Exploiting Referential Gaze for Uncertainty Reduction in Situated Language Processing

An Information-Theoretic Approach

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie der Philosophischen Fakultät der Universität des Saarlandes
Vorgelegt von
Mirjana Sekicki
aus Novi Sad, Serbien

Saarbrücken August 2019

Dekan der Fakultät: Prof. Dr. Heinrich Schlange-Schöningen
Erstberichterstatterin: Dr. Maria Staudte
Zweitberichterstatter: Prof. Dr. Matthew W Crocker
Tag der letzten Prüfungsleistung: 12 Juni 2019

Mojim roditeljima

Abstract

A large body of contemporary psycholinguistic research utilizes the information-theoretic notions related to the transmission of information in an attempt to better understand and formalize the regularities of language production and comprehension. The overarching hypothesis is that prediction is a core mechanism underlying language comprehension. Anticipating what is likely to be mentioned next based on the previous context is what is assumed to allow for smooth and effortless communication. The anticipation of linguistic units that fit the current context reduces the uncertainty about the upcoming material, which consequently facilitates the processing of that material, in a typically noisy channel.

Situated language processing allows for the integration of not only linguistic but also non-linguistic visual information that contribute to establishing the context, and facilitate the creation of anticipations regarding the upcoming linguistic material. Moreover, noticing that our interlocutor is directing her attention to a certain object, inspires a shift in our visual attention towards the same entity. Since what is relevant for our interlocutor is highly likely to be relevant for us, too, whether simply conversationally, or more importantly, even existentially (Emery, 2000). Hence, following the speaker's referential gaze cue towards an object relevant for the current conversation has been shown to benefit listeners' language processing, measured by shorter reaction times on subsequent tasks (e.g., Staudte & Crocker, 2011; Staudte, Crocker, Heloir, & Kipp, 2014; Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013, 2014).

The present thesis aimed to provide an insight into the mechanisms behind this facilitation. We examined the dynamics of combining visual and linguistic information in creating anticipation for a specific object to be mentioned, and the effect this has on language processing. To this end we used a pupillary measure of cognitive load that is robust enough to allow for free eye movements (the Index of Cognitive Activity; Marshall, 2000). This enabled us to measure not only the visual attention during language comprehension, but also the immediately induced cognitive load at various relevant points during the auditory presentation of the linguistic stimulus.

Eight experiments were conducted towards addressing our research questions. The initial three experiments established the ICA measurement in the context of our linguistic manipu-

lation. This series of experiments included reading, cognitive load during listening, and the examination of visual attention together with cognitive load in the visual world paradigm (VWP). Subsequently, we conducted five eye tracking experiments in the VWP where the linguistic context was further enriched by a referential gaze cue. All five experiments simultaneously assessed both visual attention and the immediate cognitive load induced at different stages of sentence processing. We manipulated the existence of the referential gaze cue (Exp. 4), the probability of mention of the cued object (Exp. 4, 5), the congruency of the gaze cue and the subsequent referring expression (Exp. 6), as well as the number of cued objects with equal probability of mention (Exp. 7, 8). Finally, we examined whether the gaze cue can take the role of fully disambiguating the target referent (Exp. 8).

We quantified the importance of the visual context in language processing, and showed that if a certain object from the visual context has a higher likelihood of mention given the linguistic context, its processing is facilitated, in comparison to the processing of the same sentence without the visual context. Furthermore, our results support the previous findings that the referential gaze cue leads to a shift in visual attention towards the cued object, thereby facilitating language processing. We expanded these findings by showing that it is the processing of the linguistic reference, that is the referent noun, that is facilitated by gaze-following. Importantly, perceiving and following the gaze cue did not prove costly in terms of cognitive effort, unless the cued object did not fit the verb selectional preferences. This is true regardless of the number of objects cued, or the lower likelihood of mention of the cued object.

We conclude that listeners strategically use visual information to reduce the referential uncertainty for upcoming nouns but that the visual cues, such as the referential gaze cue, do not underly the same kinds of expectations (and resulting cognitive costs) as linguistic references. We did not find evidence that the gaze cue is processed in a manner comparable to noun processing, rather, it is likely perceived as a relevant piece of information introduced in addition to the linguistic material, in order to aid language processing, but, importantly, not there to substitute it.

Zusammenfassung

Die Psycholinguistik erstrebt die Untersuchung von Vorhersagen, bzw. Wahrscheinlichkeiten mit denen sprachliche Elemente basierend auf linguistischem Kontext verwendet werden um zu erklären, wie ein komplexes Kommunikationsmittel wie Sprache so mühelos verwendet werden kann. Jedes bisherige Element führ dabei dazu die Menge der auszuwählenden folgenden Elemente zu reduzieren. Diese Information wird benutzt um vorherzusagen was wahrscheinlich auf ein aktuelles Element folgen wird. In der gleichen Weise, wird in der situierten (face-to-face) Kommunikation der visuelle Kontext verwendet um zusätzliche Informationen zu erhalten, die die Wahrscheinlichkeit einer konkreten visuell anwesenden Entität mit der Sprache referiert zu werden erhöht. Außerdem werden visuelle Verweise häufig benutzt um die Wahrscheinlichkeit der Nennung einer Entität zu erhöhen. Dabei lösen sie die Ambiguitäten und helfen Missverständnissen vorzubeugen.

Flüssige Kommunikation und müheloses Sprachverständnis sind potentiell dadurch möglich, dass wir antizipieren können was wahrscheinlich ist erwähnt zu werden. Das wird erreicht durch die Bewertung von dem was in dem aktuellen Kontext am sinnvollsten wäre. Informationstheoretische Konzepte werden in der Psycholinguistik verwendet um diese Vermutung empirisch zu testen. Zunächst gilt es das Konzept Surprisal (Shannon, 1948; Hale, 2001; Levy, 2008) zu nennen. Dieses hat Korrelationen zwischen Sätze mit Fortseztzungen mit unterschiedlicher Wahrscheinlichkeit gezeigt. Darüber hinaus, wird auch Entropie als Metrik der Schwierigkeit, mit der ein linguistisches Element perzipiert wird vorhersagt, genutzt. Meistens ist in diesem Kontext von entropy reduction die Rede. Dies bedeutet die Reduktion der Ungewissheit über das Erscheinen einer linguistischen Einheit. Die Hypothese der gleichmäßigen Informationsdichte (Uniform Information Density; Jaeger, 2010) versteht dass die optimale Verteilung des Informationsgehalts in einer Sprachquelle nah an der oberen Grenze der Kanalkapazität des Zuhörers liegt – weder zu sehr herausfordernd, noch zu langweilig (Genzel & Charniak, 2002), und mit gleichmäßiger Informationsdichte – ohne großen Unterschiede zwischen linguistische Einheiten.

In der Vergangenheit wurde gezeigt, dass die linguistischen und visuellen Kontexte zusammen benutzt werden um den nachfolgenden Inhalt zu antizipieren.Kamide, Altmann und Haywood (2003) fanden, dass die vom Verb stammende Information dazu benutzt wird,

die nachfolgenden linguistischen Argumente zu antizipieren. Die Augenbewegungen der Zuhörer zeigten eine aktive Aktualisierung von einer mentalen Repräsentation, basierend sowohl auf sprachlichen als auch visuellen Informationen (Altmann & Mirkovic, 2009; Huettig & Altmann, 2011).

In situierter Kommunikation ist es üblich, dass visuelle Verweise, wie z.B. Zeigegesten, benutzt werden um referenzierende Ausdrücke zu disambiguieren (Bangerter, 2004). Hanna und Brennan (2007) zeigten, dass die Blickrichtung des Sprechers ähnlich benutzt werden kann um die temporare Ambiguität zu lösen. Die Autoren haben die Schlussfolgerung gezogen, dass die Blickrichtung eine konversationsbasierte Informationsquelle ist, die aktiv genutzt wird um schnell eine Referenzauflösung zu erwirken. Sprecher schauen ein Objekt 800—1000ms vor dem Referieren an (e.g. Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998). Dabei zeigt sich der Fokus der Aufmerksamkeit des Sprechers (Emery, 2000; Flom, Lee, & Muir, 2007). Dieser Verweis wird vom Zuhörer verwendet um das linguistische Material leichter zu verarbeiten, weil es hilft die referierenden Ausdrücke zu disambiguieren. Das Folgen der Blickrichtung des Sprechers erleichtert das Sprachverstehen (e.g. Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013, 2014; Staudte & Crocker, 2011; Staudte et al., 2014). Allerdings sind die Ergebnisse, die zu dieser Schlussfolgerung geführt haben meist behaviorale Daten (Aufgabenausführung, Reaktionszeiten, Fehlerfreiheit, etc.). Im Rahmen der vorliegenden Arbeit, haben wir versucht diese Ergebnisse zu erweitern. Dazu verwenden wir ein Maß des kognitiven Aufwands, das robust genug ist um die simultane Untersuchung freier visueller Aufmerksamkeit, bzw. Augenbewegungen zu erlauben, welches in der Dilation der Pupille widergespiegelt wird. Dieses Maß wird häufig als ICA bezeichnet (the Index of Cognitive Activity; Marshall, 2000).

Im Rahmen der vorliegenden Arbeit untersuchen wir die genauen Zeitpunkte zu denen Informationen aus verschiedenen Modalitäten relevant werden, deren Zusammenspiel und Integration. Die offensichtliche Frage ist ob ein visueller Verweis wie der Objekt-gerichtete Blick des Sprechers, in die Satz-Interpretation integriert wird wie ein linguistisches Element. Anders ausgedrückt: Werden visuelle Verweise nur beschränkend verwendet um die Gruppe von möglichen Objekten zu reduzieren (ähnlich einem Verb), bevor eine linguistische Einheit (referierendes Nomen) das Objekt identifiziert; oder versteht man den Verweis als Identifizierung des Zielobjekts (ähnlich einem referierenden Nomen)? Unsere Hypothese war es, dass linguistischer und visueller Kontext sowie visuelle Verweise gemeinsam betrachtet werden und bei der Verteilung der Information mitwirken, sodass visuelle Verweise zu einer gleichmäßigeren Informationsverteilung während eines Satzes beitragen.

Wir stellen acht Experimente vor, mit denen wir unsere Forschungsfragen untersucht haben. Um zu überprüfen ob das ICA Maß sensitiv genug ist um die Effekte unserer

subtilen linguistische Manipulation zu erkennen, führten wir zunächst eine Reihe von Experimenten durch; Ein Lese-Experiment und ein Zuhör-Experiment in denen die kognitive Belastung gemessen wurde (ICA), und schließlich ein Experiment in dem das Visual-World-Paradigma (VWP) verwendet wurde, bei dem neben der kognitiven Belastung auch die visuelle Aufmerksamkeit untersucht wurde. Die restliche fünf Experimente benutzen weiter das Visual-World-Paradigma und untersuchen die Augenbewegungen und die gleichzeitige kognitive Belastung von Versuchspersonen, während diese Sätze in der deutschen Sprache hörten und einen passenden visuellen Kontext betrachteten.

Ziel dieser Dissertation war es, drei allgemeine Forschungsfragen zu beantworten. Im Folgenden stellen wir diese Fragen zusammen mit den darauffolgenden Ergebnissen vor.

Erstens untersuchten wir ob der Effekt von visuellem Kontext auf die linguistische Verarbeitung mit dem ICA Maß quantifiziert werden kann. Da der visuelle Kontext üblicherweise in situierter Kommunikation dauernd anwesend ist, und während des gesamten Ablaufs eines Satzes integriert werden kann, war unsere Hypothese, dass der visuelle Kontext die kognitive Belastung während der Sprachverarbeitung erhöhen würde. Die visuelle Information wird relevanter zu dem Zeitpunkt, an dem es in Zusammenhang mit linguistischem Material gesetzt wird. Darüber hinaus erwarteten wir, basierend auf der UID Hypothese, dass an den Stellen, an denen der visuelle Kontext relevant wird, eine Erhöhung der kognitiven Belastung zu erwarten ist. Demzufolge sollte auf diese Erhöhung eine Erleichterung folgen, zu dem Zeitpunkt, zu dem der referierende Ausdruck verarbeitet wird, weil die verfügbare relevante Information schon früher inkrementell in Betracht gezogen wurde.

Die drei ersten Experimente (Experiment 1–3 genannt) dienten dazu diese Frage zu beantworten. Das gleiche linguistische Material wurde in drei verschiedenen experimentellen Designs untersucht. Es wurde erst gelesen, dann gehört, und schließlich zusammen mit dem relevanten visuellen Kontext präsentiert. Unsere Ergebnisse zeigen, dass die Anwesenheit von einem visuellen Kontext zu erhöhter kognitiver Belastung führt. Außerdem führt sie auch dazu das linguistische Material leichter zu verarbeiten, wenn dieses aufgrund des visuellen Kontextes leichter zu antizipieren ist. Die Augenbewegungen der Probanden waren nach dem Verb auf jene Objekte gerichtet die mögliche Referenten sein könnten. Je weniger passende Objekte es in dem visuellen Kontext gab, desto einfacher war die Sprachverarbeitung des nachfolgenden referierenden Ausdrucks.

Zweitens: Nachdem wir gesehen haben, dass ein visueller Kontext inkrementell mit dem linguistischen integriert wird, und dass diese Verknüpfung die Sprachverarbeitung des passenden linguistischen Materials unterstützt, führten wir einen visuellen Verweis ein, nämlich einen referierendes Blickverweis (referential gaze cue). Der Verweis auf das relevante Objekt durch die Blickrichtung ist eng mit der Sprache verbunden. Dadurch ist

es auch zu erwarten, dass die Verweise im Prozess der Sprachverstehen integriert sind. Es gibt bereits Hinweise darauf, dass das Folgen der Blickrichtung das Sprachverstehen erleichtern kann, wenn die Blickrichtung auf das Objekt verweist, welches danach sprachlich erwähnt wird. Was unklar bleibt, ist die Art und Weise, auf die diese Erleichterung des Sprachverstehen stattfindet. Daher wollten wir die folgende Frage beantworten: Betrifft der referenzierende visuelle Verweis die kognitive Belastung die notwendig ist für die Verarbeitung der übereinstimmenden linguistischen Referenz?

Unsere Hypothese war es das die Blickrichtung zusammen mit dem linguistischen Material dazu führt, dass ein bestimmtes Objekt als Gegenstand des Gespräches, bzw. des Satzes, antizipiert wird. Wenn das der Fall ist, erwarteten wir, dass die sofortige kognitive Belastung die notwendig für die Verarbeitung von referierenden Ausdrücken ist, dabei reduziert wird, weil diese Information schon antizipiert wurde. Demzufolge erwarten wir, wenn die kognitive Belastung auf dem referierenden Nomen durch die Blickrichtung reduziert wird, dass dieser Erleichterung eine entsprechende Erhöhung vorangegangen ist. Präziser erwarteten wir eine höhere kognitive Belastung zu dem Zeitpunkt, zu dem die Blickrichtung des Sprechers zu sehen ist und auf ein relevantes Objekt verweist. Demzufolge, weil das relevante Objekt schon als potentielles Objekt betrachtet ist, sollte die kognitive Belastung auf dem referierenden Nomen entsprechend niedriger werden.

Wir adressieren diese Frage auf zwei verschiedene Arten und Weisen. Auf der eine Seite manipulierten wir den visuellen Verweis auf qualitativer Art und Weise. Dazu quantifizierten wir die Rolle der visuellen Verweisen die unterschiedliche Wahrscheinlichkeiten haben in dem gegebenen Gesamtkontext. Auf der anderen Seite manipulierten wir quantitativ die Information, die mit einer Blickrichtung gegeben ist. Wir untersuchten dazu die visuellen Verweise, die die Ungewissheit über den Referenten reduzierten in Bezug auf die Anzahl potentieller Referenten.

Drei Experimente (Experiment 4–6 genannt) untersuchten die qualitative Perspektive. Jedes dieser Experimente wurde im VWP durchgeführt wo die visuelle Aufmerksamkeit zusammen mit dem gleichzeitigen kognitiven Aufwand gemessen wird. Die Ergebnisse zeigten, dass der Blickrichtung gefolgt wird, was zu einer Verlagerung der visuellen Aufmerksamkeit führte, egal ob das angeschaute Objekt passend zum linguistischen Kontext war oder der Blick nicht immer zuverlässig war. Dieser Verweis hat folglich die kognitive Belastung, die benötigt wird für die Referenzbearbeitung, beeinflusst, sodass es in einer erleichterten Sprachverarbeitung des referierenden Ausdruckes resultierte. Wenn die Blickrichtung zu einem Objekt hindeutete, das zum linguistischen Kontext passte, war die danach folgende Verarbeitung des referierenden Ausdrückens erleichtert.

Nachdem Experiment 4 gezeigt hat, dass der Blickverweis auf ein Objekt zu einer Erleichterung der Verarbeitung des linguistischen referierenden Ausdrucks führt, untersuchte Experiment 5 den Blickverweis auf ein Objekt, das nicht zu dem Kontext passt, das aber nachfolgend auch genannt wird. Die Ergebnisse zeigten, dass die Zuhörer nach einer gewissen Zeit verstanden haben, dass der Blick-Verweis, obwohl überraschend und nicht passend, doch zuverlässig ist, und deutet darauf hin welches das Target-Objekt ist. Wenn das Vertrauen hergestellt wurde, sank die kognitive Belastung auf dem Nomen, das auf dieses Objekt referierte. Experiment 6 manipulierte die Kongruenz des Blickverweises mit dem nachkommenden referierenden Nomen. Interessanterweise zeigten die Ergebnisse, dass ein inkongruenter Blick-Verweis nicht erschwerend auf die Referenzverarbeitung wirkte, sondern nur, dass es hilfreich war, wenn der Blickverweis und das Nomen übereingestimmt haben.

Darüber hinaus untersuchten wir die direkten Kosten der Perzeption des Blickverweises und seine Verwendung. Die Ergebnisse zeigten, dass es hat keine höhere kognitive Belastung verursacht, wenn der Blickverweis zum linguistischen Kontext passte. Nur wenn das verwiesene Objekt nicht als Objekt des Satzes in Frage käme war es der Fall, dass der Blickverweis eine höhere Belastung verursachte. Dies deutet darauf hin, dass höhere Kosten induziert sind, wenn der visuelle Verweis nicht mit dem vorher etablierten Kontext integriert werden kann. Von der UID Hypothese inspiriert, erwarteten wir eine Distribution der kognitiven Belastung zwischen den linguistischen und visuellen Verweisen. Interessanterweise zeigten alle drei Experimente, dass eine Erleichterung der Verarbeitungslast auf dem referierenden Nomen nicht von einer Erhöhung der Kosten auf dem Blickverweis verbunden ist.

Die zwei folgenden Experimente (Experiment 7–8 genannt) nahmen die quantitative Perspektive, nämlich die Kombination von linguistischen und visuellen Verweisen, erlaubten eine schrittweise Reduktion von referenzieller Unsicherheit (referential uncertainty). Anstatt unterschiedlicher Wahrscheinlichkeit der Erwähnung in dem gegebenen Kontext, waren alle präsentierten Konkurrenten gleichmäßig wahrscheinlich, während die Anzahl relevanter Objekte allmählich reduziert wurde. Die referenzielle Unsicherheit wurde in zwei Schritten reduziert: visuell – durch einen Objekt-orientierten Blickverweis, und sprachlich – durch Objektbenennung. Über die beiden Experimente wurde die Reihenfolge der beiden Verweise verändert, sodass es entweder der sprachliche oder der visuelle Verweis war, der die Referenz auflöste.

Auf dem ersten der beiden Verweise fanden wir sofortige Verlagerung der visuellen Aufmerksamkeit zu dem relevanten Objekt aufgrund des Blickverweises oder dem referierenden Nomen. Auch wenn dieses Ergebnis zu erwarten war, ist es dennoch relevant für die

Bewertung der Art und Weise wie die visuelle Aufmerksamkeit sich auf die kognitive Belastung auswirkt. Der sprachliche oder visuelle Verweis auf eine Menge von Objekten, zu einer Unterteilung der Gruppe der potenziellen Targets war nicht in der kognitiven Belastung reflektiert. Die neue Information motivierte eine Verlagerung der visuellen Aufmerksamkeit zu den relevanten Objekten. Allerdings ist das Betrachten von mehr oder weniger neuerlich relevanter Objekte und das gleichzeitige in der Lage sein die anderen auszuschließen, nicht unterschiedlich aufwändig. Wir behaupten, dass die Ergebnisse mit surprisal zu erklären sind. Das Ausbleiben eines Effekts auf dem ersten Schritt, sowohl auf dem visuellen als auch dem linguistischen Verweis, weist darauf hin, dass keine Antizipation für ein spezifisches Objekt an dieser Stelle gemacht wurde. Der gegebene Kontext hat schließet eine Menge präsentierter Objekten die alle zum Verb gepasst haben. Daher zeigte weder das Sehen visueller Verweise auf eine Gruppe von Objekte, noch das Hören eines Nomen, das darauf referierte einen gradualen Unterschiede relativ zu der Größe der selektierten Gruppe. Diese Ergebnisse untermauern die vorherigen.

