkum et al., 2008) have argued for examining language processing in the context of social information. As the studies reviewed above (see Sections 3.4 and 3.5) suggest, all kinds of social information seem to contribute towards meaningfully interpreting our environment and language. Carminati and Knoeferle (2013) showed that visual social information, just as linguistic information, is rapidly integrated into real-time language comprehension. Their results, and also the results of Staum Casasanto (2008), Squires (2013) and Kreysa et al., (2014) cannot be interpreted using a two-step model of sentence interpretation, in which syntactic processing precedes semantic processing (e.g., Friederici, 2002). Van Berkum et al. (2008) already propose “a one-step model in which knowledge about the speaker is brought to bear immediately by the same fast-acting brain system that combines the meanings of individual words into a larger whole” (p. 581). This view also fits nicely with the emergentist perspective and the studies about language processing and learning

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

reviewed above (see Sections 2.3 and 4.3.1), as it suggests for example that emotional facial expressions and linguistic information rapidly interact (see section 3.4).

Constrained-based models such as the one by Novick et al. (2008) are insufficient in their present version to accommodate social effects because they exclude social cues and the same is true for other processing accounts, (e.g., Knoeferle & Crocker, 2006, 2007). However, we can enrich existing accounts of language processing with social information and listener characteristics, for instance with a speaker's current emotional facial expression and the comprehender's age. This will be exemplified in detail in Section 10.8.1.1 taking the results of the present studies (Sections 6-9) into account.

One suitable account for such an extension is indeed the Coordinated Interplay Account (CIA) by Knoeferle and Crocker (2006, Figure 1: 2007). The CIA accommodates the interplay of language comprehension with visual attention (Knoeferle & Crocker, 2006) and with representations in working memory (Knoeferle & Crocker, 2007). It supports the assumption that perception and action do not follow sequentially one after another, but that they rather represent a continuous cycle. The representation of a continuous cycle means that linguistic and visual input streams are continuously-flowing and partially overlapping (Anderson et al., 2011). Moreover, it already includes extralinguistic visual context (e.g., objects and events) and is based on language processing in real time. It consists of three temporally dependent processing steps (the steps are presented serially below but the account assumes they can overlap and occur in parallel, Figure 1):

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

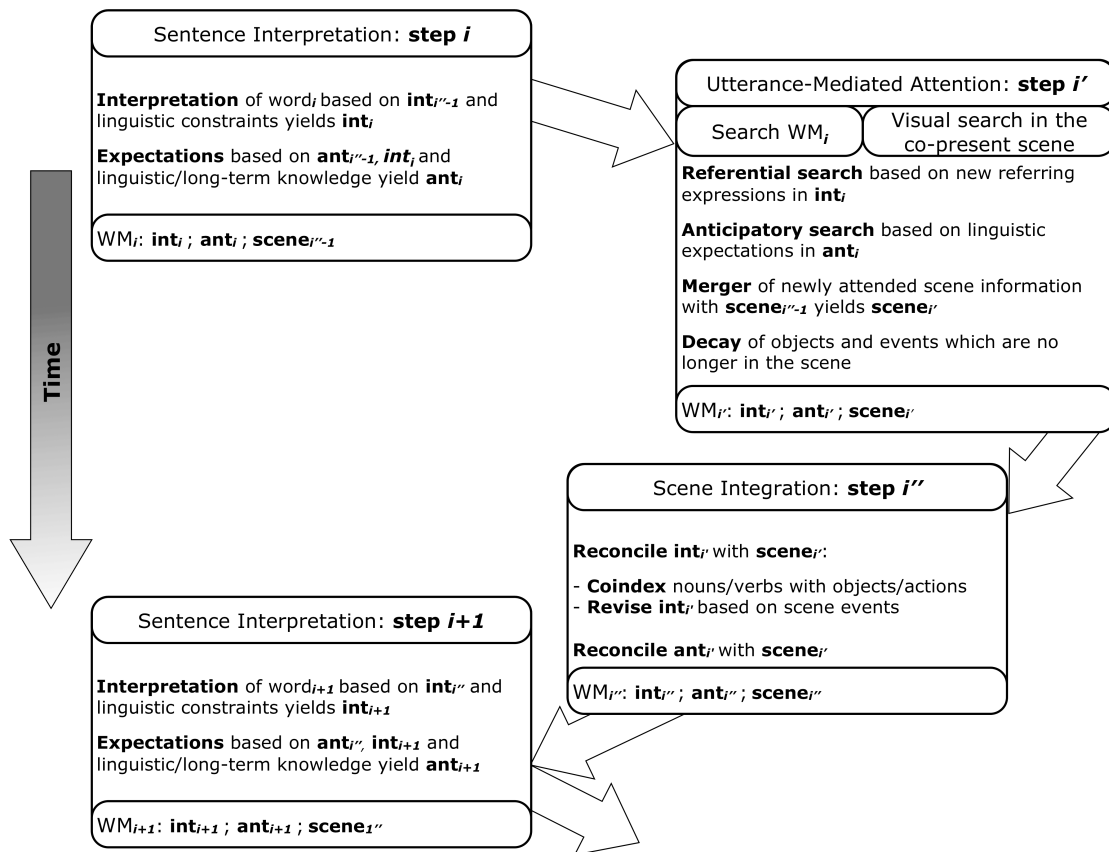


Figure 1: The revised Coordinated Interplay Account (Knoeflerle & Crocker, 2007).

Step i (Sentence interpretation) deals with the incremental interpretation of the incoming linguistic input. Here, the input, i.e., a word (word $_i$), is interpreted on the basis of an existing linguistic interpretation of this word and with respect to its linguistic constraints. This yields interpretation int_i of the word. Moreover, using int_i , linguistic and long-term knowledge as well as previously established expectations of the linguistic input forms expectations (ant_i). Furthermore, the working memory (WM) of this model comprises both scene- and utterance-based representations for each processing step of word $_i$.

In step i' of the account, attention is mediated by the linguistic input. WM $_i$ and the co-present scene are searched. Both a referential and an anticipatory search are performed. The referential search is based on new referring expressions in int_i , and the anticipatory search is based on the linguistic expectations that were set up in ant_i in step i . A merger additionally combines the information from the newly attended scene with the scene held in WM ($scene_{i-1}$). This merging process yields scene i' . Furthermore, events and objects no longer present in the scene decay.

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

In the next step (step i'') the interpretation int_i and the expectation ant_i are reconciled with the scene $_i$. Regarding int_i , this is done via coindexing of nouns / verbs with depicted objects / actions in the scene. Additionally, revision of the word interpretation int_i is done based on the events in the scene. After step i'' the processes start again when encountering the next word, taking the previous expectations and interpretations – also regarding the WM – into account (Knoeferle & Crocker, 2007).

To illustrate this for an adult listener on an example sentence and an example scene, imagine looking at a display depicting a cat, a ladybug and mouse (see Figure 2, see also Section 5.1). The cat is holding a feather in its paw ready to tickle the ladybug and the mouse is holding handcuffs ready to arrest the ladybug. Hence, the scene portrays two potential agents acting upon a patient.

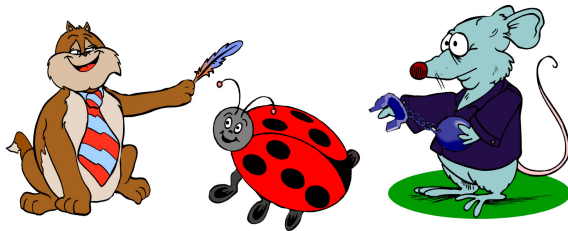


Figure 2: Two potential agents acting upon a patient.

While looking at the display, imagine listening to a sentence describing the scene in Figure 2, such as sentence ((1), see also Section 2.1)

(1) *Den Marienkäfer kitzelt der Kater*

Transl.: ‘The ladybug tickles the cat’

The (determiner, masculine, accusative case) ladybug (object) tickles (transitive verb) the (determiner, masculine, nominative case) cat (subject).

The CIA assumes that in step i , when the first word, i.e., the determiner, is heard, prior linguistic knowledge about the word *Den* and linguistic constraints are activated. In this case, these constraints indicate that the determiner is masculine and in the accusative case. The case-marking constrains the way the upcoming sentence is structured. Moreover, expectations ($ant_i''_{-1}$) based on int_i , i.e., the linguistic constraints

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

and prior linguistic experience with *Den* together with linguistic and long-term knowledge are used to refine expectations for what is coming next (ant_i). This means that upon encountering *Den* an adult listener expects the upcoming input to be syntactically congruent with the heard determiner. In this case, that a masculine noun is expected and that this noun, together with the determiner will function as an object or a patient. Moreover, the adult listener can already expect the upcoming sentence to be in OVS sentence order. All these representations are assumed to be stored in WM together with representations of the inspected scene. In step i', the linguistic input mediates attention in the scene is. WM and the visual scene are searched referentially and anticipatorily. The referential search is based on the interpretation of *Den* (int_i), while the anticipatory search is based on linguistic expectations in ant_i . This means that upon hearing *Den* the scene is searched for a reference object. However, as *Den* is a determiner, no reference with an entity in the scene can be established. Nevertheless, due to the information stored in ant_i (of an object and patient), the anticipatory search might result in fixating the ladybug prior to its mention, marking it preliminarily as the most likely character to be the carrier of the heard determiner and the object of the sentence. Moreover, the newly attended scene information is merged together with the scene information held in WM ($scene_{i-1}$). No information will decay, as in this case all the events and objects in the scene are visible throughout sentence presentation.

In the final step of the account (step i'') the linguistic interpretation int_i is reconciled with $scene_i$. In this case no nouns and verbs have been heard, yet therefore no coindexing is performed. However, the interpretation of the determiner (int_i) is revised based on the attended information in the visual scene. Moreover, the expectations (ant_i) are reconciled with the scene ($scene_i$). After this, the listener encounters the next word in the utterance and it enters the processing cycle based on the interpretations, constraints and expectations of the previous input (step i+1). Hence, upon hearing *Marienkäfer* ('ladybug') the word is interpreted based on the previously heard determiner and the linguistic constraints that both the determiner and the new word set, yielding int_{i+1} . Additionally, expectations are raised based on int_{i+1} , the expectations that were raised in step i'' (ant_i'') and the linguistic / long-term knowledge, yielding ant_{i+1} . Moreover, again all these updated representation are stored in WM. Next, similar to step i' an anticipatory search based on the previously obtained and stored information is performed. However, this time as the input is a

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

noun, a referential search can also be performed, yielding and maybe confirming looks to the ladybug in the scene. Again, old and new scene information are merged. In step $i+1$ (not in Figure 1 anymore) the interpretation of the linguistic input (int_{i+1}) is reconciled with the scene $_{i+1}$. This time, the now complete noun phrase ('the ladybug') can be coindexed with the depicted character corresponding to the ladybug. Moreover, int_{i+1} is revised based on the scene events and expectations (ant_{i+1}) are reconciled with the scene.

The next processing cycle starts when the next word (word $i+2$) is encountered. When hearing the verb *kitzelt* ('tickles') the referential search based on the verb interpretation (int_{i+2}) will probably lead the listener to search for (step $i+2$) and fixate (step $i''+2$) the object of the tickling action in the scene. The verb is used to coindex the object, i.e., the feather, with which the action denoted by the verb can best be performed (in contrast to the also depicted handcuffs). Moreover, the interpretation of *kitzelt* (int_{i+2}) is revised based on the events in the scene, i.e., that in this case, tickling is depicted as performed with a feather. Additionally to this referential search, an anticipatory search based on the linguistic expectations in ant_{i+2} is also performed in step $i+2$. This likely leads to anticipatory fixations towards the cat (vs. the mouse), as this is the character holding the action object – which has been identified in the referential search – denoted by the verb. Old and new scene information are again merged in step $i+2$ and furthermore, the linguistic expectations (ant_{i+2}) and the newly interpreted scene (scene $_{i+2}$) are reconciled in step $i''+2$.

For the adult comprehender this means that s / he has not only established reference to the characters and objects in the scene, but has moreover also already identified the correct agent of the action denoted by the verb even though this agent (the sentential subject, i.e., the cat) has not been mentioned yet. Using the events and objects depicted in the scene, sentence processing is facilitated. The processing steps for the last noun phrase of the example sentence (1), i.e., *der Kater* ('the cat'), will not be outlined here.

This example illustrates the emphasis of the CIA as an account of real-time language processing on the close (temporal) coupling between auditorily presented linguistic input and visual context information. It should be stressed again however, that even though the depiction and steps of the account are presented in serial order, the CIA assumes that the outlined processes may very well partially overlap and occur in parallel (Knoeferle & Crocker, 2007).

4. Inclusion of Social Cues into Real-Time Situated Language Processing Accounts

Nevertheless, the account does not yet take characteristics of the comprehender, such as age or the integration of social cues that must be interpreted by the comprehender, such as emotional facial expressions into account. It provides however a very suitable basis as it is one of the few accounts that model situated processes of language comprehension. It already takes extralinguistic visual cues such as depicted actions and events into account and thus explicitly models how utterance-based representations are related to scene-based representations. The CIA should hence be adapted and enriched to also accommodate social factors and listener characteristics that have been shown to clearly play a role in language processing (see Sections 3 and our own results in Sections 6-9).

Yet, before we will present this extension of the CIA in more detail, we will demonstrate the influence of direct visual (depicted actions) and indirect social visual (emotional facial expression) cues on real-time language processing and how this influence is mediated by listener characteristics (age) in our own studies (Sections 5-9).

5. Motivation and Research Questions

As Sections 2 and 3 have demonstrated, processing linguistic utterances is influenced not only by direct referential cues such as depicted objects or actions (Knoeferle et al., 2005), but also by visual social factors such as the appearance (Staum Casasanto, 2008), or the emotional facial expression of the speaker (Carminati & Knoeferle, 2013). However, psycholinguistic studies thus far have either focused only on social influences on language processing or on direct referential (visual) cues and in how far they can be used for language processing. Moreover, studies taking visual social factors into account when investigating language processing in real-time are rare (see Section 3.4). Up to date it remains elusive in how far direct (e.g., depicted actions) and indirect visual (e.g., emotional facial expressions) cues can be used together for language processing, in how far they might interact with each other, and in how far the time course of the integration of these different cues types differs depending on their nature (e.g., a schematic smiley vs. a video of a real face) and depending on the characteristics of the comprehender. Furthermore, real-time language processing accounts have thus far failed to explicitly include indirect social cues. Taking these cues into account is however essential if the overarching goal is to come to a fuller understanding of how language processing works in the mind of the comprehender (see Section 4).

The present studies hence present a first step towards answering these questions. More specifically, six visual-world eye-tracking studies investigated the effect of depicted actions as direct cues and positive emotional facial expressions as indirect cues on the real-time processing of positively valenced German OVS sentences. To also investigate how the characteristics (age) of the listener affects his / her ability to integrate these different cues into sentence processing, we tested different age groups i.e., small children, younger and older adults. Apart from investigating in detail if cue effects vary in timing depending on the age of the comprehender, testing different age groups using the same stimulus materials moreover excludes the possibility that differences in the stimulus sets account for time course differences between different age groups (Knoeferle, 2015b).

We chose children as one participant group, because they are already sensitive to direct referential cues and can use referencing from words to objects for language learning (Smith & Yu, 2008; Yu & Smith, 2011). Moreover, German children until

about 6 years of age have not yet acquired a full understanding of OVS sentences in the absence of visual information that might aid in sentence processing (Dittmar et al., 2008a). Hence, evidence of a close temporal coordination between different kinds of visual cues (direct and indirect), hearing the specific cue referenced in or drawing inferences about it from the linguistic input and inspecting the corresponding part of a visual display would inform not only language processing research, but would also contribute to accounts of language learning (Knoeferle, 2015b). Furthermore, children seem to differ from younger adults in their valence preference for emotional facial expressions (i.e., children seem to prefer positive expressions whereas younger adults show a negativity preference, cf., Isaacowitz et al., 2009; Todd et al., 2010, see Section 3.7.1.3, but see also Berman et al., 2016 for contrasting results on children's emotional preference) and will thus further contribute to the discussion about the role of social cues in real-time language processing.

We chose older adults because they also show a preference towards positive emotional material and away from negative stimuli, in contrast to younger adults (Isaacowitz et al., 2009). Hence age differences in the perception of social visual cues could influence the processing of sentences and the time course of visual context effects. Moreover, up to date real-time language processing research has not yet considered possible timing differences in visual context effects – be it regarding direct referential or indirect and more social cues – between younger and older adults.

5.1 Design and Conditions

In all six experiments, participants looked at a visual display showing three characters (see Figures 2 and 3) while listening to a related positively valenced German OVS sentence. Two of these characters were potential agents acting upon a patient. In half of the trials participants were primed with a positive emotional facial expression prior to encountering the scene. In the other half of the trials we presented participants with a non-emotional (a star, experiments 1, 2 and 4, see Figure 4) or emotionally incongruent prime (a negatively valenced human face that mismatched both sentence valence and the valence of the target agent, experiments 3, 5 and 6, see Figure 5). Moreover, the potential agents on the screen (the cat and the mouse in Figure 1) were depicted as either performing an action towards the patient (see Figure 4 or 5 conditions a) and c)) or as not performing any actions (see Figures 4 or 5, conditions

b) and d)). Participants listened to the sentence describing one of the actions performed towards the patient while looking at the display. Following this scene, an identical scene without depicted actions (regardless of condition) was presented and participants were asked to orally answer a comprehension question for ‘who is doing what to whom’ after each trial (Figure 3). In studies 1, 2 and 4 we posed the comprehension questions in the active voice in half of the trials and in the passive voice in the other half of the trials. In studies 3, 5 and 6 we posed the comprehension questions for the experimental items only in the passive voice.

We chose depicted actions as direct cues because research has suggested that even small children can reliably use these visual cues to facilitate sentence processing (Zhang & Knoeferle, 2012). We chose emotional facial expressions as indirect visual social cues because their processing varies depending on the age of the comprehender (Isaacowitz et al., 2006; Todd et al., 2012). In addition, the valence of the facial expression rapidly affects the semantic interpretation of emotionally valenced SVO sentences (Carminati & Knoeferle, 2013; Münster et al., 2014, Münster et al., 2015). We used positive emotional facial expressions in the present studies (i.e., happy faces) as critical primes and all critical sentences had a positive emotional valence. We chose the positive valence because neutral and negative facial expressions and emotions are more difficult to process for children (see Section 3.7.1.3 and Herba & Phillips, 2004 for a review). Positive emotional (facial) expressions are acquired and can be comprehended earlier than neutral and negative emotions. Moreover, positive emotions are arguably less diverse, less ambiguous and less prone to be misinterpreted (cf. Herba & Phillips, 2004 for a review and see Section 3.7.1.3). Additionally, older adults do not only show a bias towards positive emotional information, but the ability to identify positive facial expressions also remains stable across the lifespan in contrast with negative facial emotions (Isaacowitz et al., 2007 and Section 3.7.1.2).

We presented the sentences in the OVS sentence structure, because it allows investigating in how far and to what extent the different cue types can be used to facilitate sentence processing of structurally challenging yet grammatical sentences. Moreover, using non-canonical OVS instead of canonical SVO sentences sheds light on the way direct and indirect cues can be integrated into sentence processing, i.e., whether the nature of the integration is semantic or whether the cues can also be used

5. Motivation and Research Questions

to facilitate the processing of structurally challenging sentences (in this case by assigning thematic roles prior to their mention).

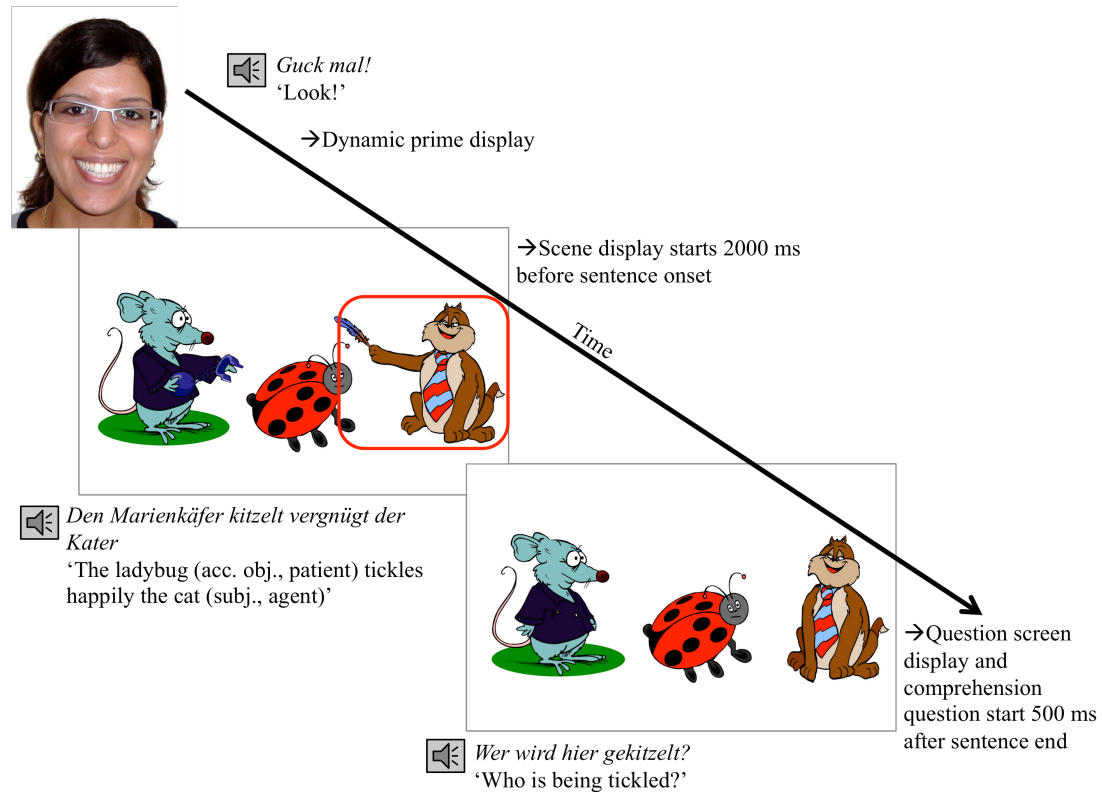


Figure 3: Study procedure for all experiments. Note that the dynamic prime face was presented as a schematic smiley in experiments 1, 2 and 4 but as a natural facial expression in experiments 3, 5 and 6 (as shown here). The exact timing of prime presentation depended on the nature of the prime. The dynamic schematic smiley (exp. 1, 2, and 4) was presented for 1750 ms, while the dynamic natural facial expression (exp. 3, 5, 6) was presented for 5500 ms. Note also that participants did not see the red square around the cat in the actual experiments. The square marks the target agent, i.e., the agent of the action referenced by the sentential verb. Moreover, the comprehension question was in the active and passive voice in experiments 1, 2, and 4 and in the passive voice only in experiments 3, 5 and 6.

Using a 2 (emotional prime vs. non- / incongruent emotional prime) x 2 (action depiction vs. no action) design yielded 4 conditions (Figures 4 and 5):

- Positive emotional prime + depicted action
- Positive emotional prime + no action
- Non- / incongruent emotional prime + depicted action
- Non- / incongruent emotional prime + no action

5. Motivation and Research Questions



Figure 4: Experimental conditions for experiments 1, 2 and 4. In conditions a) and b) a positive emotional prime face in the form of a dynamic schematic smiley (which was valence-congruent with the facial expression of the target agent and patient and with the sentence in studies 1 and 4. In study 2 it was only congruent with the target agent's facial expression and with the sentence). The prime in conditions c) and d) are non-emotional and static. In conditions a) and c) the presented potential agents performed a visually depicted action towards the patient whereas no action was depicted in conditions b) and d). The related positively valenced emotional OVS sentence ('who is doing what to whom') was identical in all conditions.

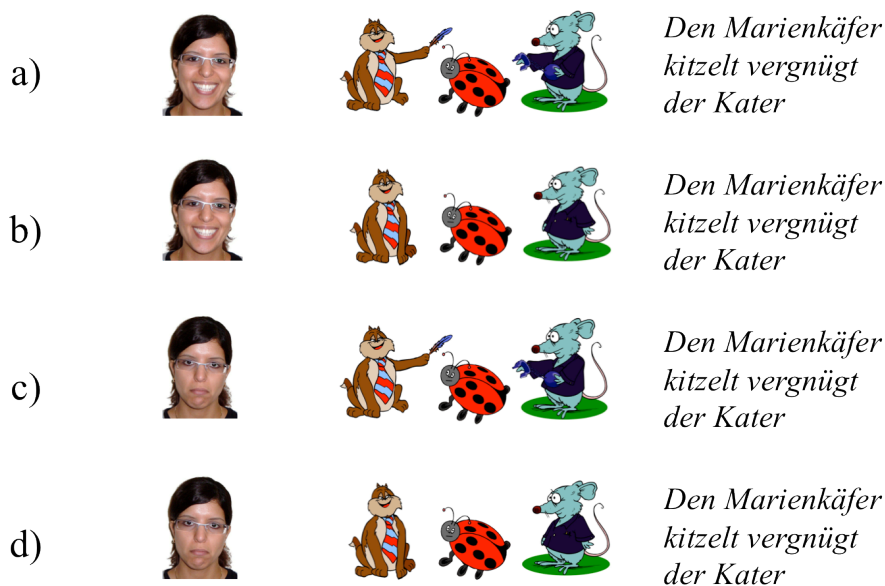


Figure 5: Experimental conditions for experiments 3, 5 and 6. In conditions a) and b) participants were presented with a (congruent) positive emotional prime face in the form of a dynamic natural facial expression of a woman. By contrast, the primes in conditions c) and d) were negatively valenced and thus incongruent with the target characters' facial expression and the sentence valence. In conditions a) and c) the presented potential agents performed a depicted action towards the patient. In conditions b) and c), the actions were absent. The related positively valenced emotional OVS sentence ('who is doing what to whom') was identical in all conditions.

5.2 Analyses

5.2.1 Regions of Interest

We analyzed all 6 experiments in an identical manner. We created regions of interest and preprocessed the fixation data using SR research's software Data Viewer (SR Research, Mississauga, Ontario, Canada). The scene displays were divided into regions of interest. The regions corresponded to the agent of the action denoted by the subject of the sentence, i.e., the target agent, the patient denoted by the object of the sentence, and the distractor character (i.e., the other potential agent, which was not mentioned in the sentence). We were only interested in fixations towards the target agent (henceforth 'agent', see Figure 3) and the distractor character (henceforth 'distractor', see Section 5.3 for the underlying reasoning). Hence, we excluded fixations on the background and the patient from the analyses. Fixations shorter than

80 ms were automatically merged with the nearest fixation (threshold distance 0.5 degrees) by the software.

5.2.2 Word Regions

We divided the experimental OVS sentences into word regions using trial-by-trial word onsets (see Appendix A.2). Hence the NP1 region extended from NP1 onset until Verb onset for each trial. Fixations starting before NP1 onset were excluded. The Verb region extended from Verb onset until Adverb onset and the NP2 region extended from NP2 onset until 500 ms after NP2 offset. The long NP2 region permitted us to capture potential late visual context effects, especially regarding possible processing differences in children and older adults. Henceforth, this region will simply be called ‘NP2 region’. Moreover, we analyzed a combined Verb-Adverb region and a region comprising the whole sentence, i.e., a Long region. We chose to analyze the Long region to capture visual cue effects on the processing of the sentence as a whole, irrespective of a specific word region. Furthermore we computed a preNP1 region ranging from the onset of the scene display until NP1 onset to capture potential effects of the emotional prime face that are solely based on prime and target scene effects, independent of any linguistic input. This preNP1 region had a duration of 2000 ms (see Figure 3).

For each word region, we divided the fixation record provided by the Data Viewer software (SR Research, Mississauga, Ontario, Canada) into consecutive 20 ms bins comprising the longest duration of a word region across experimental sentences (see Table 1). Fixation counts were then aggregated for each word region by region of interest (agent vs. distractor) and by prime (positive emotional vs. non- / incongruent emotional) and action (depicted action vs. no action). We aggregated the data separately for subjects and items.

	NP1 Region	Verb Region	Adverb Region	NP2 Region + 500 ms	
	<i>Den Marienkäfer</i>	<i>kitzelt</i>	<i>vergnügt</i>	<i>der Kater.</i>	
		Verb-Adverb Region			
		<i>kitzelt</i>	<i>vergnügt</i>		
	Long Region				
	<i>Den Marienkäfer</i>	<i>kitzelt</i>	<i>vergnügt</i>	<i>der Kater.</i>	
English	‘The ladybug	tickles	happily	the cat.’	
Mean word onset	0 ms	1228 ms	2380 ms	3720 ms	5060 ms
Longest duration	1620 ms	1540 ms	1740 ms	1180 ms	500 ms

Table 1: Analyzed German word regions, their English translation, mean word onsets and longest duration for each word region. Note that regions were computed for each item individually. We analyzed the *preNP1* region (not shown in the table), *NP1* region, *Verb* region, *Adverb* region, *NP2 + 500 ms* region (henceforth ‘*NP2* region’), *Verb-Adverb* region and *Long* region.

5.2.3 Inferential Analyses

Since our main interest lies in what happens in incremental processing of the sentence especially during the Verb and Adverb region (see Section 5.4.1), we performed repeated measure ANOVAs on the fixation data for separate word regions, namely *preNP1*, *NP1*, *Verb*, *Adverb*, *Verb-Adverb*, *NP2* and *Long* Region, with special focus on the Verb and Adverb regions. Although all word regions described in Table 1 were analyzed, the critical word regions for the cue integration effects were the verb and adverb of the OVS sentence. The verb was the first region of the sentence in which the depicted action could be referenced to the linguistic input. The adverb was the first region in which the emotional prime could be linked to the positive valence of the adverb and the target agent’s happy facial expression¹⁰.

Prime (prime vs. incongruent prime) and action (depicted action vs. no action) were used as factors in the ANOVAs. Note however, that prime was the only factor used for the *preNP1* and *NP1* analysis, since during these regions participants cannot

¹⁰ Note that because all critical sentences were positively valenced, the verb often referred to a positive but never a negative action (e.g., to kiss, to hug, but also to take a picture, see Appendix A.1). Yet, we did not strictly control for the verb valence and hence take the positively valenced adverb as the first region in which the positive prime can be linked to the positive valence of the sentence and the agent’s happy facial expression. Nevertheless, earlier effects of the positive prime face, i.e., effects in the Verb region could be due to the valence of the verb.

yet link the verb to the depicted action, as they haven't heard the verb yet. For age comparisons, we included age group as a between-subject factor in the subject analyses and as a within-subject factor in the item analyses.

We performed inferential analyses for each word region by subjects (F_1) and items (F_2). Data analyses relied on the methods used in Arai et al. (2007). For each word region the aggregated fixation counts (that fell into a word region) by subjects and items were used to calculate the probabilities of fixations towards the agent versus the distractor. Fixations towards the agent in one condition were divided by the fixations towards the distractor in the same condition to calculate the probabilities of looks towards agent and distractor per condition. A constant of 0.5 was added to the fixation counts of the agent and distractor region per condition. The constant was added because the natural log of 0 is undefined; hence the log cannot be computed if fixations to the agent region divided by fixations to the patient region equals 0.

We transformed the probabilities of looks into log-ratios because these probabilities are not linearly independent of one another, i.e., more fixations towards the agent imply fewer looks to the distractor. Log ratios on the other hand express the strength of the visual bias towards the agent relative to the distractor. If both agent and distractor were fixated equally frequently, the score of the log ratios would be 0. If the values were positive, the agent was fixated more than the distractor. Negative values, by contrast, indicated more looks towards the distractor than the agent. Log ratios are symmetrical around 0, meaning that the absolute score reflects the magnitude of an effect, whereas the direction of the visual bias is expressed by the valence. Moreover, probabilities of looks violate the assumption of homogeneity of variance and are thus not well suited for parametric tests such as repeated measure ANOVAs. Log ratios on the other hand are more suitable because they can take on values between minus and plus infinite in contrast to values between 0 and 1 as it is the case for probabilities (see Arai et al., 2007).

We subjected accuracy scores of the comprehension question for the experimental items to repeated measure ANOVAs for subjects and items with prime and action as within-subject factors¹¹. In experiments 1, 2 and 4 voice (active vs. passive) was used

¹¹ We additionally ran generalized linear mixed model analyses on the accuracy data for all 6 experiments. See Appendix E for details on the mixed model analyses and their accuracy results for all experiments. The results of the mixed model analyses mainly corroborate the results of the ANOVA analyses.

as an additional within-subject factor. For age comparisons, we included age group as a between-subject factor in the subject analyses and as a within-subject factor in the item analyses.

5.2.4 Descriptive Analyses

To descriptively analyze the eye-movement results, we computed time course graphs comprising the whole sentence, i.e., the Long region. The fixation record from NP1 onset¹² to question screen onset was divided into 20 ms bins and aggregated by region of interest, prime and action. A constant of 0.5 was added to each data point before computing log ratios of looks of agent over distractor for each condition. These log ratio of looks are displayed in time course graphs, i.e., line graphs, to visually represent the results of the inferential statistic and to inspect the changes of the effects over time as the sentence unfolds. Values above 0 indicate a preference to look at the agent while values below 0 indicate a preference to look at the distractor (unless otherwise stated). The time course graphs are visually divided into sentence regions based on mean word onsets¹³ (see Table 1). We analyzed the accuracy scores for the memory of the prime face (experiments 3, 5 and 6) descriptively only, as they yielded too few data points per condition per study (160 observations per study, i.e., 1 data point in each condition per subject).

5.3 Linking Hypothesis

Based on the assumption that our eye-movements are closely linked to our on-line linguistic processing when a visual scene is present (see Section 4.1), we are assuming that the linguistic input together with a bias of fixations to the agent (vs. distractor) suggests a relationship between the agent character and the processed linguistic input. This established relationship (or linking) indicates that visual and linguistic information have been integrated and interpreted on the basis of the participants preexisting linguistic knowledge and the visual context. After having heard the verb, a higher number of fixations towards the agent versus the distractor

12 Note that we excluded looks starting before NP1 onset. For this reason the displayed time course graphs might show a high difference in log-ratio of looks between 0 and 20 ms as there is no smooth transition from the preNP1 to the NP1 region.

13 Note that sentences varied in length, hence the displayed log-ratios at the end of the NP2+500 region in the time course graphs only stem from the longest sentences and thus only contain few data points.

indicates a preference for the agent (vs. the distractor) to be interpreted as the subject of the sentence. On the other hand, a higher number of fixations towards the distractor (vs. the agent) after verb onset would indicate a preference to interpret the distractor as the subject of the sentence.

5.4 Predictions

Recall that in the conditions in which the action (named in the sentence) was depicted in the scene, the verb was the first word region in which the agent in the scene could be unambiguously discriminated as the subject of the sentence. This discrimination should be anticipatory in nature, to the extent that existing findings replicate (Zhang & Knoeferle, 2012). This anticipatory discrimination means that the target agent should be identified in the scene before it is mentioned as the subject of the sentence.

Recall further that in the conditions in which the action (named in the sentence) was not depicted in the scene, the positive valence of the adverb was the first clue that participants could exploit for discriminating the agent of the sentence if the positive emotional prime was present. The adverb was the first clue because the positive emotional prime matched in valence with the positive emotional adverb and the positive facial expression of the agent. The agent who portrayed a positive facial expression performed this action mentioned in the sentence in a positive manner (depicted or not). The congruence could, in principle, lead to a discrimination of the correct character (the agent, i.e., cat in Figure 3) as the depicted subject of the sentence.

5.4.1 Eye-movement Predictions

We hence predicted the following anticipatory eye-movement patterns across all experiments and age groups regarding the integration of the positive emotional prime and the depicted action into real-time sentence processing. Participants' objectives were to correctly interpret the sentence:

In the single action condition in which the depicted action was the only cue to the agent (condition c), we predicted more anticipatory looks towards the agent versus the distractor compared to the no-cue condition (condition d). This fixation pattern should emerge from the Verb region onwards if participants rapidly used the depicted

action to determine ‘who does what to whom’ prior to NP2 onset. We should hence replicate the findings by Zhang and Knoeferle (2012).

In the single emotion condition in which the positive emotional prime was the only cue to the agent (condition b)), we predicted more anticipatory looks towards the agent versus the distractor compared to the no-cue condition (condition d)). This fixation pattern should emerge from the adverb onwards if participants could rapidly use the positive prime to assign thematic roles and understand the sentence.

Moreover, it could be the case that one of the two cues is stronger than the other cue and could thus be used to a greater extent or more easily than the other cue. If this is the case, we should see a greater magnitude of the respective stronger effect. Specifically, we expected more looks towards the agent (than the distractor) when the action was present in the single cue condition, than when the emotional prime was the only available cue. As the depicted action was mediated via direct reference, we predicted this cue to have a stronger effect on anticipating the associated agent than the subtler, social cue. The latter was mediated by the utterance more indirectly (via a match between the facial prime, the adverb, and the facial expression of the target agent).

In the condition in which both the positive emotional prime and the depicted action were present (condition a)), the effects could be either cumulative¹⁴ or inhibitory. The effects could be cumulative, because the two visual cues might support each other and hence sentence processing and thematic role assignment was easiest compared to the other conditions. The effect could also be inhibitory because recall that Kreysa et al. (2014) did not find an advantage in sentence processing if both the gaze-shift and the depicted object could be used together as cues (see Section 3.5). Moreover, using two distinct visual cues together (direct and indirect) in order to process a non-canonical sentence in real-time might be costly in terms of cognitive resources. If the effect was cumulative, then we should see even more and / or earlier fixations towards the agent versus the distractor compared to all other conditions. If the effect was inhibitory, then the fixation pattern could be similar to that observed in the no-cue condition (condition d)), or in one of the single-cue conditions (condition b) or c)). In the latter case, the accuracy results might indicate whether the effect was

¹⁴ The term cumulative was chosen over additive to prevent data interpretations / predictions according to Sternberg (cf. Sternberg, 2013). This kind of additivity is not assumed here.

in fact inhibitory or whether one cue was stronger than the other (see Section 5.4.2). When no cue was present (condition d)), we predicted the least difference between fixations towards the agent and the distractor.

Furthermore, it was unclear how the schematic nature of the emotional prime face would influence the visual context effects in conditions a) and b). Possible null effects could not only be due to participants inability to use an emotional prime face for real-time assignment of thematic roles, but could also be due to the nature of presentation of the emotional facial expression (see Section 3.7.2 for interpretation and processing differences between natural and schematic facial expressions).

5.4.2 Accuracy Predictions

The accuracy results for the comprehension question about ‘who is doing what to whom’ should underscore the on-line data patterns. Hence, if the single cues (conditions b) and c)) could be used to facilitate OVS sentence comprehension, we should see higher accuracy for these conditions compared to the no cue condition d). If the single action cue was more helpful than the single-emotion cue, the single-action cue should lead to a higher accuracy. Furthermore, if both cues could be used cumulatively, we would see the highest accuracy scores for condition a) compared to all other conditions. However, if the two cues rather hindered each other or competed with each other, we might observe an accuracy pattern similar to that in condition d). Nevertheless, it should be noted that the comprehension questions, especially when posed in the active voice, are relatively easy to answer, at least for adults. We could hence see a ceiling effect, concealing the facilitating effect of the depicted action and / or the positive emotional prime.

6. Experiments – Younger Adults

6.1 Procedure

The procedure was identical in all six experiments and we will therefore only explain it here. An SR Research Eye-link 1000 eye-tracker (SR Research, Mississauga, Ontario, Canada) monitored participants' eye-movements in the remote setup using a sampling rate of 500 Hz and a 5-point manual calibration. No head stabilization was necessary. Images were presented on a 15.6 inch TFT monitor with a resolution of 1680×1050 pixels. The dynamic schematic prime smiley and its static baseline (i.e., the star, experiments 1, 2 and 4) were presented with a resolution of 288×288 pixels whereas the videos of the emotional facial expressions as primes had a resolution of 1364×760 pixels and were played with a frame rate of 30 Hz. We only tracked the right eye but viewing was binocular. Participants' head movements were not restricted but a target sticker on their foreheads or cheeks adjusted calibration settings for their head movements. However, we asked subjects to move as little as possible in order to ensure high-quality data collection.

Before the experiments, we instructed participants to look at the screen during all trials and to try to understand the sentences and the scenes. After each trial we asked participants a comprehension question for 'who is doing what to whom'. There was no time limit in answering the question, and participants answered orally.

After participants had read the instructions on the screen (adults) or had been instructed verbally (children in experiments 4 and 5) we calibrated their right eye. If validation of this procedure was poor, calibration was performed again until validation was good as indicated by the eye-tracker software. Each participant completed 4 practice trials. If the participant did not have any further questions, the experiment started. After each trial, we asked participants to fixate a centrally-located dot and the experimenter performed a manual drift correction. If this failed, the eye-tracker was re-calibrated. After half of the trials participants were asked if they would like to take a break. The calibration procedure was performed again after the break. The eye-tracking experiment took approximately 30 minutes.

6.2 Experiment 1

6.2.1 Participants

40 University of Bielefeld students aged between 18 and 30 years (Mean: 25.2, SD: 2.9) took part in the experiment. All participants were monolingual native speakers of German and had normal or corrected-to-normal vision. Participants received 4 Euro for their participation, and testing took about 30 minutes. Testing took place in the eye-tracking laboratory of the Language and Cognition Group.

6.2.2 Materials

We will cover material construction in detail here. However, as the remaining 5 experiments used the same stimuli, we will only report the changes that were made to these studies for the following experiments.

6.2.2.1 Critical Items

We constructed 16 German non-canonical Object-Verb-Subject (OVS) sentences in the form: NP1 [masculine accusative case, patient] - Verb - Adverb [positive valence] - NP2 [nominative case, agent], as e.g., *Den Marienkäfer kitzelt vergnügt der Kater.* (transl.: ‘The ladybug (acc. obj., patient) tickles happily the cat (nom. subj, agent)’ (see Appendix A.1). The masculine gender of the NP1 has been chosen in order to avoid sentence ambiguity with Subject-Verb-Object (SVO) structures (see Section 2.1). The verbs are transitive action verbs. A pretest obtained ratings of sentence valence (see Section 7.1).

One female native speaker of German recorded all sentence with neutral (neither positive nor negative) prosody. The speaker emphasized the adverb of the sentence to highlight its importance. The sentences were recorded with normal pace, leaving a slightly longer pause than naturally between each phrase. The on- and offsets of the word regions for the critical sentences have been marked for later analyses of the individual word regions (see Appendix A.2).

We created the target scenes using Adobe Illustrator CS5 and commercially available clipart characters and instruments. Most of the figures were animals ($N=86$), and some were humans ($N=30$). The target scenes ($N=16$) for the experimental items

always consisted of 3 clipart characters. The critical characters were presented in a row next to each other. The two interacting characters always faced each other, while the third character also faced the middle character. The middle character was always the receiver of the action, i.e., the patient (mentioned by the NP1 of the OVS sentences). One of the other characters was the target agent and performer of the action mentioned in the sentence (by the NP2 of the OVS sentence). The third (distractor) character was also performing an action towards the patient but it was not mentioned in the sentence. Both the target agent and the patient, i.e., subject and object of the sentence, portrayed a happy facial expression, thus matching in positive valence with the prime smiley. The distractor character had a neutral facial expression.

The emotional prime consisted of a schematic yellow happy looking smiley with a broad smile (see Figure 4, conditions a) and b) or Appendix B.1). The facial expression of the smiley was dynamic in the sense that it started off with a light and subtle smile and changed into a broad smile after 250 ms. The change was accompanied by, and time-locked to, the auditorily presented phrase *Guck mal!* (transl.: ‘Look!’) to focus the participants’ attention and to prepare for the following scene. The dynamic prime smiley and the static red star were both presented for 1750 ms in total.

Regarding the action depiction, in half of the 16 experimental trials an action was depicted in the scene. In these trials both the agent and the distractor performed an action towards the patient (see Figure 4, conditions a) and c) or Appendix C). In the other half of the trials, the characters did not perform any actions mentioned in the sentence (see Figure 4, conditions b) and d) or Appendix C). However, agent and patient were still facing each other. The agent and the distractor were facing towards and looking at the patient. In the depicted action condition, the patient was similarly facing towards and looking at the agent. In the no-action condition, by contrast, the patient still faced, but no longer looked at the agent. Instead the patient looked straight ahead, towards the observer. In the no-action condition, the positioning and facial expression of the agent and distractor stayed the same as in the depicted action condition. In the no-action condition the actions were not depicted and the characters’ body posture was neutral (vs. e.g., with outstretched hands when holding a camera).

6.2.2.2 Filler Items

We constructed 24 filler sentences with a Subject-Verb-Object (SVO, $N=24$) sentence structure. Of these, 10 started with a feminine NP, 4 started with a neutral NP and 10 started with a masculine NP. An additional 4 filler sentences had an OVS structure (NP1 masculine). Twelve filler sentences had neutral adverbs (8 SVO and 4 OVS sentences) and 16 were positively valenced (16 SVO sentences, also rated for valence in the same pretest as the experimental items, see Section 7.1.2).

We constructed 28 filler scenes. The scenes also consisted of clipart animals and humans. 8 filler items showed three characters and another 20 showed two characters. The characters were either facing each other or the observer. Facial expressions of the characters were either neutral ($N=12$) or happy ($N=16$) and always matched the sentence valence. When three characters were present, the agent and distractor faced the patient or all of them faced the observer. For 14 of the filler trials one of the characters was depicted as performing the action mentioned by the sentence. For another 14 trials, no actions were depicted.

6.2.2.3 Counterbalancing

Balancing of materials was identical for all experiments. We counterbalanced the position of the three characters across the conditions, but the patient always stayed in the middle. That means for each of the two action conditions, we created two pictures (see Appendix C). The exact coordinates of the characters within each picture were held constant across the conditions.

Moreover, across all pictures (critical items and fillers) we balanced the number of actions depicted, meaning that in half of all trials the characters in the scenes were performing an action whereas in the other half of all trials no actions were depicted. Furthermore, we balanced the number of characters presented in a scene, i.e., in total, participants saw 20 scenes depicting 2 characters and 24 scenes depicting three characters. Additionally we created SVO and OVS sentences across all trials in a way that SVO sentences were slightly more frequent than OVS sentences (24 SVO sentences vs. 20 OVS sentences in total), since canonical SVO sentences are also more frequent than OVS sentence in natural speech (see Section 2.1).

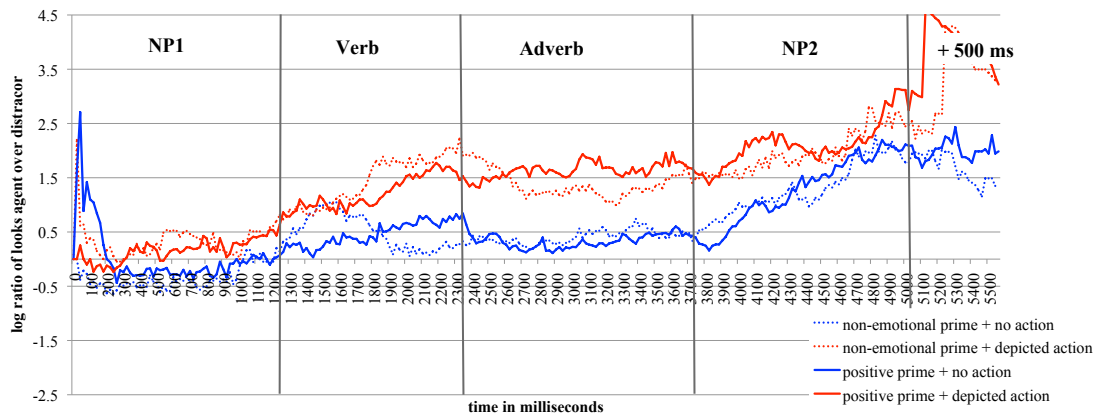
Using a Latin square design, we created 8 lists such that each subject saw each item only once and in one of the 4 conditions. Agent and patient were equally often

on the right as on the left side for critical items. Moreover, each participant saw a different pseudo-randomized list. No list started and ended with an experimental item. Additionally, two critical items never directly followed after one another, and if possible two filler trials were placed after each critical item.

