Investigating the influences of world knowledge and intelligence on the resolution of German personal and demonstrative pronouns

Bachelor Thesis
September 2007

Written by: Marlene Meyer (marlmeye@uni-osnabrueck.de)

Supervisors: Prof. Dr. Peter Bosch (pbosch@uni-osnabrueck.de)
Prof. Dr. Peter König (pkoenig@uni-osnabrueck.de)
Abstract

This study was conducted in order to clarify the distinction between the resolution of German personal and demonstrative pronouns with further interest in the influence of world knowledge and intelligence on the resolution process. By means of eye-tracking we measured fixations on Playmobil figures presented on a screen while simultaneous presenting short stories. Our findings support a difference in lexical antecedent preferences of the two pronoun types. While the personal pronoun has a preference for the grammatical subject of the proceeding sentence the demonstrative pronoun prefers the referent in object position. This effect of lexical antecedent preference was similar for both pronoun types. When a conflict between the linguistic preference and the available context information was present, a shift of fixations from the linguistically preferred to the contextually preferred referent occurred. Therefore, world knowledge prevails lexical antecedent bias of pronouns during the process of resolution. An influence of intelligence on the velocity of the integration process could not be affirmed.
Contents

1. Introduction 4
   1.1. German Personal and Demonstrative Pronouns 5
   1.2. World Knowledge and Intelligence 8
   1.3. Eye-tracking 9
   1.4. Hypotheses 10

2. Methods 11
   2.1. Visual Material 11
   2.2. Auditory Material 12
   2.3. Combination/Randomization/Balancing 14
   2.4. Intelligence Test 15
   2.5. Participants 15
   2.6. Apparatus 15
   2.7. Procedure 16
   2.8. Regions of Interest 16

3. Results 17
   3.1. Stimulus Validity 17
   3.2. Variance over Participants 18
   3.3. Temporal Processing of the Story 19
      3.3.1. Timecourse during the first 9000ms 21
      3.3.2. Timecourse during the last 5000ms split up in Conditions 25
      3.3.3. Order of Image Presentation 51
   3.4. Answers to Control Questions 53
   3.5. Differences between Demonstrative and Personal Pronouns 54
   3.6. Intelligence 55

4. Discussion 58

5. Conclusion 63

6. Acknowledgments 64

References 64
7. Appendix

A. Visual Stimuli

B. Auditory stimuli

C. Consent Sheet

D. Subject Questionnaire

E. Feedback Sheet

F. Affirmation

G. Assignment of Tasks
1. Introduction

The occurrence of pronouns in apparently simple everyday language does not cause comprehension problems for us. But how do we know the meaning of a pronoun? There seems to be a complex comprehension mechanism which enables us to pick the correct expression the pronoun is referring to which allows us to understand it’s meaning. Pronouns are linguistically defined as pro-forms substituting nouns or noun phrases. In order to avoid repetition of explicit names pronouns are used to refer back to an already introduced concept in the discourse [Almor et al., 2007]. As a form of anaphoric expressions, they are therefore coreferential with the expression they replace which is also called antecedent. For understanding a sentence containing a pronoun it is necessary to find out the correct antecedent. The task of finding this element which determines the pronoun’s interpretation is called pronoun resolution. In fact most of the time we are not even aware of this process. The effortlessness of understanding a sentence with such a function word can be shown by considering example (1). We do not need to pay special attention in order to know the pronoun’s interpretation within the example. It seems we automatically understand that ”He” is referring to the policeman. However, the mechanisms needed to resolve the pronoun’s reference do not seem to be trivial but reflect how advanced language comprehension mechanisms are. The ongoing research on the resolution of pronouns within the linguistic field arises from the question about how the mechanisms work which enable an effortless and partly unconscious interpretation of a relation between pronouns and their antecedents.

(1): A policeman watches the scene. He disagrees with the actions.

Nicol and Swinney (1999) propose that the resolution process consists of three steps, starting with the denomination of the pronoun triggering the search for an antecedent. In a second step all potential antecedents are taken into account until in the end exactly one candidate of this set is picked up whereas all others are eliminated. While we are not aware of the first two steps only the remaining choice is perceived consciously [Nicol and Swinney, 2002].

Syntactic as well as semantic theories deal with the linguistic constraints limiting the set of candidates considered as antecedents [Bosch et al., 2007]. A well-known syntactic approach can be found within the Binding Theory by Chomsky [Chomsky, 1981]. Within this theory Chomsky differentiates between three types of noun phrases: Anaphors like
"himself", pronominals like "he" and R-expressions like "the policeman". Thereon a linguistic differentiation between bound and referential pronouns is established.

Besides these syntactic restrictions of possible antecedents it seems to rely on semantic processes to actually choose a pronoun’s interpretation [Caplan and Hildebrandt, 1988]. Semantic approaches to this topic are the Discourse Representation Theory [Kamp, 1981] and the Centering Theory [Grosz et al., 1995]. In contrast to the assumptions of the Binding Theory, Kamp does not distinguish between bound and referential pronouns in his Discourse Representation Theory (DRT). Moreover in DRT the resolution of pronouns is considered across sentence boundaries which allows taking into account the unfolding discourse. As pronouns are supposed not to have an interpretation on their own it is necessary to integrate every sentence stepwise into a Discourse Representation Structure. Only thereby it is possible to determine the correct antecedent([Kamp, 1981]. Like in Kamp’s DRT the main aim of the Centering Theory [Grosz et al., 1995] is to establish a relation between the different discourse referents. The Centering Theory, however, is additionally concerned with explaining why some referents are more salient or important than others. In this context saliency is regarded as an explanation of why pronouns substitute certain discourse referents while others are mentioned explicitly.

In addition to these syntactic and semantic theories pragmatic information seem to play a crucial role in pronoun resolution [Bosch et al., 2007]. We are going into detail with the influence of world knowledge as belonging to pragmatics at a later point. Summing up, these theories are all concerned with pronoun resolution. But what about different types of pronouns? Do they all follow the same resolution process or does the difference in resolution build the basis for their discrimination?

1.1. German Personal and Demonstrative Pronouns

In German there is an interesting case of two sets of pronouns which at first glance seem to have the same function. These two types are the so-called personal pronouns and demonstrative pronouns. While personal pronouns consist of "er, sie, es", the set of demonstrative pronouns is homophonous with the definite determiner in German, namely "der, die, das". In former times both pronoun forms were considered to have nearly the same distribution and referent substitution function. According to this the noun phrase "a policeman" in example (2) can be equally well substituted with the personal pronoun ("Er") and the demonstrative pronoun ("Der"). The only pragmatic difference that was
established is the use of demonstrative pronouns for expressing devaluation or reinforced distance to the referred person [Duden, 1984].

(2): Ein Polizist beobachtet die Szene. Er/Der ist mit den Geschehnissen nicht einverstanden. (A policeman watches the scene. He personal/demonstrative disagrees with the actions.)

However, the fact that there are two different sets hints to a somehow established distinction. The question arises whether there exist discrimination factors which guide the intuitive usage and resolution of the pronoun types. How could these factors look like? In recent research potential factors are found on different levels. Recency as well as grammatical, semantic and pragmatic differences have been detected. Nevertheless, Bosch et al. could not support the theory of recency stating that personal pronouns refer to what is mentioned first in a sentence compared to demonstrative pronouns preferring secondly mentioned concepts [Bosch et al., 2007]. In contrast they found evidence for a grammatical distinction of personal and demonstrative pronouns [Bosch et al., 2007]. Personal pronouns have a preference for the noun phrase of the preceding sentence that is in subject position whereas a bias for the noun phrase as non-subject was detected for demonstrative pronouns. Accordingly in example (3) ”Er” would refer to the policeman while the demonstrative pronoun ”Der” substitutes the noun phrase ”dem Autofahrer” appearing in object position.

(3): Der Polizist redet mit dem Autofahrer. [Er/Der] regt sich fürchterlich auf. (The policeman talks to the car driver. He PERS/DEM is very angry.)

This is supported by similar observations for pronouns in Italian [Carminati, 2002] as well as for Dutch [Kaiser and Trueswell, 2007] and Finnish [Kaiser and Trueswell, 2005]. Since the occurrence of a demonstrative pronoun is rarer and thus more salient we assume the pronoun bias to have a stronger effect than for the personal pronoun. Besides this syntactic approach, the Complementarity Hypothesis [Bosch et al., 2003] takes a difference in semantics between the pronominal sets into account. Within the Complementarity Hypothesis personal pronouns are regarded as replacing an antecedent established during the discourse as main topic whereas the rheme is substituted by demonstrative pronouns. Consequently the policeman in example (3) who is the person acting would
be referred to by the personal pronoun ”Er”. On the other side the car driver has a passive function as listener and therefore, according to the theory, would be substituted by the demonstrative pronoun ”Der”.

Another factor with strong influence on the interpretation of both pronouns is world knowledge. Facts about the world are mainly gained by experience about how things in the world are organised and this background knowledge plays an important role for interpretation. In contrast to example (3) the pragmatic aspect seems to outweigh syntactic and semantic factors in pronoun resolution for example (4).

\[(4): \textit{Der Polizist redet mit dem Autofahrer. [Er/Der] ist mit der Verwarnung nicht einverstanden.} (The policeman talks to the car driver. He PERS/DEM disagrees with the parking ticket.\)]

Although there were detected different syntactic and semantic roles for personal and demonstrative sets it can be shown in (4) that no matter which pronoun type is used, both are perceived as referring to the car driver as the person who does not want to pay for a parking ticket.

If our knowledge about everyday life prevails over syntactic and semantic factors, exactly when during the time course of language comprehension does this influence prevail? If in (4) the personal pronoun is used do we then first consider the policeman to be referred to and shift our choice to the car driver only when world knowledge is integrated? Or do we integrate syntactic, semantic and pragmatic information only at the very end of a sentence so that we can use all necessary hints? By investigating the resolution of pronouns in situations in which linguistic and world knowledge information conflict we want to find out more about the mechanisms used and the hierarchy of factors contributing to the interpretation process. Especially the time course of the resolution is of interest. Hagoort et al. (2004) compared the time course of integration for semantic and pragmatic information. As a result they found that it takes about 400 milliseconds for word meaning as well as facts about the world to be processed within a sentence. The similarity in processing speed suggests a mechanism in which semantic and pragmatic integration are not separated [Hagoort et al., 2004].