Auf dem zweiten Verweis, bei dem die Referenz gelöst wird, fanden wir einen gestuften Effekt der kognitiven Belastung auf dem Nomen, relativ zu der Anzahl von direkten Konkurrenten. Das Referieren auf das Target, wenn es das einzige visuell hervorgehobene Objekt war, führte zu geringerer kognitiver Belastung. In dieser Kondition war das Target das wahrscheinlichste Objekt für eine Erwähnung, basierend auf dem vorherigen visuellen Verweis. Je mehr Objekte es in der verwiesenen Gruppe gab, desto niedriger war die Wahrscheinlichkeit, dass ein spezifisches Objekt aus der Gruppe genannt wird. Infolgedessen kann gesagt werden, dass die auf dem Nomen gemessene kognitive Belastung mit der Gruppengröße gesteigert wird. Drittens haben wir versucht die folgende Frage zu beantworten: Ist der referierende Ausdruck erst an dem Zeitpunkt lösbar zu dem man die linguistische Referenz hört, oder kann der visuelle Verweis die Rolle der Referentenidentifikation übernehmen, ähnlich einem Nomen? Die Ergebnisse zeigten, dass es, wenn der visuelle Verweis die Referenz auf dem zweiten Schritt gelöst hat, keine Auswirkung auf die kognitive Belastung gab.

Insgesamt lassen unsere Ergebnisse darauf schließen, dass sowohl sprachliche als auch visuelle Informationen Teil der intendierten Nachricht sind, die es zu verstehen gilt. Unsere Ergebnisse deuten darauf hin, dass restriktive Verben dazu benutzt werden den visuellen Kontext zu verstehen, während der referierende Blickverweis das Antizipieren von referierenden Nomen motiviert. Beide Modalitäten werden inkrementell und interagierend genutzt. Wir fanden Effekte von surprisal, speziell das Integrieren von Information in ein Situationsmodel. Je definierter das Model ist, desto größer sind die Effekte. Die Sprache scheint als definitivere und explizitere Informationsquelle als die visuellen Verweise wahrgenommen zu werden. Wir argumentieren, dass dies der Fall ist, weil die Verknüpfung referierender

visueller Blickverweise mit der Sprache so stark ist, dass der referierende Blickverweis nicht als ein Substitut für das referierende Nomen gilt, sondern als zusätzliche Kontribution zum besseren Satzverstehen.

Acknowledgements

During the last few years that led to writing this thesis, a number of people have touched my life and my work in influential ways. I will mention just a few.

Firstly, I would like to thank professor Emiel Krahmer for taking me by the hand in my first baby scientific steps, for believing in me and my research ideas, and inspiring me to pursue a PhD.

Secondly, I am most grateful to my supervisor Maria Staudte for being a constant source of support, uplifting spirit and motivation. Also, I would like to thank all the members of the MS-mit, MC-mit and IS-mit groups over the last four years who have allowed for a friendly productive working atmosphere.

Also, I am grateful for the financial support I have received from the Cluster of Excellence MMCI and SFB 1102 which allowed me to attend numerous amazing conferences around the globe.

Finally, I would like to thank my family for their unconditional love, for having been there with me throughout this experience, at every step of the way, regardless of the thousands of kilometers that were often between us.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

If the comprehension of a particular linguistic item relies on its immediate (linguistic) context, in face-to-face communication that context is further enriched by the visual information. Here, the likely referents for a linguistic expression may be among the co-present visual objects. The concrete set of potential referents can then be further reduced by using visual cues such as gestures or gaze, in order to avoid ambiguities and facilitate language processing. This thesis aims to tackle the mechanisms behind the integration of multi-modal information – the visual information, immensely important in our situated encounters, and the information from the linguistic stream.

Let us begin with an illustration. Please consider how difficult it would be to comprehend what was meant by the following sentence – *I broke my most expensive necklace* – in three different communication situations. Imagine yourself as an attentive listener in the following three scenarios.

Scenario 1: You are having a conversation on the phone with your friend in which she utters the following *Imagine what happened last night! I broke my most expensive necklace*. Unless you have a detailed inventory in your head about all the value your friend has in her jewelry box, you will not be very likely to retrieve the exact necklace she is referring to. Hence, she will continue to give you further information about the object she meant. For instance: *You know, the golden one with rubies that my husband got me a few years back.*

Scenario 2: You are sitting with your friend in her dressing room. You are going through her jewelry box while she is doing her makeup. Then, she utters the same sentence *Imagine what happened last night! I broke my most expensive necklace*. In this context, you have a specific collection of necklaces in front of you, and a much higher probability of deducing which is the most expensive one, even if you did not know which one had the highest price tag. You will, again, likely need some additional information in order to disambiguate the necklace in question. Importantly, though, at the first mention of the necklace you are already

a few steps ahead in the effort of identifying the relevant object than you were in the first scenario.

Scenario 3: Both you and your friend are sitting together in front of her jewelry box. While showing you her jewelry collection, your friend mentions the same sentence, as previously. This time, however, right after uttering *I broke* she looks at one specific necklace, before continuing the sentence (... *my most expensive necklace*). Noticing the change in her gaze direction, you could immediately spot which necklace she was focusing her attention on. Hence, already by the time she finished the sentence, and before she picked up the relevant necklace from the box, you could be quite certain which was the object in question.

In the recent psycholinguistic literature, there have been attempts to explain the intriguing ease with which people use language in terms of prediction, or rather, the probability of an element to follow the current state, firstly and foremostly, based on the immediate linguistic context. Each element constrains the choice of the subsequent one, which is the information used to anticipate what is likely to follow. In the same vein, in situated communication, the immediate visual context provides additional information that increases the probability of a concrete, visually present entity to be referred to by language. In addition, visual pointers are commonly used to further increase the probability of a particular entity, and hence, resolve ambiguities and help avoid misunderstandings.

In the present work, we examine the exact time points when pieces of information from different modalities become relevant, and look into their interplay and integration. The question that emerges on this level of granularity is whether a visual cue such as referential (i.e., object-directed) speaker gaze is processed and incorporated in the sentence interpretation on a par with a verb, or rather, like a noun. In other words, is the visual cue perceived as simply constraining the set of potential target objects (similar to a verb), before a linguistic unit (referent noun) identifies the target; or is the cue actually seen as concretely identifying the target object (resembling the referent noun)?

In the remainder of this chapter we give a short overview of the relevant literature. We present the current understanding of the issue raised, by mentioning relevant findings regarding prediction in language processing, as well as the role of visual cues, focusing on speaker gaze. Subsequently, we will list the remaining open questions that inspired us to conduct the present work, as well as the way in which we have approached them. Finally, we give a detailed outline of the whole thesis, mentioning the content of each chapter. Please note that some passages from this chapter appear in our recently published manuscript (Sekicki & Staudte, 2018).

1.1 Prediction and Situated Language Comprehension

Smooth communication and effortless language processing are assumed to be possible due to our ability to anticipate what is likely to be mentioned next, based on the assessment of what would make sense given what has been said (seen, known and experienced) up to now. In its minimal sense, prediction in language comprehension is understood in the following terms: "Context changes the *state* of the language processing system before new input becomes available, thereby facilitating processing of this new input (Kuperberg & Jaeger, 2016, p. 33)".

In order to be able to empirically examine this idea, much recent psycholinguistic research has employed information-theoretic notions. Inspired by Shannon (1948)'s notion of surprisal (see Hale, 2001; Levy, 2008), studies have connected sentences of varyingly surprising continuations with different measures of processing difficulty. Reading patterns (e.g., Rayner & Well, 1996), reading times (e.g., Demberg & Keller, 2008), and the N400 ERP component (e.g., Federmeier & Kutas, 1999; S. Frank, Otten, Galli, & Vigliocco, 2013) have shown that lower probability of a linguistic item leads to higher processing cost.

Besides surprisal, entropy is also used as a metric that predicts the difficulty with which a linguistic item is perceived, hence, typically linking language models with experimental measures of effort in language processing, such as reading times (see Hale, 2016). We usually speak of entropy reduction, that is, the reduction of uncertainty about the occurrence of a certain unit. For instance, a restrictive verb, such as *spill*, reduces the entropy of the referent to a larger extent than a less restrictive verb, such as *order*. This is the case since selectional preferences of the verb *spill* reduce the set of potential target referents more than is the case with *order*. Many spillable objects can be ordered, while from the large group of objects that can be ordered, many cannot be spilled.

An additional notion relevant for our present work is that of Uniform Information Density (UID; Jaeger, 2010). It is suggested that the optimal distribution of information content in a linguistic stream is (a) near the channel capacity of the listener, i.e., not too boring and not too challenging to comprehend (see Genzel & Charniak, 2002), and (b) of roughly the same information density (no big differences between different linguistic units). In the present work, we hypothesized that both linguistic and visual context contribute to the distribution of information, and that visual cues allow for a more uniform distribution throughout a sentence.

The Visual World Paradigm (VWP) has been widely used to examine anticipatory eye-movements towards visually present entities that give insight into what is being activated during listening. Kamide, Altmann, and Haywood (2003) showed that verb information is utilized to anticipate upcoming arguments, and vice versa (in their Japanese study). Many specific questions have been examined: for instance, if it is particular words that people

predict, and thus their phonological features (Allopenna, Magnuson, & Tanenhaus, 1998), or whether specific visual features (Dahan & Tanenhaus, 2005) such as shapes of objects are being activated (Rommers, Meyer, Praamstra, & Huettig, 2013; Huettig & Altmann, 2007), as well as object affordances (Altmann & Kamide, 2007), or even objects that are semantically related, but not associated with the target (Huettig & Altmann, 2004). It has been shown that both available linguistic and visual information are combined and utilized to predict the upcoming linguistic content. In addition, the eye movement analysis that the paradigm supports showed active updating of the mental representation, based on both linguistic and visual information (see Altmann & Mirkovic, 2009; Huettig & Altmann, 2011).

1.2 Referential Gaze as a Visual Cue

The information gathered from interlocutors' gaze has proven to be an inseparable part of situated communication. Interlocutors spend a lot of time looking at each other (Argyle & Cook, 1976), the listener being the one who spends more time looking at the speaker (Kendon, 1967), while mutual gaze is utilized to coordinate turn-taking (Duncan, 1972; Kendon, 1967). Situated communication commonly includes visual cues, such as pointing, that are used to disambiguate and ground referring expressions (Bangerter, 2004). In their seminal work, Hanna and Brennan (2007) showed that a speaker's eye gaze can similarly be used to resolve temporary ambiguity, concluding that it is a conversationally based source of information used to quickly constrain reference resolution.

Listeners inspect objects they anticipate will be mentioned next (Altmann & Kamide, 1999), and additionally fixate the mentioned object 200–300 ms after the onset of the corresponding reference (e.g., Allopenna et al., 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Moreover, speakers also attend to the relevant object 800–1000 ms before referring to it (e.g., Griffin & Bock, 2000; Meyer et al., 1998). The speaker's gaze towards an object present in the shared visual context thus provides a cue to her focus of attention (Emery, 2000; Flom et al., 2007). This cue is then actively used to facilitate linguistic processing by helping to ground and disambiguate referring expressions (e.g., Hanna & Brennan, 2007). This also holds in communication with a robot speaker (Staudte & Crocker, 2011) or a virtual agent (Staudte et al., 2014), showing that people establish basic joint attention with robots as well, though this is dependent on belief in their agency (Meltzoff, Brooks, Shon, & Rao, 2010).

The facilitatory effect of the speaker referential gaze cue has been established in previous literature (e.g., Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013, 2014; Staudte & Crocker, 2011; Staudte et al., 2014). Importantly, though, the gaze effect on the processing

cost of the linguistic referent has been measured only offline, via task performance measures such as accuracy and reaction times. Only recently, Jachmann, Drenhaus, Staudte, and Crocker (2017) examined the effect of speaker referential gaze on the processing of the linguistic reference by considering event-related potentials (ERPs). Their results suggest that gaze led to anticipation of the upcoming noun, which resulted in integration difficulty when the cue was incongruent with the sentence. Moreover, they found evidence that gaze inspires anticipation of specific word forms.

A body of research has addressed the difference between gaze and other visual pointers like arrows. Friesen, Ristic, and Kingstone (2004) found that eyes trigger an initial reflexive attention shift to the cued location, which is not suppressed in contexts where the opposite direction is the relevant one. In contrast, they found that such a reflexive attention shift was avoidable with arrow cues. The authors concluded that the attention effect for eyes is more strongly reflexive than that for arrows (Ristic, Wright, & Kingstone, 2007). Staudte et al. (2014) initially found evidence that reverse gaze cuing disrupted comprehension, while reverse arrows facilitated it. However, once gaze was made as precise as arrows, no qualitative difference was found in their influence on comprehension. This finding led to the conclusion that a beneficial effect of object-oriented speaker gaze can potentially be replicated by any other visual cue. In addition, Böckler, Knoblich, and Sebanz (2011) examined gaze following and shared attention, and found that shared attention occurred both for schematic and for real faces (photographs of humans).

For the purpose of our present examination, we consider a schematic representation of the referential gaze cue as an ideal visual cue to employ. Since we focus on the way in which visual cues are incorporated with language, we chose gaze over arrows (or alternative visual cues) due to its natural coupling with language. Arrows indicate location, similar to a referential gaze cue, however, when presented in synchrony with language production, the location of the gaze cue is conceivably naturally associated with linguistic information (by the listener). In an experimental context where arrows are used, potentially an additional effort is needed to establish the meaning behind the use of such cues. This is not required with gaze, since it is readily related to speaker intention, and thus tightly coupled with language production. In addition, we aimed to create a well-controlled experimental setting for the relatively novel pupillary measurement we used (presented in detail in Section 2.3.2). Thus, we decided to use an abstract schematic representation of the gaze cue, rather than realistic, but visually highly complex imagery. Finally, we consider our gaze cue as an abstract and strategic visual cue, not being concerned with the difference between an actual gaze and a different visual pointer, but understanding that our findings reflect the state of affairs for visual cues in general, which typically carry meaning or serve a communicative purpose.

1.3 Research Program and Addressed Questions

We focused our examination on three general research questions. After quantifying the effect of the visual context on language processing, we further looked into the role of visual cues. We manipulated the referential gaze cue in various ways, in order to assess the mechanisms of its inclusion in the interpretation of a sentence. Here, we will present the general overarching questions in more detail.

1. Can the effect of visual context on linguistic processing be quantified by measuring the induced cognitive load?

Being constantly present and available for incorporating into the sentence representation at all stages of sentence unfolding, visual context potentially leads to an overall increase in load during sentence processing. Importantly, however, the visual information becomes more relevant at certain points, when linguistic information is introduced which can be related to what is seen. Conceivably, at such points in sentence processing, the information from the visual modality may lead to an increase in cognitive load. Consequently, however, such an increase is expected to lead to a reduction of the cogntive load required for reference processing, since the available relevant information has been considered incrementally. This question is addressed in Chapter 3.

2. Does a referential visual cue affect the cognitive load required for processing the corresponding linguistic reference?

We have seen evidence that visual cues facilitate language processing, such that listeners' performance on post-trial tasks is aided. But how does this facilitation occur? Does the gaze cue together with the linguistic context inspire anticipations for a certain referent to be mentioned? If so, this should potentially lead to the reduction of immediate cognitive load required for processing the referring expression (since this information was anticipated). If it is the case that the cognitive load on the linguistic reference gets reduced due to the existence of a visual cue (cuing the mentioned referent), we would expect that such a reduction in load should be preceded by an increase in load some time earlier during the sentence (UID, explained in Chapter 2). Specifically, an increased processing effort is expected at the point of perceiving the visual cue and considering the cued object. Hence, we anticipated higher load to be induced on the gaze cue, resulting in reduced load at the point of the reference.

We addressed this question from two different perspectives. Firstly, we quantified the role of visual cues with differing probabilities of occurrence given the linguistic and visual context (how likely it is that an object will be relevant). This work manipulates the role of

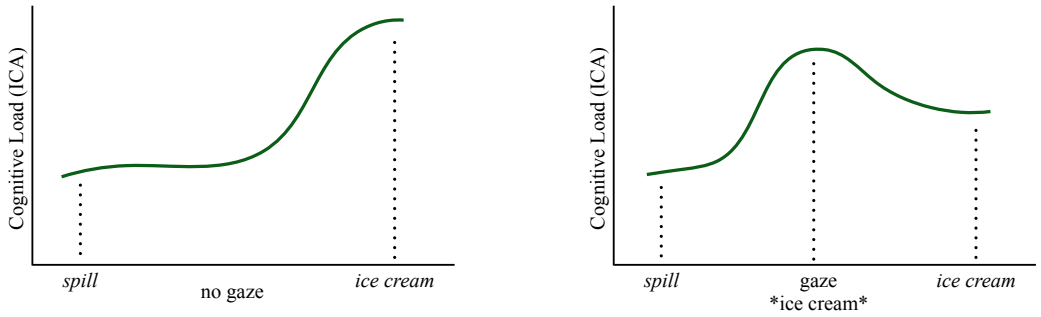
the visual cue in a *qualitative* manner (presented in Chapter 4). Secondly, we examined visual cues that reduce the uncertainty about the referent in terms of reducing the number of potential target referents. In other words, we manipulated the role of the visual cue in a *quantitative* manner (presented in Chapter 5).

3. Is a referring expression resolved at the point of hearing the linguistic reference, or can a visual cue take the role of referent identification on a par with a noun?

By swapping the roles of the noun and the visual cue, we aimed at better understanding the relationship between them and shed more light on how the visual cues are integrated. If the referent noun highlights a certain set of objects without actually identifying the target, how does the processing of the referent noun compare to the processing of the visual cue in the same role (cuing a group of objects)? And conversely, how do the noun and the visual cue compare in the role of being the only piece of information that unambiguously identifies the target referent? The work on how differently the two cues are perceived in the same role is presented in Chapter 5.

Previous research relied on tracking eye movements, as an online measure of visual attention, and post-trial task performance, as an offline indication of processing difficulty. We extend these findings by measuring processing effort online. The pupillary measure we employ, the Index of Cognitive Activity, allows for (a) inspection of free anticipatory eye movements, and (b) an online record of cognitive load at various points of interest throughout the sentence, where relevant linguistic and visual stimuli are introduced. We assessed how anticipatory eye movements connect to immediate processing effort, that is, whether anticipating objects (due to gaze-following) facilitates the processing of their reference. Unlike previous research that has examined the benefit of gaze offline, by measuring post-trial task performance, we approach the question more directly by reporting cognitive load induced when processing the linguistic reference. Moreover, we are interested in examining whether perceiving the gaze cue itself, that is, processing the cued referent, actually induces processing effort on a par with the cost of linguistic processing.

The present work assumes a direct link between cognitive load, as measured by the Index of Cognitive Activity (presented in Section 2.3.2), and the surprisal of the referent noun. Our predictions are in line with the information-theoretic notions of probability and surprisal, in that we expect that the higher the probability of an item, given its previous linguistic and visual context, the lower the surprisal on the actual referent will be. Thus, we expect lower cognitive load on the referent noun when it is in accordance with the predictions previously made based on linguistic and visual information. To that end, we manipulated the surprisal of a linguistic unit, by varying the amount of information being transmitted in terms of (a)



(a) the distribution of cognitive load without the gaze cue

(b) cognitive load distribution with the gaze cuing the target

Figure 1.1 Expected distribution of cognitive load among the visual and linguistic cues (assuming that gaze is typically informative).

probability (likelihood) of a unit, which made the less probable unit carry more information (since it was less predictable) and thus induce higher surprisal effects; and in terms of (b) the sheer number of competitors, hence manipulating the amount of information quantitatively; namely, more potential targets equal more information and thus, higher surprisal.

Moreover, inspired by the UID hypothesis, we expect that the information provided through the gaze cue affects the immediate cognitive load it induces. An informative gaze cue that leads to the reduction of load on the referent noun is thus expected to induce higher cognitive load itself. In this way, the processing of the same information content would be distributed between the visual cue and the noun. This idea is illustrated in Fig. 1.1. We assume that the noun *ice cream*, in a sentence like *The man spills the ice cream*, would induce relatively high cognitive load, since in order to be a logical argument of the verb *spill*, ice cream would have to be melted, which is not its typical state. However, if the speaker were to look at the ice cream, an object present in the shared visual context, prior to uttering the reference, both the gaze cue and the referent noun would essentially introduce the same piece of information. Thus, it would be reasonable to expect, first, a reduction of effort on the reference, and second, that it is preceded by an increase of the load at the gaze cue. In sum, we expect a re-distribution of the effort between the visual and linguistic cues.

1.4 Thesis Overview

This thesis investigates the integration of visual cues into the sentence interpretation during situated language processing. We report the results of eight eye tracking experiments that shed more light on the integration of the visual context, and more specifically, the referential gaze cue, into the sentence interpretation. Our results show the impact of the multi-modal

context on creating anticipations for the upcoming referring expressions, and create a basis for further investigations into the role of visual cues in situated language processing.

In Chapter 2 we present a review of the relevant theory, as well as the motivation for the choice of the present measurements and employed methodology. First, we give an overview of the current literature on the role of prediction and anticipation in language processing. Specifically, we consider in more detail the information-theoretic notions of surprisal and entropy, as well as the Uniform Information Density hypothesis. Then, we focus on situated language processing and anticipation in a visual context. Second, we justify the gaze cue as our visual cue of choice by providing an overview of the relevant literature on gaze cue perception. Finally, due to the not-yet-widely-used measurement of cognitive load employed in our experimental work (the Index of Cognitive Activity; Marshall, 2000), we give an overview of the relevant literature on pupillometry and give the reasoning for the specific methodology chosen for the present work.

The following three chapters present the experimental work conducted towards the conclusion of this dissertation.

In Chapter 3 we present the experimental work that lays the basis for our further investigations. Three experiments from this chapter present a basic examination of the present methodology and experimental measures. We confirmed that the Index of Cognitive Activity is a reliable measure of cognitive load in the visual world paradigm. More importantly, we quantified the role of the visual context in reference processing, by examining the cognitive load induced during sentence processing, and considering indications of target anticipation with and without a visual context. This work was conducted in collaboration with Christine Ankener and is included in a recently published journal article (Ankener, Sekicki, & Staudte, 2018).