6.2.3 Results and Discussion

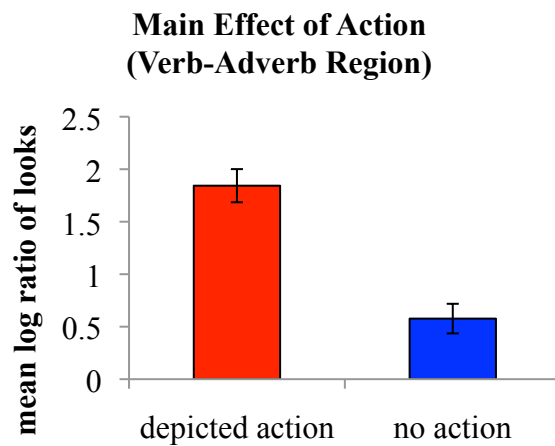
6.2.3.1 Eye-movement Results

Graph 1 (and all subsequent time course graphs) depicts the log ratio of looks towards the agent over the log ratio of looks towards the distractor on the y-axis. Looks above 0 signal a preference to inspect the agent whereas looks below 0 indicate a preference to fixate the distractor. Time in milliseconds over the course of the whole sentence – plus the 500 ms until the onset of the question screen – is shown on the x-axis. The time course graph is further visually divided into the analyzed word regions. The red lines indicate the conditions in which an action is depicted (conditions a) and c)) and the blue lines indicate the conditions in which no actions are depicted (conditions b) and d)). The solid lines indicate the presence of the positive emotional prime (conditions a) and b) whereas the dotted lines show the non-emotional (incongruent emotional) prime conditions (b) and d)).



Graph 1: Time course for the eye-movement results by condition (exp. 1). Time in milliseconds is shown on the x-axis and log ratios of looks to the agent versus distractor are shown on the y-axis. Values above 0 indicate a preference to look at the agent whereas values below 0 indicate a preference to look at the distractor. Word regions are visually marked and the conditions are color-coded as shown in the legend.

Analyses of the fixation patterns yielded clear main effects of action depiction in the Verb, Adverb, Verb-Adverb, NP2 and in the Long region (F_1 and F_2 , all $p_s < .05$). Descriptively, these strong effects of action depiction can be seen in the time course graph (Graph 1). The high deviations between the red and the blue lines especially in the Verb and Adverb region mark the main effect of action depiction and show that younger adults fixated the agent significantly more than the distractor when actions were (vs. weren't) depicted. This action effect is also displayed in Graph 2.



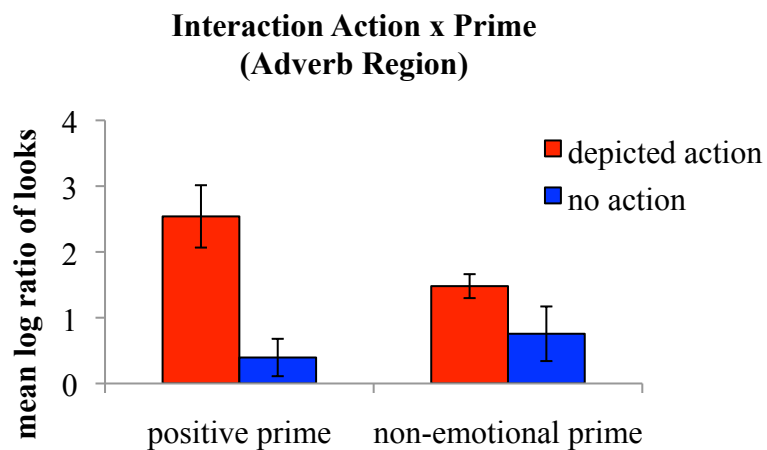
Graph 2: Mean log ratio of looks for the main effect of action in the Verb-Adverb region (F_1 , exp. 1). *Error bars show the standard error.*

This action effect is already significant in the Verb region, suggesting that younger adults could anticipate the correct role filler, i.e., the agent of the action, using the action cue. They hence used the depicted action (vs. no action) as a cue to establish reference from the verb. During this process, the visually associated agent became available permitting early thematic role assignment. In contrast with the clear action effects, the analyses of the real-time data did not reveal a main effect of emotional prime in any of the analyzed time regions. Interestingly though, we found a marginal interaction between these two cues ($F_2(1,15) = 4.156, p = .06$).

Graph 3 shows that younger adults looked more at the agent if both the positive prime and the depicted action could be used as cues compared to the single cue conditions. Although this effect was only marginal for the Adverb region in the item analysis, it indicates a cumulative effect of the cues. Participants' on-line thematic role assignment seemed to be facilitated the most when they could make use of both the direct and the indirect visual cue in order to anticipate the correct role filler.

However, the fact that this interaction was only marginal could be due to the scarcity of data points, since each participants only saw 4 trials for each of the 4 conditions, i.e., 160 data points in total across all 40 participants that could lead to an interaction. Obviously the main effects have greater power since they were based on 8 trials per participant, i.e., 320 data points in total.

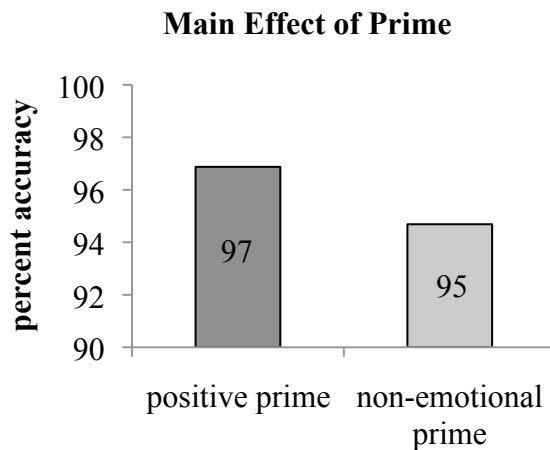
Yet, the missing main effect of the emotional prime suggests that participants could not integrate the emotional cue on-line, unless an action was also depicted. This missing effect can also be seen in the time course graph (Graph 1) since the dotted lines signaling the non-emotional prime do not deviate much from the solid lines, which show the positive prime conditions over the course of the sentence.



Graph 3: Mean log ratio of looks for the interaction between prime and action in the Adverb region (F_2 , exp. 1). *Error bars show the standard error.*

6.2.3.1 Accuracy Results

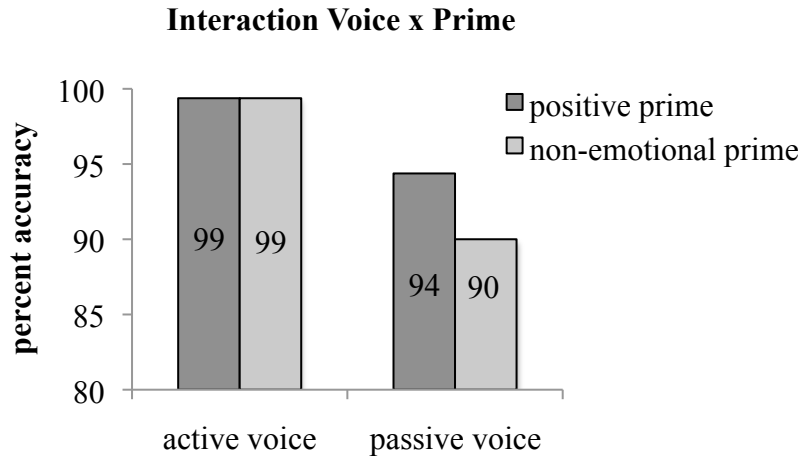
The behavioral data on the other hand did reveal a main effect of prime which was significant in the item analysis ($F_2(1,15)= 4.623$, $p < .05$), but no main effect of action. As Graph 4 shows, participants answered more trials correctly if the prime was positive (vs. non-emotional).



Graph 4: Accuracy for the main effect of prime (exp. 1). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

However, participants answered 96% of all trials correctly, regardless of condition. Hence, this ceiling effect could have overshadowed a potential main effect of action, as well as weakened the effect of prime. Younger adults did not have difficulties in answering the ‘who does what to whom’ comprehension question.

Nevertheless, when additionally including voice (active vs. passive) as a factor in the analyses, results revealed a marginal interaction between voice and prime ($F_2(1,15) = 4.286, p = .056$). Graph 5 shows that participants were facilitated by the positive prime in answering the comprehension question, but only when the question was asked in the passive voice and was hence more difficult than its active counterpart. In addition, a main effect of voice (F_1 & F_2 , all p s $< .05$) corroborated that active questions were answered more accurately than passive questions.



Graph 5: Accuracy for the voice x prime interaction (exp. 1). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

Perhaps participants only needed to rely on additional cues for answering the comprehension question when the structure of the non-canonical questions rendered the task of comprehending the non-canonical OVS sentence more difficult. Answering the comprehension question in the active voice clearly posed no difficulties for the younger adults (99 percent average); no cues are needed to facilitate comprehension any further.

6.2.4 Summary and Conclusion

To conclude, the depicted action as a direct cue could be integrated quickly into real-time sentence comprehension and was used on-line for anticipating the correct thematic role filler. The behavioral data on the other hand did not reveal any effects for the depicted action. One reason for this could be a ceiling effect overshadowing possible effects of the depicted action on sentence comprehension.

As previous research (Zhang & Knoeferle, 2012) suggested that depicted actions can also influence participants' off-line behavior, we do not think that participants only used the direct cue for on-line sentence processing facilitation. Interestingly, for the indirect cue (the positive emotional prime), we did not find any real-time effects, leading to the conclusion that younger adults did not seem to use the positive smiley on its own for real-time thematic role assignment.

However, we did find a marginal interaction between the direct and the indirect cue, which suggests that the indirect cue was only used when the direct cue was also

available. Perhaps participants only needed the direct cue to anticipate the target agent. However, if the positive prime was also available, it was perceived and held in working memory. When no action was depicted, participants did not (re)activate it in order to anticipate the agent; however when the action was also depicted, participants began to anticipate the target agent (during the verb region). At this point in time they had already assigned thematic roles using the depicted action but the decision to interpret the fixated character as the correct subject of the sentence could be corroborated by the emotional smiley. Thus, when hearing the adverb, which carried the emotional valence, the prime smiley was (re)activated and integrated with the depicted action resulting in a pronounced anticipation of the agent of the action and a strengthened sentence interpretation.

That the indirect cue, i.e., the positive prime smiley, had an effect on OVS sentence comprehension is furthermore underlined by the accuracy results suggesting that younger adults do seem to have made use of the emotional cue. However, this only seemed to be the case when the comprehension question was asked in the passive voice (answering the passive comprehension question after having heard an OVS sentence might add even more difficulty to language processing). The passive, just like the OVS sentence is grammatical, yet non-canonical in German and hence processed with more difficulty than an active sentence. Hearing a non-canonical passive comprehension question just after having processed a non-canonical OVS sentence is presumably more cognitively demanding than answering an active comprehension question (see Section 2.1, but see also Mack, Meltzer-Asscher, Barbieri, & Thompson, 2013 for adults' on- and off-line processing differences between active and passive sentences and Dittmar, Abbot-Smith, Lieven, & Tomasello, 2014 for German children's differences in active and passive sentence acquisition).

6.3 Experiment 2

6.3.1 Participants

40 further University of Bielefeld students aged between 18 and 30 years (Mean: 24.12, SD: 3.09) took part in the experiment. All participants were monolingual native speakers of German and had normal or corrected-to-normal vision. Participants received 4 Euro for their participation, and testing took about 30 minutes. Testing took place in the eye-tracking laboratory of the Language and Cognition Group.

6.3.2 Materials and Hypotheses

Materials were mainly identical to experiment 1, therefore only the changes will be reported. The changes ruled out two factors that could have prevented the indirect cue from affecting real-time sentence processing on its own.

First, in experiment 1, both agent and patient character were portraying a positive facial expression. Thus, the prime face, when positive, matched two scene characters. The multiple matches might have weakened a possible effect of positive prime for agent anticipation. In experiment 2 therefore, only the agent showed a happy facial expression congruous in emotional valence with the positive prime smiley. This change should decrease looks towards the patient after NP1 onset and increase looks towards the agent in the emotional prime conditions. Further, changing the distractor character's facial expression to slightly negative should increase the contrast in emotional valence between the agent and distractor character's facial expression. The distractor's facial expression should thus be more easily identifiable as not matching in valence with the positive prime and the positive valence of the adverb (see Appendix C.2).

Moreover, we also changed the positioning of the characters in the filler scenes such that the characters mentioned in the sentence were not always facing each other. Thus, predictions about who is interacting with whom have been rendered unreliable on the basis of the character position alone. In experiment 1, the filler trials were identical to the critical trials in character positioning, i.e., the interacting characters always faced each other. This character positioning might have led participants to develop a strategy to always focus on the outer character of the two interacting ones and could hence have made thematic role assignment easier regardless of the

experimental conditions. By rendering the positioning of the filler characters unreliable in experiment 2 this possibility was ruled out. Regarding the changes to the materials we predicted to find a stronger effect of the positive prime smiley on thematic role assignment especially in the real-time eye-movement data compared to experiment 1.

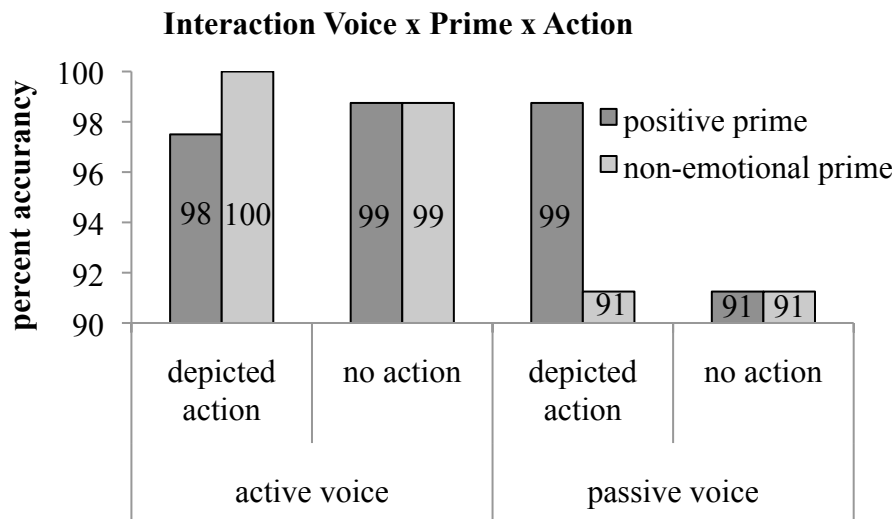
6.3.3 Results and Discussion

6.3.3.1 Eye-movement Results

The results were similar to experiment 1. We replicated the main effect of action in the Verb, Adverb, Verb-Adverb, NP2 and the Long region (F_1 & F_2 , all $ps < .05$, cf., Graph 2) in the absence of an on-line effect of the positive prime. Additionally, the interactions between action and prime were not significant. The means show however that participants inspected the target agent more when both cues were available compared to the single cue and no cue conditions (Adverb region $F_1(1,39) = 1.752$, $p = .193$, $M_{\text{cond}^a} = 3.115$, $M_{\text{cond}^b} = .599$, $M_{\text{cond}^c} = 2.309$, $M_{\text{cond}^d} = .498$).

6.3.3.2 Accuracy Results

We replicated the accuracy results. We found a significant effect of voice (F_1 & F_2 , $ps < .05$) in the analysis with voice (active vs. passive) as an additional factor. Participants answered questions in the active (vs. passive) voice significantly more accurately. Moreover, we also found a significant voice x prime interaction in the item analysis ($F_2(1,15) = 5.000$, $p < .05$). This interaction indicates again that younger adults tended to experience facilitation by the positive prime when they answered a comprehension question posed in the passive (vs. active) voice, but reached ceiling regardless of prime condition if the question was asked in the active voice (cf., Graph 5). Additionally, the accuracy analysis also yielded a marginal voice x prime x action interaction ($F_1(1,40) = 3.805$, $p = .058$).



Graph 6: Accuracy for the voice x prime x action interaction (exp. 2). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

This interaction suggests that younger adults made use of the positive prime in answering the comprehension question when the question was asked in the passive voice and when an action was also depicted. The interaction underlines the assumption from experiment 1 that participants only made use of the indirect cue when the direct cue was also available.

However, given the limited number of data points for each condition in total across all participants (80 per condition including voice as a factor), we must be careful in overinterpreting this interaction. Moreover, as can be seen by looking at the scale of Graph 6, the differences in percent are small (range: 91% to 100%). Thus, similar to experiment 1 a ceiling effect could have covered potential main effects of action and prime.

6.3.4 Summary and Conclusion

To conclude, we replicated the findings from experiment 1. However, the changes in the materials did not appear to strengthen the on-line integration of the positive emotional prime in the form of a smiley.

While participants did not seem to use the emotional prime in real time, emotional priming effects did emerge offline. First, experiments 1 and 2 have yielded effects of the emotional prime manipulation in the accuracy data. Further, the reason why participants might not have been able to use the emotional prime more extensively for

sentence processing might lie in its schematic nature (the prime faces were smiley depictions). This issue was addressed in experiment 3.

6.4 Experiment 3

6.4.1 Participants

40 further University of Bielefeld students aged between 18 and 30 years (Mean: 23.75, SD: 3.62) took part in the experiment. All participants were monolingual native speakers of German and had normal or corrected-to-normal vision. Participants received 4 Euro for their participation. Testing took about 30 minutes. Testing took place in the eye-tracking laboratory of the Language and Cognition Group.

6.4.2 Materials

Materials were identical to experiment 2 except for the changes indicated below. In experiments 1 and 2, the emotional prime consisted of a dynamic, but schematic happy looking smiley (vs. a non-emotional static star). In Experiment 3, the prime was changed to a dynamic human emotional facial expression (happy vs. sad, see Appendix B.2) presented as a short video sequence. In the video sequence, a woman naturally changed her facial expression from neutral into a broad and open smile. The change from a neutral expression into the positive expression happened after 1300 ms and the last frame of the video is the positive emotional expression (total video duration: 5500 ms).

We contrasted the positive prime face with a negative facial expression, i.e., the emotionally incongruent prime. For the negative primes, the woman's face turned from a neutral facial expression into a sad facial expression. The same woman served as a model for the positive and the negative dynamic facial expressions. We also used the same timing parameters for the negative and positive prime videos. Both positive and negative facial expressions were accompanied by an auditorily presented *Guck mal!*-phrase, time-locked to the change in facial expression from neutral to happy / sad.

The woman's face¹⁵ has been chosen on the basis of a previous 9-point scale rating study (DeCot, 2011, $N=18$, Mean: 24.7, SD: 2.74). In that study, her happy and sad facial (static) expressions were rated as one of the most recognizable among a total of 15 faces. Moreover, her happy and sad facial expression showed the greatest distance between the ratings for her neutral and the portrayed positive or negative emotion.

In Experiments 1 and 2, questions were in the active voice in half of the trials and in the passive voice in the other half of the trials. By contrast, in experiment 3 all comprehension questions were asked in the passive voice. Furthermore, on four experimental trials (once after each condition) the experimenter asked the participant post-trial to recall the facial expression of the prime face. We balanced this face-recall question across all lists (asked for each item and in each condition) over left and right character positioning across all lists.

Moreover, for 12 filler trials in each experimental list participants inspected a screen showing a picture of the prime face ($N=6$ positive and $N=6$ negative) next to a picture of a face of one of the characters from the target scene. In these trials the participant was asked to indicate how the human and the depicted character are feeling. Both pictures always portrayed the same emotional valence or a neutral and an emotional valence. The clip-art character's face was not always the face of the agent, but also the face of the patient or distractor. In order to keep the balance between active and passive comprehension questions, the remaining 16 filler trials were followed by active comprehension questions.

6.4.3 Hypotheses

If participants can make use of emotions for thematic role assignment, the natural dynamic facial expression and the task (question) emphasis on emotional faces should enhance the effect of the positive prime. For the accuracy data, we predicted stronger main effects of the emotional prime because all critical comprehension questions were now asked in the passive voice (recall that the positive prime had yielded a facilitatory effect for answering the passive questions in experiments 1 and 2).

We should further observe visual context effects of the emotional prime on-line if participants could also use the natural positive prime for real-time thematic role

¹⁵ The woman gave written consent regarding the use of her face in our experiments and publications.

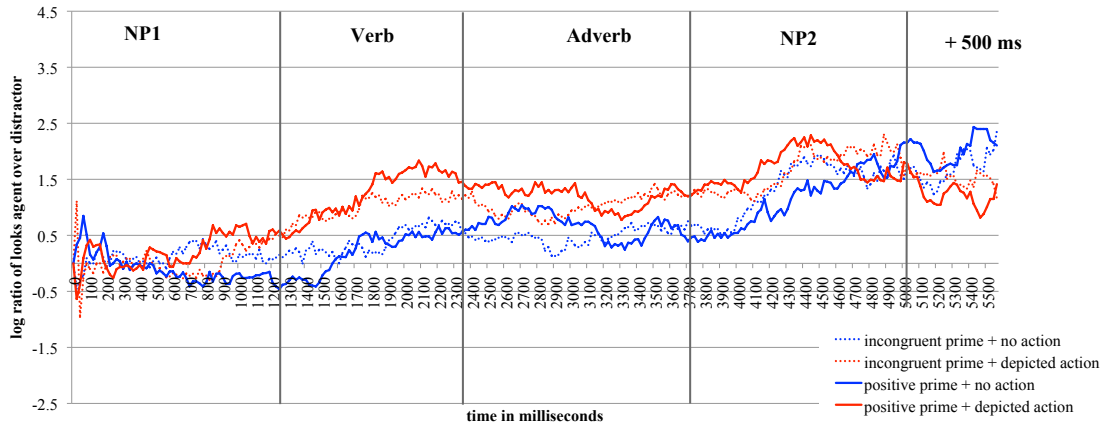
assignment. We additionally predicted to replicate the strong action effects found in experiments 1 and 2.

6.4.4 Results and Discussion

6.4.4.1 Eye-movement Results

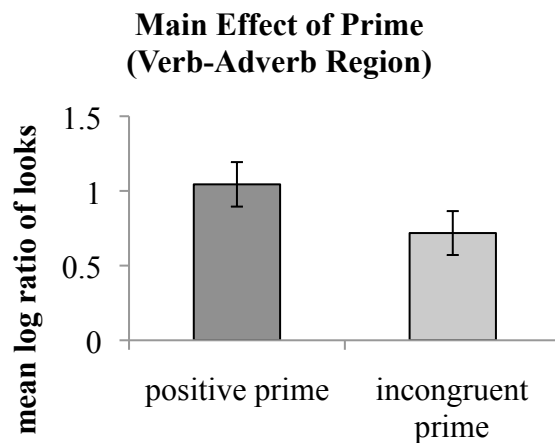
As can be seen in Graph 7 by looking at the difference between the red and the blue lines, experiment 3 replicated the main effect of action depiction in the Verb, Adverb, Verb-Adverb, NP2 and the Long region (F_1 & F_2 , all p s $< .05$). Hence participants' on-line sentence processing was facilitated when an action was depicted (vs. when no action was depicted). Thematic role-assignment was also anticipatory in nature as the significant main effects of action in the Verb and Adverb region show (cf., Graph 7: the red lines show higher log ratio of looks towards the agent than the blue lines).

Crucially, we also found a marginal main effect for the prime in the Adverb ($F_2(1,15) = 3.083$, $p = .1$), the Verb-Adverb ($F_1(1,39) = 2.984$, $p = .092$, $F_2(1,15) = 3.003$, $p = .104$) and in the Long region ($F_1(1,39) = 3.429$, $p = .072$).



Graph 7: Time course for the eye-movement results by condition (exp. 3).

Graph 8 shows that younger adults anticipated the agent, i.e., the sentential subject, more when they had been primed with the positive congruent than the incongruent natural facial expression. Hence, younger adults could quickly match the positive prime face to the positive facial expression of the target agent as soon as the adverb of the sentence revealed the positive sentence valence. This match in positive emotional valence quickly contributed to thematic role assignment independent of the action depiction.

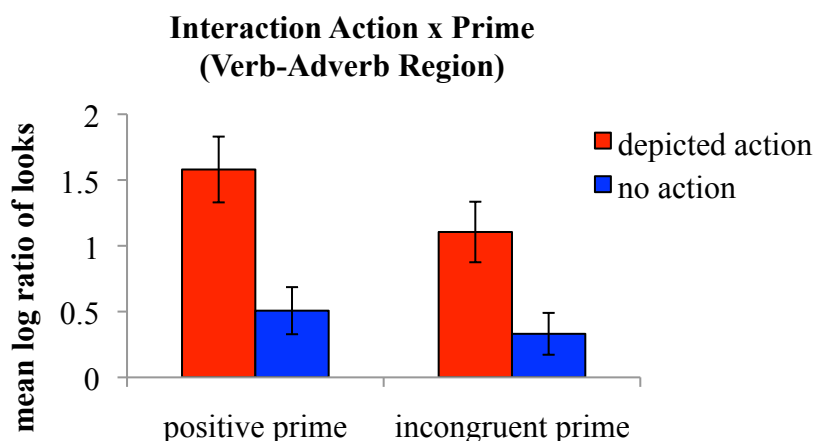


Graph 8: Mean log ratio of looks for the main effect of prime in the Verb-Adverb region (F_1 , exp. 3). *Error bars show the standard error.*

Moreover, the fact that we did not find any visual context effects of the prime in the preNP1, NP1 or in the Verb region underlines that the effect of the natural positive facial expression is not just a mere emotional priming effect due to the matching valence between the prime face and the target agent’s facial expression. Crucially, this effect is mediated by the linguistic input, i.e., the positively valenced adverb. Hence, the positive natural facial expression was integrated into sentence processing, and facilitated sentence processing of the OVS sentence by helping participants to anticipate the correct role filler before it was mentioned in the sentence.

However, we did not find a significant interaction between action and prime. Nevertheless, an inspection of the means (Graph 9) shows that similar to experiment 1 and 2 the two-cue condition elicited more anticipatory looks towards the agent (vs. the distractor) than the single-cue and no-cue condition. Thus, the natural positive prime face seemed to have a stronger facilitatory effect on thematic role assignment when an action had also been depicted.

One reason why the analyses did not reveal a significant action x prime interaction might be the low statistical power. There were only 160 observations in total, i.e., only 4 trials per subject could be used to determine an interaction effect.

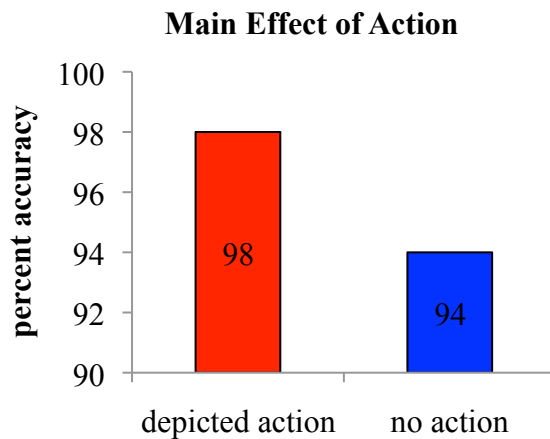


Graph 9: Mean log ratio of looks for the interaction between prime and action in the Verb-Adverb region (not sign., exp. 3). *Error bars show the standard error.*

6.4.4.2 Accuracy Results

The accuracy results underline the real-time data. We found a significant main effect of action ($F_1(1,39)= 3.857, p= .057, F_2(1,15)= 10.385, p< .05$) and a non-significant effect of prime ($p> .103$). Participants answered the passive comprehension question more accurately when an action was ($M_{F1}= 3.9, M_{F2}= 9.75$ vs. was not: $M_{F1}= 3.75, M_{F2}= 9.375$) depicted (Graph 10) and (descriptively) when the prime was positive ($M_{F1}= 3.785$ vs. an incongruent prime: $M_{F1}= 3.775$). The fact that these effects emerged in experiment 3 (but not in experiments 1 and 2) indicate that participants seemed to draw on the visual cues for answering the comprehension question only when the syntactic structure of the question was more challenging, i.e., passive compared to active questions.

However, potential stronger effects of depicted action and positive prime face might have been concealed since participants still performed at ceiling ($< 95\%$ correct answers regardless of condition). Participants answered 99.4 % of all face-recall questions correctly. The high accuracy means the subtle effect of the emotional prime cannot be due to participants ignoring the facial expression.



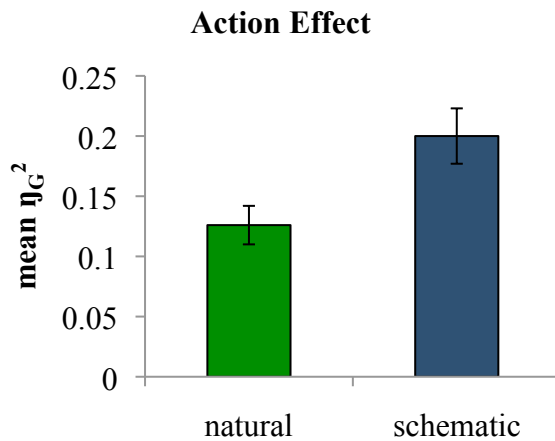
Graph 10: Accuracy for the main effect of action (exp. 3). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

6.4.4.3 Effect Size Differences: Schematic vs. Natural Prime

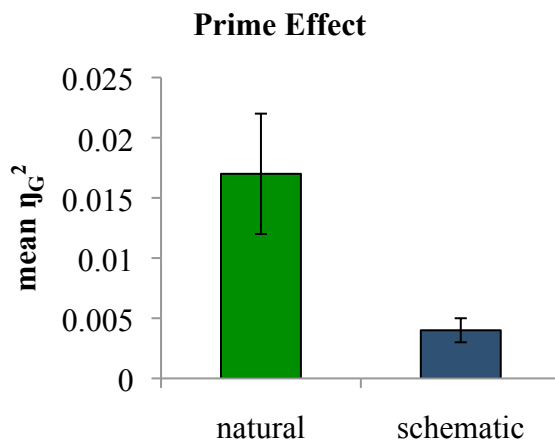
We also compared the effect sizes for the prime and action effects in the eye-tracking data between experiments 2 and 3. These two experiments mainly differed in terms of prime presentation. Whereas participants in experiment 2 were primed with a dynamic happy looking schematic smiley (see Figure 4), participants in experiment 3 were primed with a natural dynamic happy facial expression (see Figure 5). In order to directly compare the magnitude of the effect of the schematic prime with the magnitude of the effect of the natural prime, we ran repeated measure ANOVAs on the generalized eta squared effect sizes (of the F_1 and F_2 analyses) separately for the prime and for the action effects of the Verb, Adverb, Verb-Adverb, NP2 and Long region together ($N=10$), using prime presentation (natural vs. schematic) as the fixed factor.

We used Generalized eta squared (η_G^2) over other effect size measures (e.g., partial eta squared) because it is better suited for comparisons across studies and can also be used to compare effect sizes in within-study designs. It takes inter-study variations such as differing methods and analyses into account and is well suited for experiments with subtle effects. It has to be noted however, that η_G^2 cannot be interpreted in the same way as partial eta squared, meaning that the magnitude of an effect cannot be determined by looking at the value of the effect size but always should be interpreted in relation to another effect size (cf., Bakeman, 2005; Lakens, 2013).

As Graph 11 shows, η_G^2 for the action effect is higher in experiment 2 in which the schematic prime was used compared to experiment 3 which used the natural prime. This difference is significant ($F(1,9)= 15.093, p < .05$) and hence indicates that participants in the schematic prime experiment used the depicted action to a greater extent than participants in the natural prime experiment. This result becomes interesting when looking at Graph 12, which shows the difference in effect size for the prime effect.



Graph 11: Mean generalized eta squared of the action effect (of the Verb, Verb-Adverb, NP2 and Long region) for the natural face (exp. 3) compared to the schematic face (exp. 2) of the real-time data. The mean effect size is displayed on the y-axis. Error bars show the standard error.



Graph 12: Mean generalized eta squared of the prime effect (of the Verb, Verb-Adverb, NP2 and Long region) for the natural face (exp. 3) compared to the schematic face (exp. 2) of the real-time data. The mean effect size is displayed on the y-axis. Error bars show the standard error.

Graph 12 shows that the magnitude of the prime effect is significantly greater ($F(1,9)= 5.636, p < .05$) in experiment 3 (natural prime face) compared to experiment 2 (schematic prime face). Hence, participants made greater use of the natural compared to the schematic prime and at the same time reduced the use of the depicted action in experiment 3 arguably because they did not have to rely on the direct cue as much as in experiment 2 given the more accessible natural facial expression (vs. the schematic prime face).

However, effect sizes for the action effect were significantly higher than the effect sizes for the prime effect in both experiments 2 and 3 (independent two-tailed t-tests: $p_s < .05$). Given the strong and reliable main effects of action in the Verb, Adverb, Verb-Adverb, NP2 and the Long region in both studies and the present but marginal main effects of prime in experiment 3 this is not surprising. Nevertheless, regardless of effect size magnitude difference between the factors, the word region effect sizes for the individual regions for experiments 2 and 3 do not differ significantly from each other (independent two-tailed t-test: $p = .346$).

6.4.5 Summary and Conclusion

To conclude, experiment 3 replicated the strong effect of depicted action in younger adults. Crucially, when the prime was a dynamic natural facial expression, younger adults could also make use of this indirect visual social cue for OVS sentence processing and thematic roles assignment. However, although the effects of the natural positive emotional prime were present and significantly stronger than the effects of the schematic smiley, they were nonetheless less prominent than the effect of the depicted action. Moreover, the interaction means lead to the assumption that the visual social indirect cue is perhaps best used in combination with the direct cue.

Even though younger adults were able to use the indirect emotional cue for on-line thematic role assignment if the cue was natural, the effects that emerged were subtle. One reason for these weak effects might be the negativity bias that younger adults portray. The matching positive valence between the prime face and the target agent's facial expression together with the positive valence of the adverb contributed to thematic role assignment; however, young adults overall have been shown to prefer negatively valenced emotional material. Hence, testing positive social cue integration using age groups that are more biased towards positive emotions might lead to

6. Experiments – Younger Adults

stronger effects of indirect cue integration and thus further facilitate the processing of structurally challenging OVS sentences.

7. Experiments – Children

7.1 Pretest: Character, Action and Valence Recognition

We tested characters, actions and the emotional valence of the adverb in order to ensure that children between 4 and 5 years of age can reliably recognize all the critical characters and depicted actions used in the eye tracking experiment. Moreover, we tested if children can identify the valence of the adverb in the critical sentences and if they can link it to the facial expression of the depicted characters.

7.1.1 Character and Action Recognition

To test children's ability to correctly identify all the critical characters and actions named in the eye-tracking study, we presented four and five year old children who did not take part in the actual experiment ($N=10$) with the critical item scenes shown on a computer screen (see Appendix C.1). Every participant encountered every scene twice in the no-action condition (without any prime presentation or actions). Left and right positioning of the characters was balanced. The experimenter asked questions in the active voice in the form *Wer ist der Wal?* (transl.: 'Who is the whale?') and prompted the children to identify the target agent and patient separately. Additionally, every participant encountered every experimental picture in the action-depicted condition once. In these scenes the experimenter asked for the action that the sentential agent is depicted as performing (e.g., *Wer bekocht hier den Wal?*, transl.: 'Who is cooking for the whale?'). Further, we asked the children for four actions performed by the distractor character towards the patient.

Children always answered by pointing at the character they thought provided the correct answer. Each participant saw a new pseudo-randomized order of the scenes. No-action scenes, i.e., character recognition scenes, were always shown first in blocks of two followed by the same item scene in the depicted-action condition. Children were shown 52 scenes in total; 32 of these were no-action scenes for character identification and 18 depicted-action pictures for action identification. 4 depicted-action pictures were shown twice to ask for the actions of the distractor characters. We balanced left-right positioning of the characters.

7.1.2 Valence Recognition

To test if children can recognize and identify the valence of the positive adverbs used in the eye-tracking study and moreover link it to the character's facial expression, we asked another set of 10 children between 4 and 5 years of age to identify which character (out of a pair) performed a depicted action in a specific emotional manner. For this purpose, participants were presented with the agent and distractor characters used in the eye-tracking study. However, we replaced the patient of each experimental item scene with a generic figure, i.e., a girl in each item. Additionally, the agent and distractor both performed the same action towards the girl. The difference in valence was conveyed through the facial expressions of the agent and the distractor. The agent of the experimental item always had a positive facial expression, while the distractor had a neutral facial expression (see Appendix D). We showed children the scenes and asked in the active voice e.g., *Wer kitzelt hier vergnügt das Mädchen?* (transl.: 'Who tickles happily the girl?'). Emphasis was put on the adverb. Children answered by pointing at one of the characters. 22 such scenes were presented to the children; 16 of these were critical scenes (as seen in Appendix D) and 6 were filler scenes. In the filler scenes a sad adverb was used together with a neutral action (e.g., standing, telling a story, reading) to make sure children actually paid attention to the facial expression of the characters and the adverb of the question rather than always picking the character with a positive facial expression. Each child got a new randomized order of the scenes.

7.1.3 Pretest Results

The 4-5-year old children identified the characters with 96.9% accuracy. Depicted actions were identified correctly in 88.5% of cases. In 89.38% of the cases children reliably identified the character with the emotional facial expression and thus linked the meaning of the (positive) adverb to the emotionally valenced facial expression of the characters.

7.2 Experiment 4

7.2.1 Participants

24 children between 4 and 5 years of age participated in the study (Mean: 4.6, SD: 0.8). All participants were monolingual native speakers of German and had normal or corrected-to-normal vision. Children were recruited and tested in German kindergartens. Children received a small toy and a certificate for participating. Testing took approximately 30 to 45 minutes.

7.2.2 Materials and Cognitive Tests

Materials were identical to experiment 1. Prior to the eye-tracking study, participating children were tested for working, spatial and verbal memory using the digit span, the word order and the spatial memory test of the Kaufmann Assessment Battery for Children (K-ABC) for the appropriate age group.

This was done in order to ensure that on-line and off-line results of the eye-tracking study were not confounded by between-subject differences in cognitive performance due to a heterogeneous sample (e.g., children with low vs. high working, spatial and verbal memory). Administering these cognitive tests with children took about 5 minutes.

7.2.3 Hypotheses

Recall that previous research (Zhang & Knoeferle, 2012) has suggested that younger adults, but crucially also children can make use of depicted actions in order to quickly assign thematic roles and hence to facilitate OVS sentence processing. In study 4 we aimed to see whether we could replicate their results. We also wanted to investigate if children and adults differ as to when during real-time sentence processing the action cues will show an effect. Results from Zhang and Knoeferle (2012) suggested that children (vs. adults) are delayed by one word region in their action effects. Studies 1-3 have moreover suggested that adults can also make use of more social, indirect visual cues such as emotional facial expressions to facilitate sentence processing. However, the effect of social, indirect cues on children's sentence processing is

unclear. The next 2 studies thus investigated if 4-5-year old children can (like younger adults) also use emotional facial expressions to facilitate thematic role assignment.

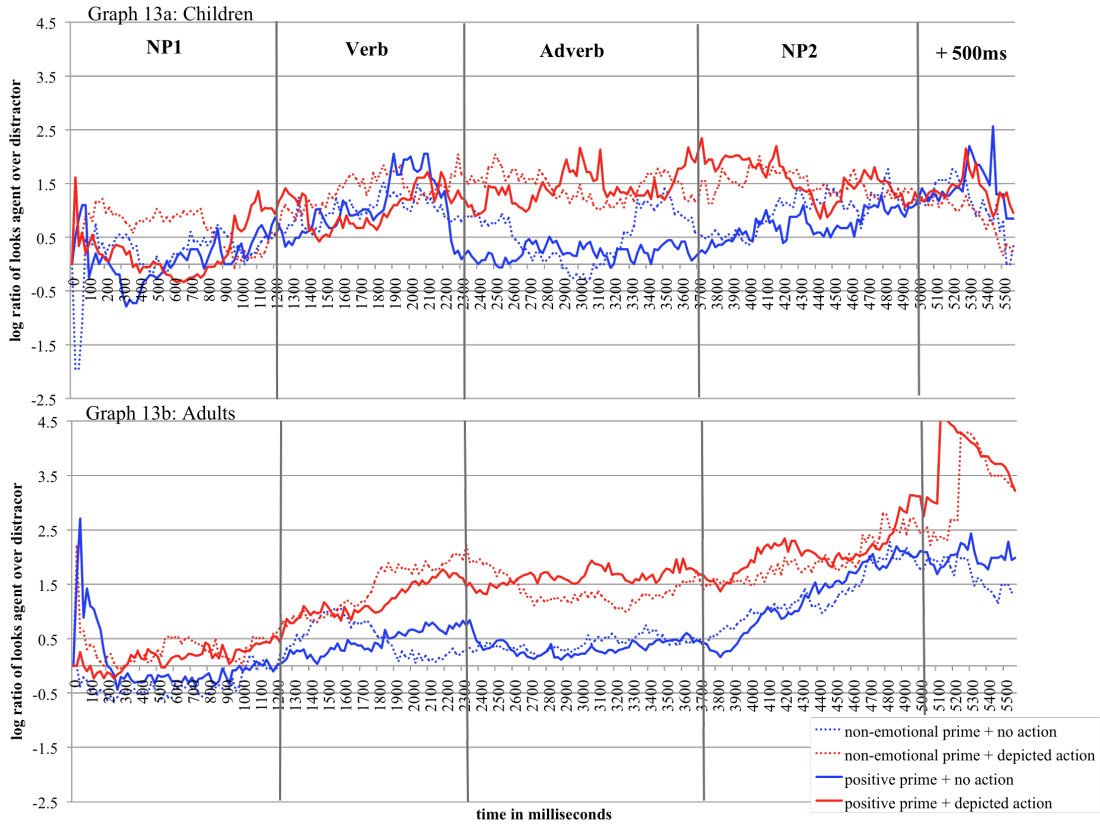
Moreover, taking children's positivity bias into account, we hypothesized that they might be able to integrate the positive prime face at least as well as (or better than) younger adults into sentence processing. On the other hand, children may not yet be able to use emotional facial expressions to boost sentence comprehension of difficult OVS sentences, since they are still in the process of developing both their skills to correctly identify and interpret emotional faces and their sentence processing competences, especially with regard to German OVS sentences. Study 4 used schematic facial expressions in order to facilitate emotion identification for children so that they might be able to quickly use it for thematic role assignment. Recall that the more complex a facial expression seems to be, i.e., for example fear or disgust, the worse children are at identifying and interpreting the emotional face (see Herba and Philips, 2004 for a comprehensive review and Section 3.7.1.3). Hence, by using a schematic smiley without complex physical features or movements we tried to simplify emotion recognition and interpretation for children.

7.2.4 Results and Discussion

7.2.4.1 Eye-movement Results

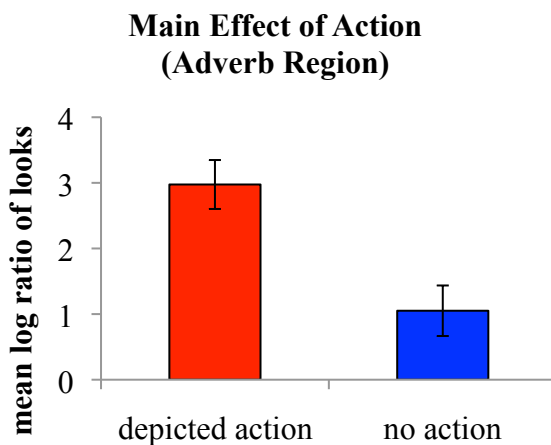
The analysis revealed a significant main effect of action in the Adverb ($F_1(1,23)=13.630$, $p < .05$, $F_2(1,15)=21.726$, $p < .05$), the Verb-Adverb ($F_1(1,23)=11.564$, $p < .05$, $F_2(1,15)=11.937$, $p < .05$) and the Long region ($F_1(1,23)=13.307$, $p < .05$, $F_2(1,16)=8.452$, $p < .05$). Recall that younger adults showed the first significant effect of action depiction in the Verb region. However, children's main effect of action only became significant in the Adverb and Verb-Adverb region. Thus, children were delayed by one word region in the effect of the depicted action on real-time sentence processing (Graph 13).

7. Experiments – Children



Graph 13: Time course for the eye-movement results by condition for children and younger adults (exp. 1, 4). Graph 13a: children’s time course (exp. 4); Graph 13b: younger adults’ time course (exp 1, see Graph 1).

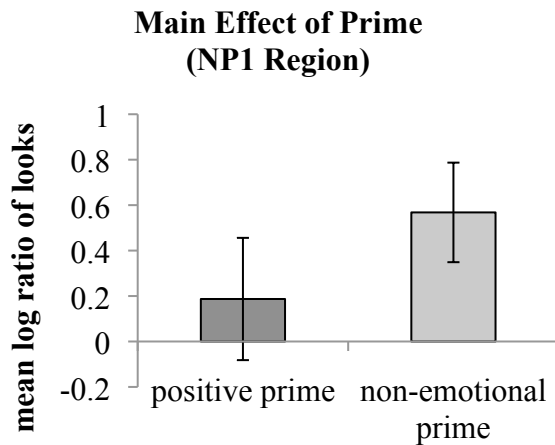
Graph 14 shows the effect of the depicted action (vs. no action) for children and demonstrates that 4-5-year old children anticipated the correct role filler significantly more when actions were depicted in the scene compared to when no actions were depicted. There were no earlier effects of action depiction.



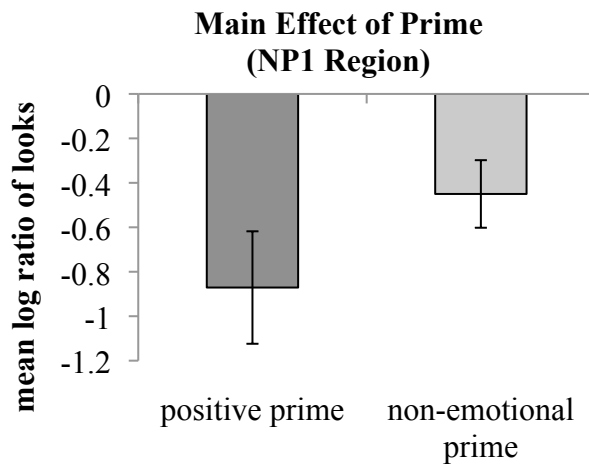
Graph 14: Mean log ratio of looks for the main effect of action in the Adverb region (F_1 , exp. 4). Error bars show the standard error.