In a next step we are concerned with the question whether this also holds for the intrinsic lexical preferences of pronouns and pragmatic information during a resolution process. As pronouns do not have a word meaning on their own they allow to investigate
the time course of linguistic and plausibility features used to assign an interpretation. Consequently we are interested in this integration process assuming that linguistic features are intrinsic whereas facts about the world have been learned [Bosch et al., 2007]. We suppose that the antecedent for a pronoun during an unfolding discourse firstly is chosen by linguistic information, namely the bias for subject respectively object position. Secondly this choice is checked against context information and what is known about the world in order to pick up the most plausible antecedent for a pronoun. In a conflict situation world knowledge outweighs the pronoun’s linguistic bias. Returning to example (4) this would imply the choice for the more plausible car driver as an antecedent for the personal as well as for demonstrative pronoun regardless of their grammatical preference.

1.2. World Knowledge and Intelligence

The discourse provides us with several information leaving a set of possible interpretations for a sentence or word. In order to pick up the most plausible interpretation we need to compare these information with our background knowledge so that other interpretations can be ruled out. In the first sentence of (4) a policeman and a car driver are introduced into the discourse. Because of our knowledge about the world we know the function of policemen who ensure that all rules are followed correctly and we are familiar with the role of a car driver on the street. The second sentence of (4) deals with someone not willing to accept a parking ticket. This implies that the person that disagrees with a punishment for a broken rule cannot be the person declaring this punishment which would be the policeman. Therefore we come to the conclusion that the referent we are looking for must be the car driver.

The process of drawing this conclusion between information about the given context and our background knowledge about the organisation in the world seems to be determined by intelligence. According to Süss intelligence is a necessary presupposition for the acquisition and utilisation of knowledge [Süss, 2007]. Moreover Horn and Cattell (1966) differentiate between two types of intelligence: the fluid and the crystallized intelligence [Horn and Cattell, 1966]. Fluid intelligence deals with the recognition and application of rules and implies solving a task in a deductive as well as inductive way. Compared to this, crystallized intelligence is defined as ability to solve problems by using the knowledge we already have [Süss, 2007]. Hence, the process of applying world knowledge to
discourse comprehension (by crystallized intelligence) enables us to resolve a pronoun’s interpretation within this discourse. The main question we are faced with is whether or not intelligence has a strong influence on pronoun resolution as a special aspect of language comprehension. Are people with higher intelligence able to integrate world knowledge more quickly with the information presented by a discourse and therefore decode a pronoun’s antecedent faster? Do differences in intelligence have an impact on what happens during a situation in which linguistic features conflict context information? As we predict a shift of focus from the linguistically to the contextually preferred referent in conflict situations and furthermore intelligence is said to be necessary for applying background knowledge we consequently assume that people with higher intelligence integrate world knowledge quicker and therefore show the predicted shift earlier than less intelligent people. Another aspect of interest is whether there is an influence of intelligence on which antecedent is chosen in the end: The lexically preferred or the pragmatically preferred one? By means of an intelligence test we want to investigate in the influence of intelligence on pronoun resolution. Additionally, in order to investigate the addressed questions, we have chosen eye-tracking as a measure.

1.3. Eye-tracking

We use eye movements in order to maintain the area of our interest in the middle of the fovea. The fovea is the spot in the eye with the highest resolution and originating here visual information are analyzed most detailed. Saccades are eye movements which we execute approximately every 250 milliseconds. The time in which no saccade is made but a focus is sustained is called fixation. It was detected that the direction of eye saccades and the direction of attention seem to correlate. Moreover Cooper (1974) established the measurement of eye movements for investigating the comprehension of spoken language. By providing a visual scene with several referents Cooper found out that people tend to direct their eye movements to those referents mentioned during an auditory presented discourse. This psycholinguistic procedure is known as the visual world paradigm which Tanenhaus et al. (1996) elaborated. The advantage for our study is the real-time measurement allowing statements about the time course of language comprehension. In addition no secondary task like in other psycholinguistic tools (as for example completion tasks) is necessary. Thus, there is no interruption of the online process of spoken language comprehension. Especially for studying the time course for pronoun resolution, eye-
tracking seems to be adequate as it allows online detecting of the participant’s choice for an antecedent. In particular the studies by [Karabanov, 2006] and [Jessen, 2006] support investigating the comprehension of pronominal expressions by means of eye-tracking. By comparing the fixations on the referents during the unfolding discourse presented auditory we can draw conclusions on the resolution of German personal and demonstrative pronouns.

Hence, for the purposes of our study we decided to record eye movements of participants who hear short stories and who are simultaneously presented visual scenes on a screen. These visual scenes contain Playmobil figures and objects in order to provide generally valid stimuli.

1.4. Hypotheses

In order to put it into a nutshell our hypotheses are:

Concerning the difference between German personal and demonstrative pronouns
1. The demonstrative pronoun "Der" will elicit higher fixation probabilities on the referent in object position than on the referent in subject position shortly after denomination of the respective pronoun.
2. The personal pronoun "Er" will elicit higher fixation probabilities on the referent in subject position than on the referent in object position shortly after denomination of the respective pronoun.
3. The difference between the fixation probabilities for the preferred referent in contrast to the non-preferred referent will be bigger for the demonstrative pronoun than for the personal pronoun.

Concerning the time course of integrating lexical preference of the pronoun and contextual information
4. In conflict situations between the lexical preference and the preference based on world knowledge for an antecedent (Conditions 2 and 3) there will be a shift in fixation from the referent with lexical bias to the referent representing the most plausible alternative on the basis of world knowledge. (in Condition 2 there will be a shift from Referent R1 to Referent R2, in Condition 3 there will occur a shift from Referent R2 to Referent R1)
Concerning the influence of intelligence on the resolution of pronouns
5. The higher the intelligence the earlier occurs the shift mentioned above.

2. Methods

2.1. Visual Material

The visual stimuli were made up of 12 photographs that showed pseudo-natural scenes of Playmobil figures and objects. We avoided drawings as visual stimuli so that the scenes were more realistic and by not using pictures of actual scenes and persons we also accounted for unintentional influences by mimic or other aspects of the pictured humans. All of the images were used with their native resolution but cropped to 2560x1600 pixels. Each scene included three human figures. Two of them were male (referents R1 and R2) and were mentioned in the auditory stimuli. In Figure 1 the two referents correspond to the policeman and the cardriver. We varied their position in a way that in some pictures R1 is positioned on the right and on other pictures on the left of R2. The third human figure was used as a (female) control which was not mentioned in the auditory stimuli. As it is known that people prefer looking at humans we wanted to make sure that people do not only look to the human figures instead of following the story that is presented auditory. We therefore placed for example the little girl in Figure 1. In the following this female control is referred to as referent R4. Furthermore there was one inanimate referent R3 that was also included in the auditory stimuli. Considering Figure 1 again, this would be the car as it comes up during the story as third referent apparent in the scene. Referents R1, R2 and R3 were arranged in a triangular position in order to exclude accidental fixations on a third referent while performing a saccade from one to the other referent. Additionally, all scenes included various objects and animals as distractors to make the scenes more interesting and natural (see Figure 1). During the experiment the 12 images were shown twice (with a distance of at least 3 images), filled with 36 images of a similar type (only the number, gender and/or visual salience of the referents varied) plus 8 images that showed a load of random Playmobil figures and objects. This was due to the fact that this experiment was conducted as a subproject together with two others (for further information about these experiments see [Bergmann, 2007] as well as [Bärnreuther, 2007] and [Kleemeyer, 2007]).
2.2. Auditory Material

The auditory stimuli consisted of 6 short stories for each of the 12 images. Each short story included three sentences. The first sentence introduced the scene (starting at 0ms after stimulus onset).

Example:

Eine Verwarnung.

The second sentence introduced the three referents R1, R2, and R3 with full noun phrases (starting at 3000ms after stimulus onset). As the third referent is always neutral in respect to its gender, it cannot function as an antecedent for the pronoun. We introduced this referent in a way that the main referents R1 and R2 of the second sentence do not immediately precede the pronoun denomination of the third sentence.

Example:

Der Polizist (R1) redet gerade mit dem Autofahrer (R2) über das falsch geparkte Auto (R3).
While the first two sentences have been the same in each of the 6 short stories, the third sentence always represented a different Condition (starting at 9000ms after stimulus onset). Condition 1 and Condition 2 had a contextual preference for referent R2. Condition 1 had a linguistic preference for referent R2 not resulting in a conflict with the following context information, while Condition 2 had a linguistic preference for referent R1 and therefore was indisputable. Condition 3 and Condition 4 had a contextual preference for referent R1. Condition 3 had a linguistic preference for referent R2, thus resulting in a conflict, while Condition 4 had a linguistic preference for referent R1 and therefore did not result in any conflict. Condition 5 and Condition 6 were neutral in regard to the contextual preference. Condition 5, however, had a linguistic preference for referent R2 while Condition 6 had a linguistic preference for referent R1.

Examples:

Condition 1: Der ist mit der Verwarnung nicht einverstanden und regt sich fürchterlich auf.

Condition 2: Er ist mit der Verwarnung nicht einverstanden und regt sich fürchterlich auf.

Condition 3: Der will 20 Euro für Falschparken kassieren und regt sich fürchterlich auf.

Condition 4: Er will 20 Euro für Falschparken kassieren und regt sich fürchterlich auf.

Condition 5: Der ist ziemlich unhöflich, schreit ganz laut herum und regt sich fürchterlich auf.

Condition 6: Er ist ziemlich unhöflich, schreit ganz laut herum und regt sich fürchterlich auf.

The fourth sentence was a control question to check if the subjects did understand the meaning of the story and was the same for all conditions.
An overview of the whole setup of a story including all parts and all conditions can be seen in Figure 2.

Since we measured fixation frequencies on defined regions of interest (the referents) we had to make sure that there was a time distance of at least 500ms between the offset of one referent and the onset of the next one, because fixations usually last about 300ms. If we did not guarantee this minimal time distance the effects on different words could not be distinguished, as the subject needs time to allocate the fixation to the next word.

For the 36 filler images there have also been different conditions (either two or four) and therefore also different short stories (compare [Bergmann, 2007], [Bärnreuther, 2007] and [Kleemeyer, 2007]). None of the auditory filler contained a control question in the end. Only the 8 images showing a load of random Playmobil figures and objects did not have any auditory stimuli.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Introduction of the scene</th>
<th>Introduction of R1, R2, R3</th>
<th>Process - linguistic preference</th>
<th>Contextual Preference</th>
<th>Neutral end - basis for question</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Eine Verwarnung</td>
<td>Der Polizist (R1) redet gerade mit dem Autofahrer (R2) über das falsch geparkte Auto (R3)</td>
<td>Der ist mit der Verwarnung nicht einverstanden</td>
<td>und regt sich fürchterlich auf</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Br ist mit der Verwarnung nicht einverstanden</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>Der will 20 Euro für falschparkten karieren</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>Br will 20 Euro für falschparkten karieren</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>Der ist ziemlich unglücklich, schreit ganz laut heraus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>Br ist ziemlich unglücklich, schreit ganz laut heraus</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Table of one story with all conditions

2.3. Combination/Randomization/Balancing

Each subject saw every image twice but each time with a different condition. Therefore, every subject could listen to 24 stories with changing condition which were balanced over three subjects. The minimal distance between the first and the second occurrence of the same image was three to avoid memory effects. We made sure that conditions with a conflict (Condition 2 and Condition 3) only appeared when the corresponding image was shown for the second time. This was to ensure that the subjects did not think we only gave them ”strange” stories. Since some of the filler images were quite similar
to the actual stimuli we also put constraints on the minimal distance between similar images. This minimal distance was again three. Eventually, the 12 actual images have been balanced equally in between the filler pictures.