Chapter 4 introduces the gaze cue in the given paradigm and examines the basic ideas of qualitative probability and surprisal in situated language processing. In addition, the experiments from this chapter examine the basic properties of the visual cue. We find evidence that the gaze cue is immediately and effortlessly incorporated in the current sentence interpretation, facilitating reference processing, its consideration becoming costly only when it cannot be incorporated with the linguistic context. This chapter addresses the question of gaze cue perception from the "nominal" perspective: that is, the cue is identifying one specific referent object, as a noun typically does. What we manipulated is the referent probability, given the previous linguistic and visual context. This work has been recently published in the Cognitive Science Journal (Sekicki & Staudte, 2018). In addition, Experiment 4 and Experiment 5 from this chapter were initially presented at the 39th Annual Conference of the

Cognitive Science Society, and have been published in the proceedings of this conference (Sekicki & Staudte, 2017).

Chapter 5 presents two experiments that manipulate referent probabilities and the resulting surprisal in a quantitative manner. By manipulating the number of potential target referents, we compared how the linguistic and visual cues behave in the role of reducing the uncertainty about the referent. This chapter examines the gaze cue in a "verb-like role" of constraining the set of potential target referents numerically to a smaller group. We also go a step further in forcing the noun into the role of constraining the set of potential targets, but not identifying a single specific target referent. This is done in order to contrast the gaze cue and the noun in the same role.

Finally, Chapter 6 concludes the present work. We first summarize the results obtained from our experimental work. Next, we interpret them in the light of the initially raised research questions and position them in the context of the current understanding of the topic. Finally, we mention the open questions that emerge from the present work together with ideas for future investigation, before concluding the thesis with our final remarks.

Chapter 2

Sentence Processing and Visual Cues

In this chapter we present the relevant theoretical background. In Section 2.1.1, we outline the current understanding of the predictive mechanisms behind human language processing. We start by introducing the information-theoretic notions widely used in present psycholinguistic research, namely, the notions of surprisal, entropy and entropy reduction, as well as the Uniform Information Density hypothesis. These notions are being used to explain both the mechanisms behind the comprehension of the linguistic material, as well as language production – that is, the choice of a certain linguistic expression over another in a given communication situation. Moreover, we include a short overview of the work conducted in the visual world paradigm (VWP), which examined the limits of prediction in language processing given a visual context.

Section 2.2 presents the relevant literature on the perception of visual cues. We present the most studied visual cues: directional cues – speaker gaze and arrows. We focus on the examination of referential speaker gaze (cuing the location of the target referent), and its use and perception during linguistic communication.

Finally, in Section 2.3 we motivate the choice of the measurements used for our experimental investigations. Due to the still largely unknown and not widely used measure of cognitive load we employed in our experimental work, we present the experimental design of our studies in more detail, with the aim of justifying the design and the measurements used. We briefly comment on the VWP and the assessment of eye movements. Moreover, the use of pupillometric measures as indicators of cognitive load is introduced, with the focus on the Index of Cognitive Activity (ICA), as the current measure of choice. We justify why the ICA was preferred to the other existing measures of cognitive load for our current experimental examinations.

2.1 Prediction in Language Processing

Much current work in the field of psycholinguistics is inspired by the idea that anticipation of what is about to happen is a basic mechanism of human cognition, and should thus also be an integral part of language processing (e.g., Clark, 2013). While some emphasize the futility of prediction in a system where sentences can continue in infinitely many different ways (Jackendoff, 2002), others see prediction as an integral and inevitable mechanism which enables quick and incremental comprehension (e.g., Altmann & Mirkovic, 2009; Pickering & Garrod, 2013).

Over the last decade, experiments observed participants' anticipatory behavior, namely, fixating objects that fit with the previously heard linguistic material but have not been mentioned yet. Also, facilitation was measured for processing of linguistic material that is more likely to follow a certain context, in comparison to the less likely option.

The most influential piece of evidence that individuals can utilize linguistic input to pre-activate representations of upcoming concrete words stems from DeLong, Urbach, and Kutas (2005)'s EEG experiment that provided evidence for highly specific preactivation on the level of a word's phonological form. They presented participants with sentence contexts such as *The day was breezy so the boy went outside to fly ...* resolved with either a highly predictable referent *a kite*, or the less expected, but equally possible, *an airplane*. Using semantically identical articles rather than content words, DeLong et al. (2005) ensured that the potential effect could not be due to the difficulty of interpreting the article itself. The brain's response to the article differed in a graded fashion relative to the contextual constraint, not only at the noun, but already at the article (N400 effect). The authors argued that listeners rapidly and incrementally integrate incoming words into their sentence representations, and form probabilistic predictions of which specific word will come next.

However, evidence was found on the limits of engaging in anticipatory behavior related, for instance, to literacy (Mishra, Singh, Pandey, & Huettig, 2012), language processing in a foreign language (Martin et al., 2013), and rate of stimulus presentation (Wlotko & Federmeier, 2015). Recently, a joint attempt has been made at nine different labs to directly replicate the findings of the seminal DeLong et al. (2005) study. Nieuwland et al. (2018) found a statistically significant effect of cloze on noun-elicited N400 activity, but, critically, no significant effect of cloze on article-elicited N400 activity. Their exploratory Bayesian mixed-effects model analyses suggested that, while there is some evidence that the true population-level effect may be in the direction reported by DeLong et al. (2005), the effect is likely too small to be meaningfully observed without very large sample sizes.

Such findings inspired a debate as to how ubiquitous, and hence how necessary, prediction actually is for natural language processing. Two recent reviews of the present literature

emphasize, on the one hand, the importance of prediction for language processing (Kuperberg & Jaeger, 2016), while on the other hand, Huettig and Mani (2016) argue that one should be cautious with sweeping conclusions, noting all the evidence against prediction as a "one-size-fits-all" mechanism behind language processing, and emphasizing its conditional occurrence.

2.1.1 Information-Theoretic Notions and Correlation with Processing Effort

If communication can optimally be understood in terms of information processing, then the amount of information that a linguistic unit conveys should be predictive of the cognitive effort required for its processing (Hale, 2006; Levy, 2008). The amount of information conveyed by a unit can be computed from probabilistic models of the language, from the values gathered from a cloze task (Taylor, 1953), or alternatively, from large corpora. The amount of cognitive load required for processing a particular linguistic unit has typically been examined in terms of reading times, where more informative words have proven to take longer to read (e.g., S. Frank, 2013; Smith & Levy, 2013). Before being correlated with the empirical values of processing effort, the amount of information a linguistic unit conveys is calculated in terms of information-theoretic notions such as surprisal and entropy.

As noted by S. Frank, Otten, Galli, and Vigliocco (2015), **surprisal** is understood as a measure of the extent to which the occurrence of an item was unexpected. Put more formally, the surprisal of the outcome of a random variable is defined as the negative logarithm of the outcome's probability given the previous linguistic context. Considering the probability of the next word w_{t+1} in a sentence, we use the following formula:

$$surprisal(w_{t+1}) = -\log P(w_{t+1}|w_{1...t}) \tag{2.1}$$

It follows that the surprisal of a word will be high when this word has low conditional probability, and vice versa. Examining the theoretical arguments that surprisal, as a measure of cognitive effort, can be used as a predictor of word reading times (Hale, 2001; Levy, 2008; Smith & Levy, 2008, 2013), many studies have found correlations between a word's surprisal and the time required for its reading. In other words, people have been found to read more slowly on words with higher surprisal value (e.g., Demberg & Keller, 2008; Fernandez Monsalve, Frank, & Vigliocco, 2012; Fossum & Levy, 2012; Smith & Levy, 2013). In addition, S. Frank et al. (2015) found a strong relation between the surprisal of a word and the amplitude of the N400 component induced while reading that word.

Moreover, **entropy** is another information-theoretic notion shown to predict the processing difficulty of a linguistic unit (e.g., S. Frank, 2010; S. Frank et al., 2013). As a measure of uncertainty about the outcome of a random variable (Shannon, 1948), the entropy of the remainder of the sentence, after having encountered $w_{1...t}$, is formalized as follows (S. Frank et al., 2015):

$$H(W_{t+1...k}) = -\sum_{w_{t+1...k}} P(w_{t+1...k}|w_{1...t}) \log P(w_{t+1...k}|w_{1...t}) \tag{2.2}$$

$W_{t+1...k}$ is a random variable with particular sentence continuations $w_{t+1...k}$ as its possible outcomes. Upon encountering the word w_{t+1}, the uncertainty about the rest of the sentence is typically decreased. Hence, $H(W_{t+2...k})$ is usually lower than $H(W_{t+1...k})$. This difference is called **entropy reduction** (ΔH). Entropy reduction quantifies the amount of ambiguity that is resolved by the current word, assuming that disambiguation reduces the number of possible continuations of the sentence (S. Frank et al., 2015). For the formalization of entropy reduction relying on syntactic structures, rather than word strings, please see Hale (2003, 2006).

In a magnetoencephalographic (MEG) study, Maess, Mamashli, Obleser, Helle, and Friederici (2016)'s participants listened to simple sentences with two types of verbs, one highly predictive of the following noun, and the other less predictive of the same noun. A reduction of the N400 response was found on the highly predicted noun and, interestingly, the opposite pattern was found for the preceding verb. Highly predictive verbs induced a bigger N400 effect. The authors argue that the effect measured on the highly predictive verb reflects the pre-activation of the expected nouns.

Finally, an additional information-theoretic notion relevant for our present work is that of **Uniform Information Density** (UID):

> Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (information density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (ceteris paribus). (Jaeger, 2010, p. 3)

In other words, language users are hypothesized to prefer distributing the information uniformly over a message, and within the bounds of the channel capacity. Hence, UID maximizes information transmission, while minimizing comprehender difficulty. The assumption is that UID inspires production choices, and is relevant for determining encoding at all levels (speech, lexical, syntactic reductions). Consequently, highly informative "units" are introduced during longer time intervals, reflected in longer production, or choice of full form

(longer) words or expressions. In contrast, highly predictable (less informative or uninformative) units are pronounced more quickly or by using a shorter form, or, if totally predictable, they can be completely elided (see Jaeger, 2006; Levy & Jaeger, 2007). The assumed goal is keeping the amount of information transmitted per time as uniform as possible.

A. F. Frank and Jaeger (2008) find evidence that the use of contractions (I'm vs. I am) in American English is influenced by their information density. Moreover, Jaeger (2010) examined UID at the syntactic level with the example of optional *that*-omission in English relative clauses. Related to this, Aylett and Turk (2004) posit their Smooth Signal Redundancy Hypothesis, suggesting that (a) there is an inverse relationship between language redundancy and acoustic redundancy (as manifested by syllable duration); and (b) prosodic prominence smooths signal redundancy by controlling syllabic duration. They indeed find evidence that the expected material is articulated with shorter duration.

2.1.2 Prediction in the Visual Context

The examination of predictive language understanding in the VWP was initiated by Altmann and colleagues. Their work has demonstrated anticipatory eye movements and shown that visual attention reflects the dynamic changes in mental representations of the events that are currently referred to. Listeners simultaneously utilize multi-modal information in order to disambiguate the current and anticipate the upcoming linguistic material. The examined eye movements are mediated by language, but reflect the constant updates of the mental representation based on both visual and linguistic information.

Altmann and Kamide (1999)'s seminal work examined the role of verb constraint in forming anticipations for the upcoming linguistic material. Their participants were presented with visual scenes containing, for instance, a boy and a group of objects, namely a cake and some toys. While observing the scene the participants would hear either *The boy will eat the cake*, or *The boy will move the cake*. Fixations of the cake were detected significantly earlier in the *eat* condition, that is, before the referent noun was uttered. Hence, Altmann and Kamide (1999) found that verb selectional preferences can be rapidly utilized to reduce the referential uncertainty, the anticipated referent being the most fixated one.

Subsequently, Kamide, Altmann, and Haywood (2003) examined the influence of the information conveyed by the combination of subject and verb on the anticipations created for the target referent. Participants were presented with scenes including two persons and two objects. A scene depicting a girl and an adult man, together with a carousel and a motorbike, was coupled with the auditory presentation of either *The man will ride ...* or *The girl will ride ....* More fixations were found to the motorbike in the first condition, and more fixations to

the carousel in the second. These results show that the information provided by the subject and the verb together can further constrain anticipatory eye movements.

A comparative study with speakers of German and English showed that the anticipatory eye movements to a potential second argument of the sentence depended on the case marking of the first mentioned argument (Kamide, Scheepers, & Altmann, 2003). The authors concluded that the syntactic information (first argument case-marking) is promptly integrated with the verb's semantic constraints, and enables the creation of predictions for the upcoming argument.

Moreover, Altmann and Kamide (2007) showed that verb tense information can also be utilized to create anticipation for the upcoming referent. Presented with an empty wine glass and a full glass of beer, upon hearing *The man has drunk ...* the participants fixated more on the empty glass, while the full glass was fixated more upon hearing *The man will drink....*

Extensive work has followed to examine the limits of prediction in a visual context, in terms of how detailed the created anticipation actually is, and under what circumstances it becomes ineffective or costly to predict. Allopenna et al. (1998) examined the activation of phonological features by presenting the participants with scenes including a named target (*beaker*), as well as a cohort competitor beginning with the same onset and vowel as the target (*beetle*), a rhyme competitor (*speaker*) and an unrelated object *carriage*. They found clear evidence for the activation of cohort competitors.

Huettig and Altmann (2004) examined whether semantic priming holds across items that are semantically but not associatively related, that is, related only by category. The visual scenes included either the target referent (*piano*) together with three unrelated distractors; or a competitor object (*trumpet*) with three unrelated distractors; or both the target and competitor referents were illustrated together with two unrelated distractors. The results show that during the processing of lexical information, visual attention is already directed towards conceptually related (but non-associated) objects, even if they represent a mismatch on other dimensions (e.g. shape, colour, conceptual detail). The authors conclude that language-mediated eye movements are a sensitive measure of overlap between the conceptual information conveyed by individual spoken words and the conceptual knowledge associated with visual objects. Moreover, spontaneous fixations can be driven by partial semantic overlap between a word and a visual object.

Finally, there is evidence that information about an object's shape is activated during the processing of its linguistic reference (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007). Rommers et al. (2013) examined the predictive activation of object shape. In the (target-absent version of) VWP they found more anticipatory eye movements to an object similar in shape to the target than to the unrelated distractors. Moreover, in an EEG study

without the visual stimuli, participants heard sentences including the target referent (*In 1969 Neil Armstrong was the first man to set foot on the moon*), its shape competitor (*tomato*), or an unrelated distractor (*rice*). They found that the N400 amplitude was significantly weaker for the semantic violation including the referent of similar shape than the unrelated distractor. The authors argue that predictive language processing is not limited to functional semantic information, but can also involve the activation of visual representations.

2.2 Visual Cues

2.2.1 The Gaze Cue

Eye gaze is a constantly available source of information that is frequently used to understand another person's intentions and emotions (see Baron-Cohen, Wheelwright, & Jolliffe, 1997). It is a very expressive social cue, commonly interpreted as such even without being intentionally used as a communicative tool. Also, it is unique as a communication cue because it is difficult to suppress and fake its expression, unlike a smile, for instance. Some aspects of its use and interpretation are culture-dependent, such as establishing mutual gaze, reading emotions from one's eyes, and the use of mutual gaze for turn-taking in communicative contexts (see e.g. LaFrance & Mayo, 1976; Greenbaum, 1985; Schofield, Parke, Castañeda, & Coltrane, 2008).

A study comparing American and Japanese participants found that in an attempt to interpret emotions, the Americans focused on the person's mouth, while the Japanese focused more on the eyes (Yuki, Maddux, & Masuda, 2007). The authors argue that this is due to cultural differences, namely, whether emotions are more readily expressed verbally, or ought to be suppressed and are understood only from subtle spontaneous involuntary signals, such as those made by the eyes.

Moreover, mutual gaze depends on the relationship between the interlocutors, in terms of intimacy and power relations. Closer physical proximity has been shown to accompany less mutual gaze, which seems to reflect higher intimacy (Argyle & Dean, 1965). Also, power relations are an important predictor of the amount of mutual gaze. The more people were looking at their interlocutor while speaking and less while listening, the higher the social power they were seen as having (Dovidio & Ellyson, 1982).

The use of mutual gaze for turn-taking in face-to-face communication is commonly such that the speaker tends to look away when planning and uttering longer utterances, but reestablishes eye-contact when approaching the end of her turn, and wishing to give the listener the opportunity to speak (Kendon, 1967). Planning an utterance requires effort;

hence, by looking away the speaker ignores the current informational input and concentrates on the output she is trying to produce. Therefore, listeners are more likely to start talking after a mutual gaze with the speaker. However, other cues are also used to establish the appropriate time to start talking, such as intonation, voice pitch and gestures (Duncan, 1972).

Importantly for our present purpose, eye-gaze is often synchronized with language production. While speaking about objects present in their visual context, speakers tend to look at a relevant object before mentioning it. Object-directed, that is, referential speaker gaze usually appears 800–1000 ms prior to uttering the linguistic reference (Meyer et al., 1998; Griffin & Bock, 2000). Meyer et al. (1998) found that people look longer at objects that require low frequency words to be labeled. The authors argued that objects are inspected not only until they are identified, but until their phonological form is retrieved. Also, objects of low codability (having multiple similarly dominant labels) are found to be inspected longer before being named (Griffin, 2001).

Listeners tend to reflexively attend to the direction of the speaker's visual attention. The understanding, from the side of the listener, that looking at an object means seeing it, and is thus directly related to visual attention, can lead to *joint attention* – the listener actively attending to the same thing that the speaker is attending to. If aware of this, the speaker may choose to use their eye gaze in order to subtly point towards a certain direction, as a communicative act in relation to the listener – *shared attention* (Emery, 2000). The ability to engage in joint and shared attention is argued to have evolutionary relevance (Baron-Cohen, Baldwin, & Crowson, 1995; Flom et al., 2007; Emery, 2000). If we understand that looking at something means perceiving it, what is relevant for our "neighbor" is potentially of importance to us too, be it food, or danger of some kind.

Moreover, listeners tend to use the direction of speaker gaze to identify the target referent prior to its explicit mention, thereby facilitating reference resolution. In their seminal study, Hanna and Brennan (2007) found that listeners utilized the direction of speaker gaze even in a condition of mirrored visual displays, where the speaker's gaze cue towards left, for them, actually meant that the relevant object was to be found on their right side. In general, findings about gaze-following suggest that listeners tend to have an initial reflexive response, after which the intention is being read from the gaze (Friesen & Kingstone, 1998; Driver et al., 1999; Langton & Bruce, 1999).

A group of studies have shown that gaze-following aids sentence processing. Aiming for a fully controlled environment, Staudte and colleagues conducted a series of experiments where they utilized robots and virtual agents as speaker figures and examined how listeners interpret referential gaze coupled with speech (Staudte & Crocker, 2011; Staudte et al., 2014). Listeners engaged in gaze-following and used the direction of robot gaze to anticipate the

upcoming referent. As measured by response times at sentence validation, congruent robot gaze facilitated comprehension (relative to neutral gaze), while incongruent gaze disrupted it (Staudte & Crocker, 2011).

Moreover, Knoeferle and Kreysa (2012) found evidence that gaze-following is robust to observing the speaker from an angle (rather than frontally). The facilitatory effect of gaze varied depending on the syntactic structure and thematic role relations in the sentence (SVO vs. SOV), and it extended to the post-trial comprehension processes. The authors conclude that the effects of speaker gaze-following contribute to visual attention and comprehension processes and should be accommodated by accounts of situated language processing.

Finally, complementing the mentioned findings from behavioral studies, based on the observation that gaze-following facilitates (or disrupts) listeners' performance on a subsequent task, Jachmann et al. (2017) examined the effect of gaze-following more directly. In an ERP study, they manipulated gaze congruency and found an increased N400 for the incongruent gaze, as well as for the neutral (uninformative) gaze. The incongruent gaze condition also showed a late (sustained) positivity that reflects the need to update the previously created situational model. This study provides further evidence that listeners exploit speaker referential gaze in order to form expectations for the upcoming referent.

2.2.2 Reflexive Cue Following

Much evidence has been gathered establishing eye gaze as a means to communicate visual attention (Frischen, Bayliss and Tipper, 2007), as well as socially relevant information, like focus of interest, thoughts and intentions (e.g. Baron-Cohen et al., 1997). Moreover, gaze-following also enables language acquisition, and the development of theory of mind (Tomasello, 1995). Inspired by such findings, it has been argued that gaze-following might represent a unique attentional process. Even though many studies have contrasted the gaze cue with other visual directional cues, such as arrows, conflicting results have been collected, leading some researchers to propose that gaze attentional effects are at least partially driven by a domain-general attentional processing (e.g. Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014).

Friesen et al. (2004) found gaze-following to be more resistant to voluntary control than the following of directional arrow cues. In the *counterpredictive cuing paradigm* (target more likely to appear in the direction opposite the cued one), better performance was found for the cued location when gaze was the examined cue, while in the case of an arrow, the cued location was not considered. However, Tipples (2008) found that both cues inspire reflexive attention shifts. Moreover, Ricciardelli, Bricolo, Aglioti, and Chelazzi (2002) found differences in overt orienting between gaze and arrow, while Kuhn and colleagues found only

small or no differences (Kuhn & Benson, 2007; Kuhn & Kingstone, 2009). Bayliss, Paul, Cannon, and Tipper (2006) found that objects inspected by other people are perceived as more likable than others, an effect which was not found for arrows. Similarly, an improvement in memory accuracy for the cued information was found with gaze, but not with arrows (Dodd, Weiss, McDonnell, Sarwal, & Kingstone, 2012; Gregory & Jackson, 2017).