The eye movement analysis furthermore revealed a marginal main effect of prime (in the item analysis only) in the NP1 region ($F_2(1,15) = 4.055$, $p = .062$, Graph 15). Interestingly, as Graph 15 shows, children seemed to fixate the target agent more in the non-emotional prime compared to the positive prime condition. However, this was the case in the NP1 region, thus when the patient, i.e., the object of the sentence, was mentioned¹⁶. Taking into account that both agent and patient portrayed a happy facial expression, children might have been primed by the happy prime smiley towards the patient (instead of the agent) when the patient was named. This could have led to reduced looks towards the (also smiling) agent in the positive prime (vs. the non-emotional prime) condition and thus resulted in the pattern displayed in Graph 15. In order to explore this assumption we additionally investigated the looks towards the patient (vs. looks towards the agent). We ran a repeated measure ANOVA for the NP1 region with prime as the fixed factor using the log ratio of looks towards the agent versus patient. Although the results did not yield a main effect of prime, the means clearly underline the assumption that children fixated the patient more than the agent when the prime smiley was presented (vs. the non-emotional prime). Graph 16 illustrates this gaze pattern. Hence, the marginal prime effect (Graph 15) was not due to children using the non-emotional prime more than the positive prime to anticipate the target agent. It rather seems to be the case that children were primed to some extent by the positive smiley towards the smiling patient mentioned by the NP1. This smiley - patient priming could explain why we see fewer looks towards the agent (vs. the distractor, Graph 15) in the positive (vs. non-emotional) prime condition and more looks towards the patient (vs. the agent, Graph 16) in the positive (vs. non-emotional) prime condition. No further effects of emotional prime were reliable. Hence, children do not yet seem to be able to use the indirect social cue to anticipate the correct role filler in real time.

16 Note that this cannot be seen in the time course graph (Graph 13) since all time course graphs depict the log ratio of looks to the agent over the distractor for all conditions from NP1 onset. However, the inferential analysis for the NP1 region only included the prime but not the action as a fixed factor (see Section 5.2.3)



Graph 15: Mean log ratio of looks (agent / distractor) for the main effect of prime in the NP1 region (F_2 , exp. 4). Error bars show the standard error.



Graph 16: Mean log ratio of looks for the effect of prime in the NP1 region (not sign., exp. 4). Displayed on the y-axis is the mean log ratio of looks to the agent over the patient. Negative values indicate a higher probability of looks towards the patient. Error bars show the standard error.

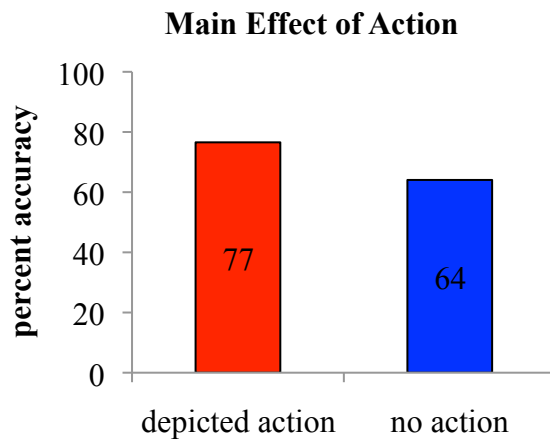
7.2.4.2 Accuracy Results

The results of the accuracy analyses underline the real-time data. The results showed a main effect of action ($F_1(1,23)= 4.793$, $p < .05$, $F_2(1,15)= 9.000$, $p < .05$, Graph 17), indicating that children answered significantly more comprehension questions correctly when an action was (vs. was not) depicted. The analyses with voice (active vs. passive) as a factor revealed a significant main effect of voice ($F_1(1,23)= 82.926$, $p < .05$, $F_2(1,15)= 117.195$, $p < .05$), and a significant voice x action interaction in the item analysis ($F_2(1,15)= 7.500$, $p < .05$). Graph 18 illustrates that the main effect of

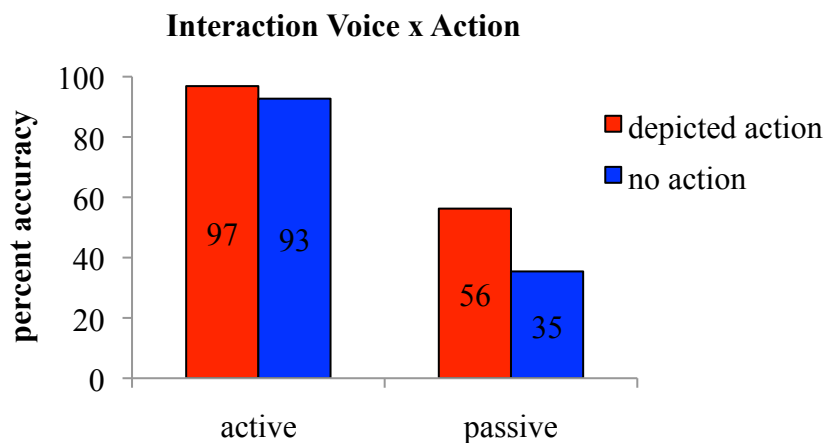
7. Experiments – Children

action was driven by the passive comprehension questions. Thus, when the comprehension questions were in the passive voice, they added difficulty to the task of assigning thematic roles and children used the direct cue to boost sentence comprehension. As Graph 18 indicates, children did not seem to have had difficulties in answering the comprehension question when it was asked in the active voice. This seemed to be the case regardless of action depiction, i.e., children answered 93% of the questions correctly in the no action condition. Yet, this ceiling effect might be explained by the way in which the active questions were asked. In contrast to the passive questions, the active questions mentioned the patient when asking ‘who does what to whom’. Thus, we cannot exclude that children simply concluded that the character-in-question must be the character that was not mentioned in the question (but that was named in the sentence, i.e., the agent). The accuracy scores for the active questions hence might have been overall lower if the patient hadn’t been named in the question.

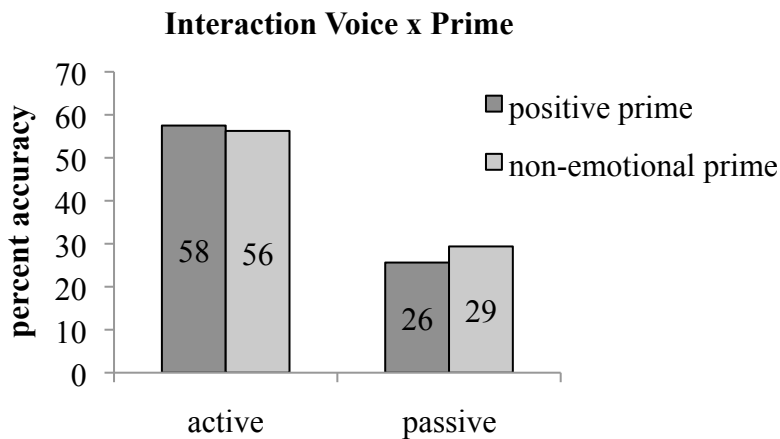
Interestingly, like in experiment 1 for the younger adults, we additionally found a marginal voice x prime interaction ($F_1(1,23)= 3.139$, $p= .09$, $F_2(1,15)= 7.737$, $p< .05$, Graph 19). However, as Graph 19 indicates, contrary to the younger adults who used the emotional prime to boost sentence comprehension in the passive voice (Graph 5), children seemed to have been hindered or confused by it as they performed worse when the emotional prime was presented and the questions were asked in the passive voice. Additionally, the results also yielded a marginal voice x prime x action interaction ($F_1(1,23)= 4.285$, $p= .05$) which was driven by the depicted action and the active voice (Graph 20) in contrast to adults voice x prime x action interaction which was driven by the active voice and crucially the effect of positive prime face in the passive questions (Graph 6) .



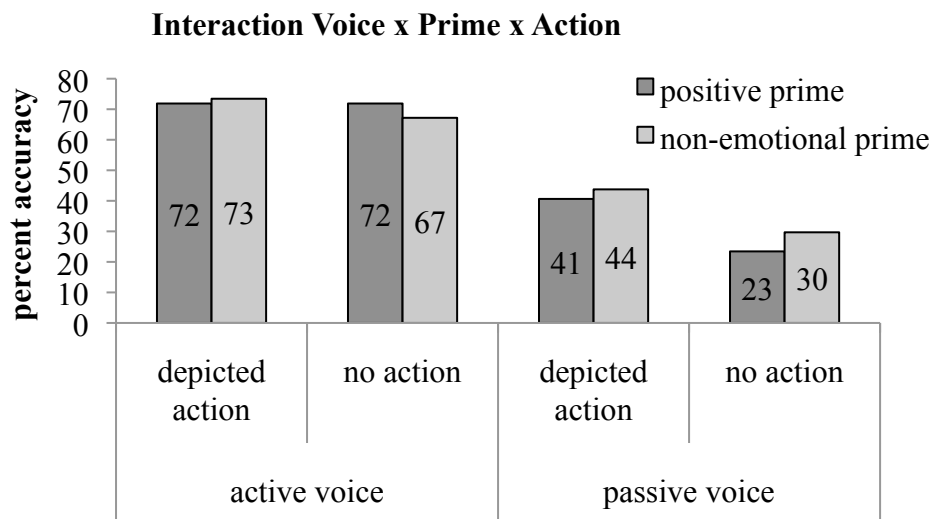
Graph 17: Accuracy for the main effect of action (exp. 4). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*



Graph 18: Accuracy for the voice x action interaction (exp. 4). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*



Graph 19: Accuracy for the voice x prime interaction (exp. 4). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*



Graph 20: Accuracy for the voice x prime x action interaction (exp. 4). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

7.2.4.3 Cognitive Test Results

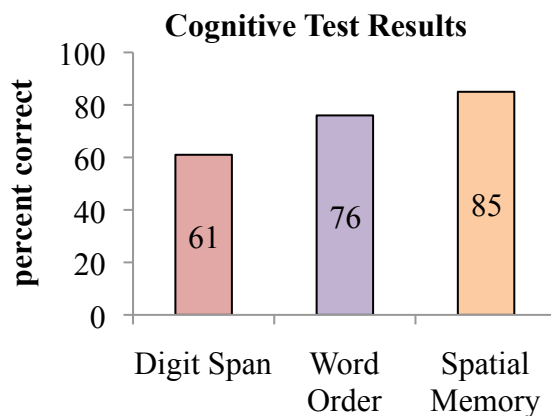
Children's test scores were highest for the spatial memory and lowest for the digit span test (see Graph 21). In order to see if their cognitive abilities correlated with the accuracy for the comprehension question, we computed a two-tailed Spearman's rho correlation coefficient¹⁷. The correlation between the K-ABC scores and the accuracy for the comprehension question was positive and significant ($r = .456$, $p < .05$).

¹⁷ K-ABC scores were normally distributed, accuracy scores were not.

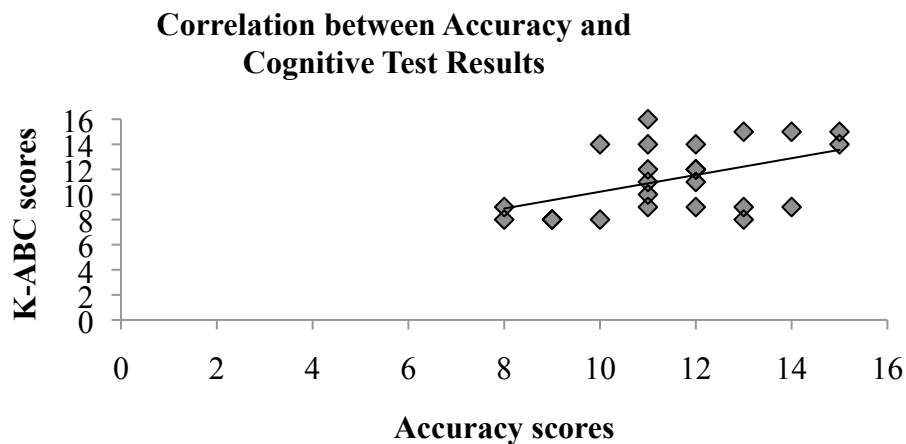
7. Experiments – Children

However, as Graph 22 shows, although the plotted trend line and the positive correlation might lead to a different assumption, the relationship between the cognitive test scores and the accuracy for the comprehension question is not entirely positive and linear. As can be seen in Graph 22 a high score in the cognitive tests does not necessarily also imply that children were more accurate in answering the comprehension question.

However, in order to test whether those children who scored high in the cognitive tests and whose accuracy for the comprehension question was high also showed stronger effects of depicted actions than low scoring children, we calculated the median for both the accuracy data and the K-ABC data separately. We entered these medians (binary variables) as between-subjects factors (low vs. high) into repeated measure ANOVAs for the Adverb region. This region was chosen because it was the first region that showed a significant main effect of action. However, both the analysis with the K-ABC median and the analysis with the accuracy median as grouping variables did not reveal any significant interactions with children's gaze patterns. Thus children's individual differences in cognitive abilities did not appear to modulate the accuracy and eye-tracking results of the experiment.



Graph 21: Cognitive test scores for the K-ABC test (exp. 4). *The y-axis displays the percentage of correct answers averaged across participants. The percentages are shown in the center of each bar.*



Graph 22: Correlation between K-ABC and accuracy scores (exp. 4). *K-ABC scores are displayed on the y-axis, accuracy scores are displayed on the x-axis. The trend line shows the line of best fit.*

7.2.5 Summary and Conclusion

To conclude, children could use the depicted action to improve OVS sentence comprehension and on-line thematic role assignment. Moreover, it seems as if they were using the depicted action in the same way as younger adults used the emotional prime, namely especially in the cases in which the structure of the comprehension question added difficulty to the task of understanding the OVS sentence. By contrast, the emotional prime in the form of a schematic smiley did not have an advantageous effect on children's on-line sentence comprehension and on their accuracy scores. It even seems that they were confused or hindered by the emotional smiley, since they performed worse when it was present (vs. absent) on comprehension questions in the passive voice. Children's confusion could also be the reason why we found a marginal main effect of prime going into the opposite-than-predicted direction in the NP1 region. Recall that the means of the agent-over-patient analysis showed that children's looks were primed towards the happy looking patient when the emotional prime was available and the patient was named. However, children did not appear to have been able to connect the positive prime face to the positively valenced adverb and the happy facial expression of the target agent to anticipate this agent and facilitate OVS sentence processing. Moreover, these results did not appear to have been modulated by individual differences in cognitive abilities.

7.3 Experiment 5

7.3.1 Participants

A further 40 children between 4 and 5 years of age (Mean: 4.37, SD: 0.49) participated in the experiment. All participants were monolingual native speakers of German and had normal or corrected-to-normal vision. Children were recruited and tested in German kindergartens. Children received a small toy and a certificate for participating; testing took approximately 30 to 45 minutes in total.

7.3.2 Materials and Cognitive Tests

Materials were identical to experiment 3 (i.e., only the depicted agent but not the distractor or patient had a positive facial expression). Cognitive tests administered prior to the eye-tracking part of the study were identical to experiment 4.

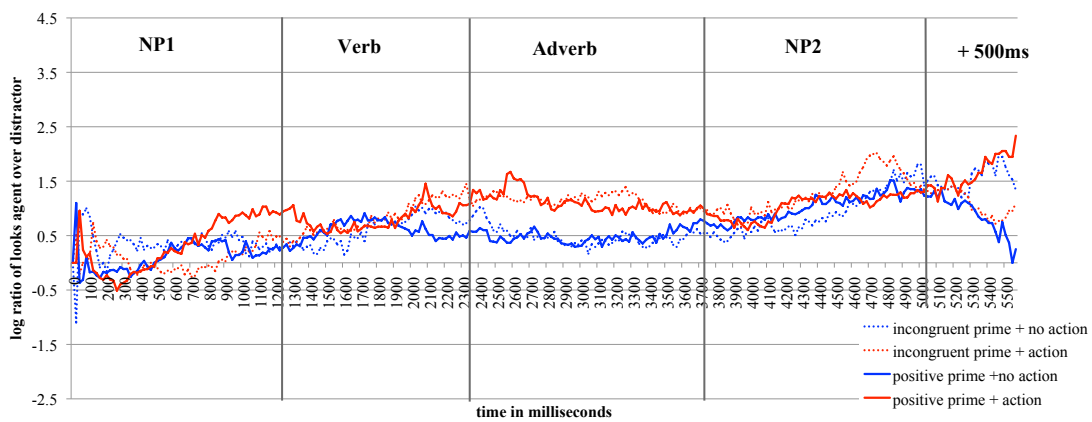
7.3.3 Hypotheses

Although experiment 4 demonstrated that children could not use the positive emotional prime in the form of a schematic smiley, this does not mean that they cannot use emotional facial expressions for thematic role assignment and OVS sentence processing at all. Perhaps the schematic nature of the emotional prime made it more difficult (vs. easier) for children to link the emotional valence of the prime to the matching happy facial expression of the agent and the positively valenced adverb of the sentence. Given the results for experiments 1, 2 and 3, we have seen that younger adults made stronger use of the positive prime if this prime was natural rather than schematic in nature. Hence, we hypothesized that even though natural facial expressions are more complex than a schematic smiley, children might be better able to use this natural emotional prime compared to the smiley, maybe precisely because it is more ecologically valid (see Section 3.7.2 and e.g., Bassili, 1979). Just like younger adults, children do interact with natural faces every day and they do so presumably more than with a schematic depiction of an emotional facial expression. As a result, their natural interaction experience might enable them to use natural positive emotional facial expressions to facilitate OVS sentence processing.

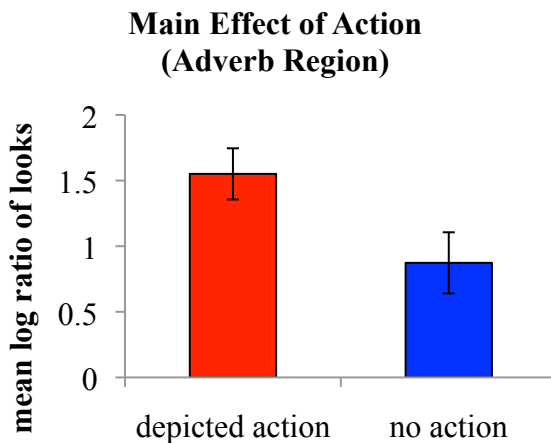
7.3.4 Results and Discussion

7.3.4.1 Eye-movement Results

As can be seen in the time course graph (Graph 23) by looking at the deviating red and blue lines, we replicated the significant main effect of action in the Adverb ($F_1(1,39)= 4.964, p < .05, F_2(1,15)= 9.159, p < .05$, see also Graph 24), the Verb-Adverb ($F_1(1,39)= 12.437, p < .05, F_2(1,15)= 29.667, p < .05$) and in the Long region ($F_1(1,39)= 15.544, p < .05, F_2(1,15)= 16.759, p < .05$).



Graph 23: Time course for the eye-movement results by condition (exp. 5).



Graph 24: Mean log ratio of looks for the main effect of action in the Adverb region (F_1 , exp. 5). Error bars show the standard error.

However, similar to experiment 4 we did not find a main effect of emotional prime in the eye-movement data, nor any significant interactions. Recall that we

observed a marginal prime effect in the opposite-than-expected direction in the NP1 region in experiment 4. We had interpreted this pattern as suggesting that children were primed by the positive smiley towards the patient; if so, more attention to the patient might have reduced anticipatory looks towards the target agent. Since we did not find this effect in experiment 5, we consider the change in materials successful (i.e., having the agent as the only smiling character likely focused effects of the positive prime on the agent).

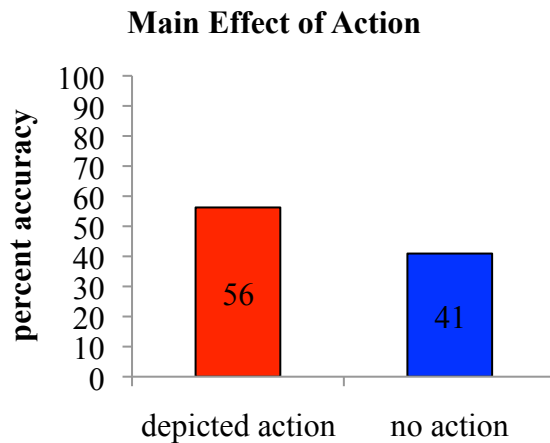
In summary, children (unlike younger adults) could not make use of the positive emotional prime face in real-time for thematic role assignment, regardless of whether it was natural (experiment 5) or schematic (experiment 4).

7.3.4.2 Accuracy Results

The accuracy analysis underlines the real-time data. The results showed a main effect of action ($F_1(1,39)= 15.714, p < .05, F_2(1,15)= 8.247, p < .05$), indicating that children answered significantly more comprehension questions correctly when an action was (vs. was not) depicted (Graph 25).

However, we did not find any effects for the emotional prime, nor any significant interactions. The absence of an emotional prime effect can moreover also not be due to participants not paying attention to, or not remembering, the facial expression since they answered 82.5 % of all face-recall questions correctly.

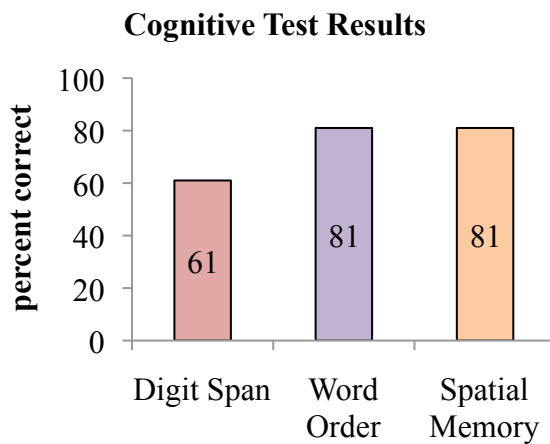
Note also that the percentages of correct answers were lower compared to the percentages from experiment 4 (see Graph 17). The reason for this is likely due to the fact that in experiment 4 the comprehension questions were asked in the active and in the passive voice, whereas they were only asked in the passive voice in this study (see also exp. 3). As the passive comprehension questions were more difficult to answer than the active questions (see Graph 17), accuracy decreased when all of the comprehension questions following a critical trial were asked in the passive voice (in contrast to both active and passive questions).



Graph 25: Accuracy for the main effect of action (exp. 5). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

7.3.4.3 Cognitive Test Results

Children's test scores were again (see exp. 4) lowest for the digit span test. However, they scored equally well in the word order and spatial memory test (Graph 26).



Graph 26: Cognitive test scores for the K-ABC test (exp. 5). *The y-axis displays the percentage of correct answers averaged across participants. The percentages are shown in the center of each bar.*

There was no reliable correlation¹⁸ between accuracy and cognitive test scores; hence no further analyses were performed.

¹⁸ Both accuracy and K-ABC scores are normally distributed, thus a Pearson correlation was performed.

7.3.5 Summary and Conclusion

To conclude, children, just like younger adults, made extensive use of the direct action cue for anticipatory thematic role assignment. However, compared to adults they seemed to be one word region delayed in integrating the depicted action cue into real-time sentence processing. While adults showed effects of action depiction already at the Verb region – the first word region in which it was possible to unambiguously discriminate the agent of the action in the scene, if an action was depicted – children only showed the first significant effects in the Adverb and Verb-Adverb region in both study 4 and in study 5. These on-line effects were underlined by the accuracy results (a 13-16 percent boost in studies 4 and 5).

Crucially though, in contrast to younger adults, 4-5-year old children did not yet seem to use the positive emotional facial prime to facilitate OVS sentence processing. That children did not use the positive prime was interestingly the case regardless of the prime's naturalness. Moreover, how, if and when the visual context affects language processing does not seem to depend on the children's individual cognitive abilities.

8. Experiment – Older Adults

8.1 Experiment 6

8.1.1 Participants

40 adults between 60 and 80 years of age (Mean: 67.1, SD: 6.1) participated in the experiment. All participants were monolingual native speakers of German and had normal or corrected-to-normal vision. Older adults were recruited by flyers distributed in the University, the city center of Bielefeld and senior centers in Bielefeld. Testing took place in the eye-tracking laboratory of the Language and Cognition Group. They received 10 Euro for their participation and testing took approximately 60 minutes in total.

8.1.2 Materials and Cognitive Tests

Materials were identical to experiment 3 (i.e., only the target agent smiled; the patient had a neutral expression and the distractor character a negative expression, the prime face was natural). For the older participants, additional cognitive tests were administered prior to the eye-tracking study. These tests were selected subtests of the Wechsler Intelligenztest für Erwachsene (WAIS, Wechsler, 1981). Cognitive performance was measured by using the picture completion, the digit-symbol mapping, the digit span, the similarities subtest (Wechsler, 1981) and a verbal fluency test (a subtest from Aschenbrenner, Tucha, & Lange, 2000). The latter required participants to name as many examples of a category (animals) as possible and as many words with an initial letter (the letter 'L') as possible within one minute. All subjects were given the same tests in the same order and with the same time limits to be completed. The cognitive tests ensured that on-line and off-line results of the eye-tracking study were not confounded by between-subject differences in cognitive performance, such as verbal, working and spatial memory. Administering the cognitive tests took approximately 15 minutes.

8.1.3 Hypotheses

Taking older adults' strong bias towards positive emotions into account (see Section 3.7.1.2), we hypothesized that this age group would show stronger or earlier effects of the emotional prime in real-time sentence processing and stronger effects on question answering. Effects of the emotional prime were hence predicted to arise earlier during sentence processing compared to the younger adults who showed the first effects of the positive prime in the Adverb region. Moreover, since the prime effects were only marginal in younger adults, we predicted stronger effects here.

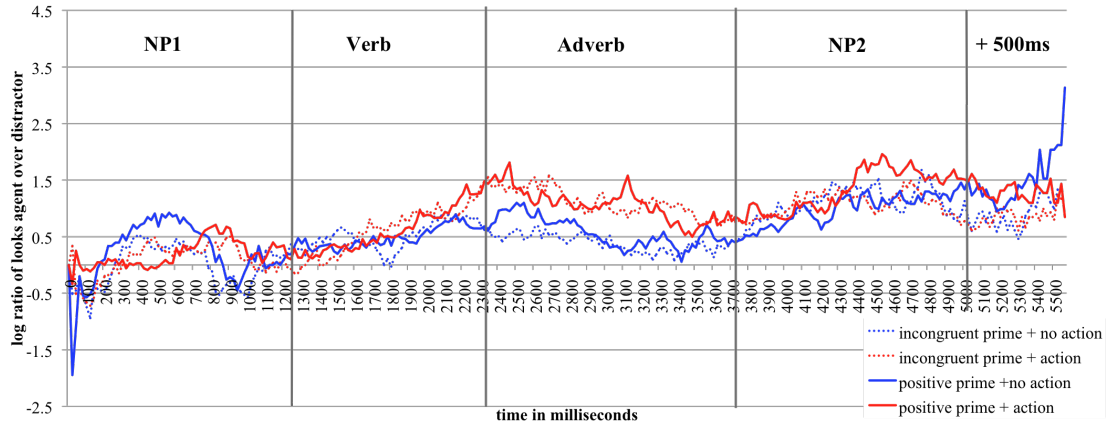
Regarding the depicted action, we did not predict differences from the younger adults in the use of the direct cue for OVS sentence processing and thematic role assignment. Nevertheless, regarding older adults general decline in cognitive functions (Salthouse, 2010), it was also possible that the visual context effects emerged later during on-line sentence processing compared to younger adults.

8.1.4 Results and Discussion

8.1.4.1 Eye-movement Results

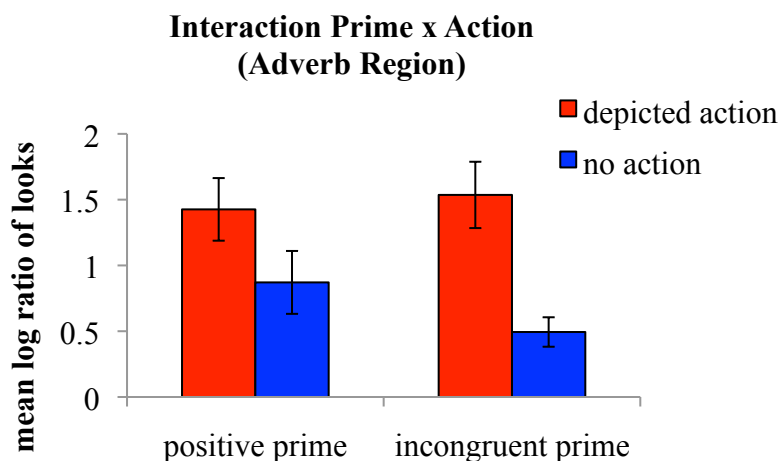
The results for the older adults revealed a main effect of action in the Adverb ($F_1(1,39)= 17.896, p< .05, F_2(1,15)= 3.166, p= .095$), Verb-Adverb ($F_1(1,39)= 16,398, p< .05, F_2(1,15)= 4.097, p= .061$) and the Long region ($F_1(1,39)= 2.797, p= .102$). This effect can be seen in Graph 27, which also shows that the action effect only emerged at the end of the Verb region, i.e., where the blue and the red lines start to diverge. Hence, older adults were delayed by one word region compared to the younger adults (see Graph 7), i.e., the effect emerged at the same time as the action effect for the children (see Graph 23).

8. Experiment – Older Adults



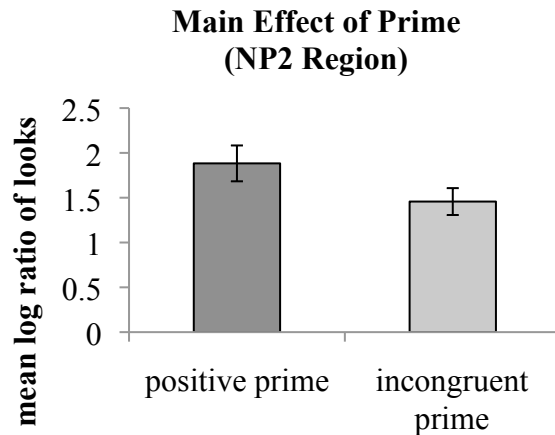
Graph 27: Time course for the eye-movement results by condition (exp. 6).

Interestingly, the interaction between action and prime was trending in the subject analysis in the Adverb region ($F_1(1,39) = 2.693, p = .109$). Graph 28 illustrates that this interaction was driven by the depicted action. However, it suggests that older adults also seem to have made use of the positive prime face even when no action was present compared to the no-cue condition. In the time course graph (Graph 27) this can be seen right at the beginning of the Adverb region, i.e., when the emotional valence of the adverb became available (where the solid blue line diverges from the dotted blue line). We have to keep in mind however, that the failure to find a significant interaction might not just be due to the weak effects of the emotional prime but also to the fact that there were only 4 data points per subject, i.e., 160 in total that could be used to yield an interaction between prime and action.



Graph 28: Mean log ratio of looks for the interaction between prime and action in the Adverb region (not sign., exp. 6). Error bars show the standard error.

Crucially, the results yielded a marginal main effect of prime in the subject analysis of the NP2 region ($F_1(1,39)= 3.785, p= .059$). Graph 29 indicates that older adults showed a preference to look at the target agent when they had been primed with a positive facial expression compared to an incongruent prime.



Graph 29: Mean log ratio of looks for the main effect of prime in the NP2 region (F_1 , exp. 6). Error bars show the standard error.

However, as this effect only emerged in the NP2 region, older adults do not seem to have made use of the positive prime face to anticipate the correct role filler. As the time course graph (Graph 27) shows, the solid lines (the prime condition) only diverge late during the NP2 region from the dotted lines (the incongruent prime condition). Hence, it rather seems that older adults kept the emotional prime face in memory and reactivated its valence again when they heard the target agent named (perhaps to match the prime face valence with the positively congruent facial expression of the target agent).

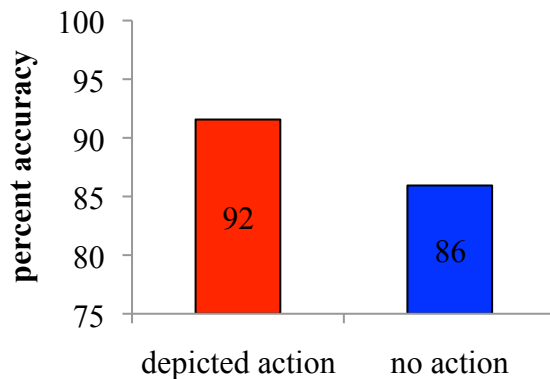
However, regarding the delay for the action effect compared with younger adults, older adults may simply integrate visual cues more slowly than younger adults. For both the action and the prime effect, they were delayed by one word region.

8.1.4.2 Accuracy Results

The accuracy results for the older adults yielded a marginal main effect of action ($F_1(1,39)= 3.162, p= .083, F_2(1,15)= 9.567, p< .05$). The means indicate that older adults answered more comprehension questions correctly when an action was (vs. was not) depicted in the scene (Graph 30). By contrast, the positive prime did not have an

effect on the accuracy results. Nevertheless, they answered 91% of the face-recall questions correctly suggesting they had no problems in recalling the valence of the emotional prime face. These results thus underline the eye-tracking results regarding the facilitative effect of the depicted action. However, the emotional prime face did not seem to improve participants' accuracy in answering the comprehension question.

Main Effect of Action

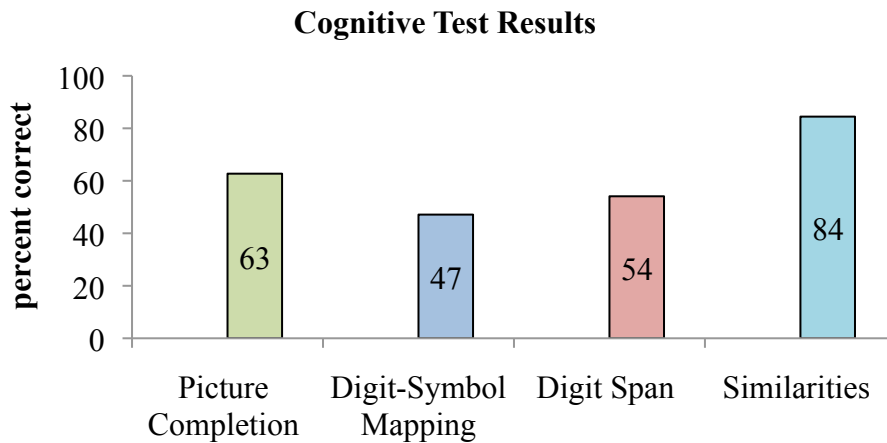


Graph 30: Accuracy for the main effect of action (exp. 6). *The y-axis displays the percentage of correctly answered trials averaged across participants. The percentages are shown in the center of each bar.*

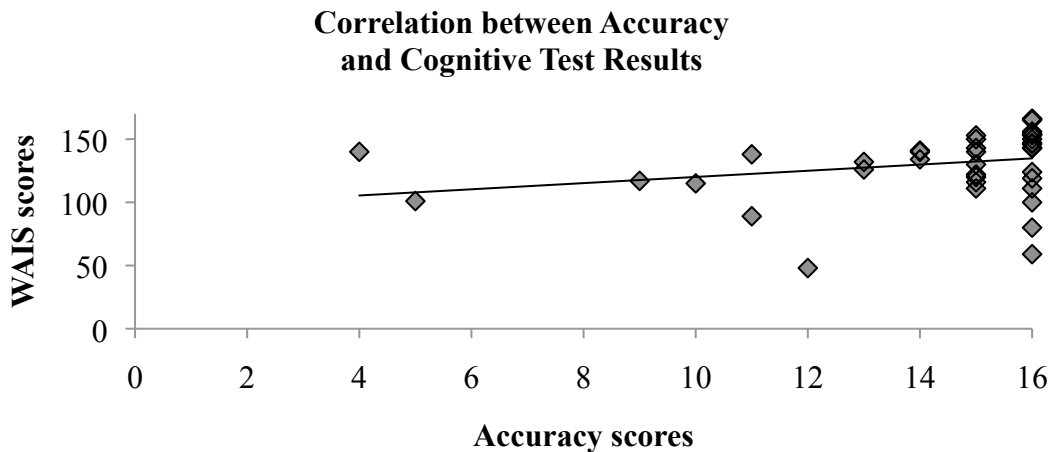
8.1.4.3 Cognitive Test Results

Older adults' cognitive test scores are shown in Graph 31. As can be seen in this graph, they performed best in the Similarities task and worst in the Digit-Symbol Mapping task. The mean for the Verbal Fluency task was 38.825 generated words. In order to test whether accuracy results and the cognitive test scores correlate with each other, we computed Spearman's rho¹⁹. The result yielded a medium-size but significant correlation ($r_s = .372$, $p < .05$). Graph 32 illustrates this correlation and suggests that although most subjects performed at ceiling in answering the comprehension question, high accuracy scores do not necessarily also have to imply a high performance in the cognitive tests.

¹⁹ Spearman's rho was chosen over Pearson's correlation because both variables were not normally distributed.



Graph 31: Cognitive test scores for the WAIS test (exp. 6). *The y-axis displays the percentage of correct answers averaged across participants. The percentages are shown in the center of each bar. Note that the verbal fluency test scores are not depicted. Since the task was free naming, there is no upper limit that can be reached.*



Graph 32: Correlation between WAIS and accuracy scores (exp. 6). *WAIS scores (including the verbal fluency task scores) are displayed on the y-axis, accuracy scores are displayed on the x-axis. The trend line shows the line of best fit.*

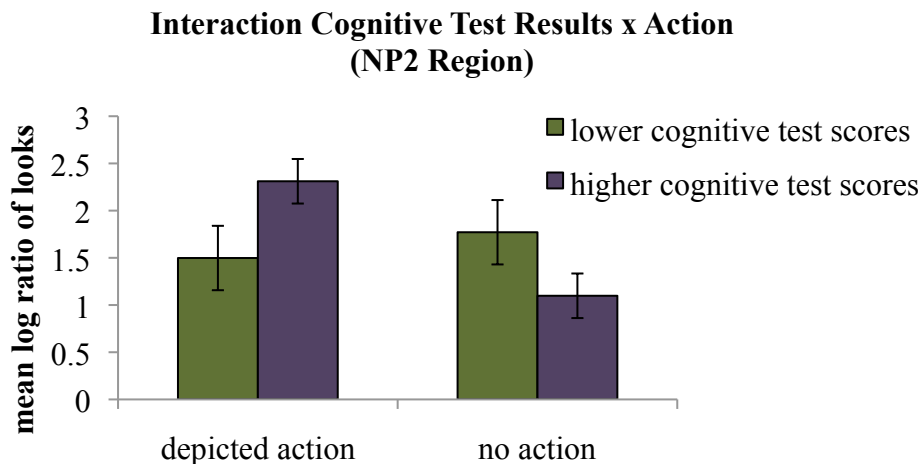
However, in order to further explore whether the cognitive test scores and the accuracy scores had an influence on the results of the real-time data, we computed the medians for the WAIS and the accuracy scores and entered the resulting binary between-subjects grouping variables ($N=20$) into the subjects analyses of the Adverb and the NP2 region. The Adverb region was chosen because it reveals the first main effect of action and could hence show an influence of individual differences in accuracy and WAIS scores on the use of the depicted action. The NP2 region was chosen because it was the only region in which a prime effect emerged. Hence, we

8. Experiment – Older Adults

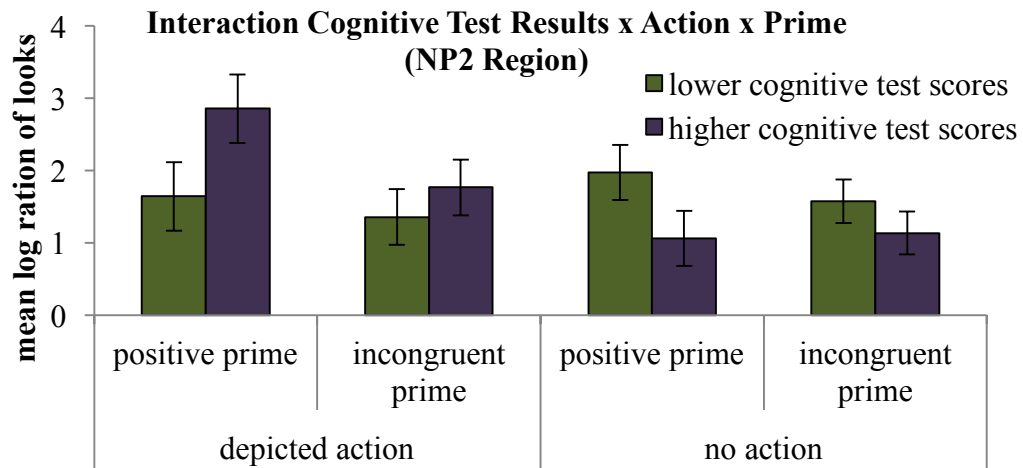
wanted to see if this effect was influenced by individual differences in cognitive abilities and accuracy for answering the comprehension question.

The results suggested that neither the WAIS scores nor the accuracy scores interacted with the action effect in the Adverb region. Interestingly, we found a significant action x WAIS interaction for the eye-movement data in the NP2 region ($F(1,39)= 5.921, p< .05$). The WAIS scores did however not interact with the prime effect. The accuracy scores also did not interact with either the prime or the action effect in the NP2 region. The significant action x WAIS interaction in the gaze data is illustrated in Graph 33. It indicates that especially older adults with higher (than the median) WAIS scores seemed to direct more attention to the agent when an action was depicted (vs. was not). Subjects with lower cognitive test scores looked slightly more at the agent when no actions were depicted (vs. depicted action). Hence, only the high scoring older adults seemed to be able to use the depicted action to direct more attention to the agent. That said, even though the three-way interaction between WAIS scores, prime and action was not significant, the means displayed in Graph 34 additionally suggest that older adults with higher cognitive abilities fixated the agent the most when both the direct and the indirect cue were available for use (compared to single and no-cue conditions).

Thus, older adults' individual differences in cognitive abilities appear to influence to which extent they use cues for on-line sentence processing. Perhaps then the higher older adults' cognitive abilities are, the more or the more distinct (direct and indirect) cues they can integrate at the same time during on-line sentence processing.



Graph 33: Mean log ratio of looks for the interaction between cognitive test results and action in the NP2 region (exp. 6). Error bars show the standard error.



Graph 34: Mean log ratio of looks for the interaction between cognitive test results, action and prime in the NP2 region (exp. 6). *Error bars show the standard error.*

However, especially when the medians of the WAIS and accuracy scores were included as a between-subject variable in the analyses, only 20 participants per group were used to determine the interaction effects, i.e., the analyses were based on 8 (scenes depicting an action) x 20 (participants per WAIS group) = 160 data points in total. This number is reduced even further to 80 in the three-way interaction. Hence, these statistical results have to be interpreted with care regarding a possible Type II Error.

8.1.5 Summary and Conclusion

To conclude, older adults could make use of the depicted action to anticipate the correct role filler. They moreover used this direct cue in answering the comprehension question. Crucially, they were also able to use the positive emotional prime face for OVS sentence processing. However, they did not seem to use this indirect cue in an anticipatory manner. Rather, it only affected their gaze pattern as the target agent was mentioned. Interestingly, the extent to which the visual direct cue could be used for on-line sentence processing seemed to depend to some extent on the older adults' cognitive abilities.

9. Age Comparisons for Experiments 3, 5 and 6

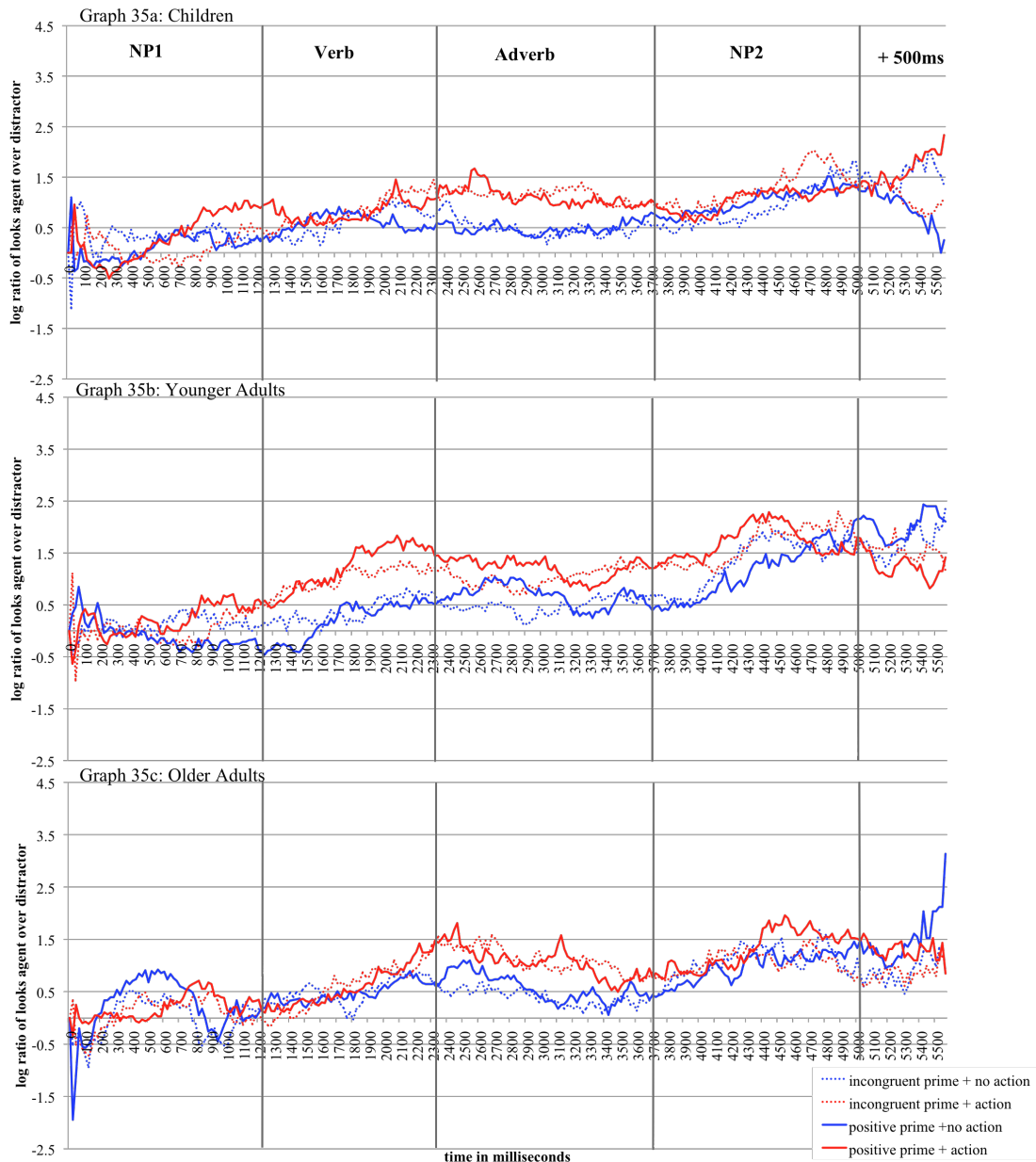
In order to compare age-related differences in direct and indirect cue effects in more detail, we conducted repeated measure ANOVAs with age as a between-subjects factor for the subject analysis and as a within-subjects factor for the item analyses for the Verb, Verb-Adverb, NP2 and Long region. A first analysis included all three age groups (children vs. younger adults vs. older adults). Follow up analyses compared the fixation pattern of younger with that of older adults (Verb, Adverb, the NP2 and the Long region) as well as of the gaze pattern of children and younger adults (Verb, the Adverb and the Long region). We also compared effect sizes for the prime and action effects, using age as a between-subjects factor. The preNP1 and the NP1 regions were not considered, since no crucial effects were found overall in these regions. In addition, age was added as a factor in the accuracy analysis.