2.4. Intelligence Test

In order to investigate the influence of intelligence on the resolution of personal and demonstrative pronouns we used the MWT-B (Mehrfach-Wortschatz-Test) [Lehrl, 1977] which is a test measuring general intelligence with a special remark on crystallized intelligence. As crystallized intelligence is also dependent on cultural background we took into account only participants with similar background namely German native speakers. This test is only meant as a measure to compare the general cognitive abilities among the subjects. We did not do any comparison with populations outside the subject pool or with what is know as the Intelligence Quotient.

2.5. Participants

60 native speakers of German (32 male; 28 female) participated in the experiment. In order to achieve a very heterogeneous field people of different educational status agreed to participate in our experiment. All of them were between 18 and 50 years old (mean age: 24.8) and had normal or corrected-to-normal vision (with one subject being colourblind). They participated voluntarily and were either paid 7.50 Euro or got 1.5 ”Versuchspersonenstunden”.

2.6. Apparatus

The images were photographed with a Sony Cyber-shot DSC-V1 5.0 Mega Pixels (Sony Corp., Japan) and were presented on a 30” TFT screen (Apple Cinema HD Display, Apple Inc., USA) with a native resolution of 2560x1600 pixels. The auditory stimuli were recorded with a Shure SM-58 (Shure Incorporated, USA) through a Tascam US-144 Audio Interface (Tascam, USA). They were edited in Audacity (www.audacity.sourceforge.org) and were presented using a Logitech 2.1 system (Z3 Style 2.1, Logitech, Switzerland) with two front loudspeakers being positioned left and right of the screen and the bass loudspeaker standing on the floor under the screen. The eye movements were measured
with an Eyelink II eye-tracking system (SR Research, USA) that had a frequency of 250Hz.

2.7. Procedure

The conduction of the experiment took place in a darkened room. Before being tracked, the participant was given a consent sheet and a subject questionnaire which he or she had to fill in and to sign. An instruction about the course of the experiment followed, but the subject was left completely naive about the aim of the study. The participant then was placed sitting in front of the screen with a spatial distance of 80cm to the eyes. A chin rest could be used if preferred by the subject but was optional. The eye-tracker was then mounted and calibrated using the 13-dot-grid procedure. Only calibration values with a mean error of less than 0.3° or less than 0.4° if the maximum was lower than 1.0° in the validation were accepted. The subject was instructed to watch the scenes attentively and to listen carefully to the short stories. The stimulus onset and offset for the images and the short stories was synchronized with the images lasting two seconds longer than the audio. Before every trial the subject was instructed to fixate a fixation point in the middle of the screen which was used for drift correction. Each trial was manually triggered by the experimenter if the drift correction value was valid. Otherwise a recalibration was done. After the presentation of 36 images there was the opportunity to make a break. Before continuing with the experiment a recalibration was forced.

After being tracked the subject was instructed to fill in the MWT-B test. The subjects were not told any results of this test. At the end the participant was informed about the aim of the study and handed out a feedback sheet if wanted.

2.8. Regions of Interest

For the data analysis we used the build-in function ”roipoly” of Matlab in order to define regions of interest (ROI) so that distractors of the scene were excluded from analysis. Firstly, we defined these regions within a picture as narrow as possible along the referents R1, R2, R3 and R4 including only fixations directly on these Playmobil figures. In a second step we considered different factors to enlarge the ROIs in order to make sure that fixations close by which are supposed to be on the referents also belong to the ROI. On the other hand the factor did not have to be too large. Consequently we scaled the regions of interest with a factor of 2.0. We used these scaled ROIs during further
analysis of the fixations on the referents. An example of the resulting regions is shown in Figure 3.

![Figure 3: Regions of Interest in Image 1](image)

### 3. Results

#### 3.1. Stimulus Validity

Since we used 12 different images during the experiment it was necessary to make sure that the material was valid. In other words we had to check whether the fixations summed over participants and conditions were equally distributed over images or whether the visual material contained outliers. Therefore we calculated the fixation probabilities for each image on the different referents during the first 3000 milliseconds because no referent was introduced since then. For each 33 millisecond time slot we accumulated the fixations on the respective referent in order to get the percentage of fixations on one referent over the timecourse of 3000ms. We used this box-car method during the whole analysis. Figure 4 shows that the distribution of the fixations on the first two referents seems more equally distributed than on the third referent. However, this is of no further interest for our analysis as it is not focus of the study and probably due to Playmobil objects differing in size. As referent R1 and R2 show nearly the same fixation probability (mean R1: 21.72% std 6.42%; mean R2: 18.37% std 4.28%) on the different
images we considered all images to be valid. Consequently we included all of them in our further analysis.

Figure 4: Distribution of fixations on regions of interest during the first 3000ms on a per-image basis (summed over all subjects and all conditions)

### 3.2. Variance over Participants

Furthermore, we wanted to find out whether there is a high variance between participants. In order to check for a normal distribution of the fixation probabilities on different referents we again took into account the first 3000 milliseconds of the scene introduction. Summing over all images and conditions shows that the distribution of fixations differed only slightly among subjects (see Figure 5). For example the minimal fixation probability during the first 3000 milliseconds for the first referent was 9.751% while it was 8.0088% for R2. This shows that the relevant characters were fixated by all participants during the first 3000 milliseconds at least to a certain extent. The distribution of
fixation probabilities of all subjects underlines also that the participants looked around in the scene which is a prerequisite for the entire design of our experiment.

Figure 5: Distribution of fixations on regions of interest during the first 3000ms on a per-subject basis (summed over all images and all conditions)

Since we also wanted to make sure that fixation probabilities for R1 and R2 were normally distributed over subjects we tested it with a Lillifors-Test confirming for both a normal distribution (p<0.05). Hence we could not identify outliers so that all participants were included into the following analysis.

3.3. Temporal Processing of the Story

By looking at the timecourse of fixations we tried to clarify our hypotheses stepwise. Figure 6 representatively depicts the fixation probabilities for the different referents and beyond region of interest over the whole timecourse summed over participants and images for the first condition. The numbers attached to the arrows indicate eight observations found. These aspects illustrate different stages of processing. While the first five effects
are independent of the presented condition we evaluated them further on over all conditions. These effects are concerned with the presupposition for our main hypotheses. By contrast observations 6 to 8 are dependent on the condition and represent the findings directly related to our hypotheses discussed.

Figure 6: Discourse of fixation frequencies for all referents in the first condition summed over all images and all subjects. The arrows entailing a circled number indicate observations visible during the timecourse.
3.3.1. Timecourse during the first 9000ms

Since the first five effects are independent of the condition they can be found even after summing over conditions (see Figure 7). The first observation is concerned with the decrease of fixations on regions which are not of interest for the story in favour of an increase of fixations on the relevant referents. Within the first 400 milliseconds the fixation probability decreases constantly from 68.42% to 38.94% for Beyond Region of Interest whereas the fixation probability increases for both referents R1 and R2. While the percentage of R1 rises from 0% to 34.67% the increase of fixations on R2 are slightly lower, namely from 0% to 13.96% (see Figure 11). It is the case that fixations on the first referent rise earlier and reaches the maximum at an earlier point in time than referent R2. This is due to the positioning of the Playmobil figures to the fixation point which can be seen in [Krause, 2007].

![Timecourse for the first 9000 ms](image)

Figure 7: Discourse of fixation frequencies for all referents in the first 9000ms summed over all images, all subjects and all conditions. The arrows entailing a circled number indicate observations visible during the timecourse.
Secondly, in the proceeding timecourse (within seconds 2 to 3) there can be seen a period of constant fixations. During this time period there is no auditory stimulus present and therefore no more information added to the introduction sentence.

Figure 8: Slopes for the discourse of fixation frequencies for each referent in the steady state between 2000ms and 3000ms after stimulus onset.

In order to estimate a slope value within a certain time span we used the build-in Matlab curve fitting tool (cftool) which enables to approximate lines for given data in a manually defined range. In the following we are using these fitted lines to indicate the increase respectively decrease of fixation probability for a given referent. Figure 8 shows the fixations in percent during seconds 2 and 3 for the different referents and illustrates that the gradient for all referents is lower than 1. This is considered to be a steady state during which no dramatic changes were recorded.
Thirdly, an increase of fixation probability for referent R1 can be observed shortly after the denomination of this referent at 3000ms (see Figure 9). The intersection between the slope of the steady state and the increase of fixation probability can be found at 3465ms after stimulus onset. Starting at this point in time the slope approximated for the fixation curve is 23.52. The curve further resulted in a maximum of 35.25% of all fixations.

![Figure 9: Slope of increase of fixation frequencies for referent R1 and intersection with slope of steady state](image)

The intersection of two fitted lines to the curve representing the fixation probability for R2 between seconds 2 to 6 is found at 4235ms after stimulus onset (see Figure 10). As referent R2 is denominated during the auditory stimuli at 4800ms after stimulus onset this shows an earlier onset of increase than for referent R1 (denomination: 3000ms; starting point of increase: 3465ms). Furthermore the fixation probability rises with a slope of 12.69 until reaching the maximum of 31.37%.

The last effect evaluated over all conditions is the fixation frequency of referent R4 during 300ms and 9000ms of the timecourse. (We are excluding the first 300ms due to the fixation on the fixation point before stimulus onset and the fact that a fixation can last up to 300ms.) Compared to the other referents R1, R2, R3 and Beyond Region of Interest, referent R4 has the lowest mean fixation frequency, namely 5.35% (see Figure 11). Also the maximal value for fixation probability of R4 is the smallest (8.89%) as depicted in Figure 12. In contrast the maximal values of R1, R2 and R3 as well as Beyond Region
Figure 10: Slope of increase of fixation frequencies for referent R2 and intersection with slope of steady state of Interest range between 33.38% (for R2) and 68.42% (for Beyond Region of Interest) as can be seen in Figure 11. The minimal value of R4 during this time period does not reach 1% whereas the next smallest minimal value belongs to referent R3 and is bigger than 5% (see Figure 11).