Importantly to our present cause, a series of experiments contrasted referential gaze and arrows as visual cues used simultaneously with speech (Staudte et al., 2014). The authors aimed at understanding whether the influence of referential speaker gaze on sentence processing goes beyond visual cuing. They found that when gaze is made as precise as arrows, no qualitative difference is to be found between the two cues' influence on the comprehension of linguistic material.

Finally, Böckler et al. (2011) contrasted gaze from schematic faces vs. gaze from photographs of humans. They set out to examine whether observing two people engaging in eye contact (or not) modulates the observer's following of their synchronous gaze cue. Evidence was found for the facilitatory effects of gaze following on participants' reaction times only for the gaze cues that followed mutual gaze between the two faces shown. Moreover, and more interestingly for our present cause, in their study Böckler et al. (2011) also compared scenes where actual photographs of human faces were used with schematic representations of faces. The abovementioned effect was present in both representation styles, showing that people engage in gaze-following even with more symbolic representations of the eyes.

2.3 Measurements

The experiments aimed towards answering our research questions were conducted in the visual world paradigm. They assessed both participants' visual attention and the accompanying processing effort. We were interested in, first, examining anticipatory considerations of the potential target referent(s). To this end, we looked into anticipatory eye movements that showed which object(s) was considered most before the linguistic reference was actually uttered. Second, we wanted to track shifts in visual attention which could happen due to visual or linguistic cues to certain object(s). Finally, our aim was to enrich the examination of the visual attention towards the relevant objects, by measuring the cognitive load induced at various points of anticipation and integration of the relevant material into the sentence representation. Hence, we employed a measure of pupillary activity that is robust to free eye movements, reporting relevant moment-to-moment changes in pupil diameter.

The effort required for processing a piece of information can be examined by using different measures with different levels of precision and relationship to the actual brain

activity. If we imagine it as a spectrum of directness and measuring granularity, temporally dependent measures such as reading, or reaction times, would be on the indirect and coarse-grained end of the spectrum, while the event-related potentials (ERPs), tapping directly into the brain's electrical activity, would be far to the other end of direct and very fine-grained measures. For the purposes of our present experimental work, we have chosen a measure that is positioned between those two on our spectrum, namely, a pupillary measure of processing effort.

Processing difficulty can translate into temporal effects, where a longer time spent to read a word or to react to a task is indicative of the greater effort required to process the previous input in order to be able to react. Such temporal measurements are indirect, since rather than measuring the actual processing effort, we are measuring its consequence on one's behavior. Longer reading times at a word suggest that the reader is not yet ready to continue acquiring new information, and indirectly suggest that she is having difficulties processing what was previously presented. Similarly, longer time needed to react to a task suggests that she does not have an answer ready, and is exerting more effort to find it. Importantly, the link between processing effort and the temporal effect on reading or question answering is indirect, and very coarse-grained.

Electroencephalography (EEG) allows for a direct inspection of the brain's response to a stimulus, even when there is no behavioral change. What is tapped into is the electrical activity in the brain, as it happens, directly during the perception of a stimulus. Hence, different ERPs are indicative of the different stages in processing that one goes through, measurable for smaller linguistic units.

The examination of pupillary response falls between the two ends of the spectrum. In comparison to a specific ERP component, pupillometry provides an indirect, more global measure of brain function. Pupils dilate in response to higher processing effort. Since this change in size is a consequence of secretion of a neurotransmitter, rather than a requirement for cognitive processing, pupillometry is an indirect measure of brain function. Higher processing effort leads to higher secretion of norepinephrine, which consequently results in pupil size fluctuations. This being said, changes in size of the pupil are not a temporal measurement, but a qualitative one, since the larger dilation indicates higher underlying effort. Although allowing for the assessment of the load required immediately for a specific stimulus, it does not give an insight into which step in the interpretation process is causing the processing difficulty. (Just & Carpenter, 1993, p. 312) noted that "the pupillary response is only a correlate of cognitive intensity, hence the marker is indirect and not causally linked". As Beatty and Lucero-Wagoner (2000) argue, this is not a problem. Rather, pupillary measures should be seen as a "reporter variable for human cognitive processes", analogous

to reporter genes that are used in molecular biology to study the cell (for an illustration of the argument see Beatty & Lucero-Wagoner, 2000, p. 143). The authors emphasize the correlation, and argue that the lack of causation is irrelevant, encouraging the use of pupillometry in the examination of cognitive processes.

With the development of new methods of analyzing pupilometric data, pupillometry has developed a great advantage over more direct measures, allowing for the measure of cognitive load to be combined with the assessment of unconstrained visual attention. The eyes being typically directed towards the object of one's thoughts, combined with the information about the immediately induced cognitive effort, the direction of one's gaze informs us which piece of information directly contributed to the processing difficulty, or rather, facilitation. The Index of Cognitive Activity, a pupillometric measure we employed, allows for just that – the examination of participants' anticipatory eye movements in combination with the immediately induced cognitive effort as indexed by the pupillary jitter. We will explain this measurement in more detail in the remainder of this section.

In what follows is a detailed description of the measures employed, presenting the reasoning behind our choice of measurement combination, and highlighting the advantages of their use.

2.3.1 Visual Attention in the Visual World Paradigm

The beginnings of the visual world paradigm (VWP) date back to the work of Cooper (1974). However, it was not until Tanenhaus et al. (1995) and Allopenna et al. (1998) that the paradigm attained wide popularity. What allows for the success of this paradigm is that people tend to look at relevant objects, even when they are not supposed to interact with them (e.g. in terms of clicking on or moving them). Listeners' visual attention is driven by what they hear (Cooper, 1974; Tanenhaus et al., 1995; Altmann & Kamide, 1999; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Knoeferle & Crocker, 2006). This is the case, presumably, as an attempt is made to relate linguistic and visual information (Altmann & Kamide, 2007), since the auditory and visual modalities tend to provide complementary information. As proposed by Altmann and Kamide (2007), the linguistic input increases the activation of the mental representation of an object, resulting in an increased likelihood of a saccadic eye movement towards the object's location.

Typically, a trial in a comprehension visual world study includes participants hearing an utterance while looking at a display and having their eye movements recorded for later analyses. The displays include objects mentioned in the utterance, as well as competitor and distractor objects that are not mentioned. What is usually of relevance is the manner in which participants interact with the scene before and after hearing relevant linguistic input. The

visual display is presented simultaneously with the linguistic stimulus, or with a slight delay to allow for familiarization with the scene. Linguistic stimuli may include an instruction to complete a task, or simply a comment, description, or a statement related to the visual display. Without a specific task, there is no need for meta-linguistic judgments, which potentially affect the way in which speech is processed. As noted by Huettig, Rommers, and Meyer (2011), such "look and listen" tasks allow for the assessment of more general effects of language-vision interaction independent of a specific task.

Finally, the paradigm is used to assess how likely participants are to look at certain areas of interest (AoI) at different relevant points in time, that is, the fixation proportions on AoIs. For a detailed review and assessment of the paradigm, please refer to Huettig et al. (2011).

2.3.2 Pupillometry as a Measure of Cognitive Load

The Pupillary System

Light enters our eyes through the circular opening of the iris called the pupil. The word *pupil* originates from the Latin *pupilla* – a little doll, since it is our own reflection that we can observe in another person's eyes (Hakerem, 1967).

As noted by Beatty and Lucero-Wagoner (2000), pupil size is determined by the activity of two opposing muscle groups within the iris. The *sphincter* (circular) muscles and the *dilator pupillae* (radial muscles) are both controlled by the autonomic nervous system, and are responsible for the constriction or the opening of the iris, respectively. In other words, the contraction of the circular muscles constricts the pupil, while the activity of radial muscles makes the pupil dilate. Dilation is attributed to the activation of the sympathetic system, while parasympathetic mechanism is inhibitory (Sirois & Brisson, 2014). Pupil dilation can happen as a consequence of changes in light, where the pupillary light reflex regulates the amount of light entering the eye. Moreover, dilation also occurs due to sensory, mental and emotional events. Different causes of dilation employ different activation and inhibition processes, the dilation due to cognitive load being shorter and more abrupt than that caused by the light reflex.

Light and accommodation reflexes that optimize vision (such as focusing objects of varying distance) are the primary cause of changes in pupil size. Bright light leads to pupil constriction, while the pupil relaxes and dilates in dim light. It takes approximately 200 ms for such a reaction to occur (Lowenstein & Loewenfeld, 1962). In humans, the variation in pupil size spans from 1.5 to 9 mm. In dim light or darkness, pupil can reach an average size of about 7 mm with a standard deviation (from this average) of about 0.9 mm (MacLachlan & Howland, 2002), while standard light conditions leave the size at about 3 mm (Wyatt, 1995).

Independent of the light and accommodation reflexes, pupil size is known to be sensitive to increased attention or cognitive load, as well as emotional or cognitive arousal (Sirois & Brisson, 2014). While the pupil response to illumination change can be even more than double its typical size, changes due to cognitive effort are much subtler and rarely greater than 0.5 mm (Beatty & Lucero-Wagoner, 2000).

A tight link is established between the pupillary response related to cognitive activity and the locus coeruleus (LC) neuron activation (Laeng, Sirois, & Gredebäck, 2012; Koss, 1986; Samuels & Szabadi, 2008). The LC is the conductor of the noradrenergic system, and the primary source of the neuromodulator norepinephrine (noradrenaline) to the neocortex (Aston-Jones & Cohen, 2005; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Joshi, Li, Kalwani, & Gold, 2016; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Nassar et al., 2012; Samuels & Szabadi, 2008). Stress activates the LC, which in turn increases NE secretion. The LC has direct inhibitory projections to the parasympathetic Edinger-Westphal nucleus – the place of origin of the pupil's constricting fibers. Hence, when the LC activity inhibits the Edinger-Westphal nucleus, it indirectly dilates the pupil (Beatty & Lucero-Wagoner, 2000; Loewenfeld & Lowenstein, 1993; Samuels & Szabadi, 2008). This correlation between LC activity and pupil dilation is found in electrophysiological and neuroimaging studies (Aston-Jones, Rajkowski, & Cohen, 1999; Murphy et al., 2014).

Measuring Cognitive Effort

Pupillometry is the study of changes in pupil diameter caused by cognitive activity. Pupillary response indicates how intensely the processing system is operating (Just & Carpenter, 1993).

The seminal work of Hess and Polt (1960) marked the modern beginning of interest in pupillometry. Besides testing the pupillary response to arousing, emotionally relevant stimuli, the authors also showed that arithmetic tasks, such as multiplication of varying difficulty, are correlated with a differing extent of pupil dilation (Hess & Polt, 1964). In addition, Kahneman and Beatty (1966) have worked with recall of digit strings of different length. Also, sustained attention processing (Bradshaw, 1968; Kahneman, 1973), working memory (Ahern & Beatty, 1979; Beatty & Kahneman, 1966; Bradshaw, 1968) and decision-making (Kahnemann & Beatty, 1967) were assessed.

Moreover, pupil size has been proven to be a reliable measure of cognitive effort induced by language processing (Zellin, Pannekamp, Toepel, & van der Meer, 2011). It was found to index speech intelligibility (Zekveld & Kramer, 2014) and the difficulty of word retrieval in bilingual people and toddlers (Schmidtke, 2014; Kuipers & Thierry, 2013), and to predict the wandering of attention during reading tasks (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013). Moreover, pupil dilation reflects the effect of the visual context on

processing load during comprehension. For instance, examining processing effort for spoken garden path sentences, Engelhardt, Ferreira, and Patsenko (2010) showed that when visual context was consistent with the correct sentence interpretation, sentence prosody had little effect on processing load. Also, they found evidence that increased effort did not result in more incorrect sentence interpretations, concluding that pupil size is a more sensitive measure of language processing than task performance. Another study in the VWP showed that when the visual context disambiguated the initial NP as the object of a sentence (OVS word order) listeners' pupils became significantly more dilated than when the visual context supported the anticipation of SVO word order (Scheepers & Crocker, 2004). For a full account of the existing literature employing pupillometry in linguistic research, please see the recent review by Schmidtke (2017).

How is it Done?

Initially, the studies of task-evoked pupillary response made use of specialized pupillometers in order to measure pupil diameter. Due to the improvement of eye tracking technology in terms of availability and measurement accuracy, eye-trackers have become widely used for the extraction of the pupil diameter record. The common way of measuring the pupillary response is by comparison to the baseline.

Task-Evoked Pupillary Responses (TEPRs), such as mean pupil dilation, peak dilation and latency to peak, are established as a reliable pupillary measure of cognitive load in psychological research (Beatty, 1982). Such a measure is obtained by a time-locked averaging of the pupillary record with respect to critical events in a task, and thus it is essential to set the baseline pupil diameter on trial onset, or during another pre-measurement interval. Importantly, though, the light reflex causes pupilar dilation larger than that induced by mental events, which is a potential confound that this technique cannot account for. Thus, care has to be taken to avoid masking smaller TEPRs by such optic reflexes (Beatty & Lucero-Wagoner, 2000). This can be accomplished by keeping constant not only the luminance level in the experimental room, but also the luminance level of visual stimuli. As noted by Demberg and Sayeed (2016) this is not an easy task, since the pupil exhibits irregular oscillation under the influence of constant light. Moreover, fixating darker or lighter objects in a scene can affect overall pupil size (Demberg & Sayeed, 2016). Importantly, there is evidence that the pupillary light reflex can be artificially induced in the condition of constant luminance by showing pictures that suggest brightness, for instance by contrasting images of the sun and the moon (Binda, Pereverzeva, & Murray, 2013; Laeng & Endestad, 2012; Naber & Nakayama, 2013). A recent study by Mathôt, Grainger, and Strijkers (2017) found that hearing a word which suggests brightness (e.g. *day*) leads to a smaller pupillary response

than a word that conveys darkness (e.g. *night*), suggesting that word meaning is sufficient to trigger a pupillary light response. Another potential confound that has to be accounted for is that pupil size can vary due to participants' gaze position (see Scheepers & Crocker, 2004; Hayes & Petrov, 2016). Due to the fixed position of the eye-tracker camera, free movement of the eye results in the record under varying angles that distorts the accuracy of the pupil diameter estimate. A manufacturer of eye-trackers mentions this problem in their manual, warning that pupil size can be affected by as much as 10% due to pupil position in relation to the camera (Research, 2008). They further suggest that if one aims to conduct research using pupil size, they should make sure that the participants do not move their eyes. Finally, it is important to note that pupil response is relatively slow to return to baseline (Schmidtke, 2017).

The Index of Cognitive Activity

In the past few years, a pupillary measurement has attracted attention by proposing to solve the previously raised issues in creating a sufficiently and reliably controlled experimental setting for the use of TEPRs. The Index of Cognitive Activity (ICA) performs a wavelet analysis of the pupil dilation record, removing all large oscillations connected to the light reflex, and keeping only the small rapid fluctuations in pupil size that are related to cognitive effort and noise. Hence, it disentangles the change in pupil size due to cognitive load from that induced by the light reflex (Marshall, 2000). Further, a denoising technique is applied that results in a record of the so-called ICA events.

The index reflects unusual increases in the pupil signal occurring due to effortful cognitive activity; it is computed as the number of times per second that an abrupt discontinuity in the pupil signal is detected (Marshall, 2007). These events of small abrupt changes in pupil size are referred to as *ICA events*. Low per-second values indicate lower cognitive effort, while high values, conversely, reflect more cognitive activity. Importantly, the measurement maintains both time and frequency information, so that the exact time at which an ICA event was observed is detectable and a fine-grained analysis is facilitated. For more on the technicalities of the ICA extraction, please refer to Marshall (2000, 2002); Demberg and Sayeed (2016).

As expressed by Duchowski et al. (2018), what makes the ICA so appealing is that it is an instantaneous measure, capturing the fluctuation of the pupil diameter, rather than its difference from a baseline. Importantly, this moment-to-moment change in pupil diameter is captured independent of the gaze position. Since it reports the frequency of occurence of the rapid pupillary jitter, it allows for comparison between individuals, as well as multiple events in the performance of an individual.

Since its appearance, the ICA has been tested in a variety of different cognitive tasks (e.g., Marshall, 2002, 2007; Schwalm, Keinath, & Zimmer, 2008). Recently, it has also been examined with cognitive load induced by linguistic processing (Demberg & Sayeed, 2016; Tourtouri, Delogu, & Crocker, 2017; Ankener, Drenhaus, Crocker, & Staudte, 2018). Demberg and Sayeed (2016) employed the ICA in a series of seven experiments, with different kinds of linguistic stimuli and different modes of stimulus presentation. The ICA proved to reflect linguistically induced cognitive load for both reading and auditory presentation of linguistic material. In addition, experiments in a driving simulator and the VWP revealed that the measurement is robust with respect to eye movements and lumination changes. Demberg and Sayeed (2016) conclude that the ICA is applicable as a measure of processing effort that does not have to account for artifacts due to eye movements or changes in screen luminosity, deeming it a promising measure for simultaneous assessment of visual attention and processing difficulty. Recently, Tourtouri et al. (2017) obtained reliable results by doing just that, employing the ICA measurement in the VWP in combination with eye movement analysis. Furthermore, Ankener, Drenhaus, et al. (2018) manipulated multimodal surprisal of referent nouns in the VWP in two experiments, contrasting the ICA with the ERPs. They found that the multimodal surprisal of a word (as modulated by the visual referential context) predicts both pupillometric (ICA) and ERP (N400) measures of online processing effort.

Having considered all of the abovementioned issues, and since the aim of the present work requires simultaneous assessment of both visual attention (i.e. free eye movements) and cognitive load, we decided to design our experiments in the VWP so as to employ the eye tracking technology together with the discussed pupillary measure of cognitive load. Hence, in the following, we report and interpret anticipatory eye movements, as well as the analysis of the ICA. For a fine-grained analysis of cognitive load we use the raw ICA workload, that is, the output including information of the exact timing of each ICA event.

Finally, we would like to draw attention to the newly presented paper that introduces the Index of Pupillary Activity (IPA), a measurement that was inspired by the ICA, but has the important advantage of not being proprietary, instead presenting an open-source algorithm (Duchowski et al., 2018). This is a promising new effort toward establishing pupillometry as a measurement of cognitive load that is easy to use in terms of being robust enough for different experimental settings.

Chapter 3

Introducing the Visual Context

Recent psycholinguistic research has established a correlation between information-theoretic concepts, such as surprisal (Shannon, 1948) and entropy (Hale, 2001; Levy, 2008), and measures of processing effort. In order to quantify the amount of information a linguistic unit carries, surprisal values are typically derived from language models, corpus data, or cloze probabilities (Cloze task, Taylor, 1953). These values are then used as predictors of the processing effort that a linguistic unit is likely to induce (e.g., Demberg & Keller, 2008; DeLong, Quante, & Kutas, 2014). A more surprising linguistic unit requires more effort to process, as measured by reading times and ERP components (e.g., Demberg & Keller, 2008; Smith & Levy, 2013; S. Frank et al., 2015). Similarly, entropy reduction, that is, the reduction of uncertainty about an upcoming linguistic unit, has also been correlated with reading times. Units that bring about higher rates of entropy reduction are in themselves more informative, and hence require longer reading times (Linzen & Jaeger, 2014). Moreover, Maess et al. (2016) have found (in an MEG study) that highly constraining verbs (e.g. *conduct*; GER original: *dirigieren*) induce higher N400 activity, in comparison to unconstraining verbs (e.g. *lead*; GER original: *leiten*), while the effect is reversed on the subsequent noun (e.g. *orchestra*; GER original: *Orchester*).

Such an approach to establishing processing effort is, however, exclusively focused on linguistic context, thereby neglecting the richness of information conveyed in the visual modality, its influence on the predictability of a linguistic unit, and ultimately the effort required for its processing. If we consider situated communication, the assumption is that the objects that are present in the interlocutors' immediate visual context carry a higher probability of being mentioned. Hence, the anticipation of an upcoming referring expression is enhanced by the visual context. This leads to lowering the surprisal value of the given expression, that is, the effort required for its processing by the listener upon hearing it.

Finally, by estimating linguistic surprisal alone, one cannot account for the processing effort of a linguistic unit in a situated communicative encounter.

With the experimental work presented in this chapter we aim at addressing the first overarching research question mentioned in Chapter 1 (p. 6), namely, *Can the effect of visual context on linguistic processing be quantified by measuring the induced cognitive load?* We quantified the influence of visual information on the probability of a linguistic reference, by measuring listeners' visual attention and the immediate cognitive effort induced at a relevant linguistic unit. We conducted three experiments focused on answering the following more concrete questions:

1. Does the overall cognitive load induced during sentence processing change due to the existence of visual context?

2. Does visual context affect cognitive load induced at those sentence parts where the visual and linguistic information can be combined or mapped to one another?

Being constantly present and available for incorporating into the sentence representation at all stages of sentence unfolding, the visual context is expected to lead to an increased cognitive effort required for processing a sentence, since it is a permanently available source of additional information. Moreover, we expect the interaction between the visual and the linguistic context to be more important at specific sentence parts. We expect the restrictive verb to lead to a shift in visual attention towards the object(s) that fit the selectional preferences of the verb (replication of Altmann & Kamide, 1999). In addition, mentioning the restrictive verb is a point of high entropy reduction; that is, the verb allows for the subsetting of the visual context into fitting potential target objects. Hence, we expect this entropy reduction to be reflected in more cognitive effort required for processing the restrictive verb (consistent with Maess et al., 2016). Finally, we expect this earlier reduction of entropy enabled by the verb selectional preferences and the visual context to result in a facilitation for the processing of the subsequent linguistic reference.