9.1 Results and Discussion

9.1.1 The Action Effect

As the time course graph shows descriptively (Graph 35), the action effect was statistically significant, robust and persistent in all age groups. Thus, the depicted action could be used to anticipate the correct role filler. More interestingly however, the above mentioned age group comparisons confirmed that children as well as older adults started using the depicted action to anticipate the agent of the scene one word region later than the younger adults, namely in the Adverb (compared to the Verb) region.

9. Age Comparisons for Experiments 3, 5 and 6

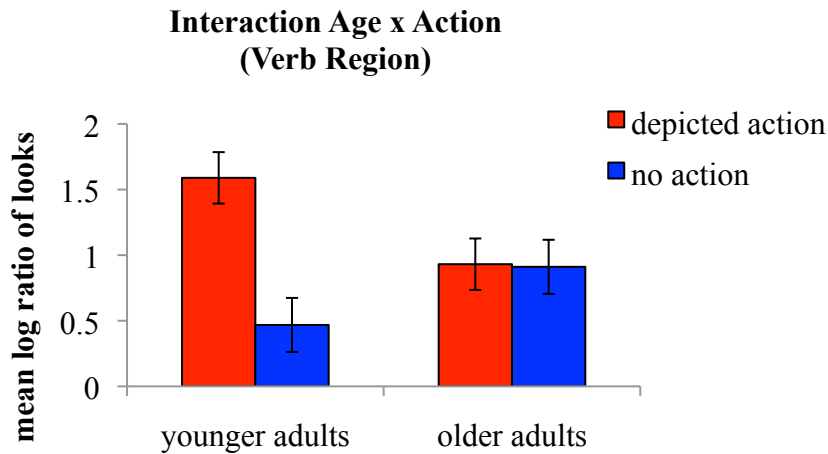


Graph 35: Time course for the eye-movement results by condition for each age group (exp. 3, 5, 6). Children's time course (exp. 5): Graph 35a; younger adults' time course (exp. 3): Graph 35b; older adults' time course (exp. 6): Graph 35c. The individual word regions are marked on the top, the conditions on the bottom.

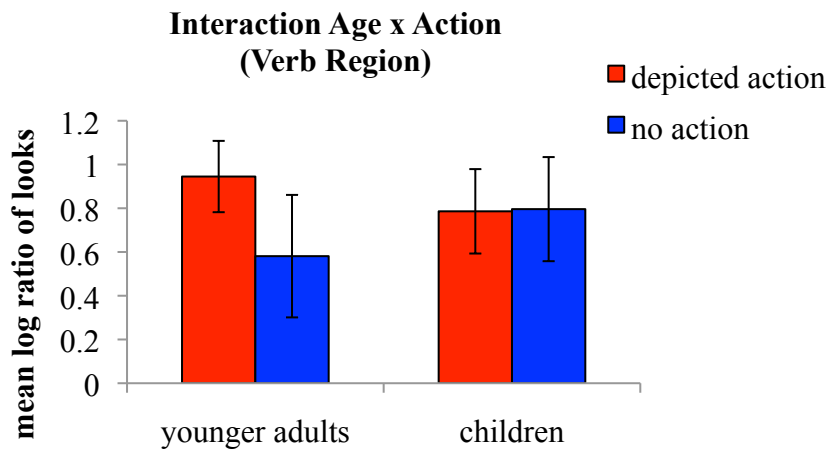
The significant and marginal action x age interactions in the Verb region for the analyses comparing younger versus older adults ($F_1(1,78)= 24.315, p< .05, F_2(1,15)= 4.756, p= .081$, Graph 36) and younger adults versus children ($F_2(1,15)= 3.19, p= .094$, Graph 38) can also be seen in the time course graph (Graph 35) where the red (depicted action) and the blue (no action) lines start to diverge only at the end of the Verb region for older adults and children. This gaze pattern is also illustrated in Graphs 36 and 37. The displayed mean log ratio of looks clearly shows that younger

9. Age Comparisons for Experiments 3, 5 and 6

adults used the depicted action to anticipate the agent whereas older adults and children did not yet use it in the Verb region. Additionally, this difference in timing is also supported by the fact that the action x age interactions were no longer significant in the Adverb region for both the analysis comparing younger vs. older adults and the analysis comparing younger adults vs. children.



Graph 36: Mean log ratio of looks for the interaction between age (older vs. younger adults, exp. 6 vs. exp. 3) and action in the Verb region (F_1). Error bars show the standard error.



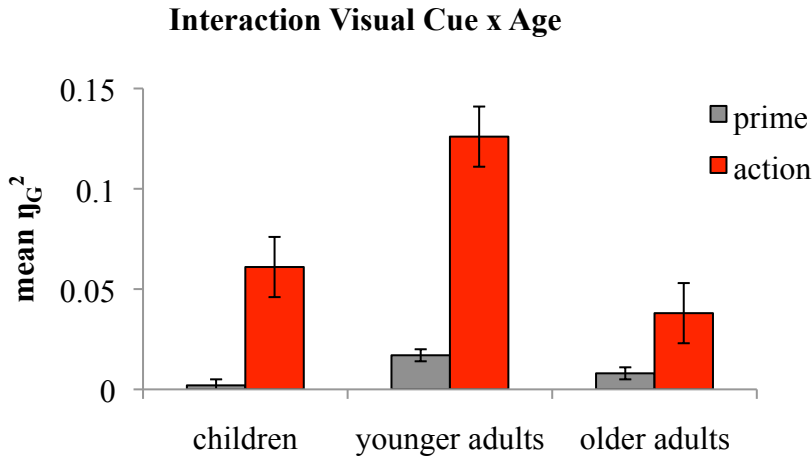
Graph 37: Mean log ratio of looks for the interaction between age (children vs. younger adults, exp. 5 vs. exp. 3) and action in the Verb region (F_2). Error bars show the standard error.

Crucially, younger adults also used the direct cue (vs. the indirect cue) more than older adults and children across the sentence, i.e., in all analyzed word regions during on-line sentence processing, as becomes evident by a significant visual cue x age

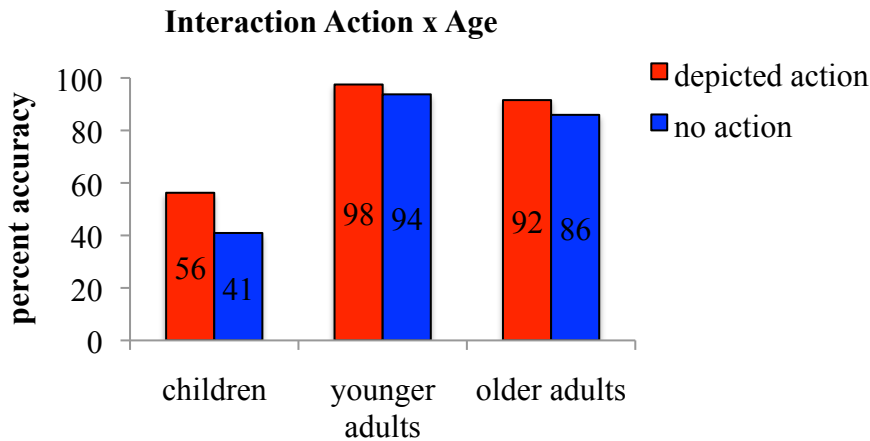
9. Age Comparisons for Experiments 3, 5 and 6

effect size interaction ($F(1,27)= 6.873, p< .05$). This interaction is illustrated in Graph 38.

Interestingly, the accuracy analysis also revealed a significant age x action interaction. Off-line, it seems that the depicted action was most helpful for the children to determine ‘who does what to whom’. This interaction is illustrated in Graph 39. A possible reason for that might be a ceiling effect for the younger and older adults, concealing the action effect.



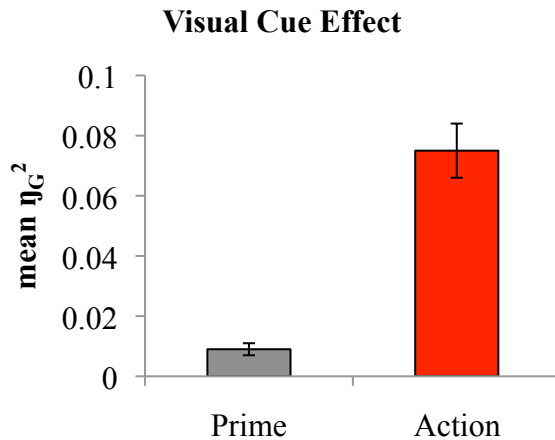
Graph 38: Mean generalized eta squared for interaction between the visual cue effect (action and prime) and age of the real-time data (exp. 3, 5, 6). The mean effect size is displayed on the y-axis. Error bars mark the standard error.



Graph 39: Accuracy for the action x age interaction (exp. 3, 5, 6). The y-axis displays the percentage of correct answers averaged across participants. The percentages are shown in the center of each bar.

Effect size comparisons of the real-time data across all three age groups moreover indicated that the action was a stronger cue than the prime ($F(1,27)= 57.454, p< .05$).

This effect is illustrated by the mean generalized eta squared effect sizes for the two cues in Graph 40.

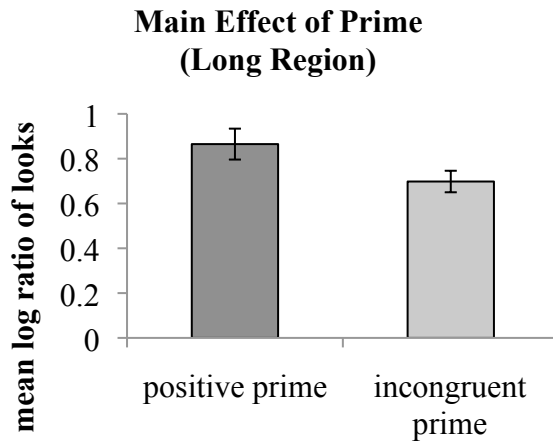


Graph 40: Mean generalized eta squared for the visual cue effects (action and prime) of the real-time data (exp. 3, 5, 6). *The mean effect size is displayed on the y-axis. Error bars mark the standard error.*

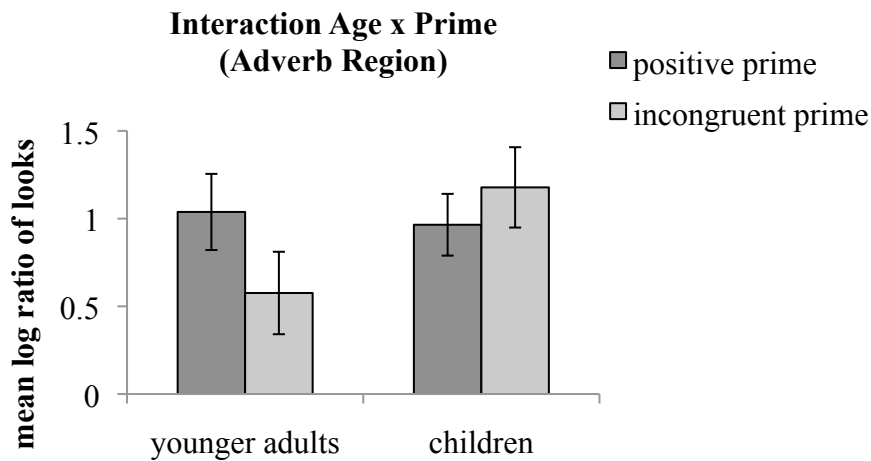
9.1.2 The Positive Prime Effect

However, even though the emotional prime seems to be a weaker cue than the depicted action, the former was also used for on-line sentence processing. Yet, this was only the case for the adults. That the younger and older adults could also use the emotional prime is underlined by the fact that the marginal main effect of prime ($F_1(1,117)= 3.571, p= .061$) for the Long region, i.e., across the whole sentence and across all three age groups became significant in the subject analysis when only the younger and older adults were taken into account in the analysis ($F_1(1,78)= 4.868, p< .05$). This main effect of prime for younger and older adults is illustrated in Graph 41. There was however no main effect of prime in the Long region for the analysis comparing younger adults with children.

Furthermore, the marginal age x prime interaction ($F_2(1,15)= 3.136, p= .097$) for the Adverb region in the analysis comparing younger adults versus children underlines the assumptions from experiment 4, namely that children seem to have been rather hindered or confused by the positive prime face. Graph 42 shows this interaction and demonstrates that both younger adults and children looked at the agent almost to the same extent in the positive prime condition. Yet, while adults fixated the agent much less in the incongruent prime condition, children fixated the correct role filler even more than in the positive prime condition.



Graph 41: Mean log ratio of looks for the main effect of prime in the Long region (F_1 , exp. 3, 6). Error bars show the standard error.

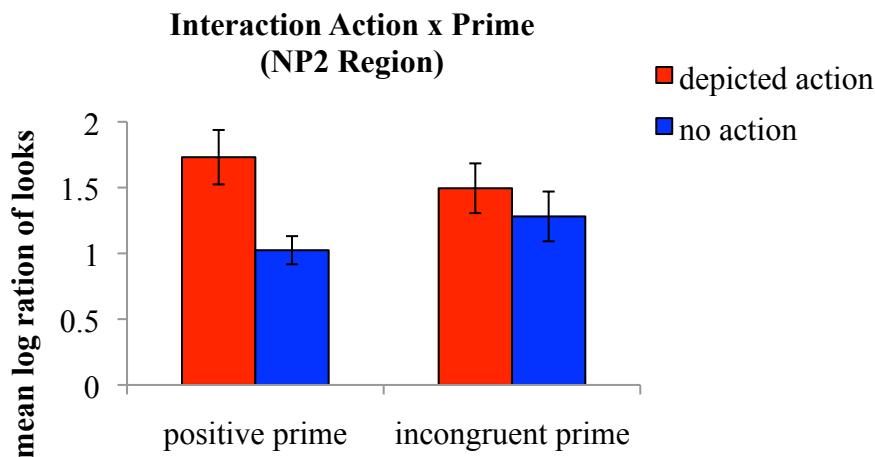


Graph 42: Mean log ratio of looks for the interaction between age and prime in the Adverb region (F_2 , exp. 3, 5). Error bars show the standard error.

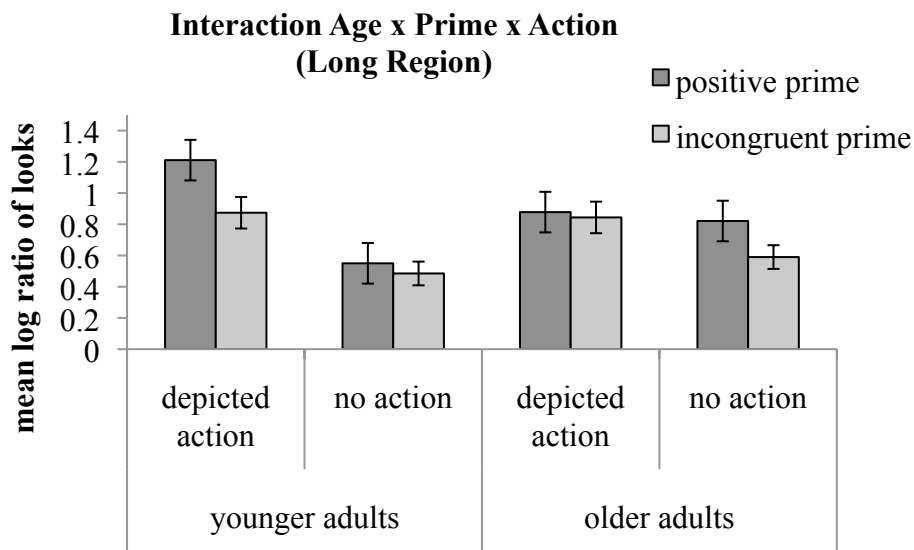
Hence, younger and older adults but not children were able to use the indirect emotional cue for on-line sentence processing. While older adults seemed to only be able to use the positive prime face for sentence processing when the link between the target agent and the positive prime was explicitly established via the linguistic input, i.e., when the agent was named, younger adults could also use the positive prime face proactively, i.e., to anticipate the correct role filler before it was mentioned in the sentence. Moreover, younger adults also seemed to be able to use the indirect cue for off-line sentence comprehension (cf., Section 6.4.4.2).

9.1.3 Cumulative Effects

Additionally, we found a marginal prime x action interaction in the NP2 region when comparing younger with older adults ($F_2(1,15)= 4.333$, $p= .055$). This interaction is displayed in Graph 43 and shows that the agent was fixated the most (compared to the other conditions) when both the direct and the indirect cues were available for use. This interaction was only marginal in the NP2 region and this could be due to timing differences in positive prime use. While younger adults used the positive prime to anticipate the agent in the adverb region, older adults used it one word region later, namely in the NP2 region when the agent of the action was mentioned (as the subject) in the sentence. Moreover, younger adults made stronger use of both cues at the same time across the whole sentence. This cumulative effect becomes evident by the marginal age x prime x action interaction in the Long region ($F_1(1,78)= 3.164$, $p= .079$) and is illustrated in Graph 44.

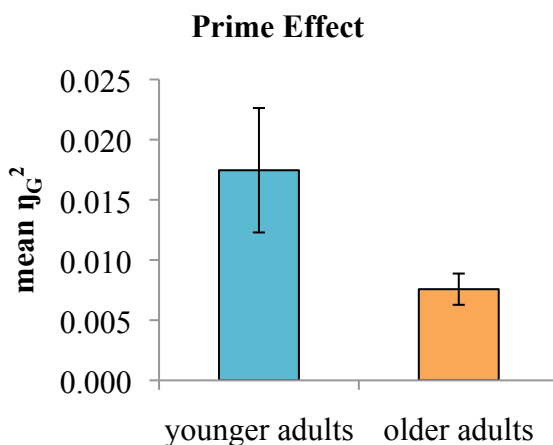


Graph 43: Mean log ratio of looks for the interaction between action and prime in the NP2 region (F_2) for younger and older adults (exp. 3, 6). Error bars show the standard error.



Graph 44: Mean log ratio of looks for the interaction between age, action and prime in the Long region (F₁) for younger and older adults (exp. 3, 6). Error bars show the standard error.

Furthermore, younger adults made more use of the indirect cue compared to older adults in real time across the sentence. Although a paired two-tailed t-test for the prime effect sizes of younger versus older adults was not significant ($t = 1.625, p = .139$), the means (Graph 45) underline this assumption. Hence, the action x prime interaction (Graph 43) presumably resulted from younger adults' stronger use of the prime across the sentence and older adults specific use of the prime in the NP2 region.



Graph 45: Mean generalized eta squared for the prime effect of the real-time data (exp. 3, 6). The mean effect size is displayed on the y-axis. Error bars show the standard error.

9.2 Summary and Conclusion

To conclude, all three age groups could use the depicted action to anticipate the correct role filler and to determine ‘who does what to whom’. Older adults and children were however, delayed by one word region compared to younger adults in using the depicted action for OVS sentence processing. Only older and younger adults could also make use of the emotional cue for OVS sentence processing. By contrast, the 4-5-year old children did not yet seem to be able to do so.

However, while younger adults could use the positive prime to anticipate the correct role filler, older adults were again one word region delayed for the prime effect. Hence, they did not use the indirect cue to anticipate the target agent in the scene but seemed to reactivate it when the subject of the sentence was named. Moreover, while younger adults could also use the positive prime face to improve the performance for the off-line comprehension question, older adults only used the positive prime during real-time sentence comprehension.

Additionally, the results yielded some evidence for the assumption that the use of both cues together led to an even stronger anticipation of the agent compared to the single cue conditions (based on higher log ratio of looks). However, the results suggested that the direct cue, i.e., the depicted action, was a stronger cue than the indirect cue, i.e., the emotional facial expression. Additionally, younger adults seemed to make more use of the indirect cue than older adults for on-line thematic role assignment.

10. General Discussion

Six visual-world eye-tracking studies assessed to which extent comprehenders of different age groups can use depicted actions and schematic / natural positive facial emotions for understanding positively valenced non-canonical OVS sentences. Moreover, we investigated whether age (children vs. younger vs. older adults) can be a predictor of time course differences for visual context effects.

Given the strong and robust action effect across all studies and age groups, we will discuss this effect first. Following this, we will address the subtler but persistent effect of the emotional prime face. We will argue in favor of the ecological validity of dynamic natural facial emotional expressions over schematic emotion depictions as experimental stimuli. Additionally, we will discuss why our results suggested that it is important to distinguish different visual context effects 1) given that we (in contrast to e.g., Kreysa et al., in prep.) found cumulative effects of the direct and the indirect cue and 2) given that the direct cue seems to be a much stronger cue than the indirect cue. Furthermore, we will address the findings that older and younger adults, but not children were able to use the emotional prime face for real-time OVS sentence processing. Moreover, the interesting time course differences for the visual cue effects between the age groups will be discussed. Crucially, following this we will relate our findings to the Coordinated Interplay Account (Knoeferle and Crocker, 2007) and will argue for an extension of the account to also include visual social aspects and listener characteristics. Using our own findings, we will demonstrate how such an extension could look like. We will conclude this thesis with a discussion about the impact of this PhD project in Section 11.

10.1 The Action Effect

Across all six experiments, participants made extensive use of the depicted action (vs. no action) for on-line sentence processing. Moreover, this effect was anticipatory in nature, meaning that younger and older adults, as well as children anticipated the correct role filler and hence facilitated OVS sentence processing in real time when the agent depicted in the scene was performing (vs. was not performing) the action mentioned in the sentence. Even though younger adults did not use the depicted action to also improve answering the post-sentence comprehension question for ‘who is doing what to whom’ (except in experiment 3), older adults’ and crucially also children’s accuracy improved

when actions were (vs. were not) depicted in the scene. We hence replicated findings by Zhang and Knoeferle (2012) who also found an on- and offline advantage of depicted over non-depicted action events for children, and on-line but no off-line effects of event depiction for younger adults. The missing off-line effect of depicted (vs. no depicted) actions for younger adults could be explained with a ceiling effect. Younger adults did not show any off-line difficulties in assigning thematic roles in OVS sentences, regardless of a supportive direct visual context being present or not. However, while children's accuracy was strongly improved by depicted (vs. no depicted) actions, older adults only showed a marginal effect. Even though older adults, like younger adults also showed a ceiling effect for answering the comprehension questions, this marginal effect demonstrates that depicted actions also had a facilitative effect on assigning thematic roles off-line for adults.

Nevertheless, children, who are still in the process of learning to interpret and comprehend OVS sentences profited strongly from these direct action cues. We have shown that 4-5-year old children can and do make use of direct visual context to disambiguate an initially role-ambiguous scene. These findings contradict Trueswell and Gleitman's (2004) interpretation of children's use of visual context and supports Zhang and Knoeferle's (2012) interpretation. Whereas Trueswell and Gleitman (2004) argue that 5-year old children still rely heavily on lexical cues instead of using the visual context for sentence processing (see Section 2.3 and Knoeferle, 2015b for a related discussion), our own and Zhang and Knoeferle's (2012) results suggest that children at that age can indeed use the visual context to facilitate sentence processing. They used the visually depicted action, which was mediated by the verb to anticipate the correct role filler of an OVS sentence – and hence to correctly assign thematic roles – which they would probably have interpreted as an SVO sentence if the direct visual cue wasn't present. However, Trueswell and Gleitman's (2014) arguments are based on sentences with syntactic ambiguities and a disambiguating visual context. In contrast, our sentences did not contain syntactic ambiguities, but a grammatical structure that 4-5-year old children have not fully acquired yet. Our visual scenes however, were initially role-ambiguous. In order to test whether our findings can also hold for syntactically ambiguous sentences, further studies are needed.

Additionally, Choi and Trueswell (2010, see Section 2.3) argue for a lack of inhibitory control as a reason for children's difficulties to recover from garden-path

sentences. Now, recalling that Dittmar et al. (2008a) demonstrated a strong reliance on word order instead of case marking in 4-5-year old children for (unambiguous) OVS sentence processing, this word order, i.e., SVO preference might as well be due to their lack of inhibitory control. SVO is the more frequent and canonical sentence structure. 4-5-year old children readily interpret OVS sentences as having SVO structures maybe only partly because they cannot yet make use of case marking on its own as a cue (see Dittmar et al., 2008a and Section 2.3, see also Choi & Trueswell, 2010 for a related discussion). If this is the case then the direct visual cues in the form of depicted actions do not “simply” facilitate real-time sentence processing of structurally difficult German OVS sentences but they moreover improve children’s use of their executive functions. That is, depicted action events reduce their lack of inhibitory control regarding agent - patient ordering and hence help children to arrive at the correct sentence interpretation, namely the patient - agent ordering. This reasoning is additionally supported by Dittmar et al. (2008b) and Gertner et al. (2006), who demonstrated that already 2 year-olds are sensitive to thematic roles in canonical SVO sentences (see Section 2.3). Children are hence aware of the different roles characters can take in a sentence, but might have stored them in an ordered prototypical set in the sense that linguistically, agents are followed by patients (see also Ferreira, 2003). Upon encountering OVS sentences, the case-marked NP1 determiner respectively, inhibitory control is needed in order to overcome this prototypical set. Since 4-5-year old children’s executive functions are not yet fully developed, their lack of inhibitory control might prevent them from overcoming the default ordering in this prototypical set of thematic roles, i.e., changing the order to patient - agent in OVS sentences. This inhibitory control might be improved by visually depicted actions allowing for a quick determination of the correct thematic roles. However, the argument that visual context can improve children’s executive functions regarding (real-time) sentence comprehension is only speculative and needs further empirical evidence for stronger support.

10.2 The Positive Prime Effect

The positive prime face was, similarly to the depicted action, also used for on-line sentence processing. Younger and older adults but not children used the positive emotional prime to facilitate the processing of the positive emotional OVS sentence.

Crucially, this effect cannot be seen as a pure priming effect, i.e., the positive facial expression primes looks to the happy looking character. Since we did not find effects of the indirect cue before the onset of the sentence, but only once the emotional valence could be linked to the linguistic input, we can assume that the emotional facial expression was mediated by the emotionally valenced adverb. Younger adults hence use it in the same way as they use the depicted action, namely to anticipate the correct role filler.

Nevertheless, compared to the effects of the direct cue, i.e., the depicted action, the effects of the indirect cue, i.e., the emotional facial expression were less strong. One reason why the effects for the indirect cue were weaker than the effects for the direct cue could be the differences in when during a trial the cues became available for the listener. While the depicted action was already present as the listener started hearing the sentence and was also displayed on the scene during sentence presentation, the emotional facial expression was used as a prime and hence removed before the onset of the target scene and the linguistic input. Thus, whereas the listener had access to the depicted action throughout sentence processing, the emotional prime face had to be held in working memory from the end of the prime presentation onwards until the end of the trial. The effects of the emotional prime might hence be weaker because establishing reference between or making inferences about visual and linguistic input is arguably easier when both are presented together than when they are presented separately. This assumption is indirectly supported by Glenberg and Robertson's (1999) indexical hypothesis, which states that referencing linguistic input to a visual context is essential for language comprehension. Their indexical hypothesis assumes that the referencing (or indexing) process is easier when the visual context and the linguistic input it is referring to are co-present compared to when they are presented separately.

Furthermore, the finding that the effects of the indirect cue were weaker compared to the effects of the direct cue are not surprising taking the way the listener had to link them to the linguistic input into account. While the depicted action was directly

mediated by its referring linguistic expression, i.e., the verb; matching the emotional valence of the prime face and the target character's facial expression to the valence of the adverb was much more complex. The listener first of all had to infer the valence of the emotional prime face, since no emotion labels were provided. Keeping his / her interpretation of this emotional face in memory, s / he then inspected the target scene and started to interpret the scene in the light of the incoming OVS sentence. Only when the valence of the emotional sentence became clear, i.e., during the Adverb region, the listener reactivated the previously seen emotional prime and linked the valence of the face to the emotion label provided by the linguistic input. This prime-adverb connection could now finally be used to anticipate the character on the scene whose facial expression matched the prime-adverb interpretation that the listener had just established: the correct role filler has been found. Hence, the use of the indirect cue is arguably cognitively much more demanding and involves more processing steps than the use of the direct cue for real-time thematic role assignment.

Another possible reason why the effects of the indirect cue were relatively weak could be the different processing types investigated, i.e., our emotional prime effects were weaker compared to the results by Carminati and Knoeferle (2013) and Münster et al. (2014). Recall that they investigated emotional priming effects on the processing of SVO sentences. Yet processing non-canonical structurally challenging OVS sentences and assigning thematic roles, as it was the case in our studies is arguably a cognitively more demanding task than the reconciliation of a prime face with a canonical SVO sentence. Hence linking an emotional facial expression to a corresponding emotionally valenced adverb and a character's emotional facial expression and simultaneously processing a non-canonical OVS sentence and assigning thematic roles might have resulted in subtler effects compared to the study by Carminati and Knoeferle (2013, see also Münster et al., 2014, 2015).

However, although our effects of the emotional prime face were weaker compared to Carminati and Knoeferle (2013) and Münster et al. (2014), our results demonstrated that the emotional valence of a prime face can not only be used to facilitate sentence processing semantically, but crucially also syntactically by facilitating the real-time assignment of thematic roles. This assumption is further supported by Martín-Loeches et al. (2012) who found a decreased LAN for positive adjectives (and an increased LAN for negative ones) and interpreted this decreased

LAN as facilitating syntactic processing (see Section 3.7.1.1 for more details on the study). However, unlike in our study they did not find behavioral effects of the emotionally valenced adverbs on syntactic processing. Our behavioral data showed that younger adults answered more comprehension questions correctly when they were primed with a congruent positive emotional facial expression (compared to a non- / negative emotional prime) when the comprehension questions were asked in the passive voice (but not when they were asked in the active voice). The fact that this effect was only present when the comprehension questions were cognitively and syntactically more demanding might further support the interpretation that the indirect cue was used to facilitate the syntactic processing of the sentence. Moreover, these results speak in favor of the assumption that semantic information like in this case the emotional valence of a prime face can in fact affect syntactic processing (Martín-Loeches et al., 2012; Martín-Loeches, Nigbur, Casado, Hohlfeld, & Sommer, 2006), and speak hence against assumptions stating that syntactic processing is mostly unaffected by other cognitive processes (see e.g., Pulvermüller, Shtyrov & Hauk, 2009 for a review).

10.3 Cumulative Visual Context Effects

However, even though the effects of the indirect cue were weaker compared to the effects of the direct cue, we found some evidence for an interaction between these two cues. Specifically, the indirect cue seemed to be used mostly when the direct cue was also available to facilitate thematic role assignment. This was the case regardless of the naturalness of the facial expression. A possible explanation might be that the indirect cue was best used for reassurance purposes, i.e., it corroborated the already powerful direct cue provided by the depicted action. Thus, the emotional cue might have been used to re-confirm the role relations that were assigned by the direct reference to the visual context when an action was depicted. Moreover, when the direct cue was available for the listener, this cue could be used earlier than the emotional face when the latter was also available. While the depicted action was mediated by the verb, the emotional facial expression could only be definitely linked to the positive looking character once the adverb had been heard, i.e., one word region later. Hence, the cumulative effect in the two-cue condition for an underlying “double-checking” mechanism for the correct role filler does not seem surprising.

However, this cumulative effect differs from the results found by Kreysa et al. (2014). Recall that Kreysa et al. (2014) used speaker gaze shift and depicted objects as two different cue types for on-line sentence processing of canonical SVO sentences. They found that when these two cues were both available for the listener, sentence comprehension was not further facilitated by the two-cue condition compared to when only one of the cues was available. Yet the reason for the discrepancy between our results and the results by Kreysa et al. (2014) could lie in the timing of cue presentation and the point in time the cues could be linked to the linguistic input. Whereas in our studies the two cues were not presented simultaneously and were moreover mediated by different words in the sentence, the speaker's gaze shift and the depicted object in Kreysa et al. (2014) were presented simultaneously during sentence presentation. Hence, in our case the emotional valence of the face and the depicted action could be used cumulatively. The effect of the emotional prime came into play when the effect of the depicted action was already ongoing and thus the indirect cue further facilitated OVS sentence processing and thematic role assignment, was used to re-confirm the choice of the role filler respectively. In the Kreysa et al.'s (2014) study, it could be the case that two simultaneously presented cues are too much information for the listener to be used rapidly or it could also be that one of the cues rendered the other cue redundant because they were presented at the same time and were linked to the same word in the sentence, i.e., the verb (Kreysa et al., in prep.). Hence, it seems that not only the type of visual cue that can be used for real-time language processing matters, but also the point in time when this visual information is made available for the listener.

10.4 The Influence of Cognitive Abilities

Additionally, our results for the older adults suggested that another reason for a cumulative use of both cues might be participants' cognitive abilities. Older adults showed a trend to making more use of both cues together, but only when their cognitive test scores were above the median. However, children's eye-movement or accuracy data was not meaningfully related to their cognitive test results (see also Nation et al., 2003 for similar results, but see Zhang & Knoeferle, 2012 for contrasting findings). Additionally, as we do not have cognitive test scores for younger adults, we cannot get any insights into younger adults relation between

cognitive abilities and their ability to use the direct and the indirect cue cumulatively. Nevertheless, the finding that older participants with higher cognitive abilities could use both cues cumulatively is also supported by Huettig and Janse (2015). In a visual-world eye-tracking study with spoken instructions to fixate a target on a visual display they found that higher cognitive abilities led to more target anticipation. Crucially, they tested participants between 32 and 77 years of age, but found that age in contrast to cognitive abilities had very little effect on predictive processing. Hence, in our studies we might also assume that younger adults with higher cognitive abilities would show stronger effects of a cumulative use of the two cues compared to younger adults with lower cognitive test scores.

10.5 The Naturalness of the Emotional Prime Face

Yet, there were not only differences between the visual context effects of distinct cue types. There were furthermore differences within our visual context effects for the differing presentations of the emotional prime face. Specifically, the effects of the dynamic *natural* emotional facial expressions were significantly bigger than the effects for the dynamic *schematic* emotional facial expressions. One possible reason for that could lie in our experimental design, the changes in emotional prime presentation between experiments 1, 2 and 4 and experiments 3, 5 and 6 respectively. While in the studies that used a schematic emotional facial expression in the form of a dynamic smiley (studies 1, 2 and 4) a non-emotional prime in the form of a static red star was used to contrast the positive emotional prime, in the studies that used a dynamic natural facial expression (studies 3, 5 and 6) the incongruent emotional prime consisted of the same woman's sad facial expression. Hence, in the former case we contrasted an emotional with a non-emotional prime while in the latter case two emotional expressions of opposing valences were used. Using two emotional faces of contrasting valence instead of only one emotional face might have increased the overall awareness of emotions in the study and thus lead to bigger positive prime effects compared to the schematic prime studies in which the prime was either positively valenced or did not have an emotional valence at all (see also Münster et al., 2015). This might additionally be the case because in the studies in which a natural happy (vs. a natural sad) emotional prime was used, the focus on emotions

was further increased by occasionally asking participants to comment on the characters' feelings and to recall the valence of the prime face.

However, it could also be argued that the longer exposure duration of the natural facial prime (5500 ms vs. 1750 ms for the schematic smiley) caused the better integration into sentence processing. A longer exposure to the emotional expression might have led to more in depth processing of the emotional content (see also Münster et al., 2015). Yet, studies that have varied stimulus duration (from as short as 50 ms to 10 s) of emotional facial expressions did not find differences on exposure durations in emotion perception and recognition especially for happy faces (Calvo & Lundqvist, 2008; Codispoti, Mazzetti, & Bradley, 2009). Hence, we believe that the stronger effects of the natural dynamic facial expression are rather due to its higher ecological validity and do not result from the longer stimulus exposure.

This latter factor might have contributed to a better integration of the natural (compared to the schematic) prime face into the real-time processing of the OVS sentence, as the natural prime face fitted better into the experimental context. That is, due to its greater ecological validity it could more readily be perceived to be the face of the speaker of the following sentence (which was spoken by a female voice). This setup might have rendered the experimental situation more realistic, i.e., the listener could link the natural emotional facial expression to the female voice and hence establish the link that the positively valenced scene-related sentence was uttered by the happy looking woman whose face they had just seen. That this is likely the case is also supported by 'Theory of Mind' assumptions (e.g., Premack & Woodruff, 1978). Using the facial expressions of our interlocutors we attribute our own mental states onto others and hence expect the other person to act like we would have acted with a similar facial expression. Thus, if we display a happy facial expression, we are likely to feel happy and to utter something positive. Therefore, we expect the same behavior from our interlocutors if they portray this facial expression. The natural facial prime (in contrast to the schematic, unrealistic prime face) set up this expectation and could thus be better integrated with the target agent's facial expression and the positive sentence (see also Münster et al., 2015).

Using more naturalistic real-world stimuli in experimental settings, especially those involving emotions and language is moreover strongly encouraged for instance by Adolphs (2006). He states that even though many studies on human (social)

cognition use social stimuli, these stimuli are very seldom realistic, i.e., often morphed, synthetic or static faces are used. Hence, the subjects know that these are not real people and do not interact with them accordingly. The interactive link between two humans that is established during e.g., communication cannot be established in the same way as it would be the case with (real) humans. Hence rendering our emotional prime face as naturalistic as possible by using a video sequence of a real woman, our adult participants were likely able to infer the portrayed emotions from the face and could hence better link it to the following scene, the target agent respectively.

Moreover, the stronger effect for the natural compared to the schematic facial expression underlines also the assumption that the effects of the emotional face are based on emotion effects. This means that the face effect is not just based on semantic and visual congruence between prime and target / sentence valence. If the latter was the case, we should not have observed differences in the magnitude of the face effect between the studies using a schematic smiley and the studies using the natural happy facial expression. Congruence based on visual similarity and semantics can be established in both cases and might arguably even be stronger in the case of two congruently valenced unrealistic faces (the schematic prime smiley and the target agent's facial expression) compared to the natural prime face and the target agent's face. Yet, the latter combination yielded stronger effects. We hence interpret the stronger effects of the natural prime face as based not only on visual and semantic congruence but crucially also on the integration of emotions into real-time sentence processing of emotional sentences.

10.6 Age Differences in Emotional Prime Use

Even though the visual context effects were stronger for the dynamic natural emotional prime face compared to the dynamic static emotional prime face, these face effects differed also by the age of the comprehender. While all three age groups could use the direct cue, i.e., the depicted action to determine the correct role filler on-line, only older and younger adults could also use the emotional prime face. Children on the other hand did not show any facilitative effects of this indirect social cue.

10.6.1 Younger Adults' Positive Prime Effect

Although younger adults' positive prime effects were present in the on-line as well as in the off-line data and although they could moreover use the prime face to anticipate the target agent, the magnitude of the positive prime effects were still relatively weak compared to the effects of the direct cue. Taking younger adults negativity bias into account (e.g., Isaacowitz et al., 2009), see also Section 3.7.1.1) these effects might have been stronger if we had used negatively congruent prime - character / sentence trials compared to only positively congruent trials. Yet, our findings suggest that emotional facial expressions, even if they are not in line with the specific emotional bias of an age group can facilitate the real-time comprehension of structurally challenging sentences of a matching emotional valence.

10.6.2 Older Adults' Positive Prime Effect

Even though the positively matching valence between the prime face, the target agents facial expression and the sentence was in line with older adults positivity bias, they also did not show stronger real-time effects of the indirect cue. Similar to the younger adults, older adults did make use of the positive emotional prime face for on-line sentence processing. Nevertheless, their positive prime effects were as well weaker as their depicted action effects. One possible reason for these relatively weak effects might lie in the comprehension and recall tasks that participants had to perform. As discussed in Section 3.7.1.2, Reed et al. (2014) found a strong influence of task demands on older adults' positivity bias. They report that the positivity bias was strongest in studies in which emotion processing happened unconstrained. In our studies, even though participants were not asked directly to provide a label for the emotional prime face, in the natural face version (studies 3, 5 and 6) they were occasionally asked to recall what the prime face looked like. They were moreover occasionally asked to state how the displayed characters might be feeling (only in filler trials) and they were asked to indicate 'who was doing what to whom' after each trial. Thus this accumulation of slightly different tasks that had to be performed throughout the experiment might have weakened older adults use of the positive prime face.

Moreover, the fact that the woman whose positive and negative facial expressions were used as dynamic natural primes had approximately the same age as our younger

adult participants might have also played a role for older adults subtle effects of the prime face. We might in contrast have seen stronger positive prime effects for the older adults if we had, in addition, used older adults emotional faces as primes. Several studies have reported a higher decoding accuracy, longer fixation durations, better expression identification and better memory for own-age compared to other-age faces (Ebner, He, Fichtenholz, McCarthy, & Johnson, 2010; Thibault, Bourgois, & Hess, 2006; and see Rhodes & Anastasi, 2012 for a meta-analysis on memory of own-age vs. other-age faces). On the other hand, that emotions in own-age faces are more accurately recognized by in- compared to out-group members seems to be especially the case for negative but not so much for positive emotional facial expressions. Plus, emotion identification appears to be overall more difficult in older compared to younger faces, regardless of the identifier's age (Riediger, Voelkle, Ebner, & Lindenberger, 2011). Hence, using in-group instead of out-group faces in accordance with our participants' ages might have strengthened our positive prime effects, specifically for the older adults. On the other hand, we would not expect a drastic increase in the magnitude of the face effect due to the age of the prime face, as emotion recognition for happy faces is not as strongly influenced by the age of the perceiver as it is for negatively valenced faces (cf., Riediger et al., 2011).

10.6.3 Children's Positive Prime Effect

Even though the advantage for own-age compared to other-age faces in face recognition has also been suggested for children (Hills & Lewis, 2011; Rhodes & Anastasi, 2012), we do not think that the positive emotional prime face would have had an effect on children's sentence processing if we had used children's instead of adults' faces as primes for testing children. Rather, the missing effect of the positive emotional prime face – regardless of its degree of naturalness – might be due to an accumulation of cognitively demanding and especially still developing skills in 4-5 year old children. At this age, children are still in the process of acquiring their native language, i.e., they have not yet learned to assign thematic roles in OVS sentences when no supportive visual context is co-present (see Section 2.3). Moreover, they are also still in the process of learning to correctly identify and interpret emotional facial expressions (see Section 3.7.1.3). Hence, using this indirect social cue (which is itself still in development) in order to facilitate the processing of a linguistic structure that

children also have not fully mastered at that age might be too cognitively demanding. Additionally, recall that Scherf et al. (2007, see Section 3.6) demonstrated in a functional Magnetic Resonance Imaging (fMRI) study that 5-8-year old children did show significantly less face-specific activation for video sequences of natural faces in the ventral visual cortex compared to adults. However, no differences in activation were found for places or objects compared to adults. Hence, the missing effect of the indirect cue for children could further be due to the fact that they are not yet able to process faces (irrespective of their emotional valence) in the same way as adults.

Additionally, the positivity bias in children is not yet as well established as it is in older adults (see also Section 3.7.1.3). However, recent behavioral findings by Picardo, Baron, Anderson and Todd (2016) strongly support a positivity bias in 5-7-year old children. In two studies using different emotional facial expressions and different participants, Picardo et al. (2016) suggested that children in contrast to young adults rated happy facial expressions as more arousing and intense than negative facial expressions. Moreover, as they controlled for children's ability to correctly identify all facial expressions, they excluded the possibility that this positivity bias in ratings is due to difficulties in emotion identification. Their studies hence, support neuroimaging findings that demonstrated similar amygdala activation patterns in children for positive faces compared to older but not to younger adults (Todd et al., 2012; Todd et al., 2010, see also Section 3.7.1.3). A positivity bias in children is further supported by Sato and Wakebe (2014). They asked 5-year olds and younger adults if an agent who destroyed a block castle acted intentionally or unintentionally. The agent either hit the block castle, thereby destroying or missing it; or the agent fell into versus next to the castle, i.e., destroying or missing it. The castle could hence, be intentionally (hitting) or unintentionally (falling) destroyed; or the agent did not destroy the castle but intended to; versus the castle was not destroyed because the agent fell next to it (no intention to destroy the castle). Children in contrast to adults judged the agent's intention based on the result of the action, i.e., if the castle was destroyed or not. Crucially, children in contrast to adults also judged the agent's intentions as more positive, meaning that they did not think that the agent intended to destroy the castle. Hence, this recent evidence supports a positivity rather than a negativity bias in 4-5-year old children (for further discussion see also Section 3.7.1.3). However, Berman et al. (2016) and Berman, Graham, Callaway and

Chambers (2013) found contrasting evidence supporting a negativity bias, at least for vocal affect. Additionally, Vaish et al. (2008) argue for a hard-wired negativity bias based on ontogenetic mechanisms that is present from early infancy. However, recall that this contrasts with neuroimaging studies demonstrating a higher amygdala activation for positive (versus negative) emotional faces for both children and older adults but a higher activation for negative (versus positive) emotional faces for younger adults (Mather et al., 2004; Todd et al., 2012; Todd et al., 2010), hence demonstrating a positivity bias for children and older adults but a negativity bias for younger adults. This bell-shaped development in emotional preferences directly contrasts with arguments for a hard-wired negativity bias.

Thus, whereas there is considerable evidence supporting the positivity bias in older age, there are at least some contrasting findings regarding the positivity bias in young children. Our findings that 4-5-year old children are not yet able to benefit from positive emotional faces for thematic role assignment and the processing of positively valenced OVS sentences might thus additionally be due to a weaker influence of positive emotional facial expressions on children compared to the influence on older adults.

10.7 Time Course Differences

All three age groups could use the direct cue, i.e., the depicted action, to facilitate OVS sentence processing and thematic role assignment. However, only younger and older adults could also use the indirect cue, i.e., the positive emotional facial expression, for the same purpose. Yet, we have seen that the age differences regarding the visual context effects did not only affect which cues could be used, but crucially also when the cues could be used in real-time sentence processing.

Children, as well as older and younger adults could likewise use the depicted action to anticipate the correct role filler. However, even though this effect was anticipatory in nature for all age groups, both children and older adults were delayed by one word region compared to younger adults. In contrast to younger adults, older adults and children showed the visual context effect for the depicted action only in the late Verb and in the Adverb region. However, unlike for young adults, no effects emerged in the Verb region. Zhang and Knoeferle (2012) found the same time course

effects as our studies showed, i.e., a delay in the depicted action effect for children compared to younger adults when an action was (vs. was not) depicted.

Children's eye-movements to a visual display can be mediated by the linguistic input with a very similar time course, as is the case for adults in canonical SVO sentences. Thus we do not believe that a different time course with which children link visual to linguistic input is the reason for children's delayed visual context effects (see e.g., Mani & Huettig, 2012; Nation et al., 2003). Moreover, in a Visual-World-Paradigm study measuring children's and adults' eye-movements, Borovsky, Elman and Fernald (2012) suggested that children are just as quick as adults when it comes to real-time sentence comprehension of canonical sentences and the anticipation of an expected target object depicted in a visual display and mentioned in the sentence. However, even though anticipation speed did not vary between children and adults, when controlling for age-specific vocabulary size, participants with higher vocabulary size looked earlier at the target than participants with lower vocabularies (Borovsky et al., 2012, see also Mani & Huettig, 2012). Hence, as we did not test children's and younger adults' vocabularies, we cannot exclude that children's delay in target agent anticipation might be driven by the majority of them having relatively low vocabulary scores.