<table>
<thead>
<tr>
<th>Distribution over the timecourse of 300ms (first saccade) and 9000ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referent 1</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>Min</td>
</tr>
</tbody>
</table>

Figure 11: Distribution over the timecourse from 300ms to 9000ms with maximum, minimum and average fixation probability for the different referents.
3.3.2. Timecourse during the last 5000ms split up in Conditions

The observations concerning the last 5000ms are different over conditions. Observation 6 is concerned with whether fixation data evince an effect triggered by pronoun denomination within about the first 1000ms after pronoun onset (at 9000ms). The next observation deals with whether in the ongoing timecourse a shift in fixation can be found. Finally, the eighth observation considers the remaining time period until offset of auditory stimuli at 14000ms (Further on the onset and offset is only used to refer to auditory stimuli. Exceptions are mentioned explicitly.). In the following the three aspects 6 - 8 (Figure 6) are described for each condition.

Condition 1 contains linguistic and contextual information both referring to the second referent. Figure 13 depicts the fixation probabilities for this condition differentiated by referents during the last 5000ms starting at the onset of the denomination of the demonstrative pronoun (at 9000ms). Observation 6 deals with the increase of fixation probability for referent R2 between 9000ms and 10000ms.

Figure 12: Distribution of mean, maximum and minimum of fixation probabilities for all referents as well as Beyond Region of Interest between 300ms and 9000ms after stimulus onset.
While the slope of the fixation curve for R2 is 0.83 before 9400ms (in the range of 8400ms to 9400ms) the fixation frequency increases between 9400ms and 10000ms with a slope of 27.57 (see Figure 14). The intersection of the fitted lines can be found at 9536ms depicted in Figure 14. The maximal fixation probability for referent R2 is 31.39% during seconds 9 to 10 whereas the maximal value for referent R1 during the same time does not reach 25% (22.67%).

In order to compare the fixation frequencies for referents R1 and R2 we used bootstrapping as statistical means. Hereby we tested for significant fixations on referent R1 over those on R2 and vice versa for all participants on a single-subject level as well as for all images (single-image level). This was done for each condition during the whole timecourse in steps of 33ms. We used a sampling time of 1000 and a significance level of $\alpha=0.5$ throughout the analysis for all conditions.
Consequently we are able to compare the number of participants (or images) having significant values for one or the other referent. As the results for the first 9000ms are not part of the main hypotheses of our study we are not going into detail concerning these values. Regarding Condition 1, observations like 3 and 4 (compare Figure 6) can also be seen in the results of bootstrapping. We found that fixations on referent R2 were significant over those of R1 for 29 participants between 9999ms and 10032ms (see Figure 15). The fixation probability of only 9 participants was significant for R1 in contrast to R2 for this time slot.

On a single-image basis similar results were obtained. While at the same time slot 4 images reach a significant level for fixations on R2 over R1 no image shows significance for R1 (see Figure 16).
Figure 15: Number of subjects with significant results for referent R1 or referent R2 during the whole timecourse for Condition 1 on a per-subject basis summed over all images.

Figure 16: Number of images with significant results for referent R1 or referent R2 during the whole timecourse for Condition 1 on a per-image basis summed over all subjects.
The growing difference between fixations on R1 and R2 in favour of the second referent supports the observation described above (see Figure 17). In the ongoing timecourse between 10 and 11 seconds there can be seen that distances of fixations on R1 and R2 first approach 0 and then magnify again (see Figure 17). While between 10000ms and 10500ms the biggest difference between R1 to R2 is -16.14 it maximizes to only -1.36 and extends between 10500ms to 11000 again to -16.52. This seventh observation can also be seen in milliseconds 10000 to 11000 for significance of R2 on single-subject level (see Figure 15) as well as on per-image basis (see Figure 16).

Figure 17: Distance between the fixation frequencies for referents R1 and R2 between 9000ms and 14000ms in Condition 1.

After this gradual change, observation 8 deals with the relatively constant course of fixations which is maintained during the last 3000ms (see Figure 13). While the fixation probability for referent R2 is relatively high with a peak of 40.19%, the peak of R1 illustrates the low fixation frequency on the first referent during the end of stimulus presentation (25.23%). Within this time span the difference between R1 and R2 does not come closer to zero than -8,12 (see Figure 17). Moreover observation 8 can be seen with respect to the significance test on per-subject respectively per-image basis. For this last period the number of participants significant results for R2 goes up to 30 (out of 60) at 11682ms and furthermore ranges between 16 to 30 in contrast to the results for R1 which were significant only for minimally 5 to maximally 13 subjects during this period.
Moreover findings on a single-image level support this observation as can be seen in Figure 16.

Figure 18: Fixations between 9000ms to 14000ms after stimulus onset summed over images and participants for Condition 2.

Condition 2 as a conflict situation contained the same contextual information as Condition 1 but the demonstrative pronoun was replaced by a personal pronoun. Therefore observation 6 shows an increase of fixation frequency for the opposite referent (R1) compared to Condition 1 during seconds 9 to 10 (see Figure 18).

The slope of fixation frequencies on R1 increases from 1.16 (between 8400ms and 9400ms) to 28.31 (in the range of 9500ms to 10000ms) which is illustrated in Figure 19. The intersection of both fitted lines at 9486ms occurs slightly earlier than in Condition 1 (9545ms). The rising fixation frequency for referent R1 peaks in 33.64% within seconds 9 to 10 whereas 18.55% is the maximal fixation frequency for R2 in the same time interval.
Figure 19: Fixations between 8000ms to 12000ms after stimulus onset summed over images and participants for Condition 2. Slope of fixation frequencies for referent R1 and intersection with the state before.

Bootstrapping on a single-subject basis for Condition 2 results in significant values for the fixation on referent R1 for 23 participants at 10098ms (see Figure 20). For the same slot R2 is only significantly fixated by 11 subjects. Figure 38 illustrates the increase of significant slots for R1 in contrast to the decreasing slots for R2 starting shortly after 9000ms.

Taking bootstrapping results for images into account reveals that between 10098ms and 10131ms 4 participants fixated significantly more on referent R1 than on referent R2 (see Figure 21). In the range of 9537ms to 10428ms no image results contain significantly higher fixations on referent R2 than on R1.
Figure 20: Number of subjects with significant results for referent R1 or referent R2 during the whole timecourse for Condition 2 on a per-subject basis summed over all images.

Figure 21: Number of images with significant results for referent R1 or referent R2 during the whole timecourse for Condition 2 on a per-image basis summed over all subjects.
Observation 7 is expressed by the change of the course of fixation curves for R1 and R2 which occurs between 10500ms and 12000ms (see Figure 22). While fixation frequencies for referent R1 decrease during this time, referent R2 is fixated increasingly more often. Figure 22 illustrates the descent of the curve belonging to fixations on R1 (slope: -6.83) in contrast to the rising fixation frequency on referent R2 (slope: 19.83) in this interval.

Figure 22: Decrease of fixation frequency for referent R1 and increase for fixation frequency for referent R2 after the local maximum of fixation frequency for R1 in condition 2.

This change is also clearly visible in Figure 23 which shows the difference between the fixations on the two referents. First the difference is higher than zero indicating higher fixation frequencies for referent R1 and shifts then to negative values representing more fixations on referent R2 after 11204ms.
Figure 23: Distance between the fixation frequencies for referents R1 and R2 between 9000ms and 14000ms in Condition 2.

Looking at Figures 20 and 21 confirms the shift of fixation. The number of participants who fixated R1 significantly over R2 decreases from 23 to 5 (from 10098ms to 11517ms) while the number for R2 increases from 11 to 18 (see Figure 20). The same holds for the tendency concerning significant slots on single-image basis (R1: 4 to 0; R2: 0 to 5 in the range of 10098ms to 11946ms).

During the last 2000ms the fixation frequency remains higher for referent R2 than for R1 (see Figure 23). However, an approximation of both curves can be seen (compare Figures 18 and 23). An additional observation in this condition (as well as in the following conflict condition) is the fact that for more than 2000ms in the end the fixation frequencies of referent R2 (respectively R1 for condition three) are constantly higher than for Beyond Region of Interest (see Figure 18) which is not the case for non-ambiguous conditions.

Like Condition 2 also Condition 3 creates a conflict between lexical preference and contextual antecedent preference. After the denomination of the demonstrative pronoun which has a preference for the second referent the further discourse contains a preference for the first referent. Remarkably, observations 6 and 7 differ from those made for all other conditions. Fixation frequencies depicted in Figure 24 show a different pattern than in other conditions because already before pronoun denomination there is an increase of fixation frequencies for referent R2.
In the range of 8500ms to 9500ms there is the highest ascent compared to slopes present in other conditions for the referent biased by the pronoun (9.09). Figure 25 illustrates furthermore the even lower increase of fixation frequency for the second referent between 9500ms and 10000ms (3.45). The peaks observed for other conditions between 9900ms and 10200ms show in this condition an even higher result for the non-biased referent R1 (30.57%) than for referent R2 (27.48%).

Figure 24, however, makes clear that although fixations on referent R2 do not increase essentially after pronoun denomination (which is the case for the other conditions) but fixations on referent R1 decrease shortly after the pronoun is named.
Figure 25: Slopes of increases of fixation frequency for referent R2 in Condition 3.

Although there is no increase observable, testing for significance reveals a predominance of referent R2 (see Figure 26) which confirms higher fixation frequency for referent R2 during this time span. Within the first second after pronoun presentation in the auditory stimuli the number of subjects having significantly fixated referent R2 over R1 is constantly higher than the amount of participants with significant results for the other referent. The number of participants in favour of R2 ranges from 16 to 23 whereas maximally 14 and minimally 9 participants fixated referent R1 significantly more often (between 9000ms and 10000ms).

Bootstrapping on a per-image basis shows a similar pattern in Figure 27. Only one slot contains more significant images for R1 than for R2 (within 9471ms and 9504ms). All other slots within seconds 9 to 10 are at least significant for equally many images for both referents but more often in favour of R2.

The lowering values during this period depicted in Figure 28 confirm the observation visible on Figure 24 implying the increase in difference of fixation frequency in favour of R2.
Observation 7 deals further with the shift in fixation frequency for the two referents during the unfolding discourse. As can be seen in the increase of distance values after 10000ms which are exceeding zero (shown in Figure 28) the fixation frequencies on referent R1 rise such that they become higher than those for the second referent.
Figure 28: Distance between the fixation frequencies for referents R1 and R2 between 9000ms and 14000ms in Condition 3.