In what follows, we report on the three experiments conducted in order to compare the values of cognitive load induced during language processing with and without the relevant visual context. First, we conducted an eye tracking reading study (Experiment 1). This study aimed at assessing the processing effort required for comprehending the linguistic stimuli in isolation (without visual context) in the written modality. Second, we moved the same stimuli into the auditory modality and examined participants' pupillary response during language processing (Experiment 2). Finally, Experiment 3 is an eye tracking study conducted in the VWP, presenting the same linguistic stimuli in the same auditory manner as in Experiment 2, but with the addition of the corresponding visual stimuli. We examined participants'

eye movements together with their pupillary response. Since all three experiments made use of the same set of linguistic stimuli, in what follows, we first describe the process of stimuli creation and pre-testing. Subsequently, the three experiments are presented separately. Finally, the findings are discussed as a whole and presented together with the interpretations.

3.1 Linguistic Stimuli Creation and Validation

3.1.1 Stimuli Creation

We created 36 items making use of simple independent German main clauses of the following structure: Subject – Verb – Adverb – Object. See Table 3.1 for an illustration. The subject noun phrase was kept identical within an item. Another sentence part that was not part of the manipulation was the adverb, which was kept constant within all items (*now*, GER original: *gleich*). The adverb was included to add more time between the two crucial sentence parts, namely, the verb and the object, without introducing any additional relevant information.

We manipulated verb constraint and included highly restrictive (*spill*, GER original: *verschütten*) and relatively non-restrictive (*order*, GER original: *bestellen*) verbs. The difference in verb constraint made the verb argument (subsequent referent noun) more or less predictable.

The other manipulation concerned the referent noun. An established method of stimuli creation, when examining lexical predictability, is collecting target nouns of differing probability by employing the cloze task (Taylor, 1953). Native speakers of a language in question are presented with a linguistic context and a gap, which they should complete with a word that first comes to their mind. In this way, one can estimate the values for the most predictable word in a certain linguistic context. Predictability is thus defined as the probability that a certain word will be provided as a sentence continuation. Note, however, that such a method does not account for the differences in frequencies among different target words that are collected.

Word frequency being a potential confound in the examination of probability effects (e.g., Schilling, Rayner, & Chumbley, 1998), we have decided against the canonical use of the cloze task in the process of stimuli creation. Rather, in order to be able to control for the word frequency of the critical words, we first created the stimuli sentences we wanted to work with, and only subsequently acquired the cloze probabilities, as well as a measure of noun plausibility.

First, we used the DeReWo word lists of the German Reference Corpus[1] to extract inanimate concrete nouns with approximately same frequency. These nouns were used to create experimental items with constant noun frequencies within an item. In combination with the restrictive verb (*spill*) one highly predictable target object was chosen (*water*, GER original: *Wasser*), and one less predictable but still plausible referent (*ice cream*, GER original: *Eis*). Both referent nouns were equally plausible arguments of the non-restrictive verb (*order*). This allowed for the creation of four experimental conditions that were used in the planned experiments.

An additional potential confound, namely unequal length of the critical words, was controlled for by either including the information about the word length as a factor in the analysis, or by choosing the start of the relevant analysis window relative to the word's length. These strategies will be explained in more detail in the following sections.

Table 3.1 An example linguistic item. Note that conditions 1c and 2c are used only in Exp. 1 (Exp. 2 and Exp. 3 include only four conditions). All items are translated word-for-word, maintaining the original word order.

Condition	Sentence	Original
1a	The man spills now the water.	*Der Mann verschüttet gleich das Wasser.*
1b	The man spills now the ice cream.	*Der Mann verschüttet gleich das Eis.*
(1c)	The man spills now the book.	*Der Mann verschüttet gleich das Buch.*
2a	The man orders now the water.	*Der Mann bestellt gleich das Wasser.*
2b	The man orders now the ice cream.	*Der Mann bestellt gleich das Eis.*
(2c)	The man orders now the book.	*Der Mann bestellt gleich das Buch.*

In the context of *spill*, *water* is a very likely continuation; that is, it is expected to be highly predictable. *Ice cream*, in contrast, needs to be melted in order to be spillable, and since that is not its typical state, it is a possible referent, though, not as predictable as *water*.

Additionally, only Experiment 1 includes *book* as an impossible referent of *spill*. This additional condition is included in order to create a baseline comparison, where a difference between the highly probable *spill water* and the anomalous *spill book* should induce a clear effect of processing load. In addition, this comparison allows us to observe the positioning of the second condition, unpredicted but still possible, in relation to the other two conditions. Finally, the three objects are equally plausible in combination with the non-restrictive verb *order*.

[1] DeReKo (Deutsch Referenz Korpus) contains more than 28 billion words and is the largest linguistically motivated collection of German texts (DeReWo, 2012). It includes scientific, belletristic and newspaper texts from past and present.

Such a design enabled us to examine the differences between the referents in combination with the restrictive versus the non-restrictive verb. Being less predictable, the less plausible referents carry more surprisal. This is expected to reflect on the eye tracking measures of reading and the cognitive load induced upon hearing the target referent.

It should be mentioned that our process of stimuli creation has left us without the possibility to later gather predictability values for these particular word combinations from the corpus, since not all verb-object combinations could be found. Hence, we conducted two offline studies, a cloze task and a plausibility rating questionnaire, in order to assess the plausibility of our intuitively created word-object combinations.

3.1.2 Stimuli Validation

As mentioned previously, in order to control for the frequency of different referent nouns, we first chose a group of concrete inanimate nouns of roughly the same frequency derived from the German Reference Corpus (DeReWo, 2012). We then created the stimuli sentences and subsequently conducted two pretests in order to check whether our manipulation was successful.

First, in order to asses the strength of verb constraint, a cloze task was conducted. Participants were presented with our stimuli sentences truncated prior to the referent noun. They were asked to spontaneously complete the sentences with a noun they find best fitting to the sentence context. All items were presented in one list. The same amount of filler sentences were added. 17 German native speakers took part without reimbursement (18 to 55 years of age). The results showed that in the highly constraining context of *spill*, plausible objects such as *water* were more predictable (cloze probability of 13.67%) than the possible objects such as *ice cream* (cloze probability of 0.16%).

Second, the noun's plausibility given the previous context was validated. We conducted a plausibility rating questionnaire, where our stimuli sentences were rated on a seven-point Likert scale, with 1 set to "highly plausible" and 7 to "not plausible at all". 14 German native speakers completed the questionnaire (18 to 58 years of age). For the analysis of the obtained data, the dependent variable Rating Score was treated as a count variable. Hence, we used generalized mixed effects models of Poisson type. Maximal converging random structure was included (Barr, Levy, Scheepers, & Tily, 2013). Also, independent variables Constraint (restrictive vs. non-restrictive) and Plausibility (plausible vs. possible) were contrast coded before the analysis. The full model[2] revealed a significant Constraint:Plausibility interaction (β = -1.034, SE = 0.101, z = -10.271, $p < 0.001$), as well as a main effect of Plausibility (β

[2]RatingScore $\sim$ Const*Plaus + (1+ Const + Plaus || Subject) + (1 + Const*Plaus || Item), family = Poisson (link = "log")

= -0.654, *SE* = 0.058, z = 11.355, $p < 0.001$). Further comparisons[3] showed a main effect of Plausibility in the subset of the restrictive verb (β = -1.093, *SE* = 0.127, z = 8.616, $p < 0.001$), suggesting that *spill water* (*M* = 1.043, *SD* = 0.346) was rated as more plausible than *spill ice cream* (*M* = 3.421, *SD* = 2.247). In addition, in the subset of the non-restrictive verb we found no effect of Plausibility ($p = 0.167$), suggesting that *order water* (*M* = 1.679, *SD* = 1.583) was not rated differently from *order ice cream* (*M* = 1.964, *SD* = 1.860).

In conclusion, the results from our two stimuli validation tests confirmed that our experimental manipulations were successful. The same stimuli sentences are used in all experiments presented in this chapter, and serve as a basis for moderate alterations in the experiments of the subsequent chapters, depending on the experimental method used (noted in detail at relevant points throughout the text). An exhaustive list of all 36 item sentences presented in four conditions each is given in Appendix A, Table A.1.

3.2 Experiment 1

First, we conducted an eye tracking reading study assessing participants' eye movements and examining the effect of the purely linguistic context on the processing effort for the critical words. Inspired by Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005)'s finding that expectation-inconsistent adjectives result in longer reading times, we expected to see a graded effect on the referent nouns, relative to their fit to the verb.

3.2.1 Method

This experiment made use of 2×3 experimental design. The independent variable **Constraint** (restrictive vs. non-restrictive) was manipulated by verb restrictiveness (*spill* vs. *order*). In addition, the independent variable **Plausibility** (plausible vs. possible vs. impossible) concerned the referent noun fit with the restrictive verb. It was either a plausible continuation of the restrictive verbal context such as *spill water*, an implausible but still possible combination such as *spill ice cream*, or a completely anomalous combination such as *spill book*. As already presented in Table 3.1, the additional impossible condition 1c, as well as its baseline equivalent 2c, were included only in this experiment.

[3] RatingScore $\sim$ Plaus + (1 + Plaus | Subject) + (1 + Plaus | Item), family = Poisson (link = "log")

Participants

24 Saarland University students (17 female) took part in this experiment and were monetarily reimbursed for their participation. The participants' age ranged from 18 to 32 (M = 22.71). They were all native speakers of the German language.

Items and Fillers

This experiment made use of 36 item and 36 filler trials. The stimuli used as items were presented in the previous section *Stimuli Creation and Validation*. In addition to the described structure, this experiment included a sentence final prepositional phrase denoting location, to serve as a spillover region after the referent noun, for example: *The man spills soon the water at the restaurant*. We used the Latin square design to create 6 trial lists, in such a way that each item would appear in only one condition per list. In addition, all filler trials were presented in each list. Fillers were all plausible sentences that differed from the items in their structure. Illustration: *Mit dem Fahrrad transportiert die Mutter das frische Baguette* (literal translation: *With the bicycle transports the mother the fresh baguette*). Fillers were followed by simple yes/no content questions regarding the sentence content. Moreover, a practice session with three trials preceded the experimental session (36 items + 36 fillers).

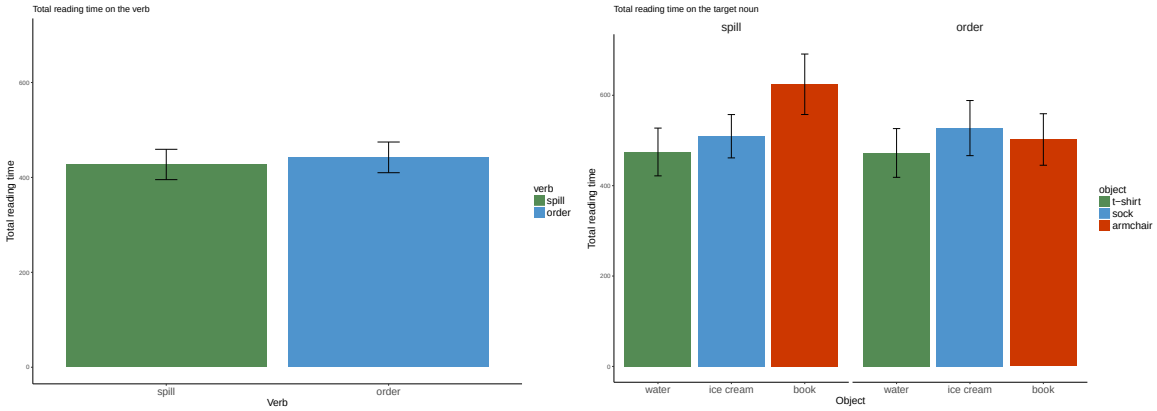
Procedure

The experiment was run using the EyeLink 1000+ eye-tracker (SR Research Ltd.; Mississauga, Ont., Canada). Only the participants' dominant eye was tracked. Head movements were minimized by using the appropriate chin rest. Participants were instructed to read for comprehension, at their own pace, and warned that some of the sentences might seem strange.

Sentences were presented as a whole, located centrally on the screen (Times New Roman, 20 pt). The fixation dot, presented prior to each trial, was located at the top left corner in order to avoid initial fixations of the sentence before participants actually started reading. Advancing to the next sentence was done by pressing a button on the button box. An experimental session lasted up to 30 minutes.

3.2.2 Analysis and Results

Two relevant interest areas (AoI) were considered for the analysis, namely, the verb AoI and the referent noun AoI. In addition, the respective spill-over regions were also examined. Time measures were log-transformed and included as dependent variables in linear mixed



(a) Total dwell time on the verb.

(b) Total dwell time on the referent noun.

Figure 3.1 Exp. 1 – Total reading time of a word across the six experimental conditions (given in miliseconds).

effects models. The independent variables Constraint and Plausibility were contrast coded and included as fixed factors, together with the scaled word length (measured in characters). All models included maximal converging random structure justified by the experimental design (Barr et al., 2013). The analyses were conducted in the R programming environment (version 3.4.0; R Core Team, 2017) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015).

The model fitted for the total dwell time on the referent noun showed, in orthogonal comparison, a significant difference between the possible and the impossible referent nouns in the restrictive condition ($p < .01$; *spill ice cream*: $M = 509.401$, $SD = 288.98$ vs. *spill book* $M = 624.350$, $SD = 396.110$). The two possible conditions did not differ significantly (*spill water* vs. *spill ice cream*). The results are illustrated in Figure 3.1b. The model fitted for the total dwell time on the verb showed no significant difference (see Figure 3.1a for illustration). Further, the analysis of regressions to the pre-target region showed the same pattern as the total dwell-time. The analysis of the first-pass measurement, as well as the dwell time on the spill-over region, did not yield significant results.

3.2.3 Discussion

The eye tracking measures correlated with cognitive effort induced while reading, showed that (in the absence of visual context) processing verbs of different strength of constraint did not induce differing cognitive effort. Moreover, the time required to read the referent noun was significantly increased only in the condition when the noun did not fit the previous

context. We did not find a difference in reading times for the two possible nouns of differing plausibility.

Several studies have found effects of predictability and plausibility during reading and listening. Rommers et al. (2013) found a graded N400 effect in response to target words of differing probability. Rayner, Juhasz, Warren, and Liversedge (2004) found an effect of implausible words on eye movements in reading. However, for less severe violations of plausibility, they found only delayed effects of smaller magnitude. The authors conclude that less extreme plausibility violations are not detected until later stages of processing. They further suggest that in contrast to visual-world studies, reading does not provide strong set restrictions, and thus arguments are less predictable. Without a visual context, readers are less likely to anticipate the argument based on its plausibility, which then leads to later plausibility effects in reading than in the visual-world studies.

It is important to note that our stimuli consisted of simple isolated sentences, not embedded in a larger discourse context. Hence, no additional information apart from the verb was available to inspire a prediction for a specific referent noun. In addition, the verb constraint was comparatively low.

We suggest that in the context of our experimental stimuli, verb constraints alone did not inspire concrete lexical expectations about the target nouns. In addition, the lack of an effect measured on the verb, inconsistent with Maess et al. (2016), may be explained by the lower degree of restrictiveness of the verbs used in our study. Alternatively, the measure of reaction times may not be sensitive enough to capture the difference between the conditions. Hence, in the following experiment, we aim to reassess the subtle differences between the referent nouns with a potentially more sensitive measurement. Instead of a time-based measure of effort, we employ a pupillary measurement which captures the amplitude of cognitive effort.

3.3 Experiment 2

As mentioned previously, there are two explanations to the results obtained from Experiment 1. First, it may be that no significant differences were measured between the plausible and possible referent noun because of our (deliberately) subtle predictability manipulation. Second, the lack of expected effect may be due to the written modality and the time-dependent measure used to assess participants' cognitive effort.

Thus, this experiment aims at shining more light onto this issue by examining the same linguistic stimuli as in Experiment 1, but now in the auditory modality. We assess the effect of our manipulation on cognitive effort as correlated with a pupillary measure. Moreover, this

experiment establishes a baseline for processing effort for the critical words in the absence of the visual context.

This is the first of a series of seven experiments we ran using the Index of Cognitive Activity, a pupillary measure that is robust to changes in illumination and eye movements, and which allows for an instantaneous assessment of cognitive effort at any point in time without the need for baseline comparison. The measurement was presented in detail in Section 2.3.2 (p. 26).

3.3.1 Method

We used a 2×2 experimental design. As in Experiment 1, **Constraint** (restrictive vs. non-restrictive) was manipulated by verb restrictiveness (*spill* vs. *order*). However, the independent variable **Plausibility**, which concerned the referent noun fit with the restrictive verb, was used in only two relevant levels (plausible vs. possible).

Participants

36 Saarland University students (20 female) participated in the study and were monetarily reimbursed for their contribution. Their age ranged from 19 to 46 years (M = 24.72). The participants were all native speakers of the German language with normal or corrected-to-normal vision. Importantly, none of the participants were familiar with Experiment 1.

Items and Fillers

20 experimental items were used for this experiment. From the same stimuli sentences presented previously and used in full in Experiment 1, 20 sentences with the most representative scores from the pretests were selected for this study. The sentences were used in the same S-V-Adv-O structure as in Experiment 1, just without the final spill-over region. An exhaustive list of all item sentences is given in Appendix A, Table A.2. The Latin square design was applied to form 4 lists in such a way that each item was presented in only one condition per list. 26 filler items were used (selected from the larger list of filler sentences from Exp. 1). As before, fillers were sentences of different syntactic structure from the items. They were followed by simple yes/no comprehension questions presented written on the screen and answered using a key press.

The linguistic stimuli were recorded in an audio format read aloud by a German native speaker at a constant natural pace. Since the target noun was the final word of the sentence, a 2000 ms silent break was introduced at the end of the sentence, to allow for a larger time-window for measurement.

Procedure

This was a comprehension experiment where participants were asked to listen carefully to the audio stimulus. Even though no visual stimulus was presented, participants' eyes were tracked in order to extract ICA values from the pupil jitter.

The experiment was created and presented using SR Research Experiment Builder (version 1.10.1630). The Eye-Link II eye tracking system (SR Research, Ltd.; Mississauga, Ont., Canada) was used to record eye movements binocularly, with a sampling rate of 250 Hz.[4] The eye tracking data were collected using SR Research Data Viewer software (version 1.8.605). The wavelet transformation and calculation of rapid small dilations was conducted in the EyeWorks Workload Module software (version 3.12).

Prior to the experimental session, three practice trials were presented, to allow for the participants to familiarize themselves with the experimental setting and the task. The experiment lasted for approximately 15 minutes.

Variable Coding and Data Analysis

The ICA events were extracted from the pupil jitter, summed over a duration of a relevant time-window and statistically analyzed. Two time-windows were of relevance: (a) the Verb time-window: 600 ms from the middle of the verb, showing the potential cost of predicting the target referent; and (b) the Reference time-window: 600 ms from the middle of the referent noun, showing the cost of processing the linguistic reference.

In their VWP experiment, Demberg and Sayeed (2016) establish a time-window taken from 600–1200 ms from the onset of the critical word to be an appropriate window size and timing for the analysis of ICA events. For the sake of comparison with the subsequent experiment (conducted in the VWP) we use the same size analysis window. Since our critical words differ in length across items and since there is article variation among items, we corrected this potential confound by taking a time-window that starts from the middle of a word and considers the following 600 ms. The word middle was calculated by taking the audio duration of the whole word and using half of this as the starting point, for each word individually.

The ICA events were extracted using the EyeWorks Workload Module software (version 3.12) for both eyes separately. Since there is no clear theoretical reason why differences should be expected for the two eyes, we combined the two datasets by summing the ICA events for corresponding time-windows and conducted the analyses on the combined data.

[4]This is the required setup for the extraction of the ICA events from EyeWorks Workload Module software (version 3.12) when using the Eye-Link II tracker.

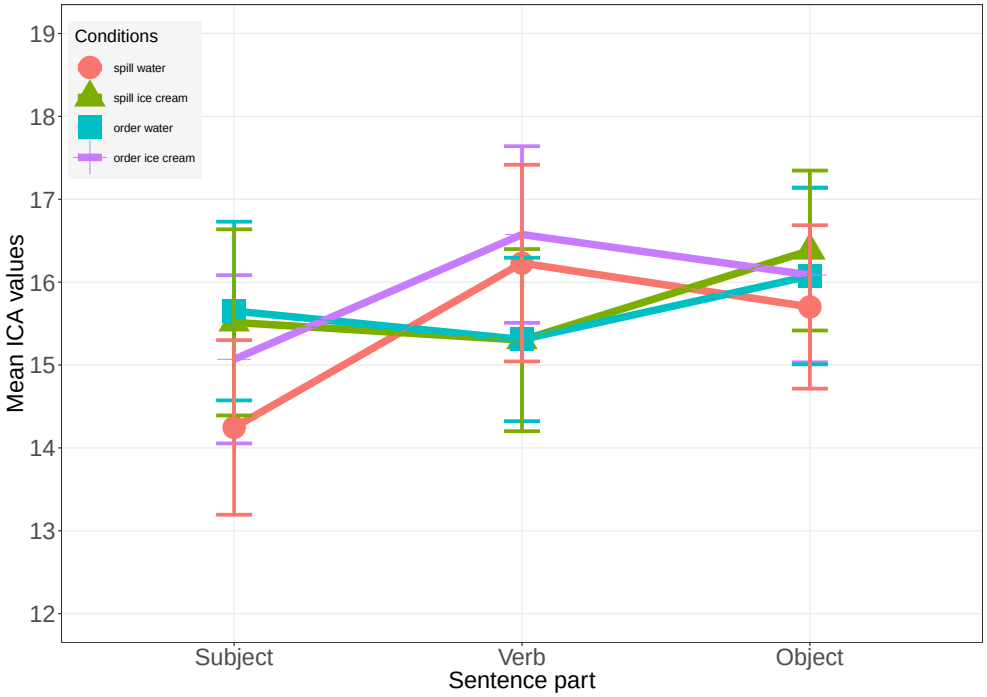


Figure 3.2 Exp. 2 – Mean ICA values in the three time-windows of a sentence. Points marked as *Verb* (Verb window) and *Object* (Reference window) are relevant for the analysis (95% CI error bars).