Another reason for the delayed action effect in children might however be the OVS sentences. As discussed in Section 2.3 4-5-year old children cannot yet make use of case marking as a cue to assign thematic roles in OVS sentences. Thus they initially interpret an OVS sentence as a canonical SVO sentence. The later use of the depicted action to assign thematic roles compared to younger adults might be due to some kind of repair effect. Children looked at the scene depicting three characters and heard what they believed was an SVO sentence starting with the subject as the agent. They interpreted the first mentioned character as the agent (and not the patient) and might hence have expected him to perform an action. Once they heard the verb denoting the action they expected the first mentioned character, i.e., the agent in their interpretation, to be the performer of this action. However, in contrast to seeing their expected agent performing the mentioned action (which he did not do because he was in fact the patient), they searched for the action denoted by the verb and found it "attached" to another not yet mentioned character, i.e., the real agent of the action in the OVS sentence. They hence revised their initial SVO interpretation on the basis of

this visual cue and anticipated the target agent, thus interpreting the acting character as the true agent. Yet, this repair based on the depicted action took time and the anticipation happened later than it did for the younger adults.

Younger adults in contrast did not need a repair mechanism because they knew already when encountering the first determiner of NP1 that it was an OVS sentence. They expected the agent to be mentioned in NP2 of the sentence and could very quickly use the depicted action to anticipate the target agent without an additional repair step.

Yet, older adults showed the same time course effects as children since they were one word region delayed for the integration of the depicted action into sentence processing. Likely, older adults have, like younger adults, no problems in comprehending OVS sentences and using case marking. Hence, they were unlikely to be delayed in visual cue use because they needed to use a repair mechanism. Still older adults seem to predict, i.e., generate expectancies about, upcoming linguistic input less well than younger adults (DeLong, Groppe, Urbach, & Kutas, 2012; Federmeier, McLennan, Ochoa, & Kutas, 2002). However, Federmeier et al. (2002) and DeLong et al. (2012) measured linguistic prediction skills of older and younger adults using ERPs. Participants had to read (DeLong et al., 2012) or listen to (Federmeier et al., 2002) sentences ending in a more or less expected word. ERP responses (in terms of different N400 effects) showed that younger but not older adults could use the already processed linguistic input to predict what will come next in the sentence. However, our studies provided extra-linguistic visually depicted cues that facilitated language comprehension in real-time. Hence, it might be the case that older adults' prediction / expectancy generation skills are overall worse than younger adults', but as our results suggest they can still anticipate / generate expectancies about upcoming context given a supportive visual context. Still, they were delayed by one word region compared to younger adults.

This delay extended to the use of the indirect visual cue, i.e., the positive emotional prime face. Older adults only seemed to be able to use the indirect cue once the target agent had already been named and not to anticipate the correct role filler during the adverb region, as it was the case for the younger adults. Older adults could however integrate emotionally valenced faces into emotionally valenced sentences in

real-time and with the same time course as younger adults (Carminati & Knoeferle, 2013).

Yet there might be two alternative explanations for this delayed effect of the indirect emotional cue. First of all, given that older adults cannot predict upcoming input as quickly / in the same way as younger adults and given that the emotional valence of the positive prime face was mediated by the adverb, older adults might not have had a chance to use the indirect cue to anticipate the correct role filler since the target agent was mentioned directly after the adverb. Hence, we might have seen an anticipatory use of the emotional prime face if the sentence structure was different, such that the mediating emotionally valenced word was not directly followed by the to be anticipated target agent, i.e., the subject of the sentence.

Alternatively, older adults' late use of the positive emotional prime face for real-time sentence processing might be related to the underlying mechanisms of their positivity bias. As discussed in Section 3.7.1.2, older adults' fixation towards positive emotional faces (vs. negative faces) only started 500 ms after stimulus onset and moreover increased over time (Isaacowitz et al., 2009). Isaacowitz et al. (2009) see the reason for this in older adults cognitive control, i.e., the shift towards positive material is costly in cognitive resources and takes time. It could hence be the case that older adults could not use the positive emotional prime face for thematic role assignment as quickly as younger adults because they did not "simply" link their interpretation of the positive prime face to the positive valence of the adjective and the positive facial expression of the target character due to a congruency in positive emotional valence. Rather because their positivity bias is based on a top-down, cognitive control process (assuming this is the case, but see the discussion in Section 3.7.1.2), older adults needed more time to use the positive emotional facial expression in real-time sentence processing compared to younger adults.

10.8 Towards an Extended Coordinated Interplay Account

These studies have suggested that not only visual context that is directly mediated by the linguistic input can be used for incremental sentence processing. Crucially, they have furthermore provided good evidence for the assumption that indirect social visual context also has an effect on real-time sentence processing. Moreover, we have demonstrated that not all types of information in the visual context can be used to the

same extent. Additionally, these context effects are mediated by listener characteristics, i.e., the age of the comprehender.

Thus, to our knowledge the presented studies are among the first to demonstrate that not all visual context effects can be treated alike in real-time sentence processing. Plus, our studies underline the importance to investigate the effects of social, more indirect cues on language comprehension and highlight the need to take listener characteristics such as age into account.

Visual social information is, as we have seen (see Section 3.4 and 3.5), represented in a variety of ways. The next crucial step is to adapt and extend extant accounts of real-time language comprehension, so that they can accommodate our findings. In Section 4.3 and Section 4.3.1 we have argued for the importance of including social information such as emotional facial expressions of the speaker and listener characteristics such as the age of the comprehender into language processing accounts. We have further argued for the use of the Coordinated Interplay Account (Knoeferle & Crocker, 2007) as a starting point, as it already takes a rich non-linguistic visual context into consideration and outlines how visual contextual information can be used in real-time sentence processing. Section 4.4 demonstrated in detail how the CIA functions. In the following we will present how such an extended social version of the CIA, the sCIA (social Coordinated Interplay Account) could look like. We will describe the changes that could be made and illustrate them using the findings from Sections 6-9.

10.8.1 The Social Coordinated Interplay Account

The social Coordinated Interplay Account (sCIA, Figure 6) features two major extensions compared to the original CIA (Knoeferle & Crocker, 2007, Figure 2). First, it introduces ProCom. ProCom encompasses the **properties** of the **comprehender**, i.e., listener characteristics such as age or literacy (see Section 11 for further discussion on effects of literacy on real-time language comprehension and its integration into the sCIA, Mishra Singh, Pandey & Huettig, 2012). ProCom can be described as a set of listener-internal more or less stable features that can affect on-line language comprehension²⁰. These properties of the comprehender can be seen as

²⁰ Note that at this point we're refraining from strictly defining and delineating which listener characteristics ProCom should encompass and what the reasons are for including a specific feature

background against which the incoming linguistic input is processed and interpreted. ProCom is present in every step of the sCIA and hence can affect utterance interpretation at any time during language processing.

Second, the sCIA extends the ‘ant’ parameter of the CIA (Figure 1) and turns it into ‘ant_s’. Recall that in the CIA ‘ant’ comprises the linguistic expectations of the listener and is now enriched by the listener’s social knowledge, his / her extra-linguistic (social) expectations and / or (social) stereotypes, yielding ant_s. Listener expectations are hence not purely linguistic anymore. Moreover, ant_s is a probabilistic parameter, meaning that a weighting factor between 0 and 1 indicates the strength of activation regarding contextually relevant (social) expectations. This is indexed by the subscript _p in the ant_s parameter (see Figure 6). More concretely, a listener with strong visual (and phonological) expectations regarding the ethnicity of a person using t-/d-deletion (cf., Staum Casasanto, 2008 and Section 3.4) could for instance get a weight of 1 assigned to the ant_s parameter when encountering the word [mas] in a context in which <mast> is meant. The listener would hence assume that the speaker is a person of color (and not a Caucasian) person and would interpret [mas] as a dialectal variant of <mast> instead of interpreting it as a lexical mistake. Note however, that the weight assigned to ant_s is not fixed and can vary in strength depending on when during or before the incoming (non-) linguistic (visual) contextual input the social expectations of the listener are activated.

In order to exemplify how the sCIA could work, we will use two crucial findings from Sections 6-9, namely the visual context effects of the depicted action and the positive prime face on real-time language processing. We will also address the finding that these effects are mediated by the age of the comprehender. We will demonstrate the sCIA using the item sentence from Section 5.1 *Den Marienkäfer kitzelt vergnügt der Kater* (see Figure 5). Because we have outlined and exemplified the processing steps of the CIA in detail in Section 4.4 using (1) and Figure 2, we will focus on the new features ProCom and ant_s of the sCIA here. Note also that (1) and the item sentence from Section 5.1 are identical apart from the additional positively valenced adverb *vergnügt* (transl.: ‘happily’). Additionally, we will describe how the above mentioned item sentence is processed on the basis of the sCIA focusing on the verb

of the listener into ProCom. The same holds true for ant_s and the specific (social) expectations it includes and with which strength this is done.

kitzelt (transl.: ‘tickles’) and the following adverb *vergnügt* (transl.: ‘happily’), as these were our critical sentence regions (see Section 5.2.2).

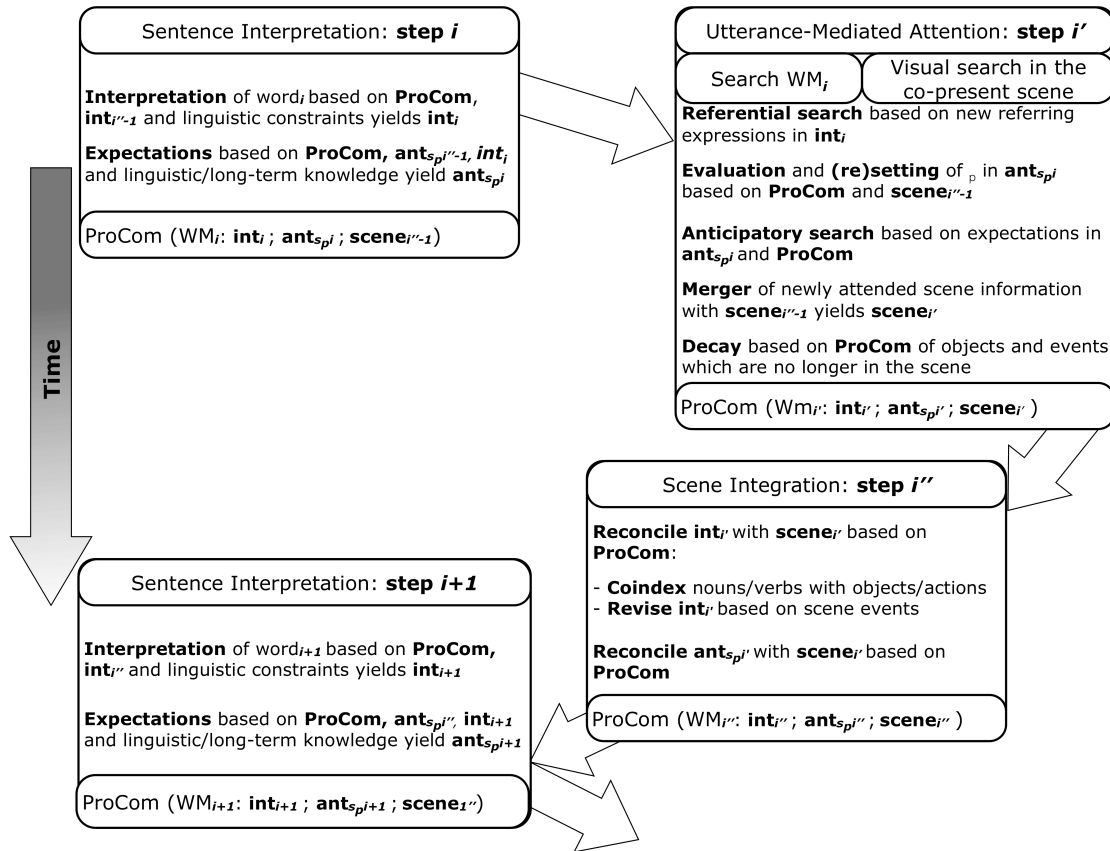


Figure 6: The social Coordinated Interplay Account. *The sCIA presents a possible extension of the CIA (Knoeferle & Crocker, 2007) and accommodates newly added listener characteristics (ProCom) such as the age of the comprehender as well as social information (ant_s) that can influence sentence processing in real-time.*

10.8.1.1 The sCIA – an Example

Taking the results from the presented studies (Sections 6-9) into account, we have seen that the age of the comprehender modulated the use of the depicted action and of the positive emotional facial expression during real-time sentence processing. Moreover, we have discussed that extra-linguistic real-world (social) knowledge, expectations and preferences, such as the emotional bias can depend on the age of the comprehender.

Hence, ProCom in our example is comprised of the age of the comprehender. ProCom moreover influences and can modulate access to the working memory of the model, which comprises scene- ($scene_{i'-1}$) and utterance-based (int_i) representations as well as the (non-) linguistic social expectations of the comprehender (ant_{si}).

The interpretation of *kitzelt* in step i of the sCIA is based on ProCom, the interpretation of previously encountered linguistic input ($int_{i'-1}$) and linguistic constraints. This yields int_i . The generated expectations in step i are based on ProCom, i.e., the age of the comprehender. Additionally, they are based on the (non-) linguistic social expectations of the comprehender prior to encountering *kitzelt*, i.e., $ant_{s_{i'-1}}$ and on his / her linguistic and long-term knowledge. These expectations then yield ant_{s_i} . Specifically in the case of the prime condition, this means that children and older adults in contrast to younger adults might set up expectations in line with their positivity bias, because they have encountered the positive prime face. Hence, their weighting factor for ant_{s_i} regarding their expectations / preferences towards positive emotional context should be higher than younger adults weighting factor. The weighting factor regarding the age groups' emotional biases should not change in the single-cue action condition, as their differing emotional biases would not come into play.

In step i' , a referential search based on *kitzelt* is performed. Crucially, the probabilistic value p of ant_{s_i} is evaluated using ProCom and the scene information from $scene_{i'-1}$. Upon encountering the verb, p should not change regardless of age group and condition. All three age groups perform an anticipatory search, in this case for a potential object mediated by the verb based on ant_{s_i} and merge the newly attended information in the scene with the old scene information to yield $scene_{i'}$. Moreover, visual information that is no longer present in the scene decays based on ProCom. In all conditions the positive or incongruent facial expression has to be kept in memory since it is presented prior to encountering the target scene. In the single-cue action condition the incongruent prime face cannot be used to facilitate thematic role assignment. However, as participants do not know this, they still attend to the prime and hence arguably keep it in memory for later use if necessary.

In step i' the scene is reconciled with the interpretation of the verb *kitzelt* based on ProCom. For the younger adults, this means that they coindex the verb with the depicted action in the action condition and hence fixate the target agent, i.e., the correct role filler. The interpretation of the verb is revised based on the scene events if necessary and the expectations in $ant_{s_{i'}}$ are also reconciled with $scene_{i'}$ based on ProCom. As our results have shown, older adults and children are one word region delayed when it comes to integrating the scene information, in this case the depicted

action with the linguistic input, i.e., the verb. Hence based on the properties of the comprehender, step i' happens later compared to younger adults when an action is depicted. In the single-cue prime condition, no actions are depicted. Thus no coindexing of verbs with actions is possible.

In step $i+1$ the next word, i.e., the adverb, is encountered and the processing cycle starts anew taking the previously derived interpretation into account. Upon hearing the adverb, participant's ant_{si+1} and its weighting factor should be evaluated and (re)set against the already interpreted linguistic input and the information stored in ProCom. Older adults' and children's p score for ant_{si+1} should increase due to their positivity bias. Yet, as older adults but not children were able to integrate the emotional prime face, the weighting factor might be even higher for older adults compared to children in the positive prime condition. Younger adults' weighting factor should, however, not change or even slightly decrease, arguably due to the mismatching emotionality between their negativity bias and the positive emotional valence of the encountered adverb in addition to the positive prime face that is kept in WM in the positive prime condition. Yet, based on ProCom and ant_{si+1} the matching emotional valence between the positive prime face and the positive adverb led younger adults but not children and older adults to an anticipatory search for the correct role filler in the positive prime condition. As discussed in Section 8.1.4.1, older adults did not use the positive prime face to anticipate the correct role filler. Yet, they could integrate it into sentence processing as the correct role filler was named. However, as our results from the depicted action condition show, older adults can indeed use visual information to initiate an anticipatory search for the target agent. Hence, the finding that they, in contrast to younger adults, do not do so in the prime condition despite their positivity bias suggests that the failure to initiate an anticipatory search is due to the differences in ProCom (i.e., the age of the comprehender) and not solely due to the differences in weighting p in ant_{si+1} regarding older and younger adults' emotional biases. This is especially the case when taking the assumption of older adults' cognitive control regarding their positivity bias into account (see Sections 3.7.1.2 and 10.6.2).

In step $i'+1$ younger adults reconcile the positive adverb *vergnügt* and their expectations from $ant_{si'+1}$ with the scene. In addition, they coindex the adverb with the emotional representation of the speaker. This coindexing leads to the anticipation

of the valence matching (smiling) target agent and they revise their linguistic interpretation based on the visual scene. While older adults seem to skip step $i'+1$ and seem to reconcile the scene with the linguistic input one word region later compared to younger adults in the positive prime condition, children did not show any visual context effects of the positive prime at all. The properties of the comprehender seem to outweigh the social expectations / preferences and the associated weighting factor (ant_{si} is nested within ProCom, see Figure 6). That is, although younger adults' weighting factor for ant_{si} should be lower than older adults' and children's p score, the younger adults are the only age group that could use the positive prime face to anticipate the correct role filler despite their negativity bias (note that this should not indicate that the emotional biases are not determined by age, it rather suggests that other factors, such as processing speed or executive functions might be a stronger influencing factor than the emotional bias).

Finally, in steps $i+2$ (*der*, transl.: 'the') and $i+3$ (*Kater*, transl.: 'cat') the correct role filler is mentioned in the sentence and hence fixated in the scene by all participants. The positively valenced OVS sentence has been processed and interpreted. Older adults (but not children and younger adults) however fixate the target agent more in the positive prime (vs. the depicted action) condition. The reason for this finding might, as just discussed (see also Section 10.6.2), be either their delay in integrating the positive prime together with the emotional valence of the adverb due to their age, i.e., ProCom, or the weighting factor of ant_{si} is evaluated and reset in step $i'+3$ after the referential search for the depicted happy looking character (the subject of the sentence) has been performed. The positive prime face held in WM is reactivated due to the matching positive valence between the already processed adverb, the target agent's happy facial expression and the prime face. Hence the p score for ant_{si} is increased and older adults fixate the correct role filler more in the positive prime condition (vs. the depicted action condition) due to their positivity bias. More so, as the nesting of ant_{si} in ProCom shows, a combination of both age related differences in sentence processing realized in ProCom and older adults top-down positivity bias realized in ProCom and ant_{si} might be a likely reason for the late visual context effect of the positive emotional prime face. To summarize, we have suggested that the extension of the Coordinated Interplay account with listener

10. General Discussion

characteristics and visual social aspects can accommodate our findings from Sections 6-9.

11. Final Conclusions and Impact

To conclude, we argue for the integration of social aspects and listener characteristics into accounts of real-time language processing. Our studies have demonstrated that visual contextual (social) information is used on-line during language processing. More so, we have shown that listener characteristics modulate the extent and the time course with which different social and non-social visual cues can be integrated into real-time sentence processing. Until now, psycholinguistic language processing accounts are underspecified regarding the integration of social and listener information. We demonstrated how these recent findings could be integrated into the Coordinated Interplay Account (Knoeferle & Crocker, 2007) – an account that already takes visual cues, such as depicted objects and action events into account. Ignoring the effect social aspects and listener characteristics have on language processing would ultimately lead to the development of incomplete and isolated language processing accounts. However, language processing is not an isolated phenomenon. How utterances are processed and interpreted depends on the properties of the comprehender and on which social aspects seem relevant for the comprehender at the time of utterance comprehension. Thus, linguistic input is not solely processed on the basis of its semantic and syntactic features. Integrating social aspects and listener characteristics as outlined here into the CIA is only the first step to acknowledging the impact of “the social” on language comprehension.

We have focused primarily on visual social cues (i.e., emotional facial expressions), yet the social information that can influence language processing does not have to be represented visually. In a series of word-by-word self-paced reading tasks, Kaschak and Glenberg (2004) and Kaschak (2006), for instance, investigated how exposure to an unfamiliar dialect affects grammatical processing. They suggested that even dialectal variants we might not be consciously aware of can have an impact on the way we process language. Using a dialect feature of the US northern midlands (the *needs* + past participle construction), they investigated whether participants unfamiliar with this construction can learn it and generalize it to a new verb and grammatical structure. Although participants had difficulty understanding the *needs*-construction at first, this difficulty disappeared within the first few trials. Moreover, they were also able to generalize this newly learned structure to another verb.



11. Final Conclusions and Impact

Additionally, Kaschak (2006) investigated whether learned dialectal constructions rely on abstract grammatical knowledge or surface word order. Participants learned the needs-construction and were tested on their learning success in a different sentential context. Those participants who had learned the needs-construction were faster in processing the needs-constructions in a different sentential context than the group that had not encountered the needs-construction during training. This was not just the case for verbs they were trained on but also for new verbs. In conclusion, Kaschak and Glenberg (2004) and Kaschak (2006) demonstrated that people can learn to comprehend new constructions and that this knowledge neither depends on the ordering of words nor on a specific context. It is likely more abstract knowledge that in turn also affects how we process familiar grammatical structures.

Furthermore, using ERPs, Van Berkum et al. (2008) investigated at what point in time and how participants integrate prosodic social information during sentence comprehension. Participants listened to sentences by different speakers. The content of some of these sentences was consistent with some speakers, but inconsistent with others (e.g., *If only I looked like Britney Spears in her latest video.* spoken with a male voice, or *Every evening I drink some wine before I go to sleep.* in a young child's voice). The inconsistent sentences contained a word that violated the "probabilistic inferences about the speaker's sex, age, and socio-economic status, as inferred from the speaker's voice" (Van Berkum et al., 2008, p. 581). The authors showed that sentence meaning and the identity of the speaker are integrated within 200-300 ms after the onset of the critical word (*Britney*, or *wine*). Mean amplitude N400s were larger to incongruent than congruent trials, an effect which has also been observed for semantic interpretation in strictly linguistic contexts (with the same topography). These results suggest that social cues behave similar to lexical semantic cues in linguistic contexts.

Additionally, Labov et al. (2011) used the non-standard, informal apical ING-form to suggest in a series of experiments that speakers make use of this information in a rapid and incremental fashion to evaluate the suitability of an interviewee as a news-broadcasting agent. Participants from different American regions listened to ten news readings of the same speaker and simultaneously moved a slider on a continuous scale indicating the speaker's professionalism throughout the reading. Labov et al.'s work suggests that adults are highly sensitive in detecting and socially judging this

11. Final Conclusions and Impact

linguistic information. This effect seems to be more pronounced in women than in men, as their results suggested that women reacted more negatively to deviations from the standard use of the ING-form than men. Their results thus also demonstrate the rapid and incremental integration of social (non-visual) cues into language processing.

Mishra et al. (2012) demonstrated that variation in listener characteristics (concerning literacy) also shapes language processing. A group of high and a group of low literates listened to spoken sentences (*There is going to be a high door.*) containing a target word, e.g., *door*, and either a restrictive adjective (*high*, restricting attention to the door) or a non-restrictive adjective. The visual display showed the target (door), together with 3 distractor objects. Participants inspected the display and listened to the sentence. High literates anticipated the target object (the door) before the target word in the restrictive condition only. This was, however, not the case for the low literates who shifted their gaze to the target only as it was named. Thus, factors related to listener experience such as literacy modulate visual anticipation of depicted objects in response to the linguistic input.

Yet, research on the integration of social visual and non-visual cues into adults' language processing is already scarce, and the same holds true for studies on child language processing. Although the studies reviewed below all focus on language internal (social) cues, i.e., intonation, prosody and speaker voice (and not on visual social cues), they underline the importance of the integration of indirect (social) cues into language processing from early on (see also Section 4.3.1).

From a usage-based point of view, while acquiring language, children have to determine the reason for why adults utter a particular word or sentence in a specific situation. More so, they have to determine why adults used a specific intonation and prosody, maybe even together with a particular facial expression. Acquiring this knowledge is essential in order to correctly interpret the linguistic input. Hence, the child needs to engage in a process of joint attention and pragmatic inference with her interlocutor, i.e., she needs to behave socially to acquire language (Tomasello, 1992, see also Section 4.3.1). This in turn suggests a very tight coupling between social behavior and language processing. Nevertheless, only a few studies to date adopt this view and moreover investigate the influence of (non-visual) social cues on children's language comprehension.

 eel (2012) investigated 3-5-year old children's (and adults') use of talker

11. Final Conclusions and Impact

characteristics, asking if children can use the knowledge of what sort of person is talking to them to facilitate spoken language understanding. In an eye-tracking study, participants learned the favorite colors of two characters. Afterwards these characters asked the participants to help them select colored shapes. Crucially, the characters asked the children to help select a colored shape for themselves, i.e., for the character in half of the trials and for someone else in the other half of the trials. The study indicated that children (as well as adults) indeed used their knowledge of which character preferred which color in that they fixated the talker's preferred color before the character asked for the object with this color. More so, children fixated the shape with the preferred color of the character when the character asked to select a shape for themselves but not when the character asked to select a shape for another character. These results demonstrated that even at a fairly young age (3-5 years) children can make use of social, talker specific cues, in this case acoustic cues, to constrain on-line language processing.

In another series of studies Creel and Jimenez (2012) demonstrated that children between 3 and 6 years of age (and adults) can use gender and age to link voices to cartoon characters. They tested whether participants could distinguish voices by timbre (female voices that differed in timbre), by gender (male vs. female voices), and by age (same gender but different speaker age). Participants were exposed to the cartoon characters on a screen and listened to sentences spoken by these characters. Children and adults learned which voice belonged to which character in a training phase. In the test phase the characters were presented next to each other on the screen and a sentence that was not presented in the training phase was played. The sentence belonged to one of the two characters and participants had to decide which character the voice belonged to. The results are clear in that children are still in the process of learning to map different acoustic cues to talker identity. However, even though they performed worse than adults in relating female voices to female talkers when the differentiating feature was timbre, they could already map voices differing in gender and in age to characters.

Additionally, even though intonation and prosody, just like gestures (as mentioned in Section 3.3) need not be categorized as social cues, they can still be classified as indirect (non-visual) cues in the sense that the information value they carry cannot be directly mediated by linguistic input but has to be inferred by the comprehender.

11. Final Conclusions and Impact

Grünloh, Lieven, and Tomasello (2011) for example suggested that 5-year old children can use intonation to overcome their sentence processing difficulties with OVS sentences to a certain degree. In a video-pointing task they presented children with unambiguous (see (1)) or ambiguous (see (4)) OVS sentences either with or without a prosodic cue that is indicative of OVS sentences for adults (Weber, Grice, & Crocker, 2006). Additionally, in a follow-up study instead of presenting the sentences independently, they integrated them into a wider discourse context. Grünloh et al. (2011) found that indeed children could use the prosodic cue to correctly assign the patient role to the first mentioned character. Without providing an additional discourse context, this was however only possible for unambiguously case-marked sentences. Hence, in independent sentences, children could only make use of the prosodic cue if both case-marking and intonation cues supported each other. In the condition in which case marking was ambiguous, they still relied on word-order (cf., Dittmar et al., 2008a) as a cue and interpreted the ambiguously marked sentences as SVO sentences, regardless of intonation. However, when sentences were embedded in a wider discourse context, children no longer relied on word order for the interpretation of both initially role-ambiguous and unambiguous sentences if they could use intonation as a cue to determine participant roles (Grünloh et al., 2011). It seems hence, that children between 3 and 5 years are to some extent already able to use some (but as our results in Section 7 suggest not all) indirect (social) cues to facilitate language understanding.


Obviously, the ultimate goal of psycholinguistics is to form a relatively holistic account of language use. The integration of visual social aspects and listener characteristics into the CIA (See Section 10.8.1.1) is only a starting point. Yet, we suggest that ProCom and ant_s cannot only be used to account for differences in listener characteristics such as age and visual social expectations or biases such as the positivity bias. Recall for example that Van Berkum et al., 2008 manipulated speaker voice and word congruence (e.g., *Every evening I drink some wine before I go to bed* spoken by a child) in a sentence comprehension ERP study. In this case, ant_s would comprise listener's social expectations regarding the content of a sentence spoken by a child. In Mishra et al. (2012) high vs. low literates were tested for their visual anticipation and linguistic prediction skills in a Visual World Paradigm. In this case, lit_s would be encompassed in ProCom as it is a property of the comprehender. Thus, the sCIA

11. Final Conclusions and Impact

presents a first step towards a dynamic real-time language processing account that can accommodate language processing in rich social contexts taking listener characteristics into account. Nevertheless, the CIA and the sCIA are accounts of language processing and the present thesis discusses the integration of social aspects solely from the viewpoint of the comprehender.

We acknowledge that more work has to be done, in order to also come up with an account of real-time language *production*, which takes the importance of social aspects into account. The overarching goal would then be to bridge real-time comprehension and production accounts and integrate them into and adapt them to a higher-level situation framework, such as the one by Zwaan (2014) or by Pickering and Garrod (2009). Pickering and Garrod indeed already suggest that communicative alignment between interaction partners need not be limited to linguistic processing. Their framework has moreover already been linked with social perspectives on cognition (Garrod & Pickering, 2009; Pickering & Garrod, 2009, see also Section 4.4). Zwaan (2014) suggests in his cognitive framework an interaction between abstract and grounded symbols. The interaction of these symbols varies depending on the linguistic situation. Objects in the real world and their referring expressions never occur in isolation but are encountered in spatio-temporal settings. These settings feature objects, agents and events, which can in turn be addressed by linguistic input. Hence, we need to take both the communicative situation and the linguistic context into account if we want to investigate how language is embedded and processed given the situation in which it is encountered (Zwaan, 2014).

Higher-level situation frameworks such as situation models could accommodate a lower-level real-time language processing account, such as the sCIA. Pickering and Garrod's frameworks and the sCIA are highly dynamic and can account for interindividual and situational (social) variation while still focusing on the linguistic situation. Hence, the presented extension of the Coordinated Interplay Account (Knoeferle & Crocker, 2007), the social Coordinated Interplay Account is a crucial step towards a more holistic account of language use that begins to consider social factors.

Furthermore, this PhD project has underlined that language processing is not modulated in the same way by all kinds visual information. Rather we have suggested that  the type of information in visual context is highly relevant for language processing. More

11. Final Conclusions and Impact

so, our findings demonstrate that more indirect, social visual context can also be used for language processing in real-time. Additionally, our studies highlight the importance to take listener characteristics such as the age of the comprehender into account when investigating on-line language processing. Taking listener characteristics into account is even more relevant since our studies demonstrate that the type of information in the visual context that is available during language comprehension interacts with the characteristics of the listener.

Put into a wider perspective, this PhD project moreover has an impact on both language acquisition and language and aging research. Certainly, more research is needed in order to investigate in more detail how and which different visual cues can be used to facilitate children's on-line language processing, especially regarding the acquisition of challenging syntactic structures. Additionally, even though we know now that social aspects can also improve older adult's real-time sentence processing, we still know relatively little about the influence of different age-related changes on situated language comprehension. This PhD project highlights the relevance and the need for psycholinguistics to move from rather language-centric studies, assumptions, theories and accounts to more interdisciplinary and socially motivated research. After all, language use likely has been shaped, among others, by the human desire to communicate with other social beings.

12. German Summary

Die vorliegende Doktorarbeit hat mithilfe von Augenbewegungsmessungen den Einfluss von emotionalen Gesichtsausdrücken und visuell dargestellten Aktionen auf das Echtzeit-Sprachverständnis von Objekt-Verb-Subjekt (OVS) Sätzen und die thematische Rollenzuweisung bei Kindern, sowie jüngeren und älteren Erwachsenen untersucht.

Vorherige psycholinguistische Studien haben gezeigt, dass Sprache inkrementell verarbeitet wird und wir den visuellen Kontext schnellstmöglich nutzen um uns die Sprachverarbeitung zu erleichtern (cf., Kamide et al., 2003a; Tanenhaus et al., 1995). Die meisten dieser Studien konzentrierten sich allerdings auf visuell dargestellte Aktionen und Objekte, auf die *direkt* durch den linguistischen Input verwiesen wird. Nur wenige Studien haben sich hingegen mit dem Einfluss von *weniger direkten, sozialen* visuellen Hinweisen, wie zum Beispiel von emotionalen Gesichtsausdrücken, auf die Sprachverarbeitung beschäftigt (Carminati & Knoeferle, 2013). Des Weiteren haben außerdem kaum psycholinguistische Studien das Alter des Rezipienten als Einflussfaktor auf den Gebrauch visueller direkter und indirekter (sozialer) Kontexte bei dem Sprachgebrauch betrachtet (siehe z.B., Trueswell et al., 1995 zum Einfluss direkter visueller Kontexten auf die Sprachverarbeitung syntaktisch ambiger Sätze bei Kindern und Erwachsenen).

Allerdings findet die menschliche Kommunikation oftmals in einem sozialen Rahmen statt. Wir interagieren mit unseren Mitmenschen und interpretieren ihre Aussagen mithilfe der uns zur Verfügung stehenden linguistischen und nicht-linguistischen, visuellen und sozialen Hilfsmittel, wie beispielsweise dem emotionalen Gesichtsausdruck unseres Gegenübers. Um also zu einem umfassenderen Verständnis unser Sprachverarbeitungsprozesse zu gelangen, müssen wir auch diese sozialen visuellen Hinweise mit einbeziehen. Es ist zudem bis dato unklar, inwieweit diese beiden unterschiedlichen visuellen Hilfsmittel (direkte und indirekt soziale) hinsichtlich auf ihren Einfluss auf die Sprachverarbeitung und die thematische Rollenzuweisung miteinander interagieren. Können sie gemeinsam genutzt werden, um das Sprachverständnis noch weiter zu erleichtern (verglichen mit einer voneinander unabhängigen Nutzung)? Ist einer der Hinweise stärker als der andere bzw. erleichtert einer der Hinweise die Sprachverarbeitung stärker als der andere? Wie muss ein emotionaler Gesichtsausdruck als Hilfsmittel dargestellt werden, um als

solches auch für die Erleichterung des Sprachverstehensprozesses zugänglich zu sein? Das heißt, ist eine schematische Darstellung beispielsweise in Form eines Smileys ausreichend oder benötigen wir natürliche, menschliche Gesichter um Emotionen als soziales Hilfsmittel nutzen zu können? Des Weiteren ist, wie weiter oben bereits erwähnt, die Frage wie sich das Alter des Rezipienten auf den Gebrauch von direkten und indirekten sozialen Hilfsmitteln für die Sprachverarbeitung auswirkt noch ungeklärt.

Das Ziel dieser Dissertation war es daher empirische Untersuchungen mithilfe von Augenbewegungsmessungen durchzuführen, um auf Grundlage der Resultate die gerade genannten offenen Fragen zu beantworten. Die Ergebnisse unserer Studien bereichern zudem nicht nur die psycholinguistische Sprachverstehensforschung, sondern können weitergehend auch existierende Sprachverstehensmodelle informieren. Existierende Modelle, die die Sprachverarbeitung in Echtzeit darstellen, beziehen nur selten auch den außerlinguistischen visuellen Kontext mit ein (siehe Kapitel 4.2). Ein Modell, welches allerdings den Einfluss direkter visueller Hinweise (wie beispielsweise visuell dargestellte Aktionen) auf die Sprachverarbeitung berücksichtigt ist der Coordinated Interplay Account (CIA) von Knoeferle und Crocker (2006, 2007). Allerdings fehlen auch hier explizite Komponenten, die den Einfluss visueller sozialer Kontexte und Merkmale des Sprachverstehers (wie das Alter) auf den Sprachverstehensprozess modellieren. Unter Berücksichtigung unserer eigenen Forschungsergebnisse legen wir daher schlussendlich auch eine Möglichkeit dar, wie der Coordinated Interplay Account (Knoeferle & Crocker, 2007) durch die Inklusion eines sozialen Parameters und eines Parameters für den Rezipienten erweitert werden kann.

Um unsere Studien zu motivieren diskutieren wir in Kapitel 2 zunächst die inkrementelle Verarbeitung der Sprache. Wir zeigen anhand vorheriger Studien, dass sowohl Kinder als auch Erwachsene kanonische Subjekt-Verb-Objekt (SVO) Sätze schnell und problemlos verarbeiten können (Kapitel 2.3). Zudem zeigen wir, dass schon 2-jährige Kinder die thematischen Rollen eines SVO Satzes problemlos zuweisen können (Dittmar et al., 2008a). Dies scheint deshalb der Fall zu sein, da in SVO Sätzen der Agens – das Subjekt des Satzes – vor dem Patiens – dem Objekt des Satzes – erwähnt wird. Nichtkanonische OVS Sätze hingegen sind sowohl für Kinder als auch für Erwachsene schwieriger zu verarbeiten (siehe Kapitel 2). Vor allem für

Kinder bis zum 7. Lebensjahr ist es in OVS Sätzen zudem besonders schwierig zu erkennen, wer der Agens und wer der Patiens eines Satzes ist, da in OVS Sätzen der Patiens, vor dem Agens, genannt wird. Erwachsene nutzen für die thematische Rollenzuweisung in Deutschen nicht-ambigen OVS Sätzen den kasusmarkierten Artikel (wie beispielsweise in *Den Marienkäfer kitzelt vergnügt der Kater*). 4-5-jährige Kinder scheinen dies allerdings noch nicht zu können und weisen daher auch in OVS Sätzen dem vorangestellten Objekt des Satzes fälschlicherweise die thematische Rolle des Agens zu, wie es in prototypischen kanonischen SVO Sätzen korrekt wäre (Dittmar et al., 2008a). Kapitel 2 zeigt dementsprechend, dass die inkrementelle Sprachverarbeitung sowohl von der syntaktischen Struktur, als auch von der Zuweisung der thematischen Rollen im Satz abhängt und zusätzlich von dem Alter des Rezipienten beeinflusst wird.

Aufbauend auf Kapitel 2 diskutieren wir in Kapitel 3 den Einfluss des visuellen Kontextes auf die Sprachverarbeitung. Wir zeigen, dass Sprachverarbeitung und thematische Rollenzuweisung insbesondere in nicht-kanonischen OVS Sätzen durch das Vorhandensein direkter visueller Hilfsmittel, wie beispielsweise visuell dargestellten Aktionen, erleichtert werden kann. Relevant sind hier vor allem Ergebnisse von Studien, die zeigen dass auch 4-5-jährige Kinder Agens- und Patiens-Rollen in OVS Sätzen deutlich häufiger korrekt benennen können, wenn sie während der Verarbeitung eines OVS Satzes die im Satz durch das Verb beschriebene Aktion als visuell dargestellt Handlung zwischen Agens und Patiens nutzen können (vs. wenn keine Handlungen dargestellt sind), um das Sprachverstehen zu erleichtern (Zhang & Knoeferle, 2012). Außerdem zeigen wir, dass auch indirekte soziale Hilfsmittel, wie beispielsweise der Blick des Sprechers, einen positiven Einfluss auf die in Echtzeit ablaufende Sprachverarbeitung haben (Kreysa et al., 2014). Wir diskutieren die Unterschiede direkter und indirekter visueller Hilfsmittel und beziehen uns im Speziellen auf die Unterschiede zwischen visuell dargestellten Aktionen und Objekten (direkte Hilfsmittel), sowie die Erkennung und Verarbeitung von Gesichtern und emotionalen Gesichtsausdrücken (indirekten Hilfsmitteln, Germine et al., 2011; Scherf et al., 2007). Dabei wird die Komplexität von emotionalen Gesichtsausdrücken hervorgehoben. Wir diskutieren weiterhin, dass besonders für die Erkennung, Verarbeitung und Interpretation eines emotionalen Gesichtsausdruckes das Alter desjenigen, der den Gesichtsausdruck wahrnimmt, entscheidend ist für die weitere

Interpretation (Herba & Philips, 2004; Isaakowitz et al., 2008). Es wird deutlich, dass jüngere Erwachsene im Gegensatz zu Kindern und älteren Erwachsenen negative emotionale Gesichtsausdrücke bevorzugen. Kinder und ältere Menschen hingegen zeigen eine Präferenz für positive emotionale Gesichtsausdrücke. Diese Präferenz hat zudem einen Einfluss auf die Sprachverarbeitung. Obwohl sich bisher nur sehr wenige Studien mit dem Einfluss von Emotionen auf die Sprachverarbeitung beschäftigt haben, zeigen Carminati und Knoeferle (2013), dass jüngere Menschen negative Sätze besser verarbeiten können, wenn sie zuvor einen negativen Gesichtsausdruck gesehen haben. Dies galt nicht aber, wenn der emotionale Satz positiv-kongruent mit einem positiven Gesichtsausdruck war. Umgekehrtes galt für ältere Erwachsene entsprechend ihrer Präferenz für positive Emotionen (Carminati & Knoeferle, 2013). Diese Studie betrachtete allerdings nur das Verständnis von kanonischen SVO Sätzen und bezog keine Kinder als Probanden mit ein. Eine thematische Rollenzuweisung in grammatikalisch anspruchsvolleren OVS Sätzen und die Frage wie und ob auch Kinder schon in der Lage sind emotionale Präferenz für die Sprachverarbeitung zu nutzen, blieb daher ungeklärt. Abschließend diskutieren wir in Kapitel 3 außerdem die Art der Darstellung eines emotionalen Gesichtsausdruckes, da es bis dato nicht eindeutig ist, ob emotionale Gesichtsausdrücke besser als indirekte soziale Hilfsmittel genutzt werden können, wenn die Emotion mithilfe einer schematischen Darstellung oder mithilfe eines natürlichen menschlichen Gesichtes interpretiert werden muss (Bassili, 1979).

Kapitel 3 hat ausführlich dargelegt und diskutiert, dass sowohl direkte, als auch indirekte soziale visuelle Hilfsmittel genutzt werden, um unsere Sprachverarbeitung, besonders in grammatikalisch anspruchsvollen Strukturen, zu unterstützen und zu erleichtern. In Kapitel 4 gehen wir daher zunächst näher darauf ein, wie visueller Kontext und linguistischer Input sich gegenseitig unterstützen. Wir zeigen, dass sowohl unsere Augenbewegungen, und der visuelle Kontext, den wir fixieren, als auch der sprachliche Input, den wir erhalten, zeitlich sowie thematisch sehr eng miteinander verbunden sind und sich gegenseitig bedingen (z.B. Tanenhaus et al., 1995). Darauf aufbauend diskutieren wir Theorien und Modelle der Sprachverarbeitung und legen insbesondere ihre Defizite im Bezug auf die Einbeziehung des reichhaltigen visuellen Kontextes – und hier speziell des indirekten sozialen Kontextes – dar. Wir argumentieren weitergehend für die Relevanz und

Unabdingbarkeit der Inklusion reichhaltiger visueller direkter und indirekter sozialer Kontexte in bestehende Modelle der Echtzeit-Sprachverarbeitung. Abschließend stellen wir in Kapitel 4 den Coordinated Interplay Account (Knoeferle & Crocker, 2007) mit seiner Funktionsweise vor und argumentieren für die Einbettung sozialer visueller Hilfsmittel und Charakteristika des Sprachverstehers in dieses Modell. Der CIA ist hierfür besonders geeignet, da er neben der Sprachverarbeitung in Echtzeit auch bereits den Einfluss direkter (aber nicht explizit indirekter sozialer) visueller Hilfsmittel auf die Sprachverarbeitung modelliert.

Bevor wir in Kapitel 6 bis 9 unsere eigenen Studien und deren Ergebnisse ausführlich beschreiben und diskutieren, befasst sich Kapitel 5 mit dem Aufbau der Experimente, den genutzten Materialien und der statistischen Datenanalyse. In allen Experimenten sehen unsere Versuchspersonen zunächst einen positiven Gesichtsausdruck. Dieser variiert zwischen den Experimenten in der Art der Darstellung und ist demnach entweder ein dynamischer schematischer fröhlicher Smiley oder ein dynamischer natürlicher menschlicher lachender Gesichtsausdruck (indirekter sozialer Hinweis). Nachdem die Versuchspersonen diesen emotionalen Gesichtsausdruck (oder je nach experimenteller Bedingung einen nicht emotionalen Stern bzw. negativen Gesichtsausdruck) gesehen haben, folgt eine visuelle Szene in der drei Figuren (zwei potentielle Agens und ein Patiens) dargestellt sind. Die beiden Agens führen jeweils in der Hälfte der Versuche eine visuell dargestellte Aktion (direkter visueller Hinweis) in Richtung des Patiens aus. In der anderen Hälfte der Versuche sind keine visuellen Handlungen zu sehen. Während die Versuchspersonen diese Szenen betrachten, hören sie einen positiv emotionalen OVS Satz, der die Aktion (in der Szene abgebildet oder nicht) des Ziel-Agens zu dem Patiens beschreibt (bspw. *Den Marienkäfer kitzelt vergnügt der Kater*). Der zweite potentielle Agens ist der Distraktor und wird nicht im Satz benannt. Während die Versuchspersonen die Szene betrachten und den Satz hören, werden ihre Augenbewegungen aufgezeichnet. Die Aufgabe der Versuchspersonen ist es, den Satz zu verstehen und am Ende jedes Versuches eine Frage zu den im Satz beschriebenen thematischen Rollen zu beantworten. Um den Einfluss des Alters auf den Gebrauch direkter und indirekter sozialer visueller Hinweise zu untersuchen, haben wir 4-5-jährige Kinder, jüngere Erwachsene (18-30 Jahre) und ältere Erwachsene (60-90 Jahre) getestet.

Unserer Ergebnisse (Kapitel 6-9) zeigen, dass alle drei Altersgruppen das direkte visuelle Hilfsmittel, d.h. die dargestellten Aktionen nutzen können, um die Satzverarbeitung in Echtzeit zu vereinfachen und die thematischen Rollen korrekt zuzuweisen. Hierbei zeigte sich sowohl in den Augenbewegungsmessungen, als auch in den Antworten auf die Satzverständnisfragen, dass sowohl Kinder als auch beide Erwachsenenengruppen die dargestellten Aktionen zur Antizipation und akkuraten Rollenzuweisung des Ziel-Agens nutzen können, d.h. mithilfe der dargestellten Aktionen konnten alle Altersgruppen den korrekten Rollenfüller identifizieren noch bevor er durch das Subjekt am Ende des Satzes benannt wurde. Im Gegensatz zu den jüngeren Erwachsenen integrierten Kinder und ältere Erwachsene die visuell dargestellten Aktionen allerdings eine Wortregion später, nämlich erst als sie das Adverb hörten. Jüngere Erwachsene konnten die dargestellten Aktionen hingegen schon in der Verbregion nutzen, um den korrekten Rollenfüller zu identifizieren.