Figure 29 illustrates this observation by fixation frequencies summed over participants and images. Lines fitted on these curves make clear that there is a continuing ascent of fixations on referent R1 (9.27) while fixations on referent R2 decrease over time (-2.88).

This shift in favour of referent R1 can be observed also by looking at bootstrapping results. For 20 out of 60 participants referent R1 is significant over R2 with an increasing number up to 23 in the interval of 10098ms to 10329ms (see Figure 26). In the same time span 14 decreasing to 12 subjects fixated significantly the second referent. This holds true also for significance tests on per-image basis (compare Figure 28).

During the last 3000ms of the discourse participants fixated referent R1 constantly to a higher percentage than referent R2 (see Figure 24). This eighth observation can be seen equally good on other levels of analysis as bootstrapping results and distances between the referents.

The maximal number of participants with significant results for referent R2 of this last period is 13 out of 60 but almost the half of all participants (28) maximally fixated simultaneously more the first referent R1 than R2 (compare Figure 26). Furthermore this effect can be observed in Figure 29 which illustrates that referent R1 has a higher percentage of fixations than has referent R2 throughout the last 3000ms. During this time fixation frequency for the first referent is at least 5% higher than the percentage of fixation on the second referent.
Other than Condition 3 the fourth condition is non-ambiguous in the sense that the pronoun’s lexical bias as well as the context refer to referent R1. Consequently during the last 5000ms the fixation frequency on referent R1 was higher than for referent R2 (see Figure 30). In contrast to the maximal fixation frequency of referent R2 (17.86%) the peak of fixations on referent R1 in the interval of 9000ms to 10100ms exceeds 25% (26.67%).

Figure 31 illustrates observation 6 which is the increase of fixation on referent R1 (with a slope of 20.64) between 9700ms and 10100ms that follows a relatively low gradient before 9700ms (0.86). The dotted line indicates the still relevant but lower increase of fixation frequency on R1 after 10100ms (11.46). Compared to Conditions 1 and 2 the first increase after pronoun denomination starts relatively later during the timecourse, namely at 9749ms (condition 1: 9536; condition 2: 9486).
Figure 30: Fixations between 9000ms to 14000ms after stimulus onset summed over images and participants for Condition 4.

Figure 31: Fixations between 8000ms to 12000ms after stimulus onset summed over images and participants for Condition 4. Slope of fixation frequencies for referent R1 and intersection with the state before.
Testing for significant slots revealed on single-subject level that at least 16 but maximally 21 subjects fixated referent R1 more than R2 between seconds 9 to 10 (see Figure 32). Fixations on referent R2, however, were only significant for maximally 14 participants within this time span.

Figure 33 illustrates that also on a single-image basis significant slots for referent R1 during 9000ms to 10000ms are found for more (or at least equally many) images than for the second referent.

Figure 32: Number of subjects with significant results for referent R1 or referent R2 during the whole timecourse for Condition 1 on a per-subject basis summed over all images.

Best depicted are observations 6 (the increase of fixations on R1) and 7 (decrease followed by another increase of fixations on R2) in Figure 34 containing the difference between fixation frequencies on R1 and R2 for the single time slots. After the steep increase of the difference between the referents in favour of referent R1 there is a decline of distance between seconds 11 to 12 (Figure 34). Comparing this to the results of bootstrapping on the single-subject level (Figure 32) underlines observation 7. The amount of participants with significant slots for R1 decreases from 28 (11000ms) to 19 (11319ms) and rises afterwards again to 32 (12045ms).

During the last 2000ms the fixation frequency of R1 decreases slightly whereas fixations on R2 increase (see Figure 30). Nevertheless, fixations on R1 remain higher than fixation frequency for R2 which can be seen by regarding distance values (Figure 34) as well as bootstrapping results both on single-subject ((Figure 32) and single-image
Figure 33: Number of images with significant results for referent R1 or referent R2 during the whole timecourse for Condition 4 on a per-image basis summed over all subjects.

Figure 34: Distance between the fixation frequencies for referents R1 and R2 between 9000ms and 14000ms in Condition 4.

...basis (Figure 33). Distance values in favour of R1 do not undercut 6 (compare Figure 34). Moreover at least 17 participants significantly fixated R1 over R2 within the last 2 seconds while maximally 16 subjects evinced significant slots for R2. Finally the maximally amount of images with significant slots for R2 during the last period is the lowest value for significant slots for R1, namely 1. In addition, more than half of all
images (7 images) were maximally summed up in significant slots for this interval for referent R1 (see Figure 33).

Conditions 5 and 6 were neutral conditions which implies that there is no clear contextual preference for one of the referents but both could be meant by the unfolding discourse after the denomination of the pronoun (with onset at 9000ms). In contrast to Condition 6 the demonstrative pronoun was presented in the auditory stimuli of the fifth condition. Observation 6 pinpoints the increase of fixation frequency for referent R2 visible around seconds 9 to 10 (see Figure). The maximal value of fixations on referent R2 in the range of 9000ms to 10200ms is 29.87% whereas the peak of fixations for referent R1 is still below 20% (19.05%).

![Timecourse of the last 5000ms for condition 5 summed over all images and all participants](image)

Figure 35: Fixations between 9000ms to 14000ms after stimulus onset summed over images and participants for Condition 5.

The increase of fixations on R2 in this time interval can also be described by means of the slope of a fitted line to the fixation curve pertaining to this referent. Before 9500ms the slope for fixation frequency on R2 is 1.76 and therefore relatively low in contrast to
the slope starting at 9525ms which goes up to an increase of 26.93 until second 10 after stimulus onset (see Figure 36).

Figure 36: Fixations between 8000ms to 12000ms after stimulus onset summed over images and participants for Condition 5. Slope of fixation frequencies for referent R2 and intersection with the state before.

Not only on the basis of comparing slopes but also via taking into account results for testing significance on single-subject and single-image level substantiate observation 6. Figure 37 illustrates the increase in the respective time window for the amount of participants with significant slots for R2 over R1. Hereby the number of subjects rises from 7 to 23 whereas during the same timecourse the amount of participants falls from 22 to 14 for the first referent (see Figure 37).

On per-image basis the same observation of increase in favour of referent R2 can be made, although it seems to occur slightly later. Nevertheless, the ascent can be seen in Figure 38 around 10000ms after stimulus onset.

Finally observation 6 can be supported by the analysis of difference between fixations made on R1 and those on R2. In Figure 39 the steep decrease within 9000ms and 10000ms indicates that referent R2 was fixated to an increasing extent.
Figure 37: Number of subjects with significant results for referent R1 or referent R2 during the whole timecourse for Condition 5 on a per-subject basis summed over all images.

Figure 38: Number of images with significant results for referent R1 or referent R2 during the whole timecourse for Condition 5 on a per-image basis summed over all subjects.
Besides the sudden increase of fixation frequency for referent R2 and decrease of R1 Figure 39 illustrates also observation 7 which deals with the approximation and again parting away of fixation frequencies for both referents. This change occurs around second 11 as visible in Figure 39. Taking the number of participants into account which indicates significant results for fixations on referent R2 over R1 (see Figure 37) it can be found that at 10989ms only one participant fixated significantly more R2 than R1 (R1: 16; R2: 17). At the exact same timeslot there is only one image for each of the referents showing a significant result for one respectively the other (see Figure 38).

Results for the further course in time support observation 8 stating that referent R2 remains center of fixation compared to a low fixation frequency for referent R1 until only during the last 300ms also referent R1 was fixated slightly more than R2 (see Figure 35). Fixation probabilities for referent R2 get even higher after 11000ms than between 9000ms and 10200ms. While frequencies go up to 38.86% for referent R2 in this last period the peak belonging to fixations on referent R1 also rises but remains below 30% (29.65%). Figure 37 additionally demonstrates that besides in milliseconds 12540 to 12606 the number of participants with significant slots for R2 is constantly higher than for significant results in favour of R1 on single-subject basis.
In the end there is Condition 6 in which the personal pronoun is presented auditory at 9000ms after stimulus onset. Further the unfolding discourse in this condition does not contain any contextual bias to one or the other referent but is supposed to be neutral. Whereas fixation probabilities for referent R1 and R2 differ only slightly around the first 600ms after onset of pronoun denomination there is a clear dominance of fixations on referent R1 up to the end of stimulus presentation after about 9600ms (see Figures 40 and 41).

Figure 40: Fixations between 9000ms to 14000ms after stimulus onset summed over images and participants for Condition 6.

The increase of fixation probability on referent R1 between 9000ms and 10000ms (observation 6) seems to start at 9516ms which can be seen in the intersection of slopes fitted to the curve of R1 in Figure 41. Before the ascent of fixation frequency on referent R1 there is a fall off for the fixations on R1 within 8400ms to 9500ms (-3.26) illustrated also in Figure 41. The slope observed after 9500ms reaches an increase of 19.07 resulting in a local maximum of 25% fixations on this referent (between 9900ms and 10200ms). The highest percentage of fixations in the same local interval for referent R2, however, goes only up to 15.93% (see Figure 40).
Comparing seconds 9.5 to 10.5 after pronoun onset with respect to significance of R1 over R2 on a single-subject basis reveals that up to 25 out of 60 participants show significant results on a 0.5 significance level for R1 (see Figure 42). In the same period the amount of participants with significant slots for referent R2 decreases from 19 to 11.

Figure 42 shows an even clearer preference for referent R1 as illustrated by the peak representing the amount of images around 10000ms that indicate significance of R1 and the simultaneous absence of significance on single-image basis for R2 over the first referent.

Consistently with the analysis of slope and significance for one referent effect 6 can also be observed in the course of difference between the fixation frequencies of referents R1 and R2 (see Figure 44). Figure 44 demonstrates that distances between the referents increase visibly during the first 1000ms after pronoun denomination.
Figure 42: Number of subjects with significant results for referent R1 or referent R2 during the whole timecourse for Condition 6 on a per-subject basis summed over all images.

Figure 43: Number of images with significant results for referent R1 or referent R2 during the whole timecourse for Condition 6 on a per-image basis summed over all subjects.

Further on in the timecourse between seconds 11 to 12 after stimulus onset observation 7 is ascribed to the sudden decrease followed by another steep incline of fixation probability for referent R1 (see Figure 44). The distance in this time interval between the two main referents appears to diminish down to 1.3 and to magnify again (up to
Furthermore, this observation can also be made on the results of bootstrapping on a single-subject level as well as single-image level. Figure 42 demonstrates the observation manifested in a temporary descent of significant results for the fixations on the first referent on per-subject basis (down to only 9 out of 60 participants showed significant results at 11781ms). The analysis on single-image basis reveals that between 11000ms and 12000ms there are several slots containing only 2 images to be significant whereas 1 became significant for referent R1 and simultaneously another shows significant results for R2 (e.g. within 11220ms and 11253ms, see Figure 42). Summed over participants and images the observation is finally supported by the percentage of fixations on referents R1 and R2 depicted in Figure 40 (minimum of 19.91% fixation frequency for R1 between seconds 11 and 12).