The independent variables were contrast coded for the statistical analysis. The ICA being a count variable, we used generalized mixed effects models with Poisson distribution. All models included maximal converging random structure justified by the experimental design (Barr et al., 2013). The analyses were conducted in the R programming environment (version 3.4.0; R Core Team, 2017) using the *lme4* package (Bates et al., 2015).

3.3.2 Results

Figure 3.2 illustrates the mean ICA values for the three relevant points during a sentence. Note that the points labeled as Verb and Object illustrate the two time-windows relevant for the analysis, namely the verb time-window and the reference time-window, respectively.

Table 3.2 presents the results of the full models used for the analysis. The results clearly show no significant differences measured as a result of our manipulation in either of the two relevant time-windows.

Table 3.2 Exp. 2 – Results of the main models fitted for the ICA analysis.

	a) Verb time-window				b) Reference time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.721	0.048	57.22	<2e-16 ***	2.741	0.040	68.14	< 2e-16 ***
CONSTRAINT	0.035	0.039	0.88	0.378	-0.006	0.035	-0.17	0.869
PLAUSIBILITY	-	-	-	-	0.021	0.037	0.57	0.570
CONST:PLAUS	-	-	-	-	-0.037	0.069	-0.54	0.588

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICA ~ Constraint + (1 + Constraint | Subject) + (1 + Constraint | Item), family = Poisson (link = "log")
b) ICA ~ Constraint * Plausibility + (1 + Constraint * Plausibility || Subject) + (1 + Constraint * Plausibility || Item), family = Poisson (link = "log")

3.3.3 Discussion

The null results of Experiment 1 for the plausibility manipulation were replicated in the auditory modality by using a pupillary measure correlated with cognitive effort. Hence, we conclude that the lack of difference between our conditions was due neither to the presentation modality, nor to the measurement used. Rather, as suggested previously, by not using the verb-noun combinations of the highest cloze probability, but those on the lower end of the scale, not enough information was available for participants to create a concrete lexical expectation of the referent noun.

In addition, the results obtained from Experiment 2 form a baseline for the cognitive effort that our linguistic stimuli induce in the absence of a visual context. A weakly constraining context does not carry enough information for lexical expectations to be beneficial (in accordance with Wlotko & Federmeier, 2015). In what follows we examine how the visual context can influence the creation of concrete expectations.

3.4 Experiment 3

Conducted in the VWP, Experiment 3 provides additional constraining information (in addition to the verb) by including a visual context. The visual context reduced the set of possible targets to the visually presented objects, potentially inspiring more concrete anticipations for a specific object. (Dis)confirming these anticipations is expected to lead to bigger differences in cognitive load than measured on the linguistic context alone.

By including the visual context, we assessed its immediate effect on creating expectations for specific target referents. We examined whether the reduction of referential entropy brought about by the verb's restrictiveness is detectable in the shifts of visual attention (as found by Altmann & Kamide, 1999). Moreover, we examined if such a consideration, that

is, the subsetting of a visual scene, is reflected in immediate cognitive effort induced at the verb. Specifically, when verb constraint allows for the exclusion of more presented objects, we expected higher cognitive effort at the verb than the condition in which all presented objects fit the verb selectional preferences. In other words, higher reduction of referential uncertainty is expected to induce more cognitive effort immediately. Also, we examined the surprisal-based processing effort required for the integration of the subsequently uttered referent noun. We expected to find evidence that the integration of the more predicted target referent requires less cognitive effort than the less plausible, but still possible referent noun. Finally, in comparison to Experiment 1, we expected that the addition of the visual context should induce higher processing load in general, since more information is constantly available during a trial.

3.4.1 Method

We used the same a 2×2 experimental design as in Experiment 2. Hence, we manipulated verb restrictiveness, **Constraint** (restrictive vs. non-restrictive), and **Plausibility** of the referent noun to follow the restrictive verb (plausible vs. possible). In addition, visual context was introduced by presenting four potential target objects. See Fig. 3.3 for an illustration. In the context of *spill*, two out of four objects were possible referents. Moreover, the target object (water) was more plausible than the competitor (ice cream). In the context of *order* all four presented objects were equally probable (water, ice cream, suitcase, coat). A new factor to be considered in this study is the change in entropy reduction at the point of the verb, which is due to the inclusion of the visual context. In addition to the ICA results, this study allowed for the analysis of the participants' eye movements. Hence, both cognitive load and the visual attention are assessed, showing which potential targets are considered and how this affects the load at the final processing of the referent noun.

Participants

36 Saarland University students took part in this study. Due to technical issues, two participants were excluded and the data from 34 participants (28 female) were analyzed. All participants were native speakers of the German language and had normal or corrected-to-normal vision. Their age ranged from 19 to 38 years old (M = 23.25). Participants were monetarily reimbursed for their participation. None of the present participants were familiar with the previous experiments.

Figure 3.3 Exp. 3 – Example of the visual display corresponding to the same item as the linguistic stimuli given in Table 3.1.

Items and Fillers

The same 20 experimental items from the previous study were also used here. In addition, visual displays of four objects were presented. As illustrated in Fig. 3.3, the four presented objects included the target object (*water*), the competitor (*ice cream*) and two additional distractor objects (*suitcase* and *coat*). The same visual scene was used for all four conditions of an item. The distractor objects were selected in such a way as not to be a possible referent for the restricitve verb (*spill*), but to be equally plausible for the non-restrictive *order* as the target and competitor. None of the presented objects illustrated the referents with the highest cloze probability for a given sentence. The images used within one item were of the same size, similar visual complexity and uniformly salient in terms of color. The positions of target, distractor and competitors were balanced throughout the experiment. An exhaustive list of all visual displays used in this experiment is given in Appendix A, Table A.2. Visual displays were presented for 1000 ms before the onset of the sentence. The same audio material from Experiment 1 was used. Images used (for all the experiments presented in this book) were taken from open-source databases[5] and were pretested for naming.

As in Experiment 2, 26 filler sentences of differing syntactic structure were used. The filler visual displays had the same layout, but included more variation in the number of objects that were compatible with the mentioned verb. Illustration: *Im Laden am Eck kauft die Frau gleich eine Vase* (literal translation: At the shop at the corner buys the woman now

[5] www.openclipart.org and www.pixabay.com

a vase). This sentence was presented with a visual display showing a vase, a chair, a can and a (military) tank. A simple yes/no comprehension question followed, presented in written form. Questions were answered with a key press.

Again, the Latin square design was used in order to create four lists with 20 item and 26 filler trials each. Each list included each item in only one of the four linguistic conditions. The trials were pseudo-randomized manually and an additional four lists were created with the opposite order of presentation.

Procedure

As in Experiment 2, the study was created and presented using SR Research Experiment Builder software (version 1.10.1630). The Eye-Link II eye tracking system (SR Research, Ltd.; Mississauga, Ont., Canada) was used to record eye movements of both eyes, with a sampling rate of 250 Hz. The eye tracking data were collected using SR Research Data Viewer software (version 1.8.605). The wavelet transformation and calculation of rapid small dilations was conducted in the EyeWorks Workload Module software (version 3.12). Since the target was the final word of the sentence, a 2000 ms silent break was introduced sentence finally, to allow for a long enough measurement window.

Participants were presented with both audio and visual stimuli and were instructed to interact naturally with the scenes, and listen for comprehension. The actual experimental session was preceded by three example trials of the practice block, in order to allow the participants to familiarize themselves with the study. The experiment lasted for approximately 15 minutes.

Variable Coding and Data Analysis

First, in order to gain insight into the patterns of visual attention, we considered the proportion of fixations to the presented objects throughout a trial. Each of the presented objects was treated as a separate area of interest (AoI).

Second, in order to statistically assess shifts in visual attention resulting from verb constraints, we analyzed new inspections, that is, the first of potentially a series of consecutive inspections of an AoI that occurred within a temporal region of interest. We consider new inspections in the verb region of interest – showing how the verb category influenced visual attention, hinting at an object which was considered as most fitting to the verb constraint.

The analysis of the ICA events was carried out in the same way as in Experiment 2 (see p. 39).

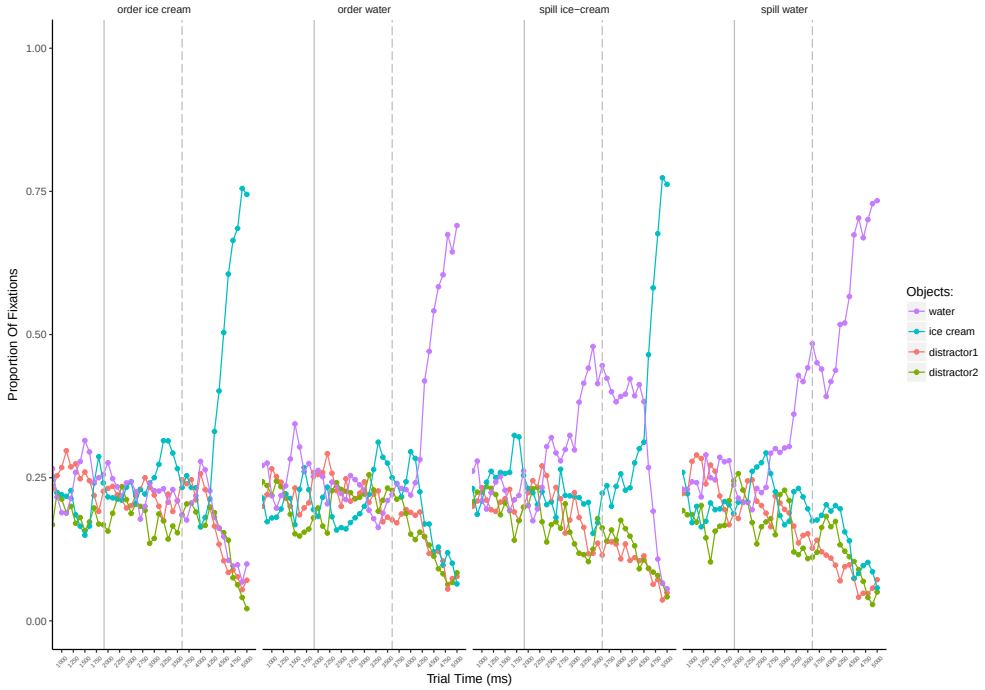


Figure 3.4 Exp. 3 – Proportion of fixations to presented objects in the four experimental conditions. The solid line represents verb onset and the dashed line referent noun onset.

All independent variables were contrast coded for the statistical analysis. New inspections, a binary dependent variable, required the use of generalized mixed effects models of binomial type. The analysis of the ICA, a count variable, required the use of generalized mixed effects models of Poisson type. All models included maximal converging random structure justified by the experimental design (Barr et al., 2013). The analyses were conducted in the R programming environment (version 3.4.0; R Core Team, 2017) using the *lme4* package (Bates et al., 2015).

3.4.2 Results

Proportion of Fixations

Figure 3.4 illustrates the proportions of fixations during a trial in all four of our experimental conditions. The solid line presents the verb onset and the dashed line the onset of the referent noun. The first two plots represent the two non-restrictive conditions, and show that *order* did not induce preferences for any of the presented objects. It was only after hearing *water* or *ice cream* that the relevant object was fixated. The other two plots show that upon hearing the restrictive verb *spill*, participants fixated water more than any other object. Moreover,

Table 3.3 Exp. 3 – Results of the main models fitted for the new inspections analysis for the verb region of interest.

	a) Target inspections				b) Competitor inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.813	0.058	-31.519	< 2e-16 ***	-1.945	0.067	-29.002	<2e-16 ***
CONSTRAINT	-0.378	0.094	-4.038	5.39e-05 ***	0.047	0.097	0.488	0.626
	c) Distractors inspections							
INTERCEPT	-0.79	0.053	-15.050	< 2e-16 ***				
CONSTRAINT	0.141	0.071	1.985	0.047 *				

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) WaterInsp ~ Constraint + (1 + Constraint || Subject) + (1 | Item), family = "binomial"
b) IceInsp ~ Constraint + (1 | Subject) + (1 + Constraint || Item), family = "binomial"
c) DistractInsp ~ Constraint + (1 | Subject) + (1 + Constraint || Item), family = "binomial"

after hearing the referent noun, if it was *ice cream*, it was only then that this object started being fixated more, and that fixations to water started to drop. Finally, if *water* was uttered, that object continued being fixated as the anticipation was confirmed.

Hence, we saw that restrictive verbs shifted the visual attention to that one object in the scene that was the most probable target referent. This preference was slowly reconsidered only upon the other possible referent being mentioned.

New Inspections

In order to statistically assess the previously mentioned patterns in fixations, we consider new inspections to a target object in the relevant time region of interest, namely upon the verb being uttered. In other words, we analyze the likelihood of an inspection to fall within an AoI upon the restrictive piece of information becoming available.

Table 3.3 shows the models and the results obtained in the analysis of the new inspections upon mention of the verb. Regarding inspections to the target object (water) the model revealed a significant main effect of Constraint ($p < .001$), suggesting that the target object was inspected more in the context of the restrictive verb ($M = 0.16$, $SD = 0.37$) than in the non-restrictive context ($M = 0.11$, $SD = 0.32$). The model fitting the inspections to the competitor object (ice cream) showed no significant differences in the contexts of the two verbs ($p = 0.626$). Finally, new inspections to the two distractor objects taken together proved to have occured more in the non-restrictive context than in the context of the restrictive verb ($p < .05$; *spill* $M = 0.3$, $SD = 0.46$ vs. *order* $M = 0.34$, $SD = 0.47$).

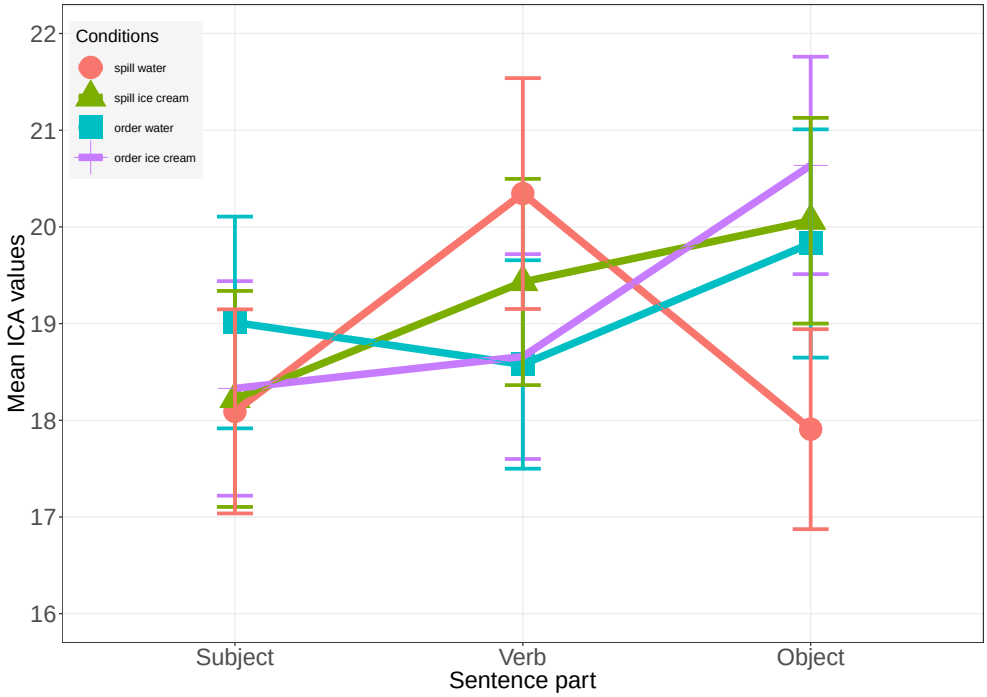


Figure 3.5 Exp. 3 – Mean ICA values in the three time-windows of a sentence. Points marked as *Verb* (verb time-window) and *Object* (reference time-winodw) are relevant for the analysis (95% CI error bars).

The Index of Cognitive Activity

Figure 3.5 illustrates the mean ICA values found at three sentence points. The points labeled as Verb and Object illustrate the two time-windows used in the analysis, namely the verb time-window and the reference time-window, respectively. We observe a potentially significant difference at the point of Object (Reference time-window).

Table 3.4 reports the results of the full models fitted for the mean ICA events in the two relevant 600 ms time-windows. Constraint did not prove significant in the verb time-window. This suggests that the processing effort induced on the verb did not differ between the two different verb contexts. In the reference time-window, however, we found a main effect of Constraint ($p < 0.05$), but more importantly, also the Constraint:Plausibility interaction ($p < 0.05$).

Further comparisons revealed that in the subset of the restrictive verb *spill*[6], *water* induced less cognitive load than *ice cream* ($\beta = 0.113$, $SE = 0.045$, $z = 2.51$, $p = 0.012$; *water*: $M = 17.91$, $SD = 6.91$ vs. *ice cream* $M = 20.06$, $SD = 7.07$). This was not the case in the subset of the non-restrictive verb *order*[7] ($p = 0.46$). Moreover, the analysis of the subset of the

[6]ICA ~ Plausibility + (1 + Plausibility | Subject) + (1 + Plausibility | Item), family = Poisson (link = "log")
[7]ICA ~ Plausibility + (1 + Plausibility | Subject) + (1 + Plausibility | Item), family = Poisson (link = "log")

Table 3.4 Exp. 3 – Results of the main models fitted for the ICA analysis.

	a) Verb time-window				b) Reference time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.932	0.036	81.58	<2e-16 ***	2.950	0.035	84.43	< 2e-16 ***
CONSTRAINT	-0.051	0.032	-1.60	0.11	0.063	0.0320	1.98	0.048 *
PLAUSIBILITY	-	-	-	-	0.081	0.043	1.89	0.059
CONST:PLAUS	-	-	-	-	-0.071	0.035	-2.05	0.041 *

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICA ~ Constraint + (1 + Constraint | Subject) + (1 + Constraint | Item), family = Poisson (link = "log")
b) ICA ~ Constraint * Plausibility + (1 + Constraint + Plausibility | Subject) + (1 + Constraint + Plausibility | Item), family = Poisson (link = "log")

plausible reference *water*[8] showed the main effect of Constraint on the ICA (β = 0.094, *SD* = 0.039, z = 2.41, p = 0.016), suggesting that *water* induced less cognitive load in the context of *spill* (*M* = 17.91, *SD* = 6.91) rather than *order* (*M* = 19.83, *SD* = 7.91). This was not the case for the possible object *ice cream* (p = 0.675; *spill*: *M* = 20.064, *SD* = 7.068 vs. *order*: *M* = 20.637, *SD* = 7.493).

3.4.3 Experiment Comparison

We compared the results obtained in the two experiments by conducting an analysis of the merged dataset.

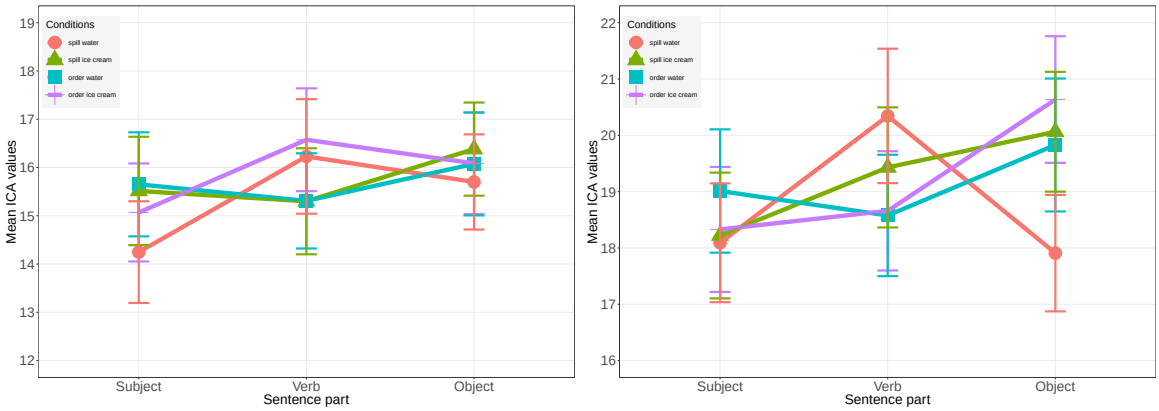
The model fitted for the ICA events measured in the verb window[9] revealed a main effect of Study (β = 0.22, *SE* = 0.021, z = 10.70, $p < .001$) suggesting that the cognitive load was overall higher in Exp. 3 (*M* = 19.19, *SD* = 7.25) than in Exp. 2 (*M* = 15.80, *SD* = 7.25). Also, we found a Constraint : Study interaction (β = -0.083, *SE* = 0.029, z = -2.82, $p < .01$). However, this interaction is carried by the opposite direction of the trend between the two verbs in the two studies.