In Bezug auf die Nutzung des indirekten sozialen visuellen Hinweises, d.h. des positiven emotionalen Gesichtsausdruckes, zeigen unsere Ergebnisse, dass beide Erwachsenenengruppen auch den positiven Gesichtsausdruck während der Satzverarbeitung des positiven OVS Satzes in Echtzeit nutzen können, um den korrekten Rollenfüller zu identifizieren. Allerdings sind die älteren Erwachsenen auch bei der Integrierung des positiven Gesichtsausdruckes in den positiven OVS Satz eine Wortregion verspätet. Während die jüngeren Erwachsenen den indirekten sozialen visuellen Hinweis zur Antizipation des Ziel-Agens schon in der Adverbregion nutzen können, zeigen die Augenbewegungsdaten, dass die älteren Erwachsenen den positiven Gesichtsausdruck erst in den OVS Satz integrieren, wenn der Ziel-Agens bereits durch das Subjekt am Ende des Satzes genannt wurde. Hier findet folglich keine Antizipation statt. Bezüglich der Frage nach der Darstellung des emotionalen Gesichtsausdruckes zeigen unsere Ergebnisse, dass der dynamische, natürlich menschliche Gesichtsausdruck signifikant besser in die Sprachverarbeitung in Echtzeit integriert werden kann, als der dynamische, schematische Smiley.

Im Gegensatz zu den Erwachsenenendaten haben unsere Ergebnisse bei den Kindern gezeigt, dass Kinder solche indirekten sozialen visuellen Hilfsmittel zur Vereinfachung der Satzverarbeitung nicht-kanonischer positiver OVS Sätze und zur thematischen Rollenzuweisung noch nicht nutzen können. Dafür zeigen die Erwachsenenendaten allerdings einen Trend hingehend zur kumulativen Nutzung der

beiden direkten und indirekten visuellen Hinweise für die Echtzeit-Sprachverarbeitung. Es scheint, dass zwei visuelle Hilfsmittel (auch wenn sie unterschiedliche Eigenschaften haben) gemeinsam genutzt werden können, um die Satzverarbeitung schwieriger OVS Sätze in Echtzeit noch weiter (als mit nur einem Hilfsmittel) zu erleichtern.

Zusammenfassend haben unsere Studien folglich gezeigt, dass alle getesteten Altersgruppen visuell dargestellte Aktionen zur Antizipation des korrekten Rollenfüllers nutzen können. Im Gegensatz dazu können anscheinend nur die getesteten Erwachsenengruppen auch positive emotionale Gesichtsausdrücke nutzen um die Echtzeit-Sprachverarbeitung von positiv emotionalen OVS Sätzen zu vereinfachen. Des Weiteren haben wir gezeigt, dass sowohl Kinder als auch ältere Erwachsene visuelle Hilfsmittel später als jüngere Erwachsene in die Sprachverarbeitung integrieren. Zudem scheint die Nutzung indirekter sozialer visueller Hilfsmittel von der Art ihrer Darstellung abzuhängen.

Diese Ergebnisse werden in Kapitel 10 unter der Berücksichtigung der in Kapitel 2 bis 4 erörterten Studien diskutiert. Des Weiteren schlagen wir die Erweiterung des CIA um indirekte soziale visuelle Hilfsmittel, sowie um Charakteristika des Rezipienten vor und demonstrieren anhand unserer eigenen Ergebnisse wie sowohl emotionale Gesichtsausdrücke als indirekte sozialen visuelle Hilfsmittel als auch das Alter des Sprachverstehers in den CIA eingebunden werden können.

Zum Abschluss dieser Doktorarbeit legen wir den Einfluss dieses Projektes auf die Psycholinguistik dar. Wir argumentieren dafür, dass dieses Promotionsprojekt zu einem umfassenderen Verständnis des Sprachverstehens insbesondere im Hinblick auf die Einbeziehung sozialer Aspekte geführt hat. Weiterhin haben die Ergebnisse unserer Studien einen Einfluss auf zukünftige Untersuchungen des Kinderspracherwerbs in reichhaltigen visuellen Kontexten. Gleiches trifft selbstverständlich auch auf die Sprachverarbeitungsuntersuchungen älterer Menschen zu, besonders im Hinblick auf die verspätete Integrierung visueller Hilfsmittel in die in Echtzeit ablaufende Sprachverarbeitung. Besonders relevant ist dieses Promotionsprojekt abschließend auch, da wir gezeigt haben, dass soziale Aspekte nicht nur eine große Bedeutung in unseren Sprachverstehensprozessen haben, sondern auch, weil wir für die essentielle Inklusion sozialer Aspekte und Charakteristika des

12. German Summary

Sprachverstehers in existierende Sprachverarbeitungsmodelle argumentieren und dies anhand eigener Ergebnisse demonstrieren.

Appendix A: Critical Item Sentences and Word Onsets

A.1 Item Sentences

The critical item sentences for all experiments are listed below. Each item has a number assigned to it. This number indicates which set of item pictures it describes (see Appendix C). The numbers for the item sentences and the item pictures coincide.

Item number	Item sentence
1	Den Wal bekocht fröhlich der Vogel.
2	Den Igel krönt begeistert der Hahn.
3	Den Hamster beschenkt heiter der Affe.
4	Den Adler füttert zufrieden der Hund.
5	Den Marienkäfer kitzelt vergnügt der Kater.
6	Den Maulwurf besingt munter der Wolf.
7	Den Schmetterling fotografiert freudig der Eisbär.
8	Den Dinosaurier begrüßt strahlend die Biene.
9	Den Delphin verkleidet liebevoll die Gans.
10	Den Bieber küsst verpielt das Mädchen.
11	Den Geparden bespaßt freundlich der Zauberer.
12	Den Elephanten umarmt lebenswürdig die Meerjungfrau.
13	Den Hirsch verwöhnt fürsorglich der Pinguin.
14	Den Fisch überrascht warmherzig der Tiger.
15	Den Helden heiratet glücklich der Spatz.
16	Den Raben bejubelt ausgelassen die Ziege.

A.2 Word On- and Offsets for each Item

The individual on- and offsets for each word in the critical items sentences are presented below.

Item	NP1 onset	NP1 offset	Verb onset	Verb offset	Adverb onset	Adverb offset	NP2 onset	NP2 offset
1	0	860	994	1694	2093	2850	3520	4229
2	0	863	1122	1756	1948	2909	3498	4204
3	0	994	1177	1898	2389	3202	3644	4402
4	0	904	1090	1795	2102	3070	3447	4112
5	0	1376	1602	2344	2558	3272	3801	4694
6	0	1016	1301	1992	2646	3430	3890	4609
7	0	1126	1264	2349	2793	3569	3881	4833
8	0	1367	1525	2264	2576	3510	3992	4710
9	0	1014	1161	2157	2507	3489	3803	4521
10	0	864	1002	1515	1934	2781	3202	4125
11	0	1102	1247	2029	2441	3261	3651	4575
12	0	1292	1358	2134	2530	3574	3884	5049
13	0	827	1171	1907	2526	3535	4031	5041
14	0	835	1049	1794	2208	3349	3932	4791
15	0	907	1099	1940	2102	2799	3289	4142
16	0	926	1332	2244	2460	3517	3775	4680

Appendix B: Primes

Appendix B shows the positive and non-emotional / incongruent primes that were presented to the participants before encountering the item images (Appendix C) and hearing the item sentences (Appendix A).

B.1 Positive Schematic Prime and Non-emotional Prime used in Experiments 1, 2 and 4.



Positive schematic prime



Non-emotional prime

Note that the positive schematic prime was dynamic in the sense that it started with a light smile and then quickly turned into the broad smile depicted here. The non-emotional prime was static.

B.2 Positive Natural Prime and Incongruent Emotional Prime Used in Experiments 3, 5 and 6.



Positive natural prime



Incongruent emotional prime

Note that both the positive and the incongruent prime were shown as short natural dynamic video sequences. The facial expression started with a neutral expression and

Appendix

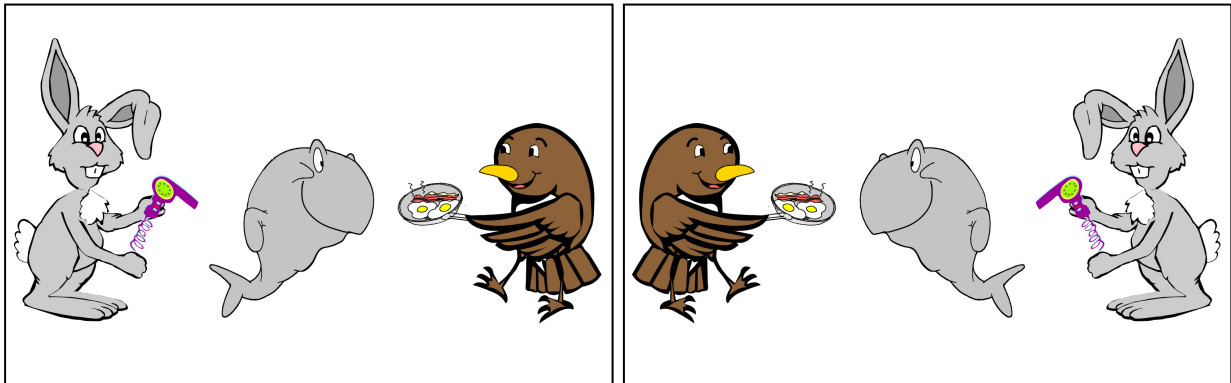
naturally changed into the happy or sad emotional expression. The last frame of the video always showed the emotional valence.

Appendix C: Critical Item Images

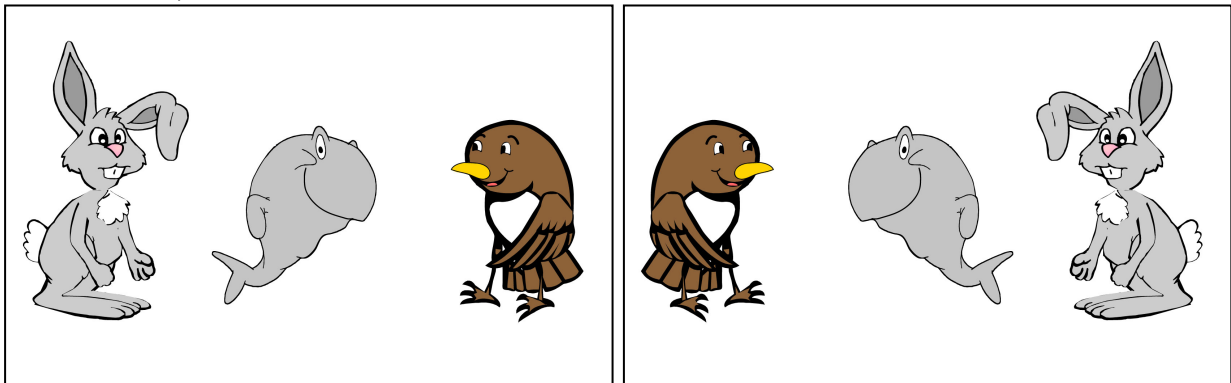
The item images depicting agent – patient – distractor characters are presented below. The image numbers correspond to the numbers of the item sentences in Appendix A, indicating which item image was related to which item sentence. In each item image set, a) refers to the depicted action condition and b) refers to the no action condition. Each image is presented in both left and right counterbalanced versions.

C.1 Item Images for Experiments 1 and 4

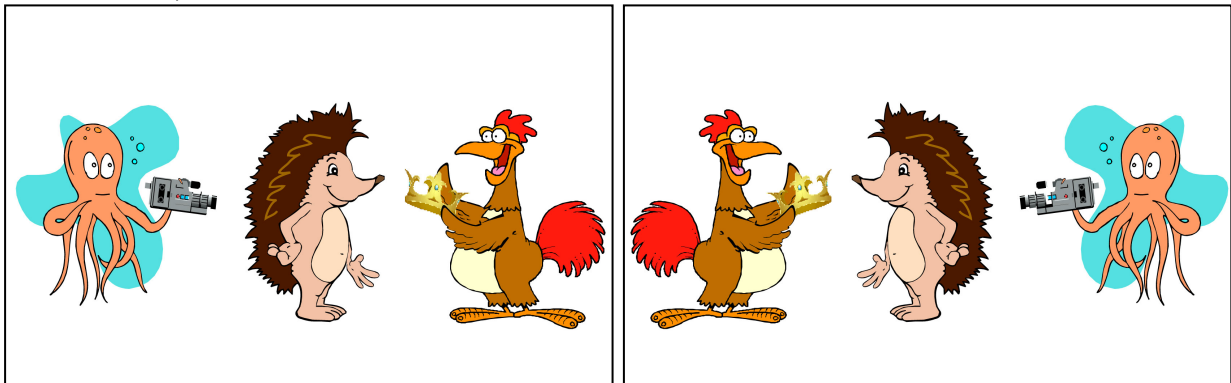
Item 1 a)



Item 1 b)

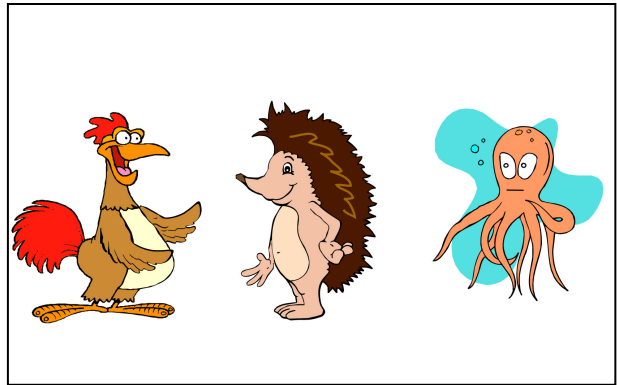
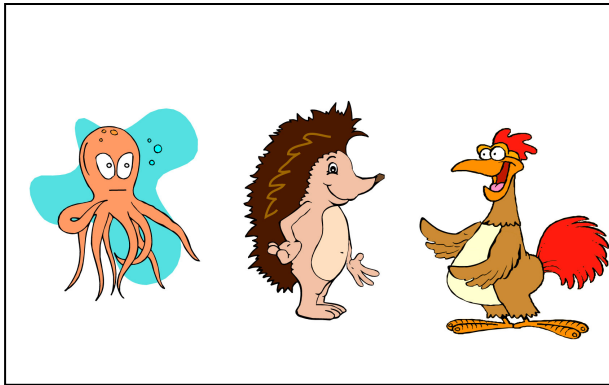


Item 2 a)

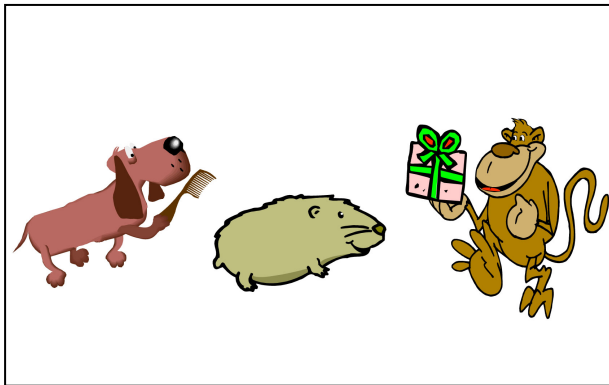


Appendix

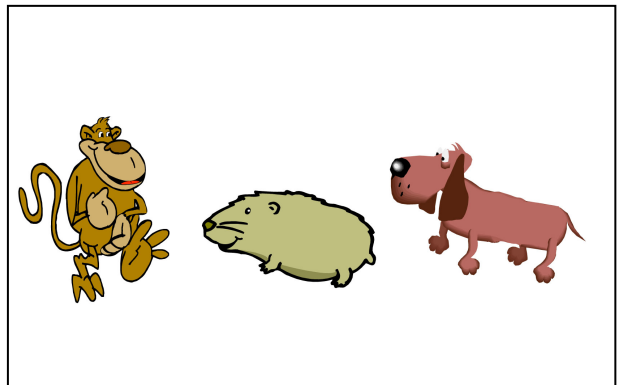
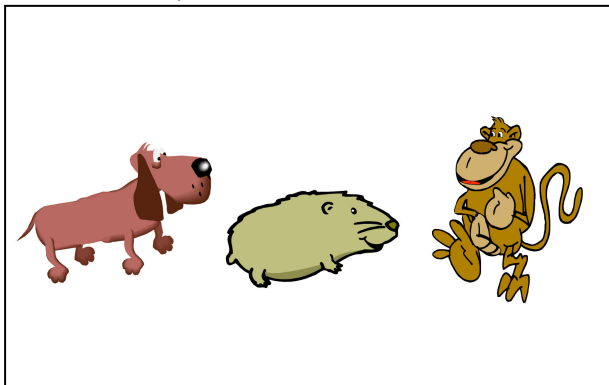
Item 2 b)



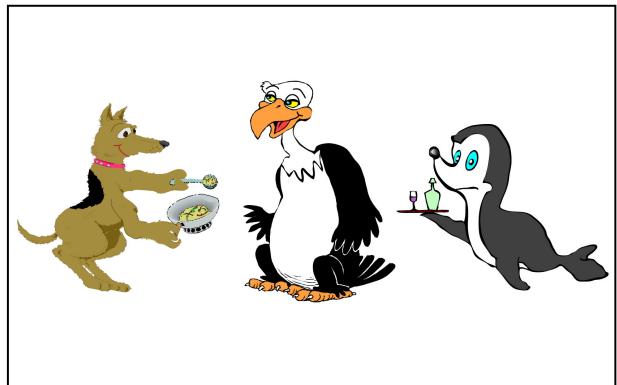
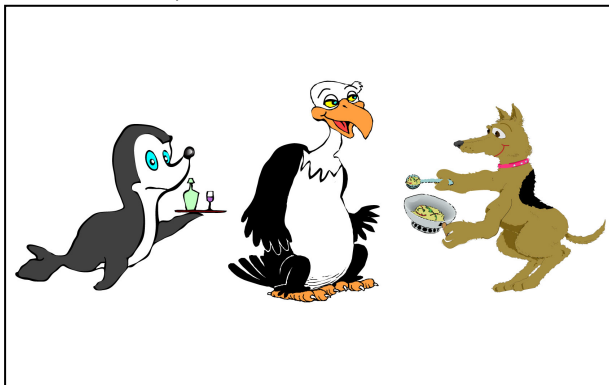
Item 3 a)



Item 3 b)

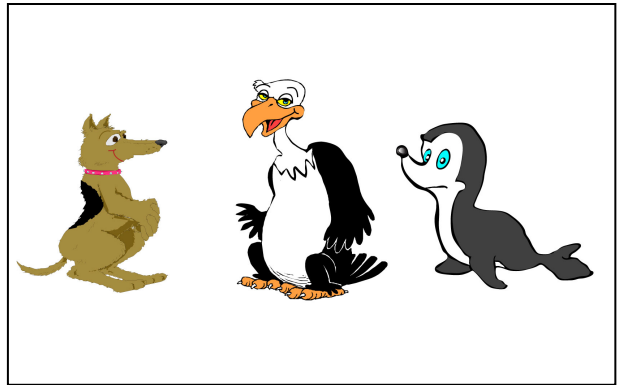
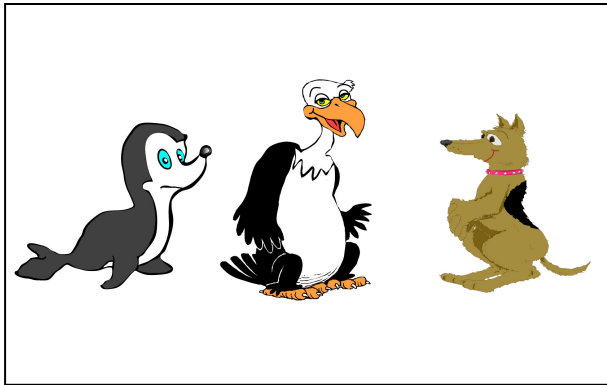


Item 4 a)

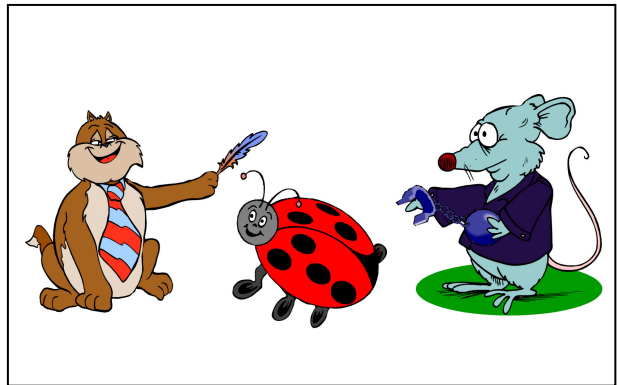


Appendix

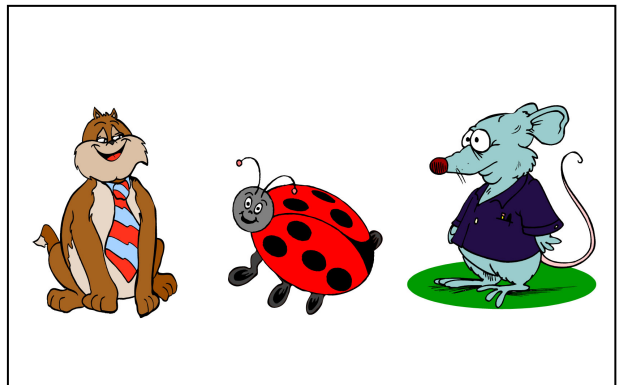
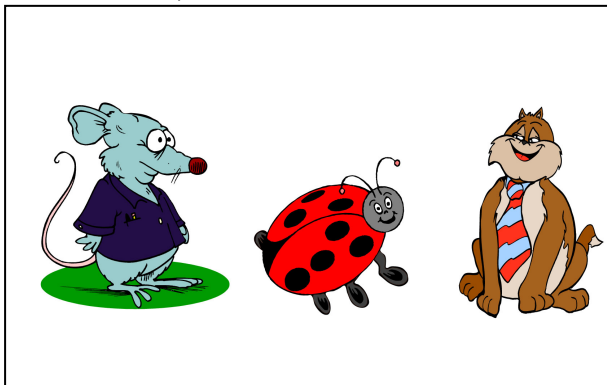
Item 4 b)



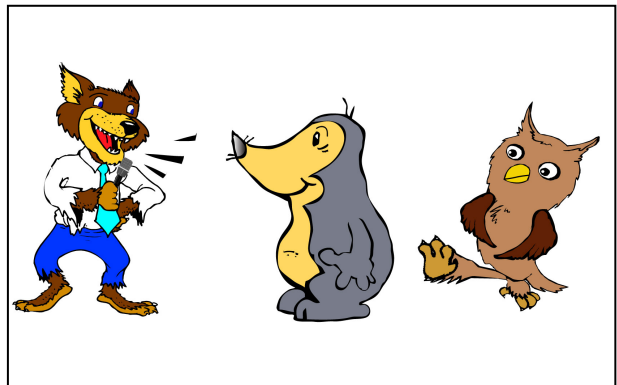
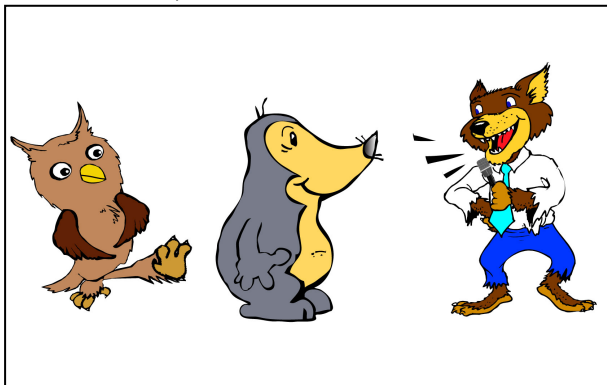
Item 5 b)



Item 5 b)



Item 6 a)

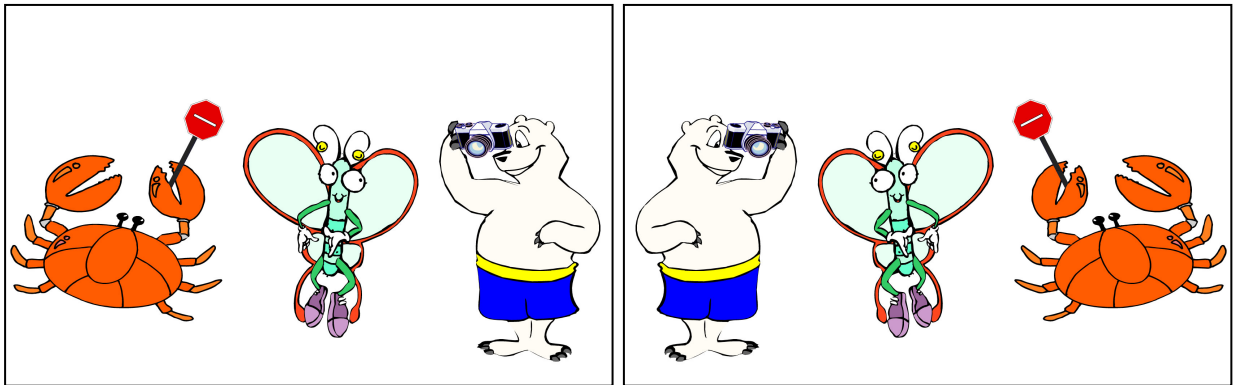


Appendix

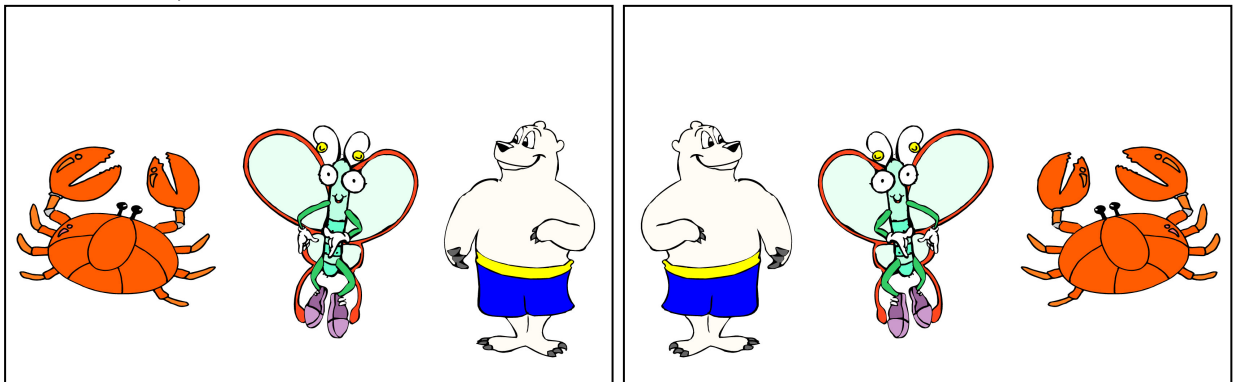
Item 6 b)



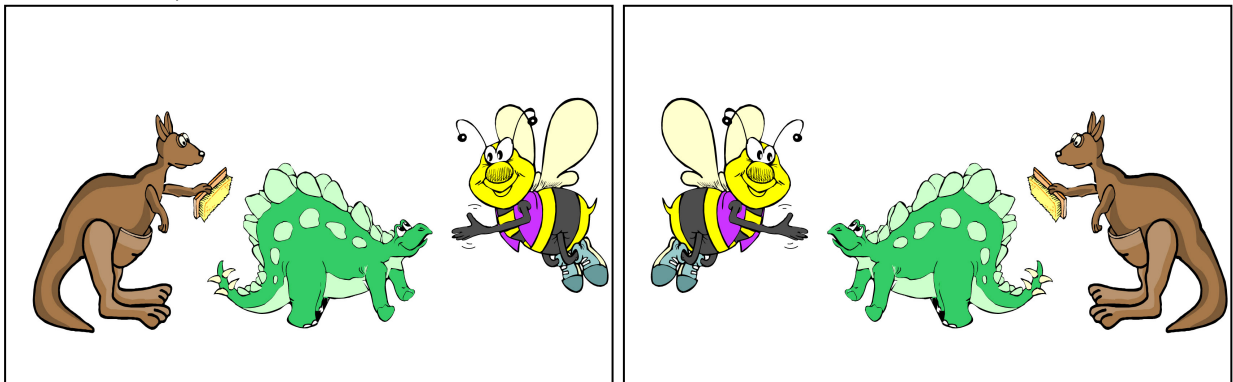
Item 7 a)



Item 7 b)

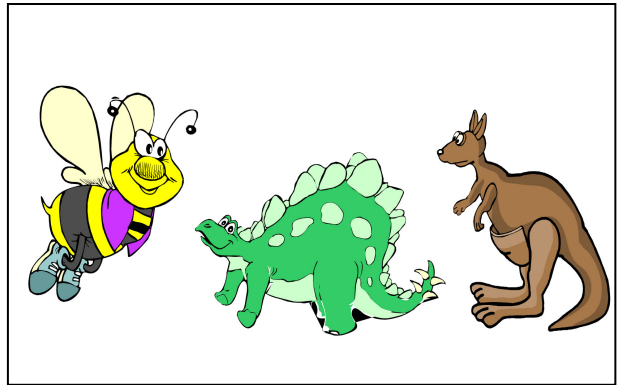
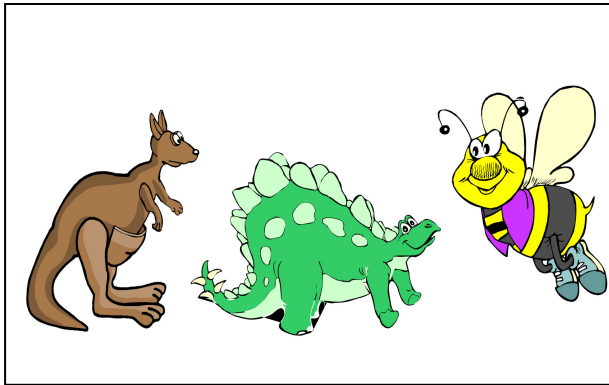


Item 8 a)

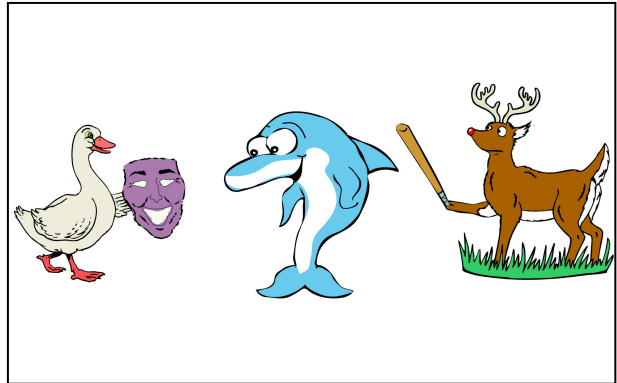
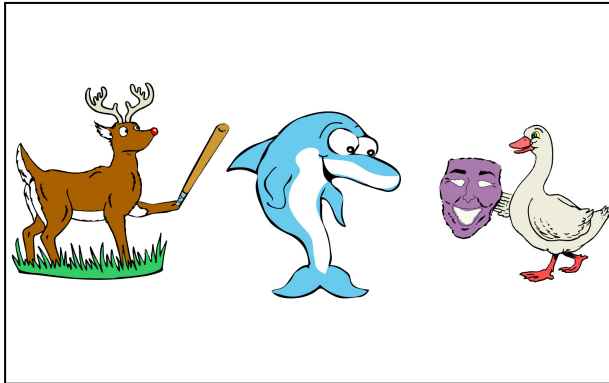


Appendix

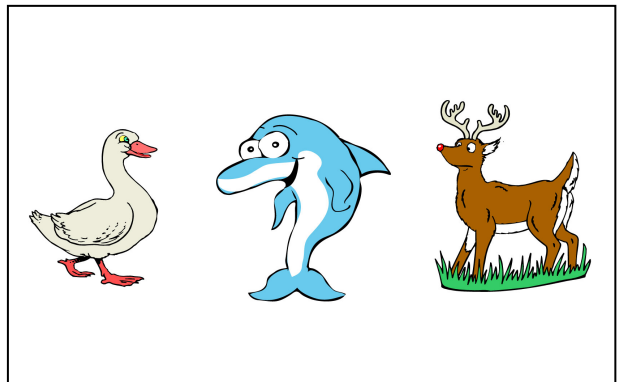
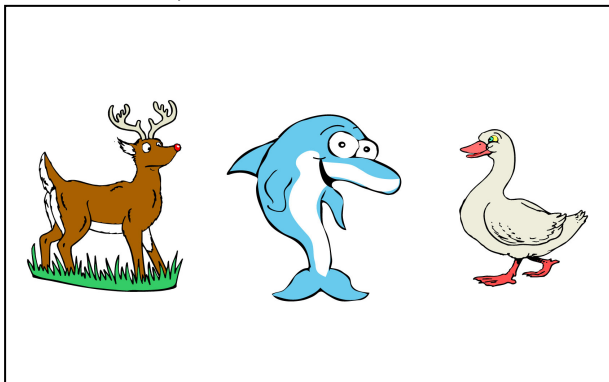
Item 8 b)



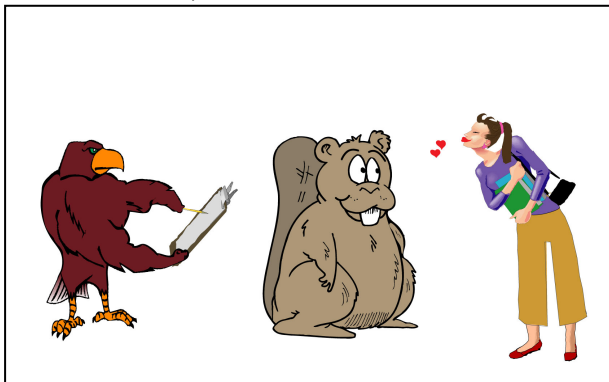
Item 9 a)



Item 9 b)



Item 10 a)

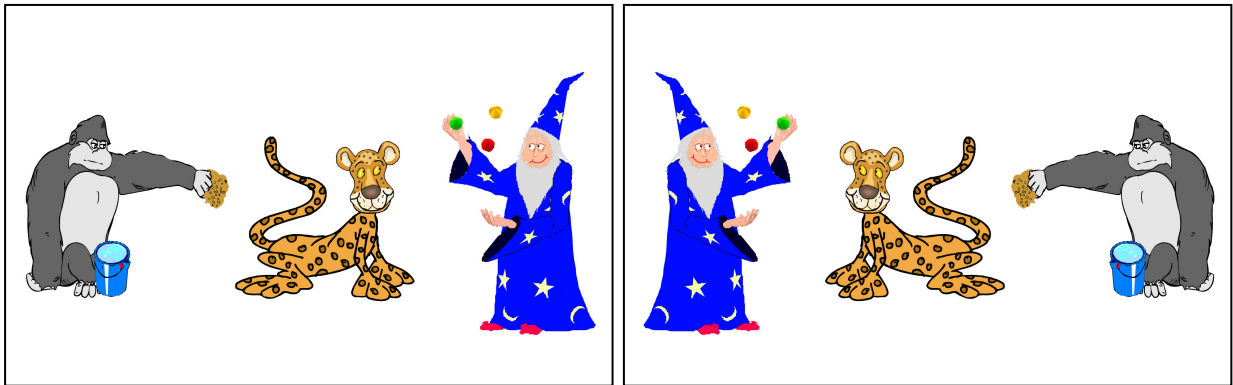


Appendix

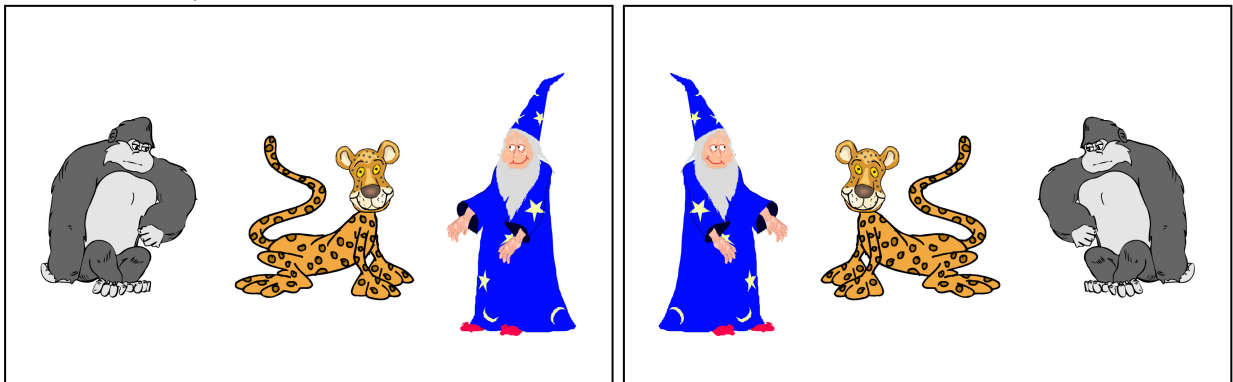
Item 10 b)



Item 11 a)



Item 11 b)

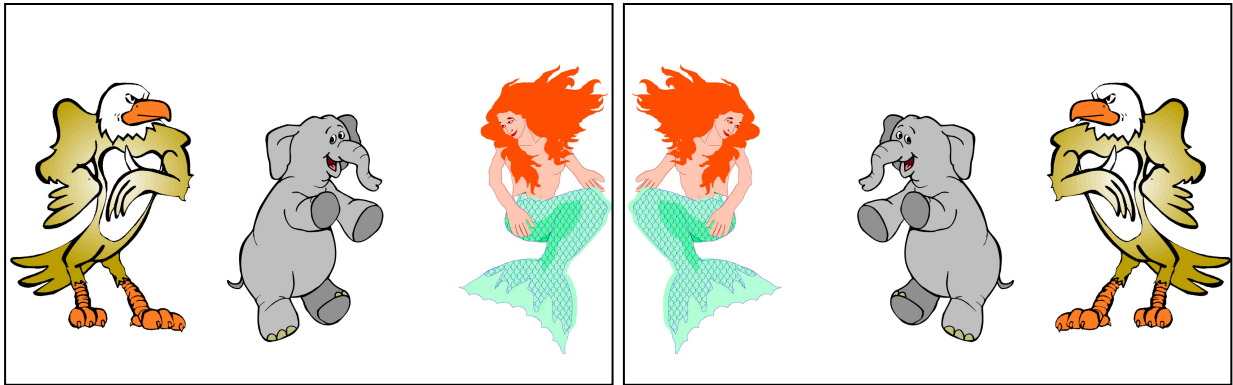


Item 12 a)



Appendix

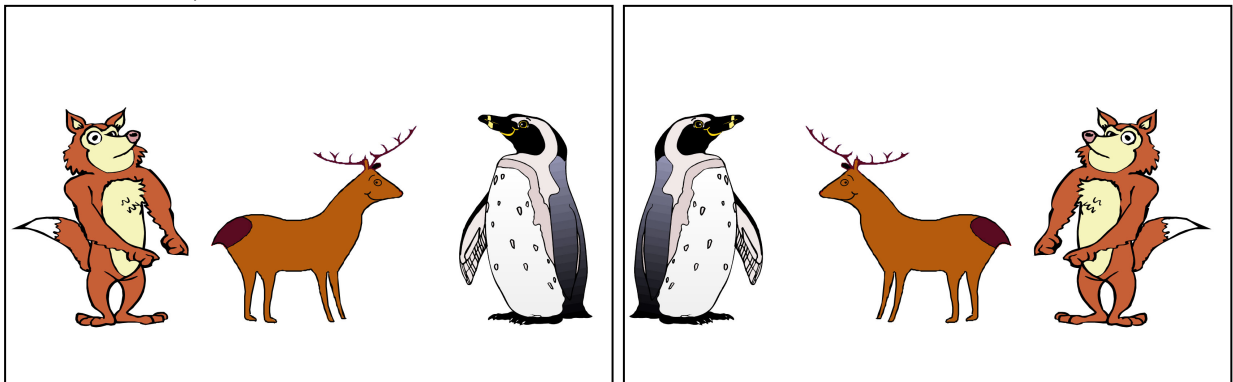
Item 12 b)



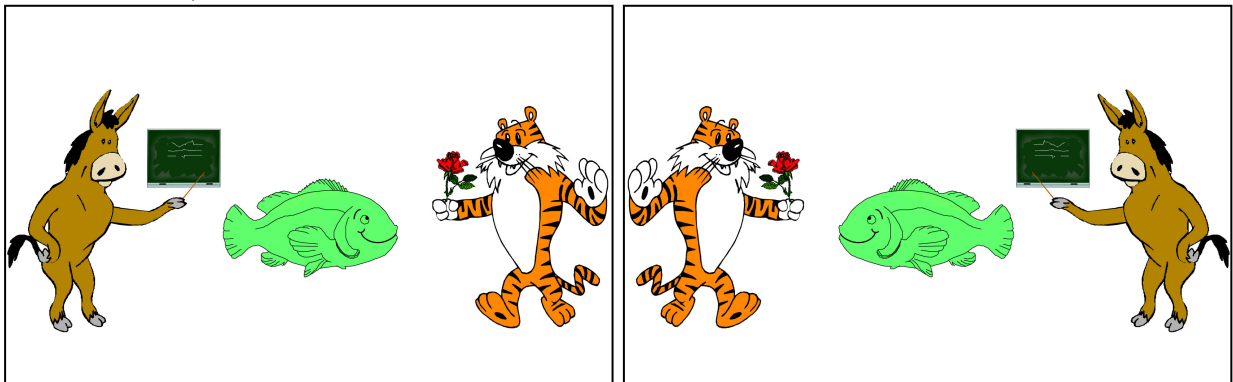
Item 13 a)



Item 13 b)

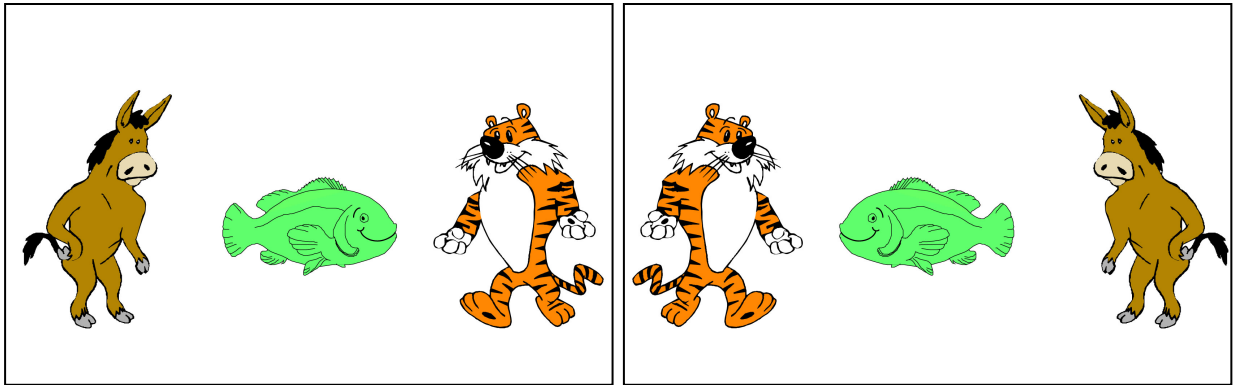


Item 14 a)

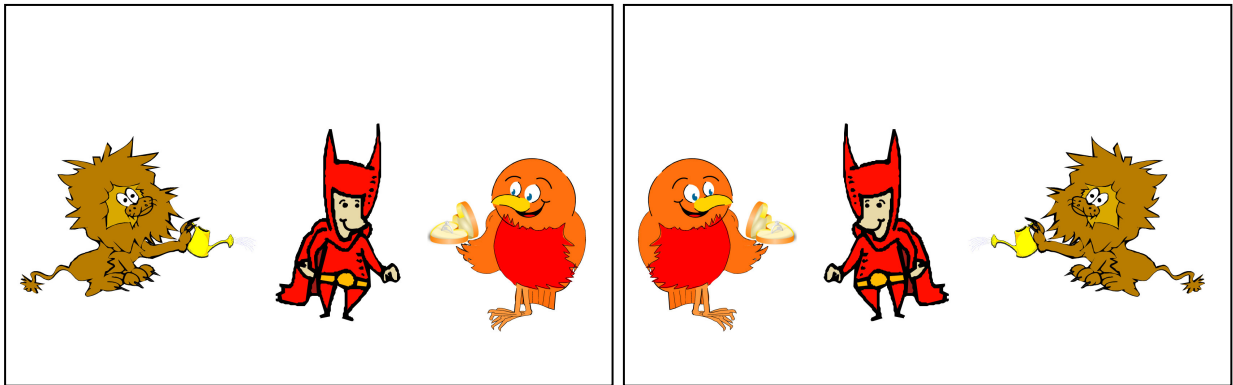


Appendix

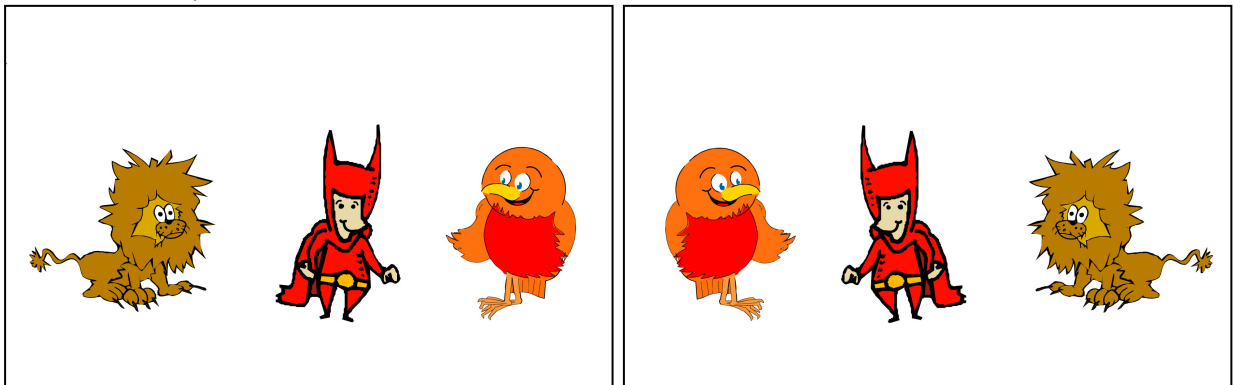
Item 14 b)



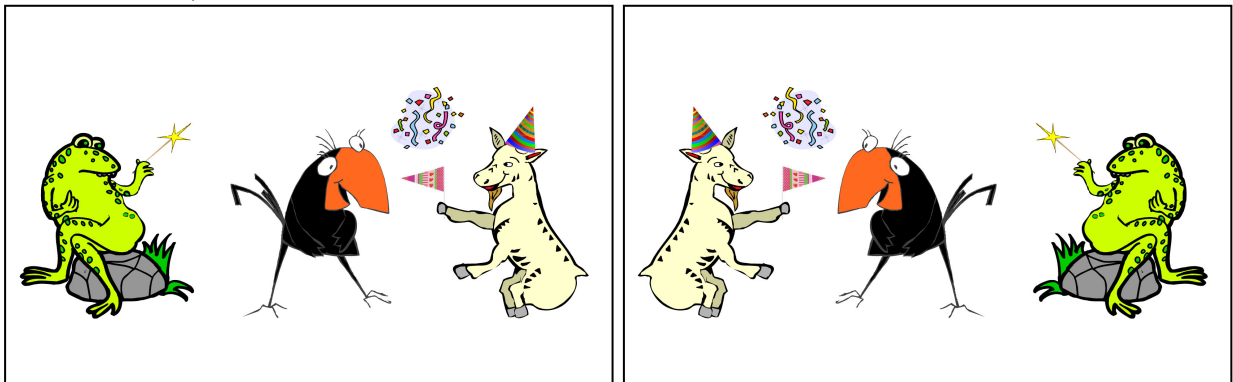
Item 15 a)