Also the last observation can be found on these levels of analysis. Fixation frequency on referent R1 maintains higher than the percentage of fixations on the second referent (compare Figure 40). The illustrated differences of the referents in Figure 44 do not fall below 5.63 during the last 2000ms still indicating more fixations on referent R1 than on referent R2. For significant results on subject as well as on image basis there are at least equally many participants (or images respectively) that are found to fixate significantly one over the other referent. Most of the time, however, there are more subjects/images in favour of R1 than of R2 (compare Figures 42 and 43).
3.3.3. Order of Image Presentation

As we presented each image twice it could be possible to induce a priming effect. Therefore we compared the fixation patterns of the first and the second image presentation. Figure 45 illustrates that the percentage of fixations summed over the entire timecourse for referent R1 are higher for the second image presentation in 11 of 12 cases. The average fixation probability is 3.57% higher for the second presentation than for the first (mean 1st presentation: 18.7%; mean 2nd presentation: 22.27%).

![Fixations on Referent 1 differentiated by time of presentation over all subjects and the whole timecourse](image)

Figure 45: Fixation probabilities for the first referent separated by order of image presentation on per-image basis.

The same holds for fixation probabilities on the second referent for the whole timecourse. Here the difference between the average fixation probabilities between the first and second presentation is slightly smaller (1.83%) but still in 9 of 12 cases higher for the second presentation (see Figure 46).
Since Figures 45 and 46 hint to a small but visible priming effect we checked whether effects could be seen also on the level of bootstrapping. It resulted that there is no remarkable difference between the bootstrapping results summed over both image presentations and the first presentation only. Representative for the other conditions we are presenting only Condition 1 in this section. While the maximal number of participants with significant results in favour of R2 is 29 at 10032ms summed over all image presentations, 27 participants showed significant results at the same slot for only the first image presentation (compare Figures 15 and 47). Furthermore the peak between 11000ms and 14000ms for summed presentations is 30 whereas it is 33 for the first presentation alone and therefore differs only slightly. All in all the bootstrapping results obtained for the first presentation do not differ significantly from those summed for both presentations. Thus, our findings are not due to priming effects.
3.4. Answers to Control Questions

After simultaneous presentation of an image and the respective auditory stimulus in one of the six conditions participants had to answer a control question. Via key-press they were supposed to choose which referent (R1 or R2) were the acting character within the last sentence. The results of this choice separated by conditions are depicted in Figure 48.

The clearest preferences can be seen in Conditions 1 and 4 in which linguistic information by the pronoun and contextual preference matched. More than 85% of all choices were made in favour of referent R2 in Condition 1 (87.03%), respectively referent R1 in Condition 4 (85.89%). Conflict Conditions 2 and 3 show higher preference for the contextually biased referent (see Figure 48). In Condition 2 68.07% of the answers were in favour of referent R2 whereas key-presses in Condition 3 indicate preference for referent R1 (75.83%). In neutral conditions participants decided more often for the referent which was biased by the pronoun denomination. This means referent R2 was chosen in 59.67% of all cases in Condition 5 while referent R1 was picked to a percentage of 75.73. The influence being smaller by the demonstrative pronoun than the personal pronoun is
mirrored also by the size of difference for the conditions with demonstrative and personal pronouns compared in the following subsection.

![Figure 48: Answers to the control questions for every condition.](image)

3.5. Differences between Demonstrative and Personal Pronouns

Our third initial hypothesis proposes a higher difference between fixation frequencies of the preferred and non-preferred referents for the demonstrative pronoun. In order to test this we determined the maximal distance of fixation curves for referent R1 and R2 between 9500ms and 10500ms (the range in which a local peak for preferred referents for all conditions was found) which indicate the subtraction of fixation frequency of the preferred minus the non-preferred referent. Furthermore, we summed up results for Conditions 1, 3 and 5 containing the denomination of the demonstrative pronoun as well as for Conditions 2, 4 and 6 in which the personal pronoun was named. As a result Figure 49 illustrates that the difference is slightly higher for conditions with personal pronouns (42.12) than for those containing a demonstrative pronoun (39.59).
Figure 49: Differences in fixation frequencies for preferred and non-preferred referent for conditions with a demonstrative pronoun as well as conditions with a personal pronoun.

3.6. Intelligence

After the conduction of the eye-tracking experiment cognitive skills of the participants were tested by means of the MWT-B test. This was to check for our last hypothesis dealing with whether intelligence has an influence on the integration of lexical and pragmatic information. The results for this task are shown in Figure 50. Performing a Lilliefors-Test on these results confirms that the data are normally distributed ($p<0.05$).

Figure 50: Distribution of the results of the MWT-B.
Based on the scores of the MWT-B we wanted to test our fifth hypothesis by evaluating whether there is a correlation between the point in time of the shift of fixation frequencies in conflict situations and the MWT-B score. For this purpose we defined the shift as the first point in time after pronoun onset at which the participant fixated the contextually preferred referent significantly more than the linguistically preferred one. We further made sure that first significance in favour of the referent preferred linguistically was present. Thus only changes in one direction (namely linguistic to contextual) were considered as shifts.

However, neither for the second nor the third condition we found a significant correlation (for condition 2: \( r = -0.02 \); for condition 3: \( r = 0.07 \)). Figures 51 and 52 make clear that there is no tendency such as the higher the scores in MWT-B the later the shift occurs.

![Figure 51: Correlation between performance in the MWT-B and point in time of shift from referent R1 to R2 in Condition 2](image)

Figure 51: Correlation between performance in the MWT-B and point in time of shift from referent R1 to R2 in Condition 2
In order to check whether intelligence has a more general impact on which referent finally is chosen by the participants (not only regarding the point in time of the shift) we correlated the answers to the control questions with the MWT-B scores. Figure 53 illustrates the percentage of answers for contextually preferred and linguistically preferred referents respectively, differentiated by scores reached in the MWT-B. The distribution shows that choices for the different referents vary without clear tendency between the participants having scores in the range of 19 to 32. However, participants with more extreme scores lying at the upper and lower boundary of MWT-B test results indicate a more distinct choice. While the participant having only 17 points in the MWT-B shows the maximum percentage of choosing the linguistically preferred referent (50%), the smallest values of answering in favour of the linguistically preferred referent can be found for subjects with the highest scores of the test (see Figure 53). Nevertheless, the correlation between answers and test results did not become significant ($r = 0.43$).
Figure 53: Correlation between performance in the MWT-B and the choice for the contextual preference in Conditions 2 and 3.

4. Discussion

Initially we had five hypotheses that we wanted to investigate. The first three assumptions dealt with the differences between German demonstrative and personal pronouns. The fourth hypothesis was concerned with the integration of pragmatic information with the pronoun’s antecedent preference provided earlier in the discourse. Finally, we hypothesized that intelligence has a certain influence on pronoun resolution. Before drawing any conclusions on the eye-tracking data we had to make sure that participants perceived and followed the auditory presented stories in a meaningful way. It was supposed that the subjects fixate referents shortly after their denomination during the short story which would then be in line with previous studies using the visual world paradigm ([Cooper, 1974] and [Tanenhaus, 1996]). Figure 6 gives an overview of the fixation frequencies over the whole timecourse of Condition 1 (summed over participants and images). As can be seen in the depicted curves, the denomination of referents (which in this example were ”Der Polizist” and ”dem Autofahrer”) triggered a rising fixation probability on the respective Playmobil figure shortly after naming the character. Observations 3 and 4 confirm this effect which was described over all conditions.
The third observation shows in Figure 9 that the fixation probability of referent R1 rises at about 3450ms up to 35.25% shortly following the denomination of the noun phrase referring to this referent (at 3000ms). This is consistent with the findings of Hagoort et al. (2004) about the velocity of processing word meaning. At 4800ms after stimulus onset the second referent was named within a full noun phrase. In contrast to the fixation behaviour elicited by introducing the first character, the increase of fixation frequency for R2 starts earlier during the unfolding discourse (at about 4200ms). This early ascent seems to be an anticipatory effect since the verb presented after introducing the first referent already hints to a second character appearing. On Figure 10 a peak of 31.7% can be observed around 6 seconds after stimulus onset. Moreover there is an increase for referent R3 shortly after denomination of the non-human object mentioned in the stimuli at 5800ms (see Figures 6 and 6). In the end, constantly low fixations on the third human Playmobil figure which is not referred to in the auditory stimuli confirm that participants did not fixate referents R1 and R2 to a larger extent because they represent human figures but due to their denomination within the story. Hence, our fixation data allow to assume that participants followed the stories presented auditory in a meaningful way while looking at the respective pictures. Thus we have a valid basis of data for further investigating our different hypotheses.

The first focus of interest was on the resolution of the German demonstrative pronoun “der”. We proposed that fixations on the second referent R2 would rise triggered by the denomination of this pronoun in Conditions 1, 3 and 5 at 9000ms. Regarding this assumption we found confirming evidence on several levels of analysis for the different conditions. For Conditions 1 and 5 the increase of fixation frequency on referent R2 starts at about 9500ms rising with a similar slope (Condition 1: 27.57; Condition 5: 26.93) up to a local maximum of about one third of all fixations (around 10000ms). Although in Condition 3 fixations reached a local maximum in the same time window of about 30%, the curve for referent R2 shows already in ascent before the denomination of the pronoun took place. We think that this might be due to the randomization. Hereby the constraint was set that conflicting Condition 3 is only used at the second presentation of the image which is shown twice to each participant. Two of the three conditions appearing in the first presentation before Condition 3 (Conditions 1 and 5) already contained a preference for referent R2 as main character of the third sentence. As shown in Figure 48 participants have chosen referent R2 to a higher percentage for both conditions. Therefore they already seemed to be primed to think of the second referent
as acting character at the end of the story. Although the priming effect should be the same for referent R1 in conflict Condition 2, we assume that a shift, which makes the non-topical character of the proceeding sentence the topic during the further discourse, is more salient than a shift to an already acting character (referent R1). Consequently subjects did already expect the second referent to be referred to by the last part of the story. Thus their fixations seem to mirror an anticipatory effect not present, though, for referent R1 in the other conflict condition. However, for Conditions 1, 3 and 5 can be said that at least 23 out of 60 participants measured on single-subject level significantly fixated referent R2 over R1 in the first second after pronoun denomination. During the same time interval the difference between the amount of participants significantly fixating R2 and the number of participants with significant results for R1 is at least 5. Hence, within this period more participants fixated significantly the second referent over the first. This finding is consistent with the difference in fixation percentage between the two referents. The difference in all three conditions is in favour of referent R2 and rises up to more than 16% higher fixation probability in Condition 1. To sum it up fixation data show that the demonstrative pronoun triggers higher fixation probability on the referent in non-subject position (R2) in comparison to the referent being in subject position (R1). This is consistent with results found by [Bosch et al., 2007].