In the noun window, we found the same results[10], the main effect of Study (β = 0.262, *SE* = 0.020, z = 13.05, $p < .001$) suggesting, again, that the cognitive load was overall higher in Exp. 3 (*M* = 19.76, *SD* = 7.41) than in Exp. 2 (*M* = 16.06, *SD* = 6.79). Also, we found a Constraint : Study interaction (β = 0.066, *SE* = 0.032, z = 2.05, $p < .05$) carried, again, by the opposite direction of the trend in difference between the two verb conditions.

[8] ICA ~ Constraint + (1 + Constraint | Subject) + (1 + Constraint | Item), family = Poisson (link = "log")

[9] ICA ~ Constraint * Study + (1 + Constraint | Subject) + (1 + Constraint | Item), family = Poisson (link = "log")

[10] ICA ~ Constraint * Study + Plausibility * Study + Constraint * Plausibility + (1 + Constraint * Plausibility || Subject) + (1 + Constraint * Plausibility || Item), family = Poisson (link = "log")



(a) Experiment 2 (b) Experiment 3

Figure 3.6 Mean ICA values at three points during a sentence obtained from the two experiments (Exp. 2 and Exp. 3). Note the difference in the y-axis for the two plots.

For the reader's convenience and easier comparison, Figure 3.6 presents the two ICA plots from the two experiments together.

3.4.4 Discussion

Experiment 3 set out to quantify the influence that the visual context has on linguistic reference processing. Overall, we found a significant difference in cognitive load between the two studies, namely, a general increase in cognitive load in Experiment 3. This is an expected result since Experiment 3 includes the additional information from the visual modality that is constantly available during the presentation of the sentence.

The examination of participants' visual attention showed that the restrictive verbs inspired more inspection of the most plausible object. It was only after uttering the alternative possible referent that this object began being inspected more. The verb constraint inspired attention towards the most plausible target referent. Hence, we see a clear indication that an anticipation for the target object was created.

Moreover, the analysis of the ICA events shows an effect of facilitated integration of the target referent noun (*water*) in the restrictive context (*spill*). The unanticipated referent noun (*ice cream*) induced the highest cognitive load in the same context. However, in the non-restrictive context, the two nouns were equally easy to process. Also, *water* was more easily integrated in the context of *spill* than the non-restrictive *order*.

3.5 Chapter Discussion

Three experiments presented in this chapter were conducted in order to connect the information-theoretic estimates of word informativity and the related processing effort to language processing in a richer context including additional visual information. We aimed at quantifying the role of the visual information in entropy reduction, and the subsequent surprisal on the reference as captured by the processing effort it induces. We found evidence that visual information has an immediate effect on anticipations formed about the subsequent linguistic reference, which is reflected in its surprisal-based processing effort.

Experiment 1 measured the difficulty processing differently predictable linguistic references in a well-established method, namely, in a reading study. We expected that the more likely linguistic reference (given the previous linguistic context) would induce less processing effort. However, only the difference between the nonsensical and the highly plausible reference was reflected in the results, showing that our manipulation was very subtle. Experiment 2, as an intermediate step, tested the same stimuli, now auditorily presented, and assessed the processing effort with the ICA measurement. The results were comparable with those of the previous study. Without the presence of a visual context, the ICA also did not pick up on the difference between the possible and plausible reference. Finally, Experiment 3 presented the same linguistic stimuli together with a relevant visual context. We saw that the existence of the visual context was reflected on the ICA, both in the overall effort induced by processing a sentence, as well as in the effort required for the processing of the linguistic reference.

Our findings suggest that the presence of the visual context increases the overall cognitive effort during sentence processing. This is an expected finding, since the visual modality introduced an additional source of information that was constantly present during a trial. Moreover, we found that the presence of the relevant visual context has an effect on the referent noun processing. Constraining verbs inspired anticipatory eye movements towards the most plausible referent objects (a replication of Altmann & Kamide, 1999). Interestingly, this did not induce higher processing load at the restrictive linguistic unit, namely the verb. However, it indeed resulted in the facilitation of the processing of the actual referent noun.

A potential reason for the lack of effect on the verb is that listeners did not actively exclude presented objects which did not fit the verb selectional preferences. Rather, the participants seem to have "only" attributed higher salience to the fitting objects. Hence, the fact that Maess et al. (2016) found constraint effects on the verb (MEG findings), and we did not, is potentially due to our verbs not being as highly restrictive compared to the verbs used in their study.

Alternatively, visually facilitated entropy reduction does not require higher processing effort, in contrast to highly surprising linguistic units. A linguistic unit that is unlikely

to appear in a certain linguisitc context could concievably induce higher processing effort upon appearing than a linguistic unit that constrains the following unit, but is in itself not unexpected given its previous context. In relation to that argument, it is possible that surprisal and entropy reduction require different congitive processes– hence, different kinds of activity – and that the ICA is sensitive to the former, but does not index the latter.

Finally, our findings allow us to conclude that surprisal, and its associated cognitive load, are not determined solely by the linguistic information, but rather, they reflect the expectations that are also derived from the additional consideration of the visual context in which the utterance is embedded.

In summary and in light of the initially raised research questions, we argue the following. We were able to quantify the role of the visual context in sentence processing, by seeing that in the presence of a relevant visual context, the cognitive effort required for processing a sentence rises on average. Importantly, however, the visual context allows for anticipation of the potential reference, which results in the facilitation of the processing of the referent noun, when the anticipated referent is mentioned. In comparison, the cognitive load induced by processing a possible but less likely reference is then respectively higher. Finally, even though the restrictiveness of a verb inspired a shift in visual attention towards the likely referent, the processing of the constraining verb did not prove more costly than the processing of a non-constraining verb that did not allow for creating anticipations about the target. Hence, the reduction of effort required for processing the anticipated reference was not compensated for by an increase in load at the point of the restrictive verb, when the anticipation was created.

The presented series of experiments shows that the ICA is a measure sensitive enough to reflect the combination of linguistic and additional non-linguistic information from the visual modality. In the following chapter, we introduce the referential gaze cue as an additional visual cue which ties both modalities even closer together. Being a cue that is related to the speaker and hence tightly connected with language production, gaze presents another link between linguistic and visual processing. We employ the direction of referential eye gaze as a visual cue tightly coupled with language, and examine how the cuing of an object in the visual scene affects the processing of the linguistic reference to that object.

Chapter 4

Qualitative Differences in Gaze Cuing

In the previous chapter we quantified the role of visual context in the processing of linguistic material. This chapter introduces a visual cue as an additional constraining element that potentially inspires referent anticipation. We aimed at quantifying its role, as well as examining the way in which it is perceived by the listener. As discussed in Chapter 1 (p. 1) and Chapter 2 (p. 11), we use the presentation of referential speaker gaze as the appropriate visual cue for the purposes of the present examination. Note that the work presented in this chapter has been published in the journal of Cognitive Science (Sekicki & Staudte, 2018).

Previous research has examined what is being activated and predicted during visually situated language processing (e.g., Altmann & Kamide, 1999), whether gaze creates a shift in viewers' attention (Friesen & Kingstone, 1998), its effect on listeners' language comprehension (Hanna & Brennan, 2007), and how beneficial such a cue is for processing subsequent linguistic material (e.g., Staudte & Crocker, 2011). A referential speaker gaze cue, as a visual pointer, directs listeners' visual attention to the cued object, which was consequently shown to be facilitatory, as measured by listeners' performance on post-trial tasks.

The present work presents a group of three studies that pioneer in measuring directly online both listeners' visual attention, as well as cognitive effort induced (a) by perceiving the speaker's referential gaze cue, and (b) by processing the linguistic reference.

Please note that when mentioning the load induced by *the gaze cue itself* (mentioned in (a)),we do not mean the load of perceiving the face, or the movement of the eyes, but rather, the load induced by considering the object cued by the gaze. We hypothesized that the information gathered from both visual and linguistic cues is incrementally combined to create anticipation for upcoming linguistic material. Since the same piece of information is conveyed by the gaze cue and the referent noun, gaze being the first to appear, we expected it to induce more cognitive load, which would consequently reduce the load on the linguistic

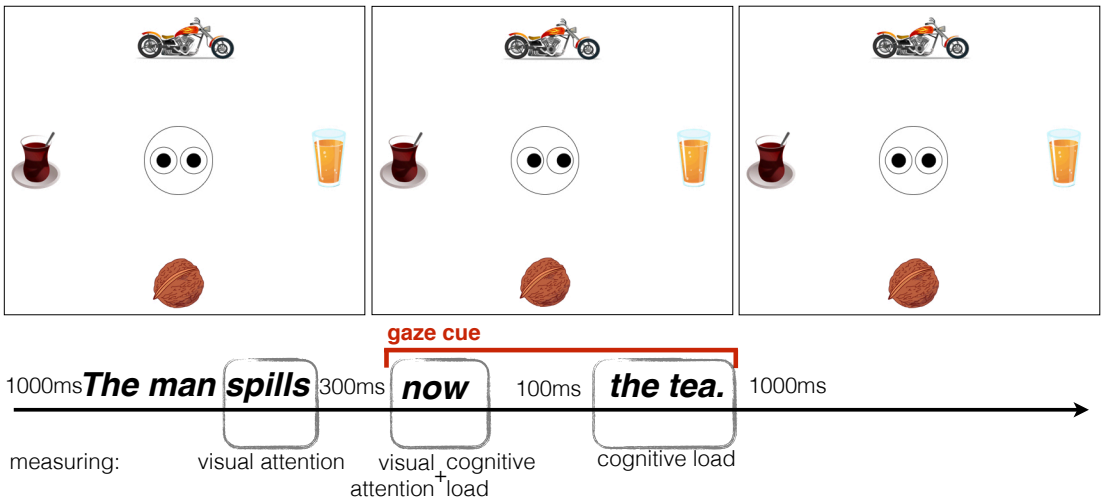


Figure 4.1 Example timeline illustrating an incongruent gaze cue (Exp. 5). Note that the sentence shown is a literal translation of the German sentence used for the experiment, and preserves the exact word order.

reference, and thus, result in a distribution of cognitive effort between the two modalities. In addition, we were interested in examining the facilitatory effect of the gaze cue in atypical contexts created through either a mismatch with verb selectional preferences, or incongruency with the subsequent linguistic reference.

The present experiments, conducted in the VWP, made use of controlled visual items paired with simple SVAdvO sentences in the German language, and with the gaze cue that was represented by a simple line drawing. The gaze cue was presented together with the adverb, that is, after the verb had been introduced and before the introduction of the referent noun. Since the gaze cue was presented simultaneously with the linguistic material, in a manner in which speakers usually use their gaze cue, we believe that this inspires the perception of the cue as being related to the speaker. Post-experimental debriefing showed that participants indeed anthropomorphized the presented drawing, perceiving it as a face.

As illustrated by Fig. 4.1, we examined visual attention at the verb – the point where verb selectional preferences allow for discarding some of the depicted objects and activating those with the highest probability to appear in the continuation of the sentence. Second, we considered visual attention at the point of introducing the gaze cue – examining whether it is being followed, causing a shift in visual attention. In addition, being interested in whether perceiving an object after such an attention shift is costly, we measured immediate cognitive load at the gaze cue. Finally, we considered cognitive load induced by the referent noun.

First, we manipulated the existence of the referential speaker gaze cue and assess its effect on referent noun processing (Experiment 4). Also, we measured cognitive load on the gaze cue itself, expecting a distribution of load between the cue and the reference, such that the reduction of cognitive load on the referent noun is preceded by its increase on the gaze cue. Thus, we examined not only the effect that gaze following may have on processing the subsequent reference, but also the cognitive load required for the shift in visual attention, where both linguistic and visual information are being considered.

Second, in order to better understand the mechanisms behind gaze perception and its integration with language processing, we manipulated the context fit of the gaze cue (Experiment 5), and its congruency with the following referent noun (Experiment 6). We expected that the gaze cue would induce significantly higher cognitive load when it did not match the previous linguistic context, but that it would nevertheless still facilitate processing of the referent noun. Finally, the incongruent gaze cue was expected to be costly for the processing of the linguistic reference.

In light of the second overarching question posed in Chapter 1 (pg. 6, namely, *Does a referential visual cue affect the cognitive load required for processing the corresponding linguistic reference*, the three experiments presented in this chapter focus on answering the following concrete questions:

1. Does a referential gaze cue affect the predictability of a linguistic reference?

2. If so, is the cognitive load (surprisal) induced at the linguistic reference consequently affected?

   a) How does a gaze cue affect cognitive load at the following reference when the cued object does not fit the previous linguistic context?

   b) Does a gaze cue that fits the previous linguistic context but is incongruent with the following reference affect the cognitive load at the referent noun?

3. Is a gaze cue costly in itself?

   a) Is the gaze cue to an object that does not match (the context) costly?

4. Is there a distribution of cognitive load between the gaze cue and the referent noun, since they both select the same target referent?

Experiment 4 aims to answer the four main questions, while the other two experiments address the more specific sub-questions, by firstly replicating the findings of Experiment 4 in a slightly different setting, and secondly, giving more detailed answers to the main research questions. Experiment 4 shows the facilitatory effect of the gaze cue on processing the

subsequent referent noun. Experiment 5 challenges this finding, and examines if facilitation of gaze holds even when it does not fit the linguistic context. And finally, Experiment 6 examines the effects of gaze congruency, that is, if gaze cuing to an object that is later not mentioned induces cost on the referent noun, even though the reference fits the context.

Please note that in the context of the present work, we use the word *gaze* to refer to the referential gaze cue presented by the stylized face (speaker). In contrast, *inspections* and *fixations* are used to refer to the listeners' visual attention.

4.1 Experiment 4

This study aimed to examine whether employing the ICA measurement supports previous findings that the gaze cue is actively considered in language processing, by quantifying online how the existence of the gaze cue modifies the cognitive load induced by the linguistic reference. More importantly, we were interested in measuring the potential cost of gaze perception and the distribution of cognitive load between the gaze cue and the linguistic reference.

The gaze cue used in this study was always **fitting** (with the previous linguistic context) and **congruent** (cuing to the object subsequently referred to by language). We manipulated the existence of the gaze cue in order to answer the four main research questions (listed on p. 55).

We expected gaze to be followed, and thus to inform anticipation of the object that is likely to be mentioned next. Consequently, this was expected to lead to lower cognitive load when the anticipated object was finally referred to. In addition, we expected the perception of the gaze cue not to be costly as such, but that higher cognitive load would be measured on the gaze cue when it is not consistent with the expectation already created based on the previous linguistic context (verb). Finally, this was expected to lead to a reduction of cognitive load on the referent noun, since the effort required to process an unexpected reference would have been distributed between the gaze cue and the linguistic reference.

4.1.1 Method

The study made use of $2 \times 2 \times 2$ mixed factorial design. The independent variable **Gaze** (no-gaze vs. referent gaze) was a between-subjects variable, that is, half of the participants were presented with the version of the experiment where all items included the gaze cue, while the other half saw the version with item trials never having the gaze cue. Fillers balanced the gaze conditions in the experiment as a whole to the ratio of 1:1, so that all participants saw

gaze and no-gaze trials. In addition, four linguistic conditions were created with two within-subjects variables, **Constraint** (restrictive vs. non-restrictive) and **Plausibility** (plausible vs. possible). As was the case in Experiment 2 and Experiment 3, Constraint was manipulated by verb restrictiveness (*spill* vs. *order*), and Plausibility by noun fit with the restrictive verb (*spill water* vs. *spill ice cream*). The language of the experiment was German.

Participants

64 Saarland University students took part in this study (45 women) and were monetarily reimbursed for their participation. Their ages ranged from 18 to 34 years old ($M = 24.16$). Participants were all native speakers of the German language with normal or corrected-to-normal vision. None of the participants was familiar with the previously conducted experiments.

Items

Each participant was presented with 20 items and 30 fillers, consisting of static visual scenes and auditorily presented linguistic stimulus. In addition, visual scenes included a face-like object forming the gaze cue.

The linguistic stimuli used for this experiment are identical to those used in Experiment 2 and Experiment 3 from the previous chapter. Hence, we will mention them only briefly here, for readers' convenience. For a detailed description please refer back to Chapter 3.

We used syntactically simple sentences with the SVAdvO structure. An item would include either a restrictive (*spill*) or a non-restrictive (*order*) verb. Next, an adverb (*gleich* – literal translation: *now*) was included as a spill-over region between the verb and the referent noun, where no relevant additional information was introduced. Rather, time was given for processing the verb and creating potential anticipations for the referent noun. The adverb was kept the same for all item sentences. In addition, this was the region where the gaze cue was introduced, making the point of the gaze cue subsequent to the constraining information introduced by the verb and prior to the resolution brought about with the referent noun. Finally, the object noun was introduced. Two linguistic conditions were created based on the noun's semantic fit in relation to the restrictive verb (*spill water* vs. *spill ice cream*).

In addition to the linguistic stimuli, the experiment presented visual displays with four concrete objects. Two of the four objects fit the category introduced by the restrictive verb (*spill*: *water*, *ice cream*), while all four fit the non-restrictive verb (*order*: *water*, *ice cream*, *suitcase*, *coat*). The position of target and competitor objects was rotated to all possible positions (individual and mutual).

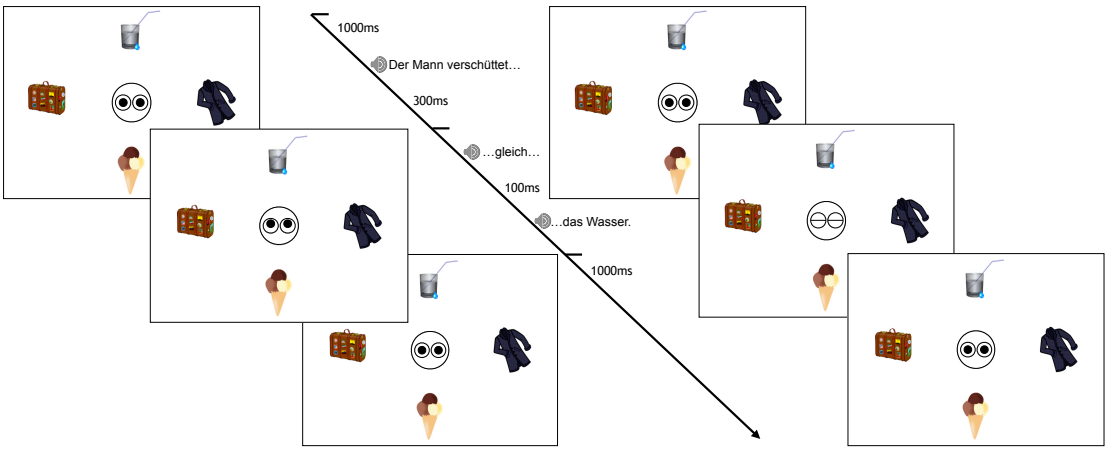


Figure 4.2 Exp. 4 – Trial timeline example: referent gaze condition (left) and no-gaze condition (right).

As mentioned, the referent noun always fit within the previous linguistic context and the gaze cue was always congruent, that is, cuing the object that was about to be mentioned. The main manipulation of the study was the presence of the gaze cue introduced before the target object was referred to linguistically.

Fig. 4.2 presents a trial timeline. The left side of the figure illustrates the referent gaze condition and the right side the no-gaze condition. The sentence and the timeline are given in the middle, since they are identical for both gaze conditions. The visual scene with open eyes was presented 1000 ms prior to sentence onset. The gaze cue (or closed eyes) was introduced 300 ms after the verb offset, that is, it was present from adverb onset to sentence end. Finally, after the end of the audio presentation, the eyes would look straight ahead for another 1000 ms.

In the no-gaze condition, the eyes closed in the relevant time-window, since we wanted to avoid introducing any additional information, but introduce the same amount of change in the visual context as was present in the referent gaze condition. The full list of visual stimuli used for this experiment is given in Appendix B, Table B.2.

Fillers

Fillers included the same visual setting, but differed in sentence structure and complexity of the linguistic stimulus, as well as the number of objects that fit the verb category. For instance, the sentence: *Bei der Großmutter gibt es immer leckere und hausgemachte Nudeln* (literal translation: *At Grandma's there is always tasty and homemade pasta*) was presented with images of a chocolate, a cookie, a muffin and a plate of pasta.

30 fillers were used, 25 of which had the opposite gaze condition as the items (5 had the same). In total, we created a ratio of 1:1 for gaze and no-gaze conditions during the

experimental session. 19 fillers were followed by simple yes/no comprehension questions that were answered by a key-press. The questions were related exclusively to the linguistic content. This was done in order not to inspire extensive inspection of the visual scene, but rather so that participants would consider the scene only optionally and freely in addition to the linguistic information.

The Latin square was used in order to create four lists with 20 item and 30 filler trials each. Each list included each item in only one of the four linguistic conditions. The trials were pseudo-randomized manually and an additional four lists were created with the opposite order of presentation. 32 participants (one half of the total number) saw the items in the referent gaze condition, while the other half of the participants saw them in the no-gaze condition.

Procedure

We used an EyeLink II head-mounted eye-tracker (SR Research, Ltd.; Mississauga, Ont., Canada) and tracked both eyes at a sampling rate of 250 Hz.[1] The tracker was manually adjusted, calibrated and validated using a 9-point fixation stimulus. Participants were instructed to listen carefully to the sentences while looking freely at the presented scenes. They would advance the experiment by a key press after each item. This allowed them to take a break whenever they needed one. Two additional keys were used to answer the yes/no comprehension questions. A practice session of three trials preceded the experimental part. After having familiarized themselves with the experimental setup, the participants would continue to the experimental session by pressing a key on the keyboard. The experiment lasted for approximately 15 minutes.