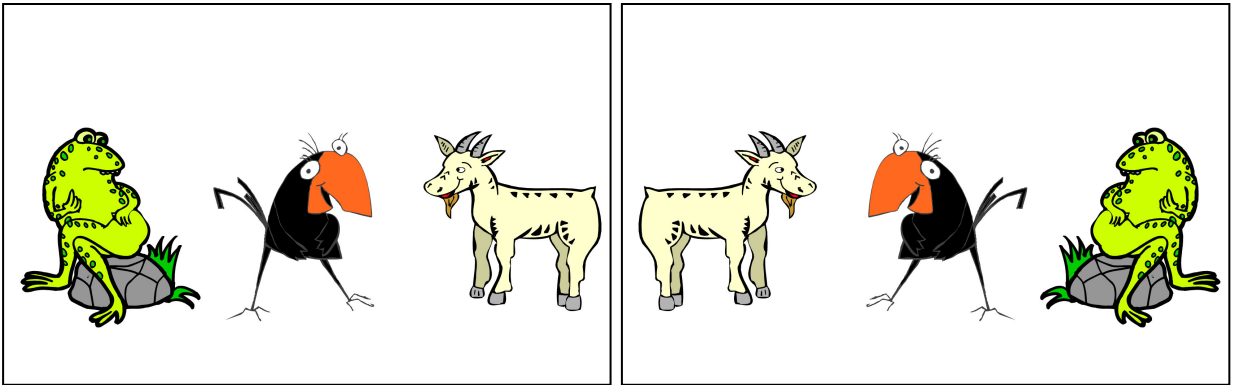
Item 15 b)



Item 16 a)

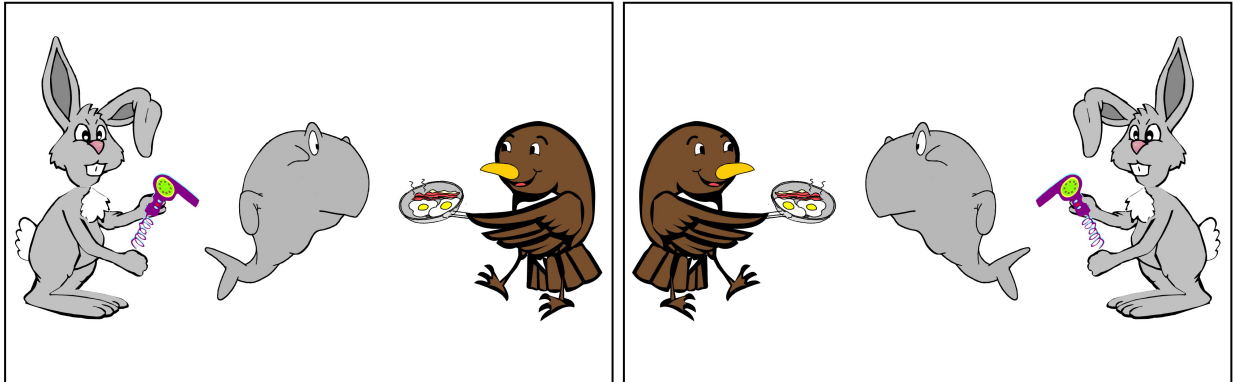


Item 16 b)

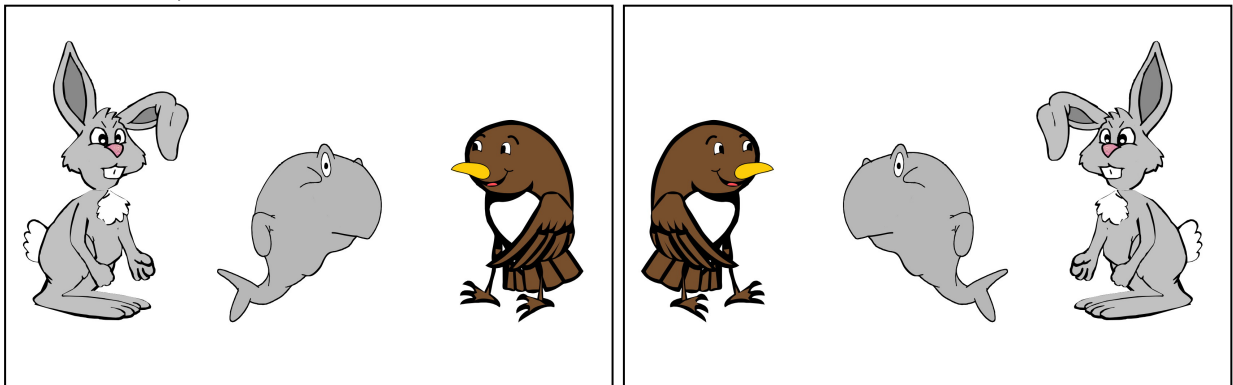


C.2 Item Images for Experiments 2, 3, 5, 6

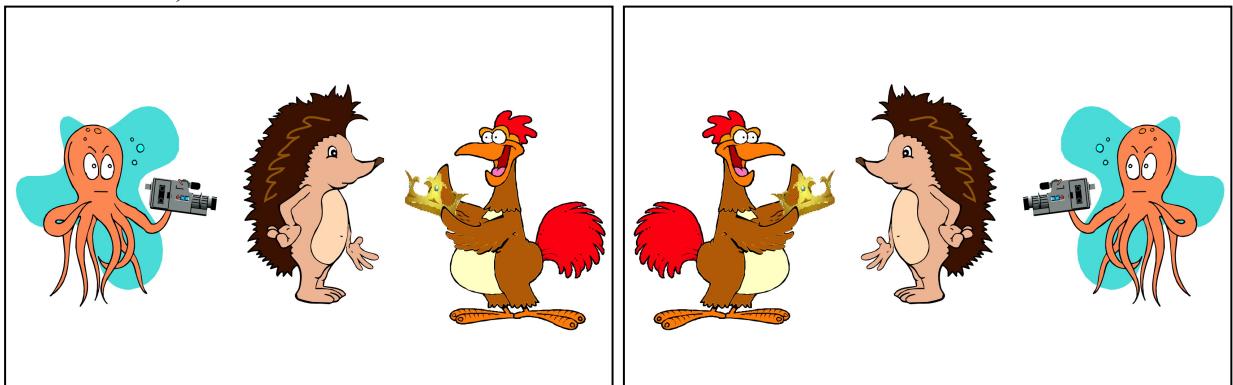
Item 1 a)



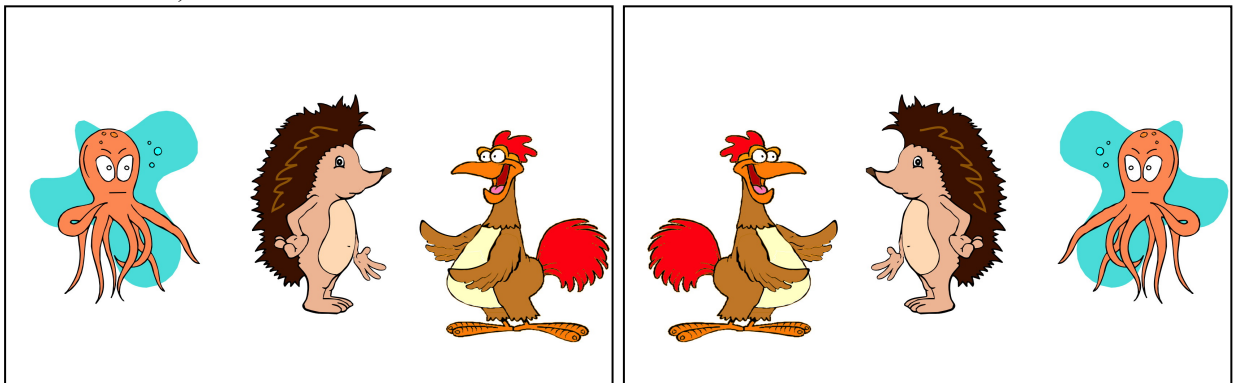
Item 1 b)



Item 2 a)

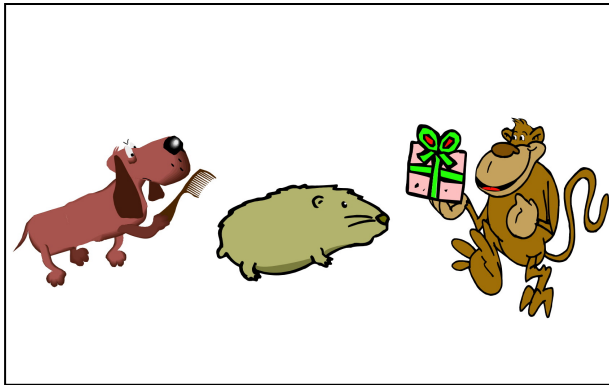


Item 2 b)

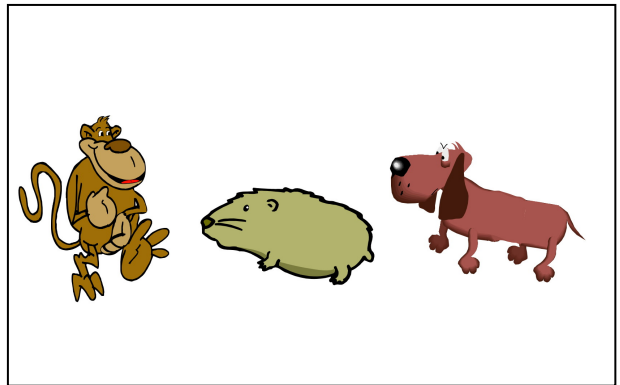
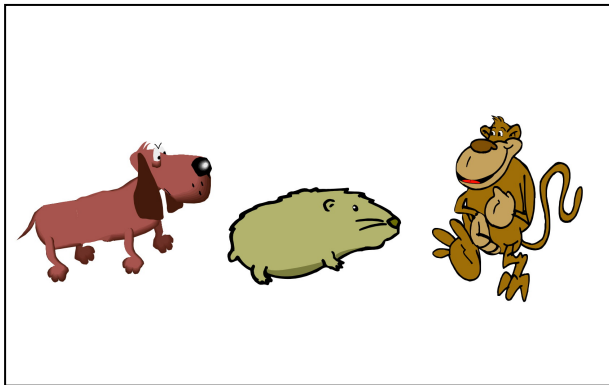


Appendix

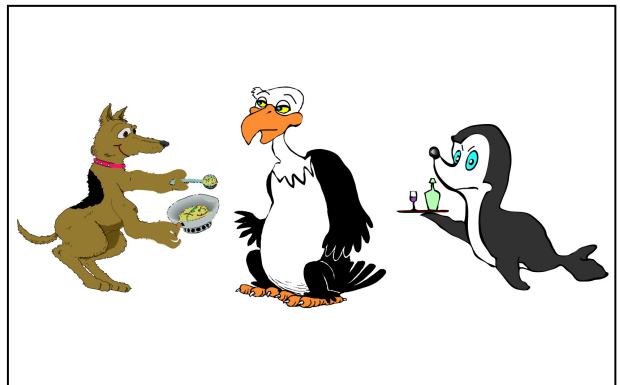
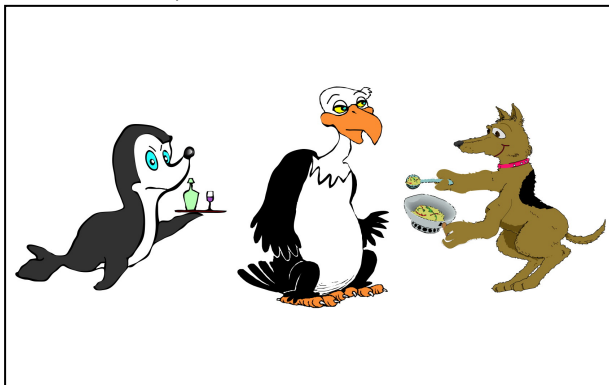
Item 3 a)



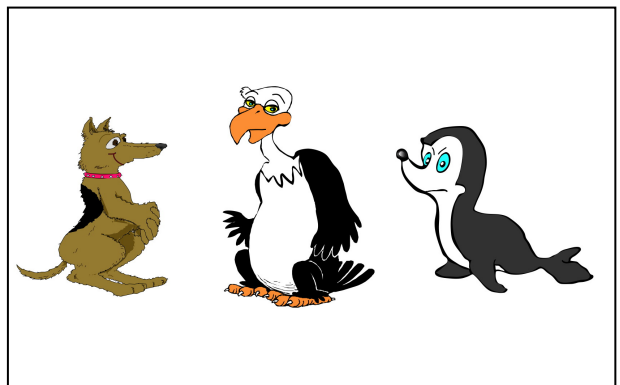
Item 3 b)



Item 4 a)

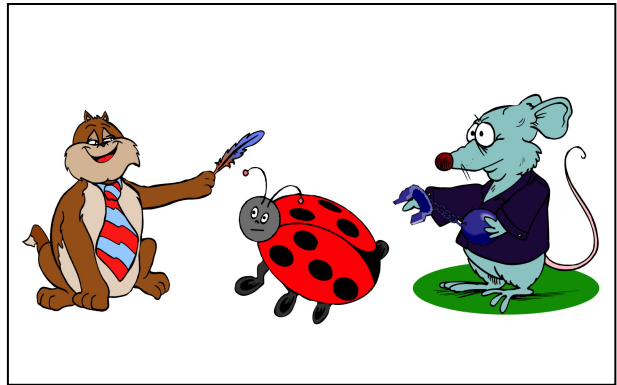
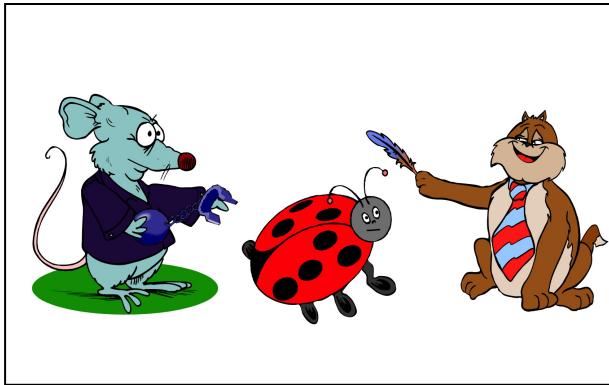


Item 4 b)

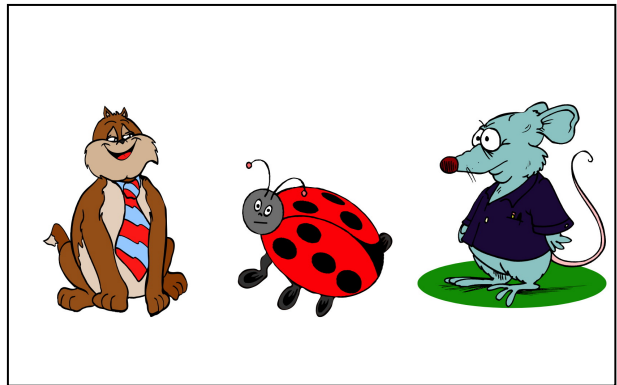
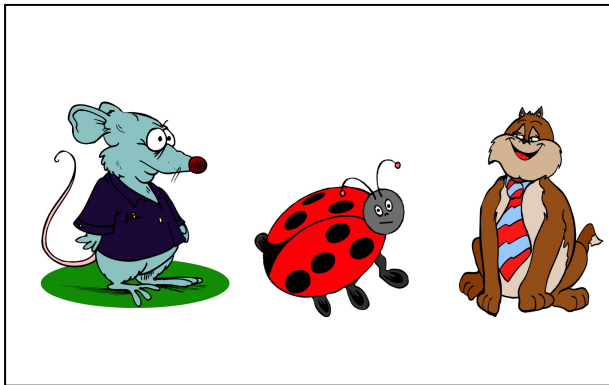


Appendix

Item 5 a)



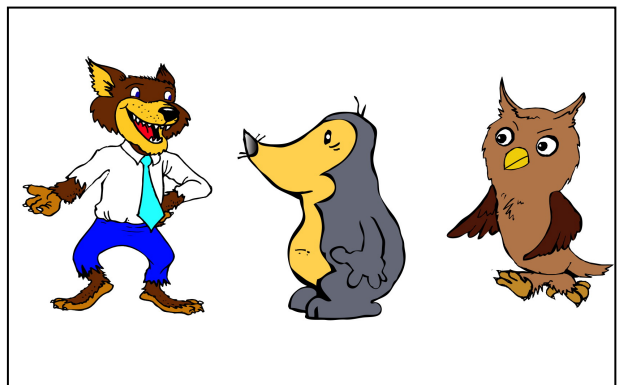
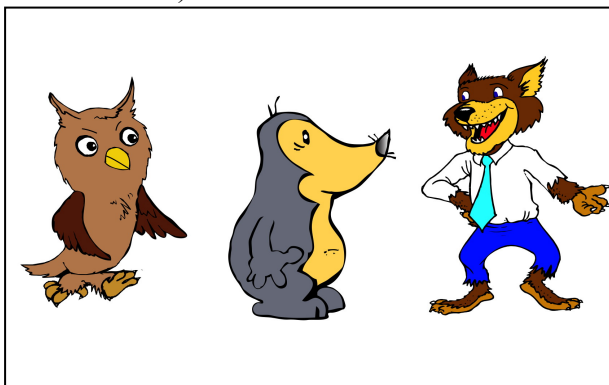
Item 5 b)



Item 6 a)

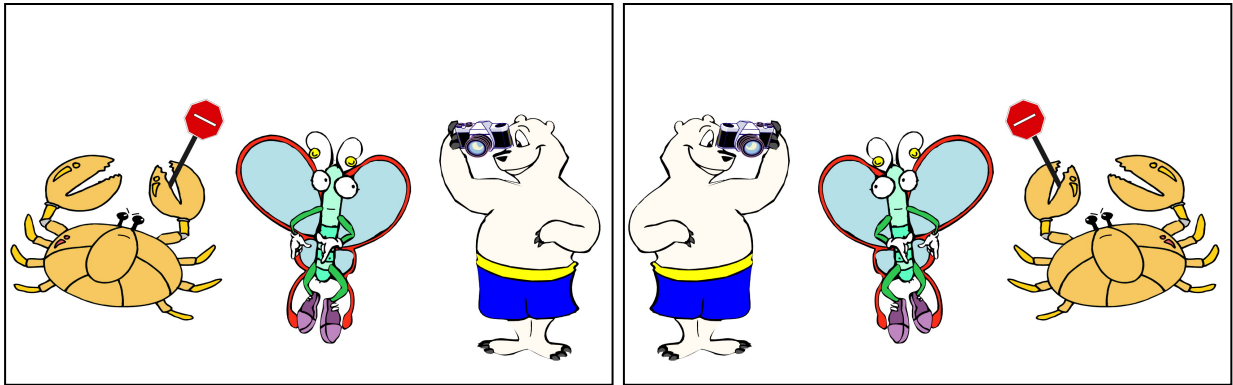


Item 6 b)

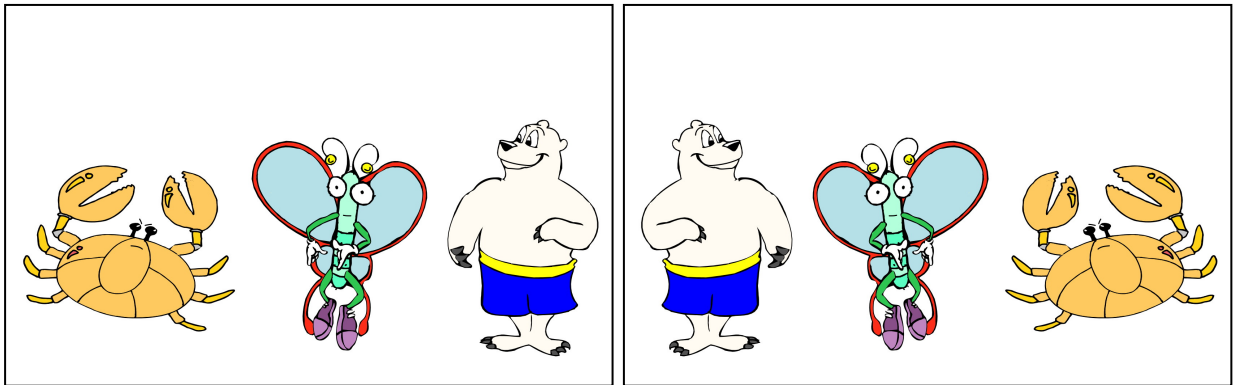


Appendix

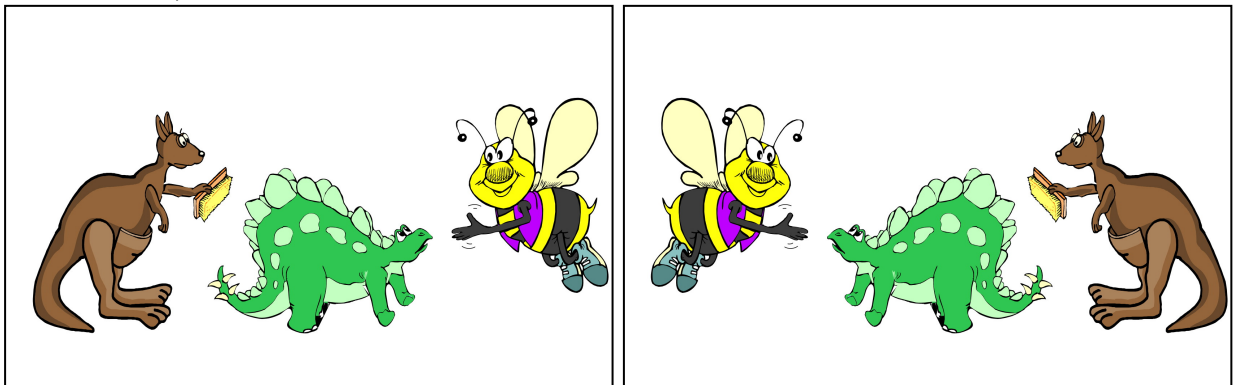
Item 7 a)



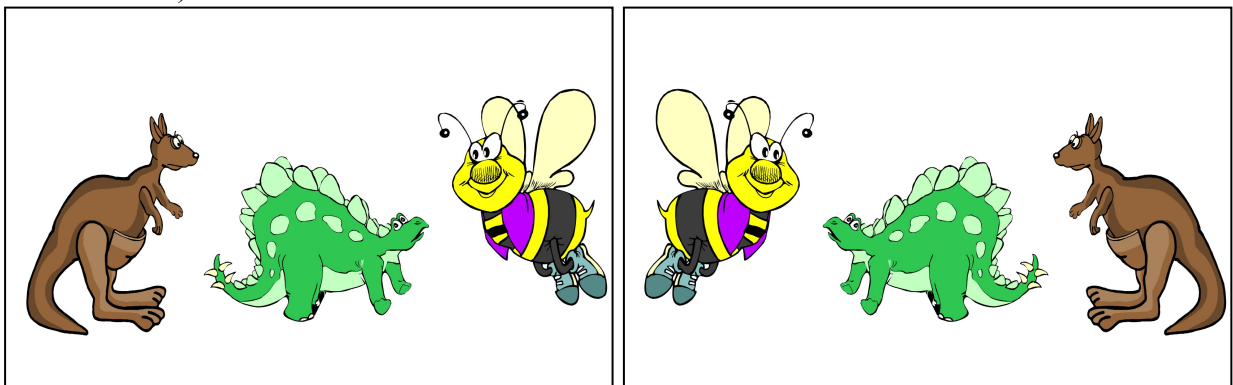
Item 7 b)



Item 8 a)

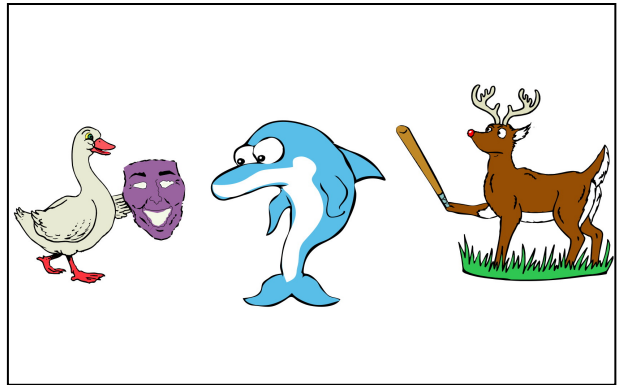
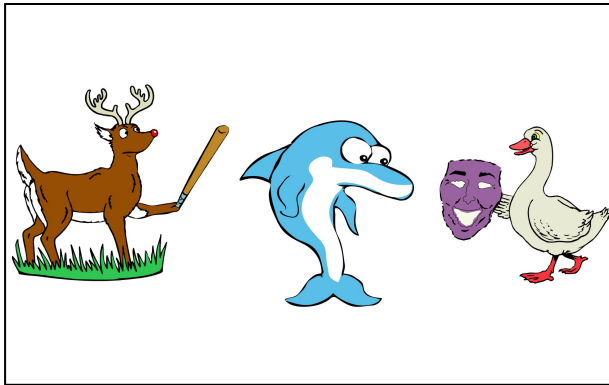


Item 8 b)

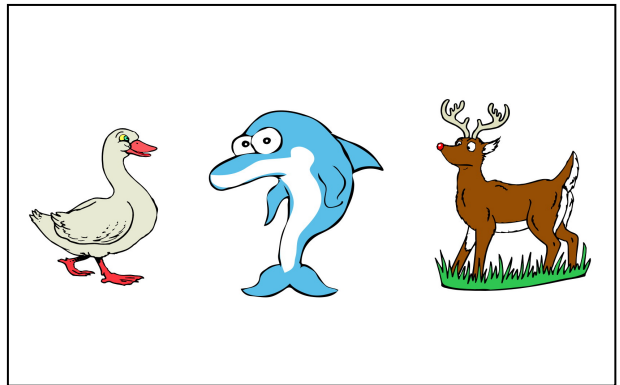
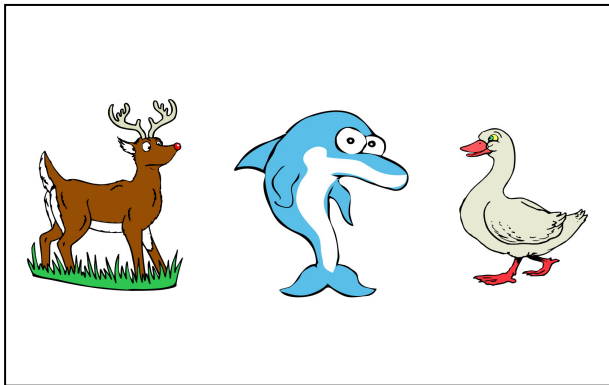


Appendix

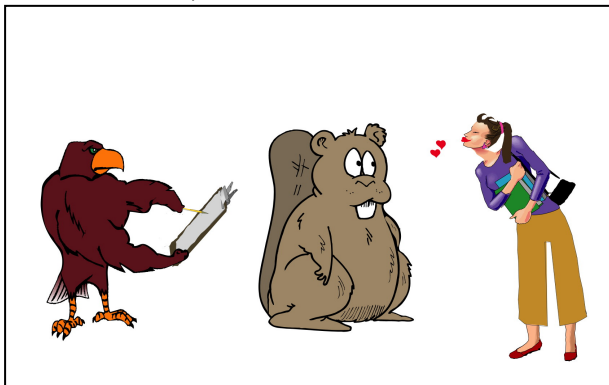
Item 9 a)



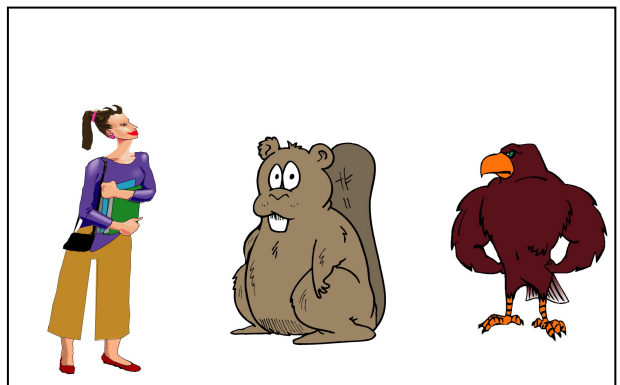
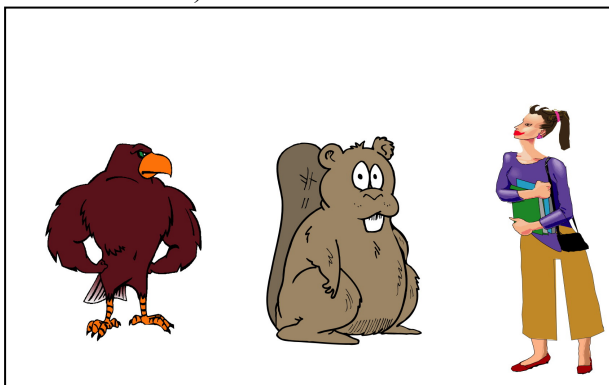
Item 9 b)



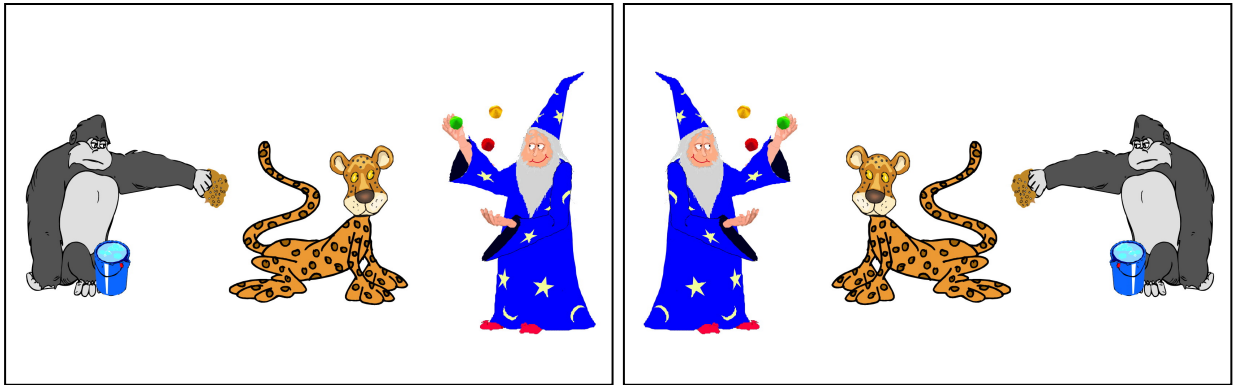
Item 10 a)



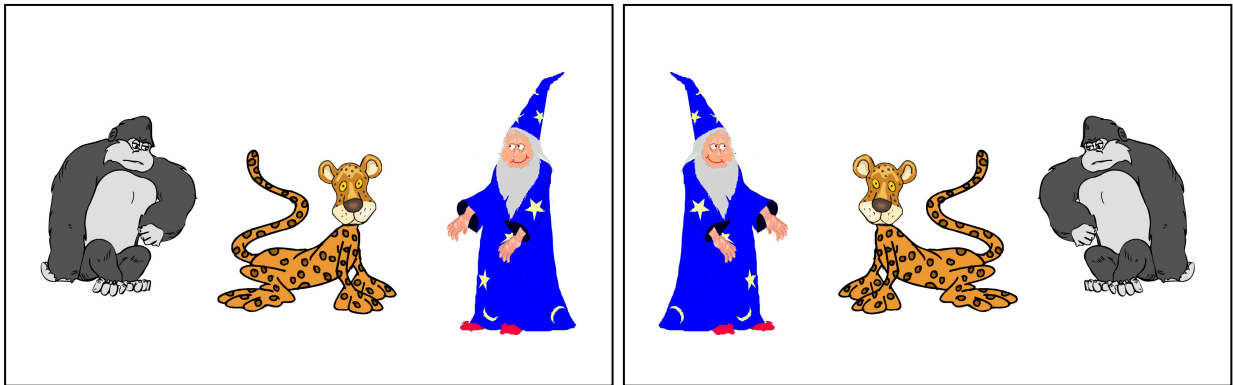
Item 10 b)



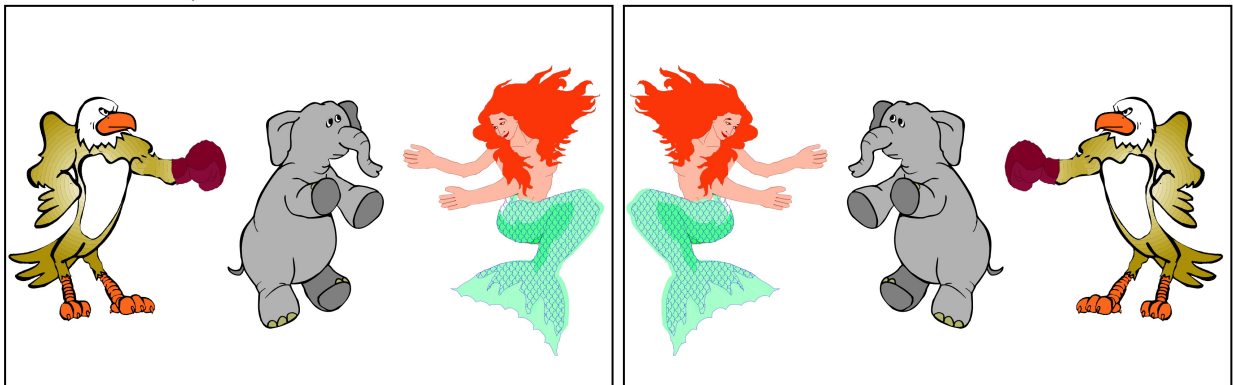
Item 11 a)



Item 11 b)



Item 12 a)



Item 12 b)

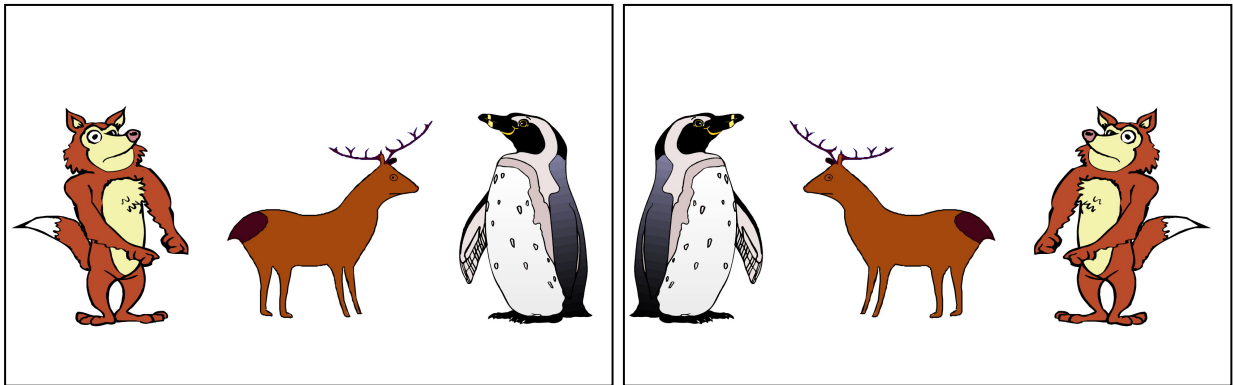


Appendix

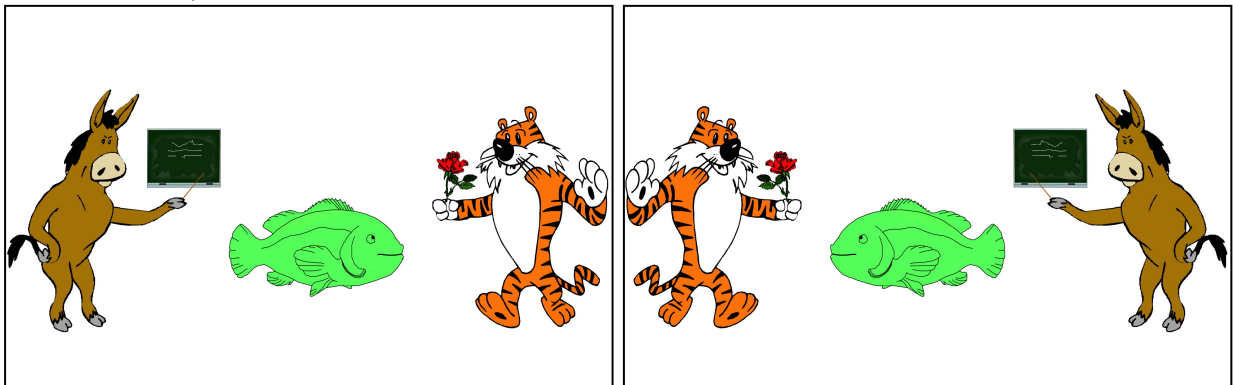
Item 13 a)



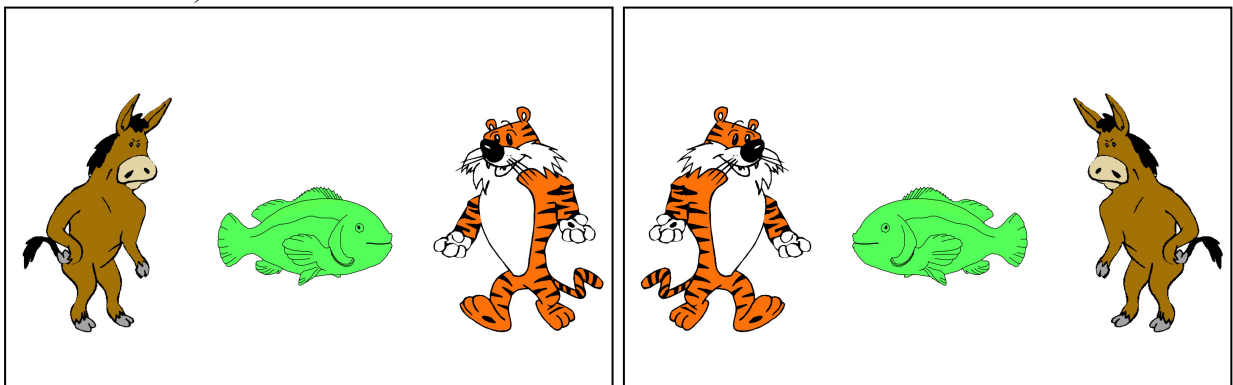
Item 13 b)



Item 14 a)

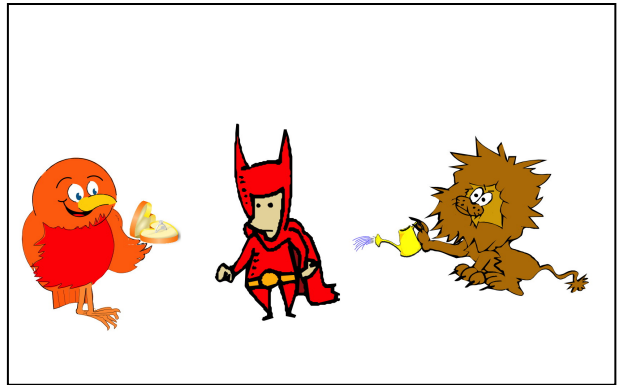


Item 14 b)



Appendix

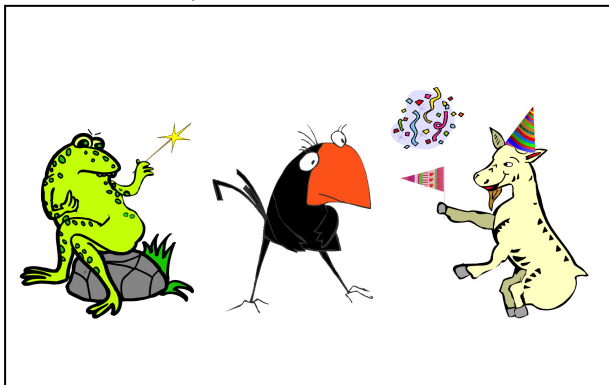
Item 15 a)



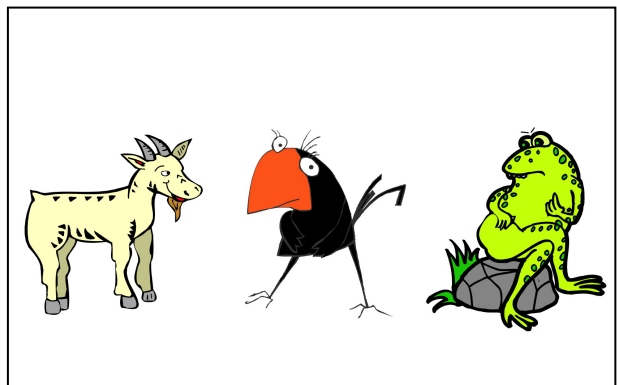
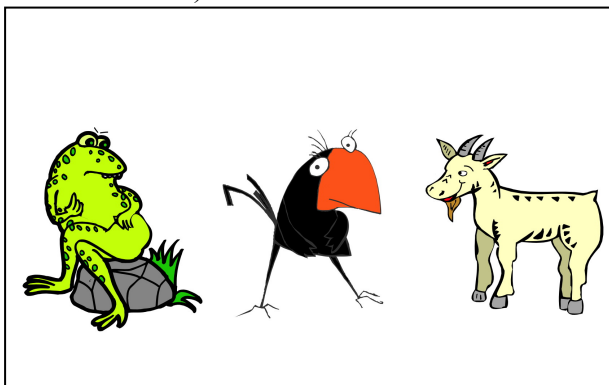
Item 15 b)



Item 16 a)



Item 16 b)

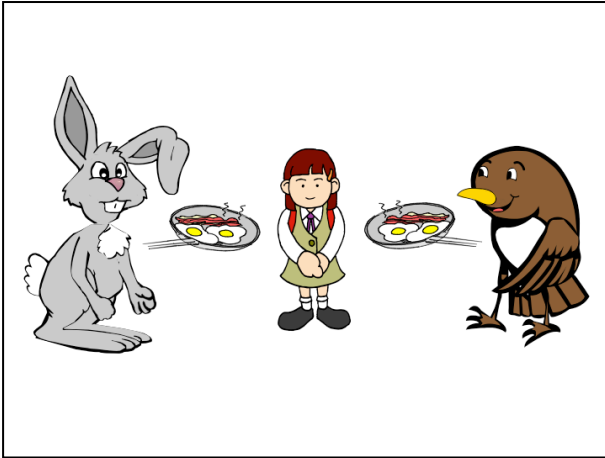


Appendix D: Valence Pretest Items

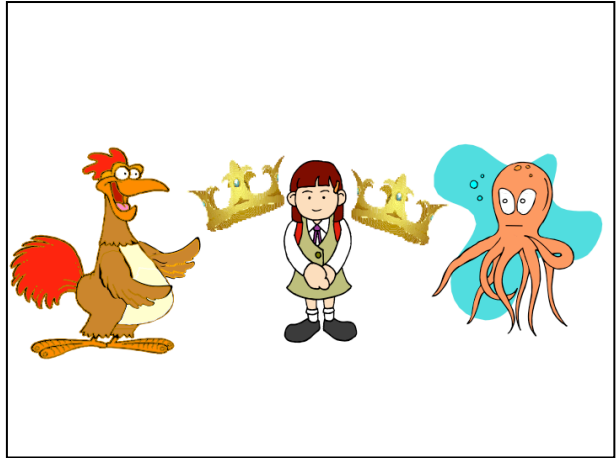
Valence Pretest – Critical Items

Appendix D shows the critical items used to pretest if children understood the positive valence of the adverb and could link it to the target agent's happy facial expression in the critical items. Children were asked questions such as 'Who is cooking happily for the girl?'. Their task was to point to the character in question.