The second initial hypothesis stated that this is also the case the other way round, namely the denomination of the personal pronoun will elicit higher fixations on referent R1 than on the second referent represented in object position during the proceeding sentence. The personal pronoun was presented in Conditions 2, 4 and 6. For all those conditions holds that after pronoun denomination the fixation frequency on referent R1 increases while the fixation level for referent R2 remains more or less the same. Confirming evidence is found by slope increase as well as by significance testing and difference calculation. In all conditions containing the personal pronoun there is a slope of at least 19 up to 28 starting between around 500 and 700 milliseconds after pronoun onset. These slopes lead up to a fixation probability of about 25% in Conditions 4 and 6 as well as a local maximum of even 33.64.% for Condition 2 around 1s after the onset of pronoun denomination. Higher fixation probability for referent R1 within this time window is supported also by the number of participants fixating significantly more R1 than R2. While maximally about one third of all participants had significant results in favour of the second referent more than two third of all 60 subjects looked significantly more to referent R1 in all three conditions. Around 1000ms after the personal pronoun
is named there are at least 6% but mostly 16% more fixations on referent R1 than on referent R2 (see Figures 23, 34 and 34). To put it into a nutshell, the evidence found for data of Conditions 2, 4 and 6 supports our second hypothesis. Thus the personal pronoun ”Er” elicits higher fixation probability on referent R1 (which is grammatical subject of the proceeding sentence) than on referent R2 being in object position of the second sentence presented in each condition. These findings are consistent with previous studies investigating the antecedent preferences for German demonstrative and personal pronouns ([Bosch et al., 2007]). Interestingly the increase elicited by demonstrative as well as personal pronouns occurred more or less 1000ms after pronoun onset.

Our last hypothesis concerning German demonstrative and personal pronouns explicitly dealt with the difference between fixation probabilities elicited by the distinct pronoun types. We proposed that the difference of fixation probability between preferred and non-preferred referent is higher for the demonstrative than for the personal pronoun which would hint to a stronger antecedent preference. This seems not to be the case comparing the results of fixation data (see Figure 49). The difference indicates that the personal pronoun triggers even slightly more fixations on the linguistically preferred referent (R1) than on the non-preferred one (R2), namely 42.12, in contrast to the difference for the demonstrative pronoun (39.59). An explanation for this finding might be the anticipation effect found for Condition 3 which consequently lowers the maximal distance between both referents in favour of R2 such that the difference becomes smaller after summing it up. The constraint of our randomization thus might be the reason for this result. In order to clarify whether this assumption is correct future studies with other randomization constraints will be necessary.

The findings for hypotheses 1 and 2 build a basis for the fourth aspect of interest which is the assumption that in conditions with competing preferences induced by intrinsic pronoun bias and pragmatic information a shift in fixation will occur away from the lexically preferred in favour of the contextually preferred referent. We found (for hypotheses 1 and 2) that the German demonstrative pronoun ”der” elicits higher fixation probabilities on the non-subject referent of the proceeding sentence whereas the fixation frequency for referent R1 is increasing after denomination of the personal pronoun. Thereby the pronoun bias for an antecedent is consistent with our presupposition which states that firstly participants process the lexical preference of the pronoun. Furthermore both conflict conditions evince a shift of fixation probability from the lexically preferred referent to the contextually preferred one. This can be seen for slope changes
as well as through changes in distance between the referents and the shifting number of participants respectively images with significant results for the referents. The rising of fixation probability for the referent which is more plausible and the simultaneous decrease of the fixation probability for the previously linguistically biased referent are depicted in Figures 18 and 22. Interestingly these figures demonstrate that the shift is slightly deferred for the conflict conditions. While it occurs between 11 and 12 seconds after stimuli onset for Condition 2 it can already be detected about one second earlier for Condition 3. Although the shifting tendency is the same for both conditions, the increase for the lexically preferred referent starts earlier in time for Condition 3 than for the second condition. As mentioned before this might be due to a mistake in the randomization causing a priming effect for Condition 3. Nevertheless, in the first place we were interested in whether such a shift does occur at all. Besides the evidence of the slopes also distance values affirm the existence of such a shift. Changes from positive to negative (Condition 2) and negative to positive (Condition 3) values clearly indicate the proposed shift. For further analysis, however, we did not want to rely on descriptive statistical means only but focused on results of bootstrapping on per-subject and per-image basis in order to define the predicted change. On per-subject and per-image basis the results for both conditions (2 and 3) show an increasing number of significant results for the contextually preferred referent exceeding the number of subjects having significantly fixated the linguistically preferred referent which meanwhile decreases (compare Figures 15, 16 and 26, 27). Hence our fourth hypothesis is supported throughout all steps of our analysis. Consequently lexical pronoun preferences and preferences based on world knowledge seem to be integrated around the time window in which this shift occurs. Additionally a remarkable finding is that given contradicting linguistic and contextual information participants decided preferably in favour of the more plausible antecedent. This can be seen for the answers indicated by key-press for the final choice of the antecedent. Moreover for conditions with no clear contextual preference subjects decided mostly for the linguistically preferred referent. Thus, available contextual information activating world knowledge outweighs the pronoun’s antecedent preference for the final pronoun resolution.

The question remains whether intelligence has an influence on how fast these contradicting information are integrated. We assumed that participants with high intelligence would be able to perform the integration earlier during the unfolding discourse. Correlating the point in time of the shift (defined as mentioned above) with the results for the
MWT-B test did not become significant (compare Figures 51 and 52). One reason for this could be the usage of MWT-B as an inappropriate measure of intelligence because although it is widely used for clinical purposes it is still matter of debate whether such a test is able to reveal the potential of cognitive abilities (see [Gould, 1996]). On the other hand it might be the case that the influence of intelligence is only significant for participants with remarkably low or high intelligence but not for those lying in between. A hint for this is the distribution of answers to the control questions differentiated by the results of the MWT-B test (see Figure 53). It can be seen that participants with very good test results decided for less than 10% in favour of the linguistically preferred referent whereas the participant with the lowest test results did choose the linguistically preferred referent in 50% of all cases after presentation of a conflict condition. This leads to the supposition that the integration of intrinsic pronoun preference and information of the context might show differences for the final choice of which character is meant by the denomination of a pronoun for participants with comparably high or low intelligence. To put it into a nutshell the fifth hypothesis could not be supported by our data, however, further research using better measures for intelligence or investigating particularly in the integration of conflicting lexical antecedent preference and contextual preference for people with very high or low intelligence would provide more insight in the role of this factor on pronoun resolution.

5. Conclusion

All in all, affirming evidence is found for the differentiation between German demonstrative and personal pronouns concluded from different fixation patterns elicited by the pronoun types. We found that fixation probability for the referent being the grammatical subject of the proceeding sentence was triggered by the personal pronoun whereas the denomination of the demonstrative pronoun induced higher fixation probabilities for the referent in non-subject position. Thus, differences in the intrinsic pronoun preferences are evinced.

Furthermore, contrasting the lexical antecedent preference of the pronouns and preferences based on world knowledge revealed a shift of fixations from the linguistically preferred to the contextually preferred referent. This suggests an integration of information obtained through pronoun bias and world knowledge necessary for pronoun resolution during the unfolding discourse. We even found that world knowledge out-
weighs antecedent preferences of the respective pronoun. However, the point in time of this shift during the discourse seems not to correlate with intelligence.

6. Acknowledgments

First of all I would like to thank Prof. Peter Bosch and Prof. Peter Ko¨ nig for their helpful support as my supervisors. I really appreciated their steady effort to facilitate our work on this study and their advices leading us throughout the whole process. Furthermore, thanks to all people in the Neurobiopsychology Department with special regard to Sonja Schall for answering patiently all kinds of questions and to Johannes Steger who provided us with all necessary technical details. A big thank you to my loving family for supporting me as they always did. Last but not least, I would like to thank my ”Mausraum”-team for being close friends and critical fellow students during this time but mostly for their unique company making this period to what it is.

References


7. Appendix

A. Visual Stimuli

Figure 54: Image 1

Figure 55: Image 2

Figure 56: Image 3

Figure 57: Image 4

Figure 58: Image 5

Figure 59: Image 6
B. Auditory stimuli

1. Eine Verwarnung.
   Der Polizist redet gerade mit dem Autofahrer über das falsch geparkte Auto.

   Condition 1: Der ist mit der Verwarnung nicht einverstanden und regt sich fürchterlich auf.
   Condition 2: Er ist mit der Verwarnung nicht einverstanden und regt sich fürchterlich auf.
   Condition 3: Der will 20 Euro für Falschparken kassieren und regt sich fürchterlich auf.
   Condition 4: Er will 20 Euro für Falschparken kassieren und regt sich fürchterlich auf.
   Condition 5: Der ist ziemlich unhöflich, schreit ganz laut herum und regt sich fürchterlich auf.
   Condition 6: Er ist ziemlich unhöflich, schreit ganz laut herum und regt sich fürchterlich auf.

2. Im Tiergarten.
   Der Wärter spricht eindringlich mit dem Besucher neben dem Löwengehege.

   Condition 1: Der findet den Eintrittspreis unverschämt hoch und verzieht grimmig den Mund.
   Condition 2: Der findet den Eintrittspreis unverschämt hoch und verzieht grimmig den Mund.
   Condition 3: Der sagt man dürfe hier keine Tiere füttern und verzieht grimmig den Mund.
   Condition 4: Er sagt man dürfe hier keine Tiere füttern und verzieht grimmig den Mund.
   Condition 5: Der sagt früher wär’s im Zoo viel schöner gewesen und verzieht grimmig den Mund.
   Condition 6: Er sagt früher wär’s im Zoo viel schöner gewesen und verzieht grimmig den Mund.
3. Auf dem Spielplatz.