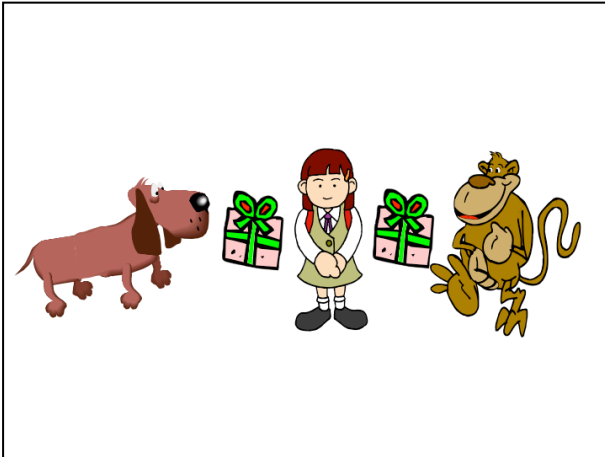
Item 1



Item 2



Item 3

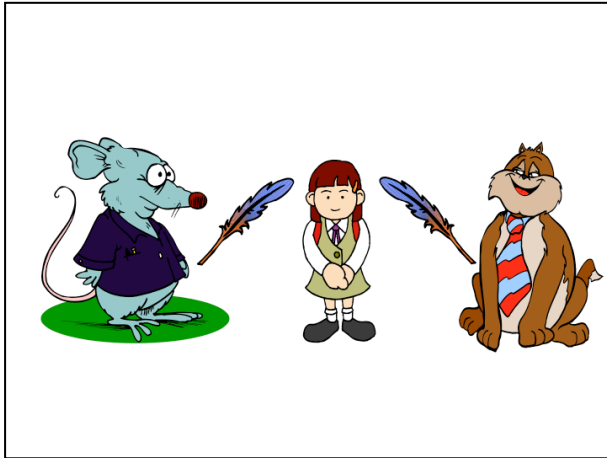


Item 4



Appendix

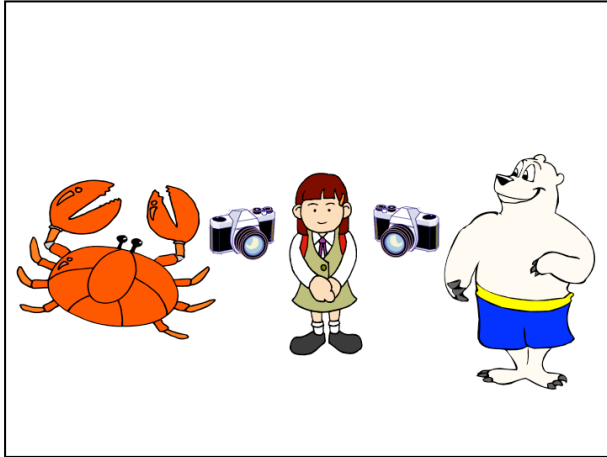
Item 5



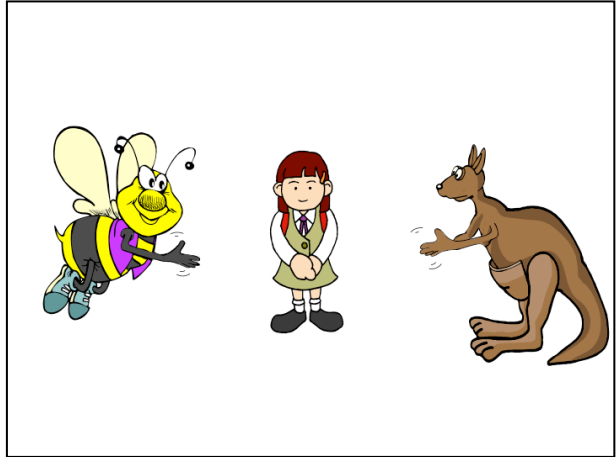
Item 6



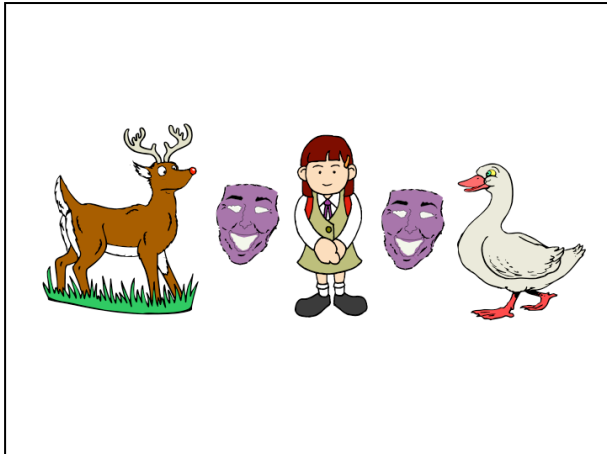
Item 7



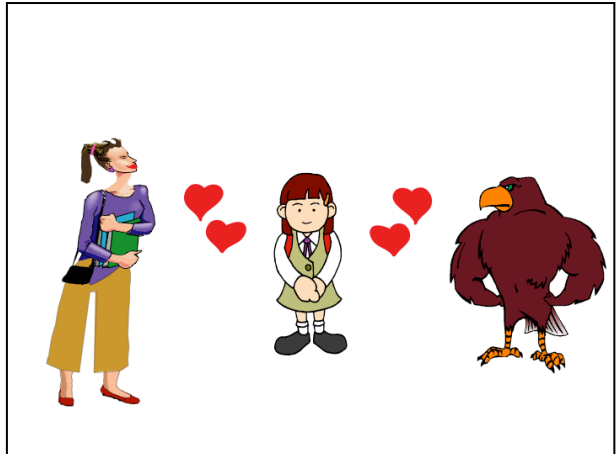
Item 8



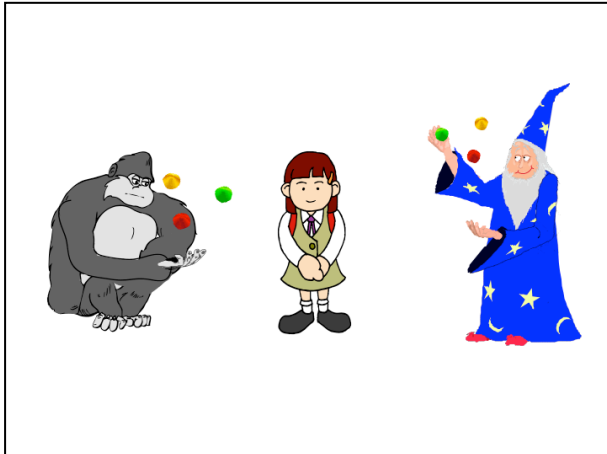
Item 9



Item 10



Item 11



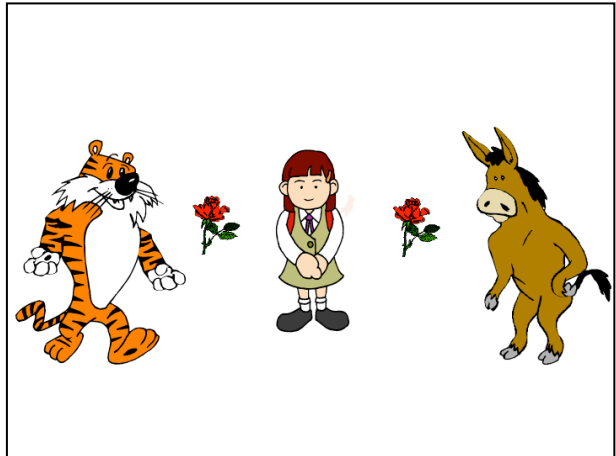
Item 12



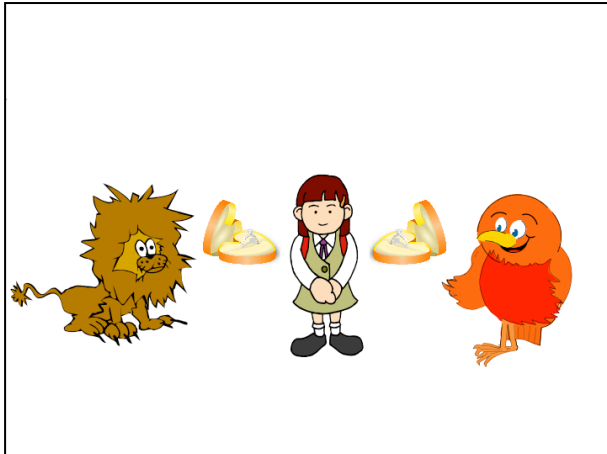
Item 13



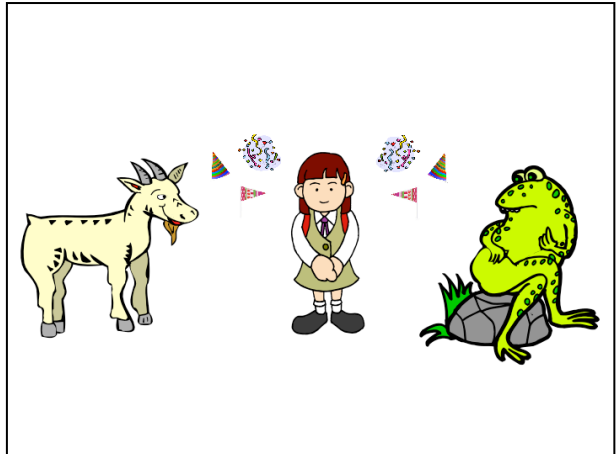
Item 14



Item 15



Item 16



Appendix E: Complementary Accuracy Analyses

Accuracy Results – Generalized Linear Mixed Effects Model

Supplementary to the ANOVA analysis of the accuracy data, we also ran generalized linear mixed effects models in R (R Core Team, 2012) using the `glmer` function in the `lme4` package (Bates, Maechler, Bolker & Walker, 2015) on the accuracy data from all experiments. Generalized linear mixed effects models are better suited for categorical data such as accuracy scores than conventional ANOVAs and provide the additional advantage of taking the variances in subjects and items into account without the need for separate F_1 and F_2 analyses.

All models used prime face and action depiction as fixed factors (with 2 levels each) and subjects and items as random intercepts. Random slopes were also included if the model converged. In experiments 1 (Section 6.2), 2 (Section 6.3) and 4 (Section 7.2) voice (active vs. passive) was used as an additional fixed factor. We transformed the fixed effects into numerical values and centered all fixed factors, so that they have a mean of 0 and a range of 1. This allows data interpretation similar to the main effect and interaction interpretation obtained from an ANOVA and hence, facilitates the comparison across analyses (Baayen, 2008). Additionally, this coding reduces collinearity (Barr, 2008). In all models “Family” was set to “binomial” due to the categorical nature of our accuracy scores. Models were all built starting with the most complex model including interactions in the random slopes and were reduced in complexity until a model converged. We report the model results for each experiment separately and briefly compare them to the ANOVA results reported in the corresponding Accuracy Results Sections of each experiment.

E.1 Complementary Analysis Section 6.2.3.1 (Experiment 1)

The most complex converging model was:

Model E.1a: $\text{answer} \sim \text{prime} * \text{action} + (1 + \text{prime} | \text{subj}) + (1 + \text{prime} | \text{item})$

The model using action as a factor in the random slopes did not converge, neither did more complex models. The results yielded no significant effects (see Fixed Effects Table E.1a). This is different from the ANOVA results, which showed a significant main effect of prime in the item analysis. Yet, our model shows that the subject

Appendix

variance for the prime in the random slope is big compared with the item variance (3.340368 vs. 0.179494 for the item variance of the prime in the random slope). As the model accounts for this variance, we do not see a statistically reliable prime effect in the mixed model.

	Estimate	SE	z-value	p
Intercept	3.2901	0.4643	7.086	1.38e-12 *
Prime	1.8257	1.6465	1.109	0.267
Action	-0.3870	0.5123	-0.755	0.450
Prime x action	0.7398	0.9039	0.818	0.413

Fixed Effects Table E.1a : The estimate, the standard error, the z-value and the p value of the Model E.1a are displayed for the intercept, each fixed factor and their interaction. Significance is marked with an asterisk in the p-value column.

Additionally, we included voice as a fixed factor in the model. The most complex converging model was:

Model E.1b: $\text{answer} \sim \text{prime} * \text{action} * \text{voice} + (1 | \text{subj}) + (1 | \text{item})$

The results yield no significant effects. They are hence, different from the results obtained from the ANOVA, which yielded a significant main effect of voice and a marginal voice x prime interaction in the item analysis. Yet, removing the interactions from the model yielded a significant main effect of voice. The model looked as follows:

Model E.1c: $\text{answer} \sim \text{prime} + \text{action} + \text{voice} + (1 | \text{subj}) + (1 | \text{item})$

Fixed Effects Table E.1b displays the results of the fixed effects of Model E.1c and shows a significant effect of voice. Yet only removing the interaction between prime and action, between action and voice or between prime and voice also yielded a significant effect of voice similar to the results obtained from the ANOVA.

	Estimate	SE	z-value	p
Intercept	5.4388	0.8731	6.229	4.69e-10 *
Prime	0.6496	0.4375	1.485	0.137580
Action	-0.1022	0.4217	-0.242	0.808608
Voice	-2.7074	0.7519	-3.601	0.000317 *

Fixed Effects Table E.1b : The estimate, the standard error, the z-value and the p value of the Model E.1c are displayed for the intercept and each fixed factor. Significance is marked with an asterisk in the p-value column.

E.2 Complementary Analysis Section 6.3.3.2 (Experiment 2)

The most complex converging model was

Model E.2a: $\text{answer} \sim \text{prime} * \text{action} + (1 | \text{subj}) + (1 | \text{item})$

More complex models with random slopes did not converge. The results of Model E.2a yield no significant effects when voice is not included in the model and hence, corroborate the findings obtained in the ANOVA analysis. Including voice as an additional factor, the most complex converging model was Model E.2b:

Model E.2b: $\text{answer} \sim \text{prime} * \text{voice} + \text{action} + (1 | \text{subj}) + (1 | \text{item})$

The model also converged using an interaction between action and prime and between action and voice. All models including one interaction yield similar results in showing a significant main effect of voice (see Fixed Effects Table E.2a for results of Model E.2b). However, in contrast to our ANOVA results, we did not find any significant interactions. Taking again into account that generalized linear mixed effects models can account for the variances in both our items and our subjects in a single model, finding no interaction in the models but marginal interactions or interactions that are only significant in either the F_1 (subjects) or F_2 (items) analyses in the ANOVA is not surprising.

	Estimate	SE	z-value	p
Intercept	5.7845	1.1488	5.035	4.77e-07 *
Prime	-1.1281	1.1830	-0.954	0.34028
Action	0.5658	1.0716	1.268	0.20469
Voice	-2.9764	0.4461	-2.777	0.00548 *
Prime x voice	1.8638	1.2849	1.451	0.14690

Fixed Effects Table E.2a: The estimate, the standard error, the z-value and the p value of Model E.2b are displayed for the intercept, each fixed factor and their interaction. Significance is marked with an asterisk in the p-value column.

E.3 Complementary Analysis Section 6.4.4.2 (Experiment 3)

The most complex converging model was Model E.3a:

Model E.3a: answer ~ prime * action + (1 | subj) + (1 | item)

The results of Model E.3a showed no significant effects. Yet, removing the interaction yielded Model E.3b:

Model E.3b: answer ~ prime + action + (1 | subj) + (1 | item)

The results of Model E.3b are shown in Fixed Effects Table E.3a and display a significant main effect of action, corroborating the ANOVA analysis. Moreover, even though the effect of prime is not significant, it is trending and thus also supporting the ANOVA analysis, which showed a marginal prime effect in the F_1 analysis.

	Estimate	SE	z-value	p
Intercept	2.8154	0.4192	6.717	1.86e-11 *
Prime	0.6407	0.4132	1.551	0.1210
Action	0.9750	0.4348	2.242	0.0249 *

Fixed Effects Table E.3a: The estimate, the standard error, the z-value and the p value of Model E.3b are displayed for the intercept and each fixed factor. Significance is marked with an asterisk in the p-value column.

E.4 Complementary Analysis Section 7.2.4.2 (Experiment 4)

The most complex model that converged was Model E.4a:

Model E.4a: answer ~ prime + action + (1 | subj) + (1 | item)

Appendix

The model including the interaction between prime face and action did not converge; neither did any more complex models including random slopes. As Fixed Effects Table E.4a shows, the results of Model E.4a yielded a significant main effect of action and hence corroborate the ANOVA findings.

	Estimate	SE	z-value	p
Intercept	0.7105	0.2498	2.844	0.00446 *
Prime	-0.1116	0.2352	-0.475	0.63504
Action	0.6645	0.2377	2.795	0.00518 *

Fixed Effects Table E.4a: The estimate, the standard error, the z-value and the p value of Model E.4a are displayed for the intercept, each fixed factor. Significance is marked with an asterisk in the p-value column.

The most complex model including voice as a factor that converged was Model E.4b:

Model E.4b: answer ~ prime * action * voice + (1 | subj) + (1 | item)

The results of Model E.4b yielded a significant main effect of voice ($p > .05$) but no other significant effects. Removing the interaction between prime and action from Model E.4b as in Model E.4c yielded also a significant effect of voice (see Fixed Effects Table E.4b) but no significant effect of action or prime. Yet, specifying an interaction between prime and action (Model E.4d) or between prime and voice (Model E.4e) yields significant effects of both action and voice in both Models (E.4d and E.4e). The Results of Model E.4e can be seen in Fixed Effects Table E.4c. The results of Model E.4d are not shown but the results are similar and yield the same effects as in Fixed Effects Table E.4c.

Model E.4c: answer ~ prime + action * voice + (1 | subj) + (1 | item)

Model E.4d: answer ~ prime * action + voice + (1 | subj) + (1 | item)

Model E.4e: answer ~ prime * voice + action + (1 | subj) + (1 | item)

Two model comparisons (Model E.4c vs. Model E.4d and Model E.4c vs. E.4e) corroborated the effects, suggesting that the interaction between action and voice (not significant in Models E.4b and E.4c but in the F_2 analysis of the ANOVA) conceals the main effects of action and voice, as both Model E.4d and E.4e provide a better fit for the data than Model E.4c (Model E.4c vs E.4d: $\chi^2(0)=0.0373$, $p < .05$; Model E.4c

vs E.4e: $\chi^2(0)=1.0748$, $p < .05$). Thus, the complementary analysis corroborated the ANOVA findings except for the interactions. Regarding the more powerful analysis using mixed models that take both subject and item variances into account it is not surprising that the weak interactions (that were only significant in either the F_1 or the F_2 analysis but not in both) present in the ANOVAs show no significant effects here.

	Estimate	SE	z-value	p
Intercept	3.2561	0.5481	5.941	2.84e-09 *
Prime	-0.1894	0.3082	-0.615	0.539
Action	0.9403	0.7231	1.300	0.193
Voice	-3.9525	0.5324	-7.424	1.14e-13 *
Action x voice	0.2291	0.8059	0.284	0.776

Fixed Effects Table E.4b: The estimate, the standard error, the z-value and the p value of Model E.4c are displayed for the intercept, each fixed factor and their interaction. Significance is marked with an asterisk in the p-value column.

	Estimate	SE	z-value	p
Intercept	2.8909	0.5481	5.274	1.33e-07 *
Prime	0.4622	0.6881	0.672	0.501797
Action	1.1312	0.3237	3.494	0.000476 *
Voice	-3.4837	0.5427	-6.420	1.37e-10 *
Prime x voice	-0.8206	0.7711	-1.064	0.287223

Fixed Effects Table E.4c: The estimate, the standard error, the z-value and the p value of Model E.4e are displayed for the intercept, each fixed factor and their interaction. Significance is marked with an asterisk in the p-value column.

E.5 Complementary Analysis Section 7.3.4.2 (Experiment 5)

The most complex model that converged was Model E.5a:

Model E.5a: answer ~ prime * action + (1 | subj) + (1 | item)

As Fixed Effects Table E.5a shows, we see a significant effect of action but no effect of prime nor a significant interaction between prime and action for Model E.5a. These results replicate the ANOVA findings.

	Estimate	SE	z-value	p
Intercept	-0.5166	0.3333	-1.550	0.121221
Prime	0.0419	0.2734	0.153	0.878212
Action	1.0558	0.2768	3.814	0.000137 *
Prime x action	-0.2641	0.3838	-0.688	0.491400

Fixed Effects Table E.5a: The estimate, the standard error, the z-value and the p value of Model E.5a are displayed for the intercept, each fixed factor and their interaction. Significance is marked with an asterisk in the p-value column.

E.6 Complementary Analysis Section 8.1.4.2 (Experiment 6)

The most complex converging model was Model E.6a:

Model E.6a: answer ~ prime * action + (1 | subj) + (1 | item)

The Model E.6a showed no significant effects. Yet, simplifying the model further, i.e., removing the interaction between prime and action yields Model E.6b:

Model E.6b: answer ~ prime + action + (1 | subj) + (1 | item)

The results of Model E.6b in the Fixed Effects Table E.6a show a significant effect of action. The ANOVA yielded a significant effect of action but this effect was only significant in the F_2 analysis and only marginal in the F_1 analysis ($p = .083$). Moreover, that the action effect for the ANOVA F_1 analysis was only marginal might be explained by the substantial subject variance that can be seen in Model E.6b (subject random effect variance: 3.6103 vs. item random effect variance: 0.3624).

	Estimate	SE	z-value	p
Intercept	2.9323740	0.5026169	5.834	5.4e-09 *
Prime	-0.0005447	0.3042044	-0.002	0.9986
Action	0.8089474	0.3149962	2.568	0.0102 *

Fixed Effects Table E.6a: The estimate, the standard error, the z-value and the p value of Model E.6b are displayed for the intercept and each fixed factor. Significance is marked with an asterisk in the p-value column.

References

- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23(3), 275–290.
- Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Research*, 1079(1), 25–35.
- Aguado, L., Garcia-Gutierrez, A., Castañeda, E., & Saugar, C. (2007). Effects of prime task on affective priming by facial expressions of emotion. *The Spanish Journal of Psychology*, 10(2), 209–217.
- Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, 33(6), 952–965.
- Alloppenna, 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, 419–439.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583–609.
- Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- 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.
- Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, 137(2), 208–216.
- Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, 54, 218–250.
- Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science*, 10(3), 249–255.
- Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). *The Regensburg word fluency test (Regensburger Wortflüssigkeits-Test)*. Göttingen: Hogrefe.

References

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Bach, D. R., Schmidt-Daffy, M., & Dolan, R. J. (2014). Facial expression influences face identity recognition during the attentional blink. *Emotion, 14*(6), 1007–1013.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379–384.
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of memory and language, 59*(4), 457–474.
- Bassili, J. N. (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37*, 2049–2058.
- Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. *Mechanisms of Language Acquisition, 157–193*.
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
- Beer, J. S., & Ochsner, K. N. (2006). Social cognition: A multi level analysis. *Brain Research, 1079*(1), 98–105.
- Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology, 3*(1), 105–135.
- Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*(2), 383–411.
- Bergelson, E., & Swingle, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America, 109*(9), 3253–3258.
- Berman, J. M., Chambers, C. G., & Graham, S. A. (2016). Preschoolers’ real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology, 142*, 391–399.
- Berman, J. M. J., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development, 84*(5), 1791–1805.
- Bernat, E., Bunce, S., & Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *International Journal of Psychophysiology, 42*(1), 11–34.

References

- Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C., & Lee, D. C. (2006). *Perception of robot smiles and dimensions for human-robot interaction design* (pp. 469–474). Presented at the IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, UK.
- Bock, J. K. (1986). Syntactic persistence in language processing. *Cognitive Psychology*, *18*, 355–387.
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: evidence from wh-questions. *Journal of Memory and Language*, *34*, 774–806.
- Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition*, *85*(2), B21–B30.
- Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436.
- Boyd, J. K., & Goldberg, A. E. (2011). Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Language*, *39*(03), 457–481.
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.
- Brennan, S. (1985). Caricature generator: The dynamic exaggeration of faces by computer. *Leonardo*, *18*, 170–178.
- Brey, E., & Shutts, K. (2014). Children use nonverbal cues to make inferences about social power. *Child Development*, *86*(1), 276–286.
- Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(01), 1–20.
- Bullock, M., & Russell, J. A. (1984). Preschool Children's Interpretation of Facial Expressions of Emotion. *International Journal of Behavioral Development*, *7*(2), 193–214.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

References

- Cacioppo, J. T., Berntson, G. G., Bechara, A., Tranel, D., & Hawkley, L. C. (2011). Could an aging brain contribute to subjective well being: The value added by a social neuroscience perspective. In A. Todorov, S. T. Fiske, D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 249–262). New York, NY: Oxford University Press.
- Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia, 41*, 195–202.
- Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods, 40*(1), 109–115.
- Carminati, M. N., & Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PLoS ONE, 8*(9), e72559.
- Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70*, 205–218.
- Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology Section A, 58*(7), 1173–1197.
- Carstensen, L. L., Fung, H. H., & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and Emotion, 27*, 103–123.
- 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), 687–696.
- Chang, Y. H. (2006). Comparing affective perception in cartoon, schematic and real faces. In *Perception, 35*, 206–206.
- Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology, 106*(1), 41–61.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Codispoti, M., Mazzetti, M., & Bradley, M. M. (2009). Unmasking emotion: Exposure duration and emotional engagement. *Psychophysiology, 46*(4), 731–738.
- Cordon, I. M., Melinder, A. M. D., Goodman, G. S., & Edelman, R. S. (2013). Children's

References

- and adults' memory for emotional pictures: Examining age-related patterns using the Developmental Affective Photo System. *Journal of Experimental Child Psychology*, *114*(2), 339–356.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge, MA: Cambridge University Press.
- Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development*, *83*(6), 2042–2056.
- Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*(4), 487–509.
- Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, *5*(5), 190–204.
- 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.
- de Bot, K. (2000). Sociolinguistics and language processing mechanisms. *Sociolinguistica*, *14*(1), 74–77.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.
- DeCot, B. M. (2011). *The role of speaker facial expression in processing emotional sentences*. (Unpublished Master's Thesis). Bielefeld University, Bielefeld.
- Del Giudice, M., & Colle, L. (2007). Differences between children and adults in the recognition of enjoyment smiles. *Developmental Psychology*, *43*(3), 796–803.
- DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*(3), 226–239.
- Denham, S., Warren, H., Salisch, M. V., Benga, O., Chin, J. C., & Geangu, E. (2011). Emotions and social development in childhood. In P. K. Smith & C. H. Hart (Eds.), *The Wiley-Blackwell handbook of childhood social development* (2nd ed., pp. 413–433). Cambridge: Blackwell.
- De Sonneville, L. M. J., Verschoor, C. A., Njiokiktjien, C., Op het Veld, V., Toorenaar, N., & Vranken, M. (2002). Facial identity and facial emotions: Speed, accuracy,

References

- and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology*, 24(2), 200–213.
- Di Domenico, A., Palumbo, R., Mammarella, N., & Fairfield, B. (2015). Aging and emotional expressions: is there a positivity bias during dynamic emotion recognition? *Frontiers in Psychology*, 6, 1–5.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107–117.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008a). German children's comprehension of word order and case marking in causative sentences. *Child Development*, 79(4), 1152–1167.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008b). Young German children's early syntactic competence: a preferential looking study. *Developmental Science*, 11(4), 575–582.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2014). Familiar verbs are not always easier than novel verbs: how german pre-school children comprehend active and passive sentences. *Cognitive Science*, 38(1), 128–151.
- Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, 298(5596), 1191–1194.
- Doost, H. T. N., Moradi, A. R., Taghavi, M. R., Yule, W., & Dalgleish, T. (1999). The development of a corpus of emotional words produced by children and adolescents. *Personality and Individual Differences*, 27, 433–451.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 547–619.
- Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, 97(1), 14–27.
- Dürscheid, C. 2000. *Syntax* (Vol.3). Wiesbaden: Westdeutscher Verlag
- 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.
- Ebner, N. C., He, Y., Fichtenholtz, H. M., McCarthy, G., & Johnson, M. K. (2010). Electrophysiological correlates of processing faces of younger and older individuals. *Social Cognitive and Affective Neuroscience*, nsq074.
- Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken

References

- language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146.
- Ferreira, F. (2003). The misinterpretation of canonical sentences. *Cognitive Psychology*, *47*, 164–203.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.
- Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516–547.
- Finn, B., Roediger, H. L., & Rosenzweig, E. (2012). Reconsolidation from negative emotional pictures: Is successful retrieval required? *Memory & Cognition*, *40*(7), 1031–1045.
- Fivush, R., Brotman, M. A., Buckner, J. P., & Goodman, S. H. (2000). Gender differences in parent-child emotion narratives. *Sex Roles*, *42*, 233–253.
- Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.
- Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465–1477.
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304.
- Gaskell, M. G. (2008). Preface. In M. G. Gaskell (Eds.). *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.
- Geiselman, R. E., & Bellezza, F. S. (1977). Incidental retention of speaker's voice. *Memory & Cognition*, *5*(6), 658–665.
- Germine, L., Cashdollar, N., Düzel, E., & Duchaine, B. (2011). A new selective developmental deficit: Impaired object recognition with normal face recognition. *Cortex*, *47*(5), 598–607.
- Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules. *Psychological*

References

- Science*, 7(8), 684–691.
- Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge University Press.
- Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, 28(1), 1–26.
- Gnepp, J. (1983). Children's social sensitivity: Inferring emotions from conflicting cues. *Developmental Psychology*, 19, 805–814.
- Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, 59(5), 675–692.
- Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology*, 40(3), 213–225.
- Gross, A. L., & Ballif, B. (1991). Children's Understanding of Emotion from Facial Expressions and Situations: A Review. *Developmental Review*, 11, 368–398.
- Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience*, 25(4), 909–919.
- Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods*, 40(2), 512–521.
- Grünloh, T., Lieven, E., & Tomasello, M. (2011). German children use prosody to identify participant roles in transitive sentences. *Cognitive Linguistics*, 22(2), 1–27.
- Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, 142(3), 880–905.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 883–899.
- Hajcak, G., & Dennis, T. A. (2009). Brain potentials during affective picture processing in children. *Biological Psychology*, 80(3), 333–338.
- Hareli, S., & Hess, U. (2012). The social signal value of emotions. *Cognition & Emotion*, 26(3), 385–389.

References

- Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal on Mental Retardation*, *104*(3), 270–278.
- Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484.
- Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. Sankt Augustin: Infix-Verlag.
- Hemforth, B., & Konieczny, L. (Eds.). (2000). *German sentence processing (studies in theoretical psycholinguistics)* (Vol. 24). Dordrecht: Kluwer Academic Publishers.
- Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology*, *50*(1), 243–271.
- Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: eye movements and the visual world* (pp. 1–58). New York: Psychology Press.
- Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. *Journal of Child Psychology and Psychiatry*, *45*(7), 1185–1198.
- Hermans, D., Houwer, J. D., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition & Emotion*, *8*(6), 515–533.
- Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion*, *17*(1), 83–99.
- Hills, P. J., & Lewis, M. B. (2011). The own-age face recognition bias in children and adults. *The Quarterly Journal of Experimental Psychology*, *64*(1), 17–23.
- Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697.
- Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2008). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience*, *21*(11), 2245–2262.

References

- Hortaçsu, N., & Ekinçi, B. (1992). Children's reliance on situational and vocal expression of emotions: Consistent and conflicting cues. *Journal of Nonverbal Behavior*, *16*(4), 231–247.
- Huetting, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.
- Huetting, 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.
- Isaacowitz, D. M., Allard, E. S., Murphy, N. A., & Schlangel, M. (2009). The time course of age-related preferences toward positive and negative stimuli. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *64*(2), 188–192.
- Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging*, *22*(1), 147–159.
- Isaacowitz, D. M., Toner, K., Goren, D., & Wilson, H. R. (2008). Looking while unhappy. *Psychological Science*, *19*, 848–853.
- Isaacowitz, D. M., Wadlinger, H. A., Goren, D., & Wilson, H. R. (2006). Selective preference in visual fixation away from negative images in old age? An eye-tracking study. *Psychology and Aging*, *21*(1), 40–48.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*, 1–19.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003a). 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. M. (2003b). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from

References

- German and English. *Journal of Psycholinguistic Research*, 32, 37–55.
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, 1148(1), 138–148.
- Kaschak, M. P. (2006). What this construction needs is generalized. *Memory & Cognition*, 34(2), 368–379.
- Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133(3), 450–467.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253–260.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25–41.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205–225.
- Kim, S. K., & Sumner, M. (2015). Effects of emotional prosody and attention on semantic priming (pp. 1099–1104) In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.). (2015). *Proceedings of the 37th annual meeting of the cognitive science society*. Austin, TX: Cognitive Science Society.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104(30), 12577–12580.
- Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, 27(4), 623–634.
- Kiss, I., & Ennis, T. (2001). Age-related decline in perception of prosodic affect. *Applied Neuropsychology*, 8(4), 251–254.
- Kissler, J., & Keil, A. (2008). Look-don't look! How emotional pictures affect pro- and anti-saccades. *Experimental Brain Research*, 188(2), 215–222.
- Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords-Early cortical responses to emotional words during reading. *Psychological Science*, 18, 475–

480.

- Knoeferle, P. (2015a). Language comprehension in rich non-linguistic contexts: combining eye tracking and event-related brain potentials. In W. Roel (Ed.), *Towards a cognitive neuroscience of natural language use* (pp. 77–100). Cambridge: Cambridge University Press.
- Knoeferle, P. (2015b). Visually situated language comprehension in children and in adults. In R. K. Mishra, S. Narayanan, & F. Huettig (Eds.), *Attention and vision in language processing* (pp. 57–75). New Delhi: Springer India.
- 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*, 95–127.
- 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., & Guerra, E. (2012). What's non-linguistic visual context? A view from language comprehension. In R. Finkbeiner, J. Meibauer, & P. B. Schumacher, (Eds.), *What is a Context?: Linguistic approaches and challenges*, (Vol. 196, pp. 129-153). Philadelphia: John Benjamins Publishing.
- Kreysa, H., Knoeferle, P., & Nunnemann, E. M. (2014). Effects of speaker gaze versus depicted actions on visual attention during sentence comprehension. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 2513–2518). Austin, TX: Cognitive Science Society.
- Kreysa, H., Knoeferle, P., & Nunnemann, E. M. (in prep.). Distinct effects of different visual cues on immediate comprehension and later recall: The case of speaker gaze versus depicted actions.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word

References

- expectancy and semantic association. *Nature*, 307, 161–163.
- Labov, W., Ash, S., Ravindranath, M., Tracey, W., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4), 431–463.
- Labov, W. (1972). *Sociolinguistic patterns* (No. 4). University of Pennsylvania Press.
- Lai, V. T., Willems, R. M., & Hagoort, P. (2015). Feel between the lines: Implied emotion in sentence comprehension. *Journal of Cognitive Neuroscience*, 27(8), 1528–1541.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
- Lamy, D., Amunts, L., & Bar-Haim, Y. (2008). Emotional priming of pop-out in visual search. *Emotion*, 8(2), 151–161.
- Lang P. J. , Bradley M. M. , Cuthbert B. N. *International affective picture system (IAPS): technical manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida, 1998.
- Lashley K. S. (1960). In search of the engram. In F. A. Beach, D. O. Hebb, C. T. Morgan & H. W. Nissen (Eds.), *The neuropsychology of Lashley – Selected papers* (pp. 478–506). New York: McGraw-Hill.
- Lavie, R.-J. (2005). *Exemplar theory in linguistics: a perspective for the cognitive subject*. Presented at the Communication to the 11th Congress of Cognitive Linguistics, Bordeaux.
- Leclerc, C. M., & Kensinger, E. A. (2011). Neural processing of emotional pictures and words: a comparison of young and older adults. *Developmental Neuropsychology*, 36(4), 519–538.
- Li, W., Zinbarg, R. E., Boehm, S. G., & Paller, K. A. (2008). Neural and behavioral evidence for affective priming from unconsciously perceived emotional facial expressions and the influence of trait anxiety. *Journal of Cognitive Neuroscience*, 20(1), 95–107.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- Mack, J. E., Meltzer-Asscher, A., Barbieri, E., & Thompson, C. K. (2013). Neural correlates of processing passive sentences. *Brain Sciences*, 3(3), 1198–1214.

References

- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*(2), 157–201.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behaviour, 23*, 127–150.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake – but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847.
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1–71.
- Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 676–684.
- Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun–adjective agreement in Spanish. *Brain Research, 1093*(1), 178–189.
- Martín-Loeches, M., Fernández, A., Schacht, A., Sommer, W., Casado, P., Jiménez-Ortega, L., & Fondevila, S. (2012). The influence of emotional words on sentence processing electrophysiological and behavioral evidence. *Neuropsychologia, 50*(14), 3262–3272.
- Mather, M. (2012). The emotion paradox in the aging brain. *Annals of the New York Academy of Sciences, 1251*(1), 33–49.
- Mather, M., & Carstensen, L. L. (2003). Aging and attentional biases for emotional faces. *Psychological Science, 14*(5), 409–415.
- Mather, M., Canli, T., English, T., Whitefield, S., Wais, P., Ochsner, K., et al. (2004). Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychological Science, 15*, 259–263.
- Matzke, M., Mai, H., Nager, W., Rösseler, J., & Münte, T. (2002). The costs of freedom: an ERP-study of non-canonical sentences. *Clinical Neuropsychology, 113*, 844–852.

References

- McNamara T. P. & Holbrook J. B. (2003). Semantic memory and priming. In R. W. Proctor & A. F. Healy (Eds.), *Handbook of psychology: Experimental psychology* (Vol. 4, pp. 447–474). New York: Wiley.
- 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.
- Mehl, M. R., & Pennebaker, J. W. (2003). The social dynamics of a cultural upheaval social interactions surrounding September 11, 2001. *Psychological Science*, *14*(6), 579–585.
- Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, *9*(5), 619–630.
- Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1), 1–10.
- Mishra, R. K., Olivers, C. N. L., & Huettig, F. (2013). Spoken language and the decision to move the eyes – To what extent are language-mediated eye movements automatic? In C. V. S. Pammi & N. Srinivasan (Eds.), *Progress in brain research* (Vol. 202, pp. 135–149). Amsterdam: Elsevier.
- Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science*, *10*, 419–422.
- Morton, J. B., & Trehub, S. E. (2001). Children’s understanding of emotion in speech. *Child Development*, *72*, 834–843.
- Murphy, N. A., & Isaacowitz, D. M. (2008). Preferences for emotional information in older and younger adults: a meta-analysis of memory and attention tasks. *Psychology and Aging*, *23*(2), 263–286.
- Murray, B. D., & Kensinger, E. A. (2012). The effects of emotion and encoding strategy on associative memory. *Memory & Cognition*, *40*(7), 1056–1069.
- 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 conference of the cognitive science society* (pp. 2675–2680). Austin, TX: Cognitive Science Society.
- Münster, K., Carminati, M. N., & Knoeferle, P. (2015). The effect of facial emotion and

References

- action depiction on situated language processing. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.). *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 1673–1678). Austin, TX: Cognitive Science Society.
- Myachykov, A., Scheepers, C., Fischer, M. H., & Kessler, K. (2013). TEST: A tropic, embodied, and situated theory of cognition. *Topics in Cognitive Science, 6*(3), 442–460.
- Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*(4), 314–329.
- Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology, 21*(4), 337–360.
- Niedźwiecka, A., & Tomalski, P. (2015). Gaze-cueing effect depends on facial expression of emotion in 9- to 12-month-old infants. *Frontiers in Psychology, 6*, 1–11.
- Novick, J. M., Thompson-Schill, S., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition, 107*, 850–903.
- Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion, 6*(2), 257–268.
- O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Routledge.
- Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381–396.
- Olguin, R., & Tomasello, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development, 8*(3), 245–272.
- Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *The Journal of the Acoustical Society of America, 119*(3), 1727–1739.
- Pagliaccio, D., Luby, J. L., Gaffrey, M. S., Belden, A. C., Botteron, K. N., Harms, M. P., & Barch, D. M. (2013). Functional brain activation to emotional and nonemotional faces in healthy children: Evidence for developmentally undifferentiated amygdala function during the school-age period. *Cognitive,*

References

- Affective, & Behavioral Neuroscience*, 13(4), 771–789.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519–526.
- Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1-2), 185–198.
- Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, 104(3), 262–269.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, 109(6), 2988–2998.
- Picardo, R., Baron, A. S., Anderson, A. K., & Todd, R. M. (2016). Tuning to the positive: Age-related differences in subjective perception of facial emotion. *PLoS ONE*, 11(1), e0145643.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169–190.
- Pickering, M. J., & Garrod, S. (2009). Prediction and embodiment in dialogue. *European Journal of Social Psychology*, 39(7), 1162–1168.
- Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curbio*, 19(4), 319–324.
- Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division.
- Pons, F., & Harris, P. (2005). Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. *Cognition & Emotion*, 19(8), 1158–1174.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, 4, 515–526.
- Průcha, J. (1972). Psycholinguistics and sociolinguistics – separate or integrated. *Linguistics*, 10(89), 9–23.
- Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110(2), 81–94.
- R Development Core Team (2008). R: A language and environment for statistical

References

- computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.
- Rayner, K., McConkie, G. W., & Ehrlich, S. (1978). Eye Movements and integrating information across fixations. *Journal of Experimental Psychology. Human Perception and Performance*, *4*, 529–544.
- Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research*, *1376*, 66–75.
- Reed, A. E., Chan, L., & Mikels, J. A. (2014). Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. *Psychology and Aging*, *29*(1), 1–15.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology*, *19*, 473–497.
- Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174.
- Richards, A., French, C. C., Nash, G., Hadwin, J. A., & Donnelly, N. (2007). A comparison of selective attention and facial processing biases in typically developing children who are high and low in self-reported trait anxiety. *Development and Psychopathology*, *19*, 481–495.
- Rickheit, G., Sichelschmidt, L., & Strohner, H. (2002). *Psycholinguistik. Die Wissenschaft vom sprachlichen Verhalten und Erleben*. Tübingen: Stauffenburg.
- Ridgeway, D., Waters, E., & Kuczaj, S. A. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology*, *21*, 901–908.

References

- Riediger, M., Voelkle, M. C., Ebner, N. C., & Lindenberger, U. (2011). Beyond “happy, angry, or sad?”: Age-of-poser and age-of-rater effects on multi-dimensional emotion perception. *Cognition & Emotion*, *25*(6), 968–982.
- Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, *32*(4), 863–881.
- Ruffman, T., Ng, M., & Jenkin, T. (2009). Older adults respond quickly to angry faces despite labeling difficulty. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64*(2), 171–179.
- Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development*, *18*(2), 139–158.
- Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*(05), 754–760.
- Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual-world studies. *Acta Psychologica*, *137*, 118–172.
- Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114.
- Sato, W., Yoshikawa, S., Kochiyama, T., & Matsumura, M. (2004). The amygdala processes the emotional significance of facial expressions: an fMRI investigation using the interaction between expression and face direction. *NeuroImage*, *22*(2), 1006–1013.
- Sato, T., & Wakebe, T. (2014). How do young children judge intentions of an agent affecting a patient? Outcome-based judgment and positivity bias. *Journal of Experimental Child Psychology*, *118*, 93-100.
- Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations. *Developmental Science*, *6*(5), 557–567.
- Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: a visual-world study. In M. Carreiras & J. C. Clifton (Eds.), *The on-line study of sentence comprehension: eyetracking, ERP and beyond* (pp. 167–185). United Kingdom: Psychology Press.
- Scherf, K. S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental

References

- trajectories. *Developmental Science*, *10*(4), F15–F30.
- Schipke, C. S. (2012). *Processing mechanisms of argument structure and case-marking in child development: Neural correlates and behavioral evidence* (Doctoral Dissertation). Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2012 (MPI Series in Human Cognitive and Brain Sciences; 139), 1–172.
- Schipke, C. S., Friederici, A. D., & Oberecker, R. (2011). Brain responses to case-marking violations in German preschool children. *NeuroReport*, *22*, 850–854.
- Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*, 227–246.
- Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, *194*(3), 465–475.
- Semin, G. R. (1995). Interfacing language and social cognition. *Journal of Language and Social Psychology*, *14*, 182–194.
- Semin, G. R., & Fiedler, K. (1992). The configuration of social interaction in interpersonal terms. In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 58–78). London: Sage.
- Siewierska, A. (1988). *Word order rules*. London: Croom Helm.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.
- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.
- Snedeker, J. (2013). Children's sentence processing. In R. van Gompel (ed.), *Sentence processing* (pp. 189–220). New York, New York: Psychology Press.
- Snedeker, J., Thorpe, K., & Trueswell, J. (2001). On choosing the parse within the scene: The role of visual context and verb bias in ambiguity resolution. In J. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 964–96). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*, 195–200.

References

- Squires, L. (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics*, *17*(2), 200–237.
- Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425–436.
- Staub Casasanto, L. (2008). Does social information influence sentence processing? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 799–804). Austin, TX: Cognitive Science Society.
- Sternberg, S. (2013). The meaning of additive reaction-time effects: some misconceptions. *Frontiers in Psychology*, *4*, 1–3.
- 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*, 1632–1634.
- Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes*, *11*(6), 583–588.
- Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Science*, *7*, 23–30.
- Thibault, P., Bourgeois, P., & Hess, U. (2006). The effect of group-identification on emotion recognition: The case of cats and basketball players. *Journal of Experimental Social Psychology*, *42*(5), 676–683.
- Thomas, K. M., Drevets, W. C., Whalen, P. J., Eccard, C. H., Dahl, R. E., Ryan, N. D., & Casey, B. J. (2001). Amygdala response to facial expressions in children and adults. *Society of Biological Psychiatry*, *49*, 309–316.
- Todd, R. M., Cunningham, W. A., Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences*, *16*(7), 365–372.
- Todd, R. M., Evans, J. W., Morris, D., Lewis, M. D., & Taylor, M. J. (2010). The changing face of emotion: age-related patterns of amygdala activation to salient faces. *Social Cognitive and Affective Neuroscience*, *6*(1), 12–23.
- Tomasello, M. (1992). The social bases of language acquisition. *Social Development*, *1*(1), 67–87.
- Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition.

References

- Cognitive Linguistics*, 11(1/2), 61–82.
- Tomasello, M., & Brooks, P. J. (1998). Young children's earliest transitive and intransitive constructions. *Cognitive Linguistics*, 9(4), 379–396.
- Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, 18(2), 121–147.
- Tottenham, N., Phuong, J., Flannery, J., Gabard-Durnam, L., & Goff, B. (2013). A negativity bias for ambiguous facial-expression valence during childhood: Converging evidence from behavior and facial corrugator muscle responses. *Emotion*, 13(1), 92–103.
- Treese, A.-C., Johansson, M., & Lindgren, M. (2010). ERP correlates of target–distracter differentiation in repeated runs of a continuous recognition task with emotional and neutral faces. *Brain and Cognition*, 72(3), 430–441.
- Trueswell, J., & Gleitman, L. R. (2004). Children's eye movements during listening: Evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (Eds.), *Interface of language, vision, and action: Eye movements and the visual world* (pp. 319–346). NY: Psychology Press.
- Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying online sentence processing in young children. *Cognition*, 73, 89–134.
- 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), 285–318.
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 155–179). Hillsdale, NJ: Lawrence Erlbaum.
- Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *NeuroImage*, 15(2), 454–461.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, 134(3), 383–403.

References

- 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 Valin Jr, R. D. (1999). Generalized semantic roles and the syntax-semantics interface. *Empirical issues in formal syntax and semantics*, 2, 373–389.
- Walker, A., & Hay, J. (2011). Congruence between ‘word age’ and ‘voice age’ facilitates lexical access. *Laboratory Phonology*, 2(1), 219–237.
- 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.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale Revised*. New York: Psychological Corporation.
- Weighall, A. R., & Altmann, G. T. M. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of Child Language*, 38(3), 579–605.
- Weist, R. M. (1983). The word order myth. *Journal of Child Language*, 10(01), 97–106.
- Wellman, H. M., Harris, P. L., Banerjee, M., & Sinclair, A. (1995). Early understanding of emotion: Evidence from natural language. *Cognition & Emotion*, 9(2-3), 117–149.
- Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39(1), 114–128.
- Wolfram, W. (1991). *Dialects and American English*. Englewood Cliffs, NJ: Prentice Hall.
- Wong, P. S., & Root, J. C. (2003). Dynamic variations in affective priming. *Consciousness and Cognition*, 12(2), 147–168.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654–667.
- Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: an ERP study. *Brain and Language*, 101(3), 234–245.
- Wyer, R. S. (1980). The acquisition and use of social knowledge: Basic postulates and representative research. *Personality and Social Psychology Bulletin*, 6(4), 558–573.
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating

References

- statistical and social cues. *Neurocomputing*, 70(13-15), 2149–2165.
- Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, 14(2), 165–180.
- Yurovsky, D., Wu, R., Yu, C., Kirkham, N. Z., & Smith, L. B. (2011). Model selection for eye movements: assessing the role of attentional cues in infant learning. In E. J. Davelaar (Ed.), *Connectionist models of neurocognition and emergent behavior: From theory to applications-proceedings of the 12th neural computation and psychology workshop* (Vol. 20, p. 58–76). World Scientific.
- Zebrowitz, L. A., Franklin, R. G., Hillman, S., & Boc, H. (2013). Older and younger adults' first impressions from faces: Similar in agreement but different in positivity. *Psychology and Aging*, 28(1), 202–212.
- Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 2593–2598). Austin, TX: Cognitive Science Society.
- Zhang, Q., Guo, C., Lawson, A., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin*, 71(1-3), 316–323.
- Ziaei, M., Hippel, von, W., Henry, J. D., & Becker, S. I. (2015). Are age effects in positivity influenced by the valence of distractors? *PLoS ONE*, 10(9), e0137604.
- Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, 18(5), 229–234.

Curriculum Vitae

Name: Katja Münster
Address: Eichborndamm 63
13403 Berlin
Email: kmuenster@cit-ec.uni-bielefeld.de

Education

- 02.2013 – 2016: **PhD Candidate**
Department of Linguistics, Language and Cognition Group,
CITEC Graduate School, University of Bielefeld, Germany
- 10.2010 – 09.2012: **Master of Arts in Linguistics**
Focus on Communication, Language and Cognition
University of Bielefeld, Germany
Thesis title: “Does Seeing a Dynamic Emotional Face Prime
the Semantic Interpretation of an Utterance? A Comparison
between Older and Younger Adults“
Supervisor: Prof. Dr. Pia Knoeferle
- 10.2007 – 09.2010: **Bachelor of Arts in English and French Linguistic and
Literary Studies**
Focus on Linguistics
University of Osnabrück, Germany
Thesis title: “Functional Aspects of English Discourse Markers:
The case of *actually*”
Supervisor: Prof. Dr. Lena Heine
- 02.2007 – 07.2007: French language classes, Vermold, Germany
- 09.2006 – 02.2007: Bury Language School, Bury St. Edmunds, Great Britain
- 2003 – 2006: CJD Gymnasium Versmold, Germany
- 1997 – 2003: CJD Realschule Versmold, Germany
- 1993 – 1997: Grundschule Bockhorst, Germany