*Der Vater diskutiert lange mit dem kleinen Sohn vor der großen gelben Rutsche.*

Condition 1: *Der ist wohl geärgert und gehauen worden und will nach hause gehen.*

Condition 2: *Er ist wohl geärgert und gehauen worden und will nach hause gehen.*

Condition 3: *Der hat gleich noch einen Termin im Büro und will nach hause gehen.*

Condition 4: *Er hat gleich noch einen Termin im Büro und will nach hause gehen.*

Condition 5: *Der hat keine Lust noch länger hierzubleiben und will nach hause gehen.*

Condition 6: *Er hat keine Lust noch länger hierzubleiben und will nach hause gehen.*

4. Es hat gebrannt.

*Der Feuerwehrmann steht erschöpft neben dem Passanten an dem Feuerwehrauto.*

Condition 1: *Der kam zufällig auf dem Spaziergang hier vorbei und wüsste gern die Brandursache.*

Condition 2: *Er kam zufällig auf dem Spaziergang hier vorbei und wüsste gern die Brandursache.*

Condition 3: *Der hat schon alles Material wieder eingepackt und wüsste gern die Brandursache.*

Condition 4: *Er hat schon alles Material wieder eingepackt und wüsste gern die Brandursache.*

Condition 5: *Der meint letzthin sei viel los in dieser Gegend und wüsste gern die Brandursache.*

Condition 6: *Er meint letzthin sei viel los in dieser Gegend und wüsste gern die Brandursache.*
5. *In der Schule.*

*Der Lehrer steht zusammen mit dem Schüler an der alten Kreidetafel.*

Condition 1: *Der hat mal wieder im Unterricht nicht aufgepasst und hat den Termin für die Arbeit vergessen.*

Condition 2: *Er hat mal wieder im Unterricht nicht aufgepasst und hat den Termin für die Arbeit vergessen.*

Condition 3: *Der ist schon ziemlich alt, denkt nicht mehr an alles und hat den Termin für die Arbeit vergessen.*

Condition 4: *Er ist schon ziemlich alt, denkt nicht mehr an alles und hat den Termin für die Arbeit vergessen.*

Condition 5: *Der war im Krankenhaus, ist jetzt erst wieder gesund und hat den Termin für die Arbeit vergessen.*

Condition 6: *Er war im Krankenhaus, ist jetzt erst wieder gesund und hat den Termin für die Arbeit vergessen.*


*Der Stürmer diskutiert energisch mit dem Schiedsrichter vor dem gegnerischen Tor.*

Condition 1: *Der droht natürlich gleich mit Platzverweis und fühlt sich voll im Recht.*

Condition 2: *Er droht natürlich gleich mit Platzverweis und fühlt sich voll im Recht.*

Condition 3: *Der hat einen sauberen Torschuß gemacht und fühlt sich voll im Recht.*

Condition 4: *Er hat einen sauberen Torschuß gemacht und fühlt sich voll im Recht.*

Condition 5: *Der verlangt eine sächliche und verständliche Erklärung und fühlt sich voll im Recht.*

Condition 6: *Er verlangt eine sächliche und verständliche Erklärung und fühlt sich voll im Recht.*

Der Bootsverleiherg geht gerade mit dem Touristen zu dem kleinen Ruderboot.

Condition 1: Der will gleich ablegen und seinen Spaß haben und achtet nicht auf den aufziehenden Sturm.
Condition 2: Er will gleich ablegen und seinen Spaß haben und achtet nicht auf den aufziehenden Sturm.
Condition 3: Der denkt nur an Tagesgeschäft und Geld verdienen und achtet nicht auf den aufziehenden Sturm.
Condition 4: Er denkt nur an Tagesgeschäft und Geld verdienen und achtet nicht auf den aufziehenden Sturm.
Condition 5: Der schwärmt nur von dem herrlichen Natursee und achtet nicht auf den aufziehenden Sturm.
Condition 6: Er schwärmt nur von dem herrlichen Natursee und achtet nicht auf den aufziehenden Sturm.

8. Gewehre für die Indianer.

Der Häuptling verhandelt ausdauernd mit dem Händler vor dem großen Zelt.

Condition 1: Der will diesmal wirklich einen guten Gewinn machen und feilscht lange um den Preis.
Condition 2: Er will diesmal wirklich einen guten Gewinn machen und feilscht lange um den Preis.
Condition 3: Der braucht bessere Gewehre für seinen Stamm und feilscht lange um den Preis.
Condition 4: Er braucht bessere Gewehre für seinen Stamm und feilscht lange um den Preis.
Condition 5: Der sagt, dass es ein faires Geschäft sein muss und feilscht lange um den Preis.
Condition 6: Er sagt, dass es ein faires Geschäft sein muss und feilscht lange um den Preis.
9. **Auf der Straße.**

Der Müllmann plaudert gerade mit dem Mieter neben der übervollen Mülltonne.

Condition 1: *Der will wissen wann wieder Müll abgeholt wird und beschwert sich über die Unordnung.*

Condition 2: *Er will wissen wann wieder Müll abgeholt wird und beschwert sich über die Unordnung.*

Condition 3: *Der hat den ganzen Tag schon Müll abgeholt und beschwert sich über die Unordnung.*

Condition 4: *Er hat den ganzen Tag schon Müll abgeholt und beschwert sich über die Unordnung.*

Condition 5: *Der findet Sauberkeit wär wirklich das Allerwichtigste und beschwert sich über die Unordnung.*

Condition 6: *Er findet Sauberkeit wär wirklich das Allerwichtigste und beschwert sich über die Unordnung.*

10. **In der Manege.**

Der Zauberer trifft gerade zufällig den Clown vor der leeren Tribüne.

Condition 1: *Der sieht echt zum Lachen aus und freut sich auf die Abendvorstellung.*

Condition 2: *Er sieht echt zum Lachen aus und freut sich auf die Abendvorstellung.*

Condition 3: *Der murmelt einige Zaubersprüche und freut sich auf die Abendvorstellung.*

Condition 4: *Er murmelt einige Zaubersprüche und freut sich auf die Abendvorstellung.*

Condition 5: *Der ist auf dem Weg zum Eingang und freut sich auf die Abendvorstellung.*

Condition 6: *Er ist auf dem Weg zum Eingang und freut sich auf die Abendvorstellung.*
11. In der Klinik.

*Der Arzt steht wieder vor dem Patienten am sauberen Krankenbett.*

Condition 1: *Der beklagt sich ber die Schmerzen und befürchtet leichte Nebenwirkungen.*

Condition 2: *Er beklagt sich ber die Schmerzen und befürchtet leichte Nebenwirkungen.*

Condition 3: *Der rät die Medizin einzunehmen und befürchtet leichte Nebenwirkungen.*

Condition 4: *Er rät die Medizin einzunehmen und befürchtet leichte Nebenwirkungen.*

Condition 5: *Der ist mit der Operation nicht zufrieden und befürchtet leichte Nebenwirkungen.*

Condition 6: *Er ist mit der Operation nicht zufrieden und befürchtet leichte Nebenwirkungen.*


*Der Klempner berät altklug den Bauern wegen der kaputten Wasserpumpe.*

Condition 1: *Der bemerkte den Schaden vor dem Melken und muss eine Lösung fr das Problem finden.*

Condition 2: *Er bemerkte den Schaden vor dem Melken und muss eine Lösung fr das Problem finden.*

Condition 3: *Der ist extra auf den Bauernhof gekommen und muss eine Lösung fr das Problem finden.*

Condition 4: *Er ist extra auf den Bauernhof gekommen und muss eine Lösung fr das Problem finden.*

Condition 5: *Der ist ziemlich gestresst, hat kaum Zeit und muss eine Lösung fr das Problem finden.*

Condition 6: *Er ist ziemlich gestresst, hat kaum Zeit und muss eine Lösung fr das Problem finden.*
C. Consent Sheet
D. Subject Questionnaire
E. Feedback Sheet
F. Affirmation

Affirmation

Hiermit erkläre ich, Marlene Meyer, die vorliegende Arbeit ”Investigating the influences of world knowledge and intelligence on the resolution of German personal and demonstrative pronouns” selbstständig verfasst zu haben und keine anderen Quellen oder Hilfsmittel als die angegebenen verwendet zu haben (siehe auch ”Assignment of Tasks”).

Osnabrück, den 28.09.2007

Affirmation

I, Marlene Meyer, hereby confirm that I composed the work at hand, entitled ”Investigating the influences of world knowledge and intelligence on the resolution of German personal and demonstrative pronouns” independently and that I did not use any other resources or auxiliary means than the ones stated (see also ”Assignment of Tasks”).

Osnabrück, 28.09.2007
G. Assignment of Tasks

- **Within the whole group:**

  - Before experiment:
    1. Preparation:
       * Python introduction by Johannes Steger: Birgit Bärnreuther, Dominique Goltz, Maike Kleemeyer, Marlene Meyer
       * AVMZ for microphone: Birgit Bärnreuther, Marlene Meyer
    2. Stimuli construction:
       * Writing stories: each group on their own
       * Taking photographs of Playmobil scenes: Birgit Bärnreuther, Christina Bergmann, Dominique Goltz, Maike Kleemeyer, Marlene Meyer
       * Audio recordings:
         - Speaker: Dominique Goltz
         - Recording: Florian Krause
       * Cutting audio files: Birgit Bärnreuther, Dominique Goltz, Maike Kleemeyer, Florian Krause
       * Cutting images: Birgit Bärnreuther
    3. Python programming: mainly Florian Krause, Maike Kleemeyer, Marlene Meyer
    4. Randomisation: Christina Bergmann, Florian Krause, Marlene Meyer
    5. Recruiting participants:
       * Flyer: Marlene Meyer
       * E-mails etc.: everybody
    6. Organizing MWT-B: Christina Bergmann, Florian Krause
    7. Documents:
       * Feedback-Sheet: everybody
       * Consent-Sheet: Birgit Bärnreuther, Maike Kleemeyer
       * Experimental plan of procedure: Dominique Goltz
    8. Eye-tracking introduction by Sonja Schall and practice: everybody

- During Experiment:
1. Eye-tracking: everybody
2. ROI definition: each group their own
3. Fixmat & Roimat construction: Florian Krause, Marlene Meyer

- After experiment:
  1. Scaling ROIs: Birgit Bärnreuther, Dominique Goltz
  2. Functions: each group on their own, exchange of those that each group needed

- Within IQ group:
  1. Analyzing Data: Florian Krause, Marlene Meyer
  2. Writing Bachelor Thesis:
     * Outline: everyone on his/her own
     * Introduction: everyone on his/her own
     * Methods: Are equal in both theses, that of Marlene Meyer and that Florian Krause
       · Visual Material: Florian Krause
       · Auditory Material: Florian Krause
       · Combination/Randomization/Balancing: Florian Krause
       · Intelligence Test: Marlene Meyer
       · Participants: Marlene Meyer
       · Apparatus: Florian Krause
       · Procedure: Marlene Meyer
       · Regions of Interest: Marlene Meyer
     * Results: everyone his/her own
     * Discussion: everyone his/her own
     * Conclusion: everyone on his/her own