Variable Coding and Data Analysis

The variables used in this experiment, their coding and the data analysis are almost identical to those from Experiment 3 (see 44). Hence, here we will only mention the specific differences for this experiment.

In the analysis of the new inspections, we consider two temporal regions of interest: (a) Verb region of interest: showing how the verb category influenced visual attention, hinting at an object being considered as best fitting the verb constraint; and (b) Gaze region of interest: showing if the gaze cue inspired an immediate shift in visual attention.

As for the analysis of the ICA events, two time-windows were of relevance: (a) Gaze time-window: 600 ms from the onset of the gaze cue, that is, while the adverb was heard,

[1] This is the required setup for the extraction of the ICA events from EyeWorks Workload Module software (version 3.12).

showing the cost of considering the gaze cue; and (b) Reference time-window: 600 ms from the middle of the referent noun, showing the cost of processing the linguistic reference.

4.1.2 Results

Proportion of Fixations

Fig. 4.3 illustrates the proportions of fixations to all AoIs during a trial. Four linguistic conditions are presented both without the gaze cue (left-hand side) and in the referent gaze condition (right-hand side). The plots are aligned to the gaze cue onset, represented by the solid line. Note that the eyes would close at this point in the no-gaze condition. In addition, the dashed line presents the onset of the article from the object noun phrase.

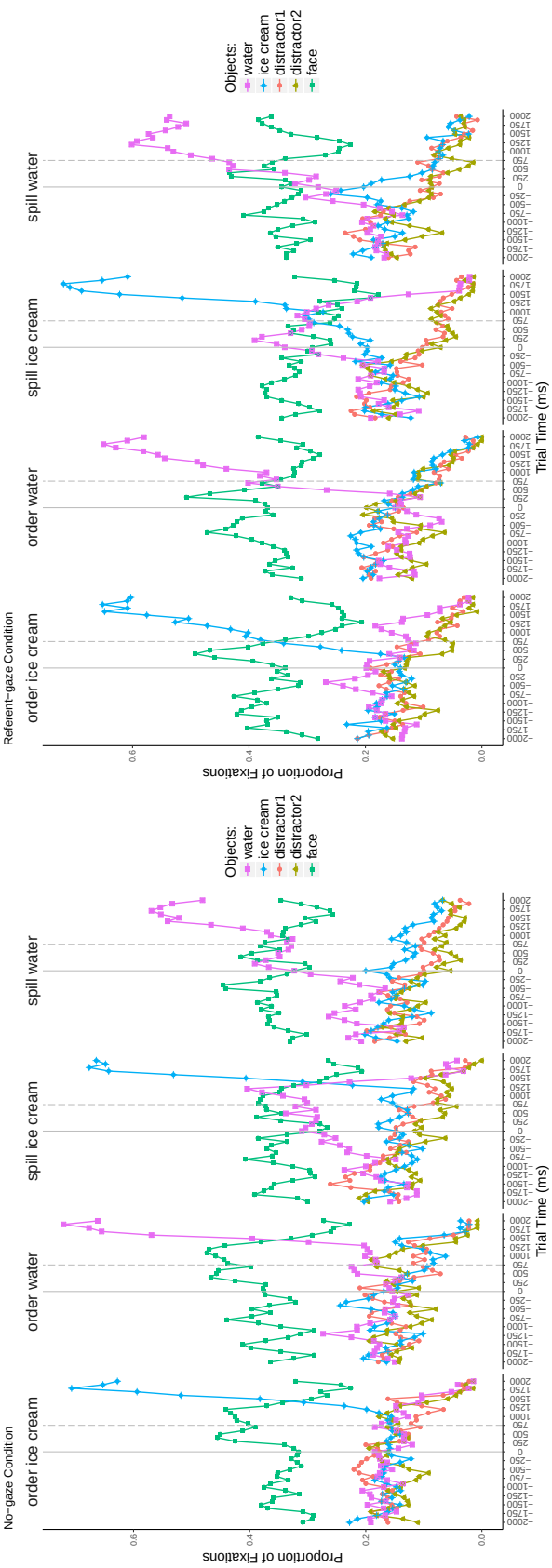
It is evident that the restrictive verb (*spill*) shifted the focus of visual attention to one particular object (*water*). The less fitting object (*ice cream*) was considered only upon being referred to linguistically (no-gaze condition), or earlier, at the point of the gaze cue (referent gaze condition). In the context of the non-restrictive verb (*order*) no preference for any of the objects was recorded until one of them was either referred to linguistically, or earlier, when it was gazed at. Thus, we found evidence that the visual attention is not only influenced by language, that is, the objects' fit with the verb category, but also by the gaze cue.

New Inspections

We considered the probability of an inspection to fall in an AoI in a region of interest (from verb onset or from gaze cue onset). All fixations that fell within one AoI and prior to a fixation outside of that AoI were grouped together and considered as one inspection. For the statistical analysis we treated new inspections as a binary variable, since trials with a new inspection to the AoI were given the value "1", and those without, a "0".

First, we conducted statistical analysis of new inspections of an object (water, ice cream, distractors) in the verb region of interest (200 ms from verb onset to gaze onset). Second, we considered new inspections of water and ice cream, in the gaze region of interest (to article onset). The main models and collected results are given in Table 4.1. Detailed information about the results of the further comparisons is given in Table B.4 (Appendix B). Also, the results are illustrated by Fig. B.7 (Appendix B).

Considering the verb region of interest, we ran separate models for the inspections of relevant objects. The analysis of target inspections (water) showed a main effect of Constraint ($p < 0.001$) suggesting that more new inspections of water occurred upon hearing *spill* (M = 0.156, SD = 0.363) than *order* (M = 0.105, SD = 0.307). Competitor inspections (ice cream) showed no such effect (p = 0.139) suggesting that ice cream was looked at with no significant



(a) Four linguistic conditions of the no-gaze condition

(b) Four linguistic conditions of the referent gaze condition

Figure 4.3 Exp. 4 – Proportion of fixations aligned to the gaze cue onset (solid line). The dashed line presents article onset of the object noun phrase.

Table 4.1 Exp. 4 – Results of the main models fitted for the new inspections analysis for both verb and gaze regions of interest.

1. Verb region of interest

	a) Target inspections				b) Competitor inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.928	0.065	-29.704	< 2e-16 ***	-2.050	0.068	-29.97	<2e-16 ***
CONSTRAINT	-0.458	0.104	-4.406	1.05e-05 ***	-0.168	0.114	-1.48	0.139

	c) Distractors inspections			
INTERCEPT	-1.446	0.074	-19.567	< 2e-16 ***
CONSTRAINT	0.433	0.095	4.578	4.69e-06 ***

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1 a) WaterInsp ~ Constraint + (1 + Constraint || Subject) + (1 + Constraint || Item), family = "binomial"
1 b) IceInsp ~ Constraint + (1 + Constraint | Subject) + (1 + Constraint | Item), family = "binomial"
1 c) DistractInsp ~ Constraint + (1 + Constraint | Subject) + (1 + Constraint | Item), family = "binomial"

2. Gaze region of interest

	a) Target inspections				b) Competitor inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.820	0.063	-28.788	< 2e-16 ***	-2.239	0.085	-26.331	< 2e-16 ***
GAZE	0.265	0.123	2.155	0.031 *	0.305	0.154	1.987	0.047 *
PLAUSIBILITY	-0.454	0.133	-3.420	0.001 ***	0.972	0.172	5.663	1.48e-08 ***
GAZE:PLAUS	-0.835	0.266	-3.143	0.002 **	0.817	0.316	2.587	0.010 **

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2 a) WaterInsp ~ Gaze * Plausibility + (1 + Gaze * Plausibility || Subject) + (1 + Plausibility || Item), family = "binomial"
2 b) IceInsp ~ Gaze * Plausibility + (1 + Gaze * Plausibility || Subject) + (1 + Plausibility || Item), family = "binomial"

Table 4.2 Exp. 4 – Results of the main models fitted for the ICA analysis.

	a) Gaze time-window				b) Reference time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.881	0.029	98.03	<2e-16 ***	2.915	0.026	110.29	< 2e-16 ***
GAZE	0.005	0.055	0.09	0.925	-0.113	0.051	-2.21	0.027 *
CONSTRAINT	0.004	0.016	0.25	0.802	0.038	0.024	1.59	0.111
PLAUSIBILITY	0.017	0.018	0.97	0.332	0.058	0.027	2.11	0.034 *
CONST:PLAUS	-0.060	0.026	-2.28	0.021 *	-0.184	0.042	-4.34	1.4e-05 ***
HALF	-0.031	0.0134	-2.31	0.016 *	-0.052	0.022	-2.39	0.017 *
HALF:GAZE	-	-	-	-	-0.027	0.043	-0.61	0.539

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICA ~ Const * Plaus + Half + Gaze + (1 + Const + Plaus || Subject) + (1 + Const + Plaus || Item), family = Poisson (link = "log")
b) ICA ~ Const * Plaus + Half * Gaze + (1 + Const * Plaus || Subject) + (1 + Const * Plaus || Item), family = Poisson (link = "log")

difference in the contexts of the two verbs. Finally, distractor inspections show a main effect of Constraint ($p < 0.001$), which suggests that there were more new inspections of the two distractors in the context of *order* ($M = 0.234$, $SD = 0.428$) than *spill* ($M = 0.169$, $SD = 0.375$).

Furthermore, we considered the gaze region of interest, showing a shift in attention inspired by the gaze cue. The model including new inspections of water reveals main effects of Gaze ($p = 0.031$), Plausibility ($p = 0.001$), and most importantly, the Gaze:Plausibility interaction ($p = 0.002$). Further comparisons show that in the subset of the plausible target *water*, there is a main effect of Gaze ($\beta = 0.648$, $SE = 0.178$, $z = 3.641$, $p < .001$), while in the subset of the possible target *ice cream* there was no such effect ($p = 0.415$), suggesting that water was inspected more when cued at (ref. gaze: $M = 0.224$, $SD = 0.417$ vs. no-gaze: $M = 0.128$, $SD = 0.334$).

The model with new inspections to ice cream also showed main effects of Gaze ($p = .047$), Plausibility ($p < .001$), and a Gaze:Plausibility interaction ($p = .010$). Further comparisons show that in the subset of *water* there was no effect of Gaze ($p = 0.688$). Such an effect was present in the *ice cream* subset ($\beta = 0.721$, $SE = 0.187$, $z = 3.853$, $p < .001$), suggesting that ice cream was more readily inspected when it was gazed at (ref. gaze: $M = 0.203$, $SD = 0.402$ vs. no-gaze $M = 0.112$, $SD = 0.315$).

The Index of Cognitive Activity

The ICA findings are illustrated by Fig. 4.4. We present the no-gaze condition on the left-hand side and the referent gaze condition on the right-hand side. The x-axis presents the 600 ms time-windows for the four sentence parts. The two time-windows relevant for the analysis are the points labeled as *Adverb* (Gaze time-window), and *Object* (Reference

window). Each point on the graph represents the mean value of the ICA events per condition for a 600 ms time-window. We clearly observe no difference at the point of the Adverb, while significant differences exist at the point of the Object both among the linguistic conditions and between the two gaze conditions. The main models and collected results are given in Table 4.2. Detailed information about the results of the further comparisons is given in Table B.5 (Appendix B).

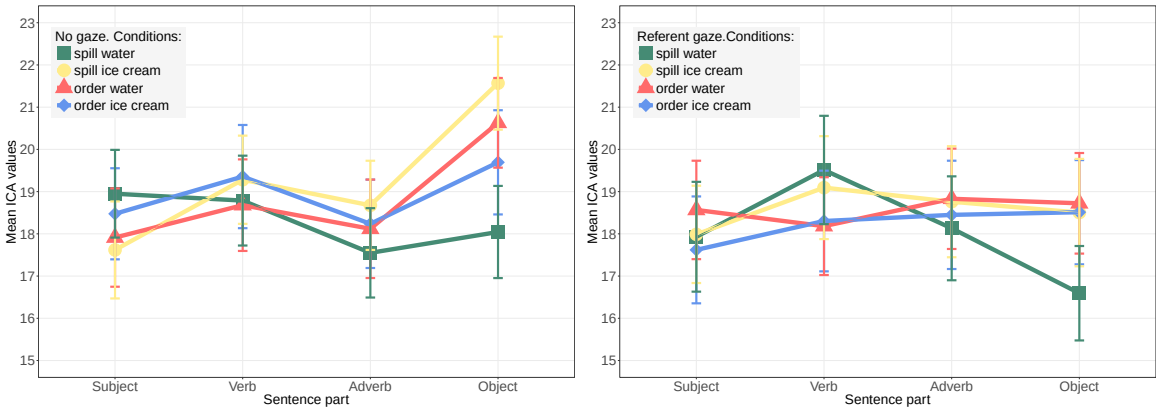
The analysis of the gaze time-window revealed a main effect of Half[2] ($p = 0.016$) and a Constraint:Plausibility interaction ($p = 0.021$). However, further comparisons show that this interaction is carried by the opposite trend of the two Plausibility levels in the subsets of the two verbs, which remain far from statistical significance (subset *spill*: $p = 0.104$; subset *order*: $p = 0.624$).

Considering the reference window, we found a main effect of Gaze on cognitive load ($p = 0.027$), suggesting that the presence of the gaze cue led to the reduction of load on the subsequent reference (no-gaze: $M = 19.983$, $SD = 7.296$ vs. referent gaze: $M = 18.08$, $SD = 7.723$). Moreover, we found a significant Constraint:Plausibility interaction ($p < 0.001$), as well as a main effect of Plausibility on cognitive load ($p = 0.034$). Further comparisons show a main effect of Plausibility in the subset of *spill* ($\beta = 0.152$, $SE = 0.035$, $z = 4.37$, $p < 0.001$), suggesting that *spill water* ($M = 17.319$, $SD = 7.102$) induced less cognitive load than *spill ice cream* ($M = 20.041$, $SD = 7.746$). No such effect was found in the non-constraining subset ($p = 0.249$), suggesting no difference between *order water* and *order ice cream*. Finally, we found no Half:Gaze interaction ($p = 0.539$), but a main effect of experimental Half on cognitive load ($p = 0.017$), since the first part of the experiment induced higher load ($M = 19.516$, $SD = 7.553$) than the second part ($M = 18.599$, $SD = 7.563$).

4.1.3 Discussion

Our results replicate the findings from Experiment 3. The eye movement data suggest that the selectional preferences of the restrictive verb *spill* inspired creating a clear prediction for *water* to be mentioned in the continuation of the sentence. Consequently, *water* was easier to process in the context of *spill* (vs. *order*). Also, in the context of the verb *spill*, *water* proved to be easier to process than *ice cream*. In addition, with the introduction of the gaze cue, this effect of verb-based anticipations changed. The gaze cue towards an object led to the inspection of the cued object, even when the verb selectional preferences did not previously put it in the focus of visual attention. Consequently, the presence of the gaze cue led to the

[2]The independent variable Half is first relevant for the analysis of Exp. 5. Hence, it will be justified in the following section (see Section 4.2.2). Here, as well as in Exp. 6, Half is included in the analyses for the sake of consistency among the experiments.



(a) No-gaze condition.

(b) Referent gaze condition.

Figure 4.4 Exp. 4 – Mean ICA values at the four time-windows of a sentence. Note that *Adverb* (Gaze window) and *Object* (Reference window) are the two time-windows relevant for the analysis (95% CI error bars).

reduction of cognitive load on the referent noun in all linguistic conditions. Moreover, it leveled out the previously mentioned differences, except for *spill water*, the only condition where a clear prediction for one object was created. This condition required significantly less cognitive effort than the other three linguistic conditions. These findings are in line with previous literature (Hanna & Brennan, 2007; Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013, 2014; Staudte & Crocker, 2011; Staudte et al., 2014) and further support it by reporting immediate cognitive effort required for processing a reference.

Moreover, we were able to expand the current understanding of gaze perception by assessing the immediate cognitive load it induces. Interestingly, even though it created a shift in visual attention to an object often not previously considered, this did not induce immediate cognitive load. Moreover, we expected the reduction of cognitive load on the referent noun to be preceded by an increased cost at the point of the gaze cue – to compensate for the lower effort on the reference. Such a distribution of cognitive load was not found, since no additional effort was measured on the gaze cue.

4.2 Experiment 5

This experiment challenged the conclusions from the first study by employing a mismatching gaze cue. In Experiment 4, we compared a plausible referent with a possible one (low plausibility), which induced differences in cognitive load on the linguistic reference, but not on the gaze cue itself. The present study builds on these findings by exaggerating the difference between the two conditions and employing an impossible object that does not fit the linguistic context. Firstly, we examined whether the gaze cue helps reduce cognitive load on the linguistic reference even when they both do not fit the previous context, and, secondly, whether the cue to such an object is more costly, since it violates the context.

We aimed to answer the following research questions:

2 a) How does a gaze cue affect cognitive load on the following reference when the cued object does not fit the previous linguistic context?

3 a) Is the gaze cue to a mismatching object costly?

We expected mismatching gaze to be surprising, and thus in itself more costly. Consequently, though, this was expected to lead to a reduction in cognitive load on the corresponding linguistic reference.

4.2.1 Method

The experiment made use of 2×2 experimental design, combining Gaze (no-gaze vs. referent gaze) and referent noun Fit (fitting vs. mismatching). Only restrictive verbs were used (*spill*), combined with either a thematically fitting (*water*) or mismatching referent noun (*sausage*).

Participants

36 participants (23 female), all students at Saarland University, took part in the study and were monetarily reimbursed for their participation. Their ages ranged from 18 to 34 years (M = 23.36). None of the participants were familiar with previously conducted experiments. Two students were excluded from the analysis due to technical issues, and two others because their mother tongue was established to be Luxemburgish. Thus, the data from 32 participants, all native German speakers, were analyzed.

Items

20 items were created. The sentence structure was the same as in Experiment 4. This time, however, only the restrictive verbs were used (*spill*). The linguistic manipulation considered

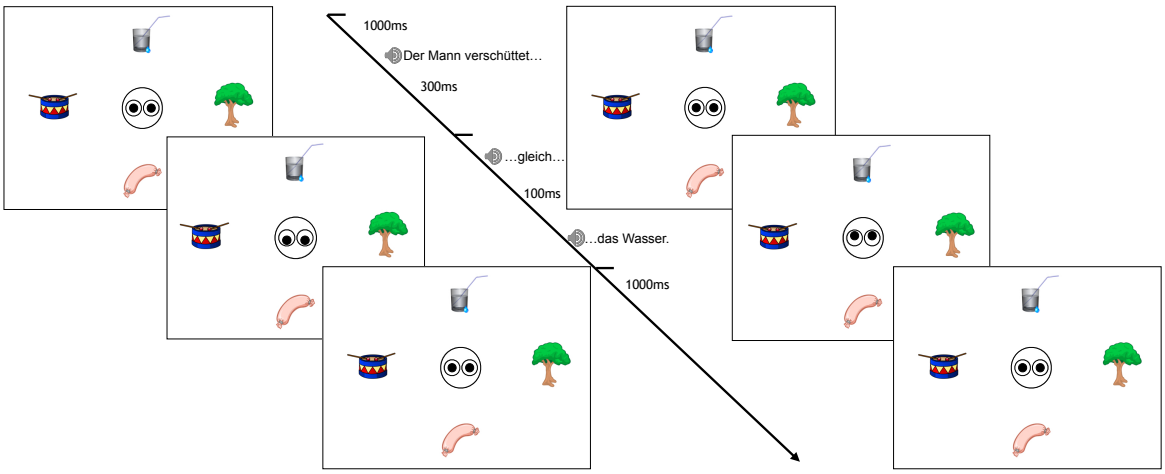


Figure 4.5 Exp. 5 – Trial timeline example: mismatching condition (left) and fitting condition (right).

the Fit of the referent noun to the category introduced by the verb. The referent could be fitting (*spill water*) or mismatching (*spill sausage*). In order to counterbalance the referent nouns, the items were run in two versions. Version A included the verb fitting to one noun in the item (*spill water* vs. *sausage*), while the verb in version B fit the other noun (*grill sausage* vs. *water*).[3]

Note that the gaze cue was, again, always congruent (cuing the object subsequently referred to); however, its fit with the linguistic context depended on the fit of the referent noun. When the referent noun did not fit the verb, that made the gaze cue to the object in question mismatching as well.

Visual scenes are of the same structure as in Experiment 4, but now only one presented object fits the verb category. A trial timeline is illustrated in Fig. 4.5, showing the referent gaze condition. On the left side of the figure is the mismatching condition, while the right side illustrates the fitting condition. Note that the no-gaze condition was also part of the experimental manipulation. As previously, the position of target and competitor objects was rotated to all possible positions (individual and mutual).[4].

Fillers

During the experiment, participants were presented with 75 trials in total, 55 of which were fillers. The fillers, again, differed from the experimental items in the complexity of the sentence structure and the number of presented objects that fit the verb category. For instance,

[3] All 20 items presented in two linguistic conditions each are given in Appendix B, Table B.1 and Table B.2 (for both experiment versions).

[4] An exhaustive list of all visual displays is given in Table B.4, Appendix B

the sentence: *Immer noch trägt der Vater seine Armbanduhr* (literal translation: *Still wears the father his watch*) presented with the images of a watch, a coat, a pocket flashlight and a clock. 35 filler trials were followed by a simple yes/no comprehension question referring only to the content of the heard sentence.

The gaze cue was present in 2/3 of all trials (10 items, 40 fillers). 10% of the total number of trials included anomalous sentences (10 items, 5 fillers). Only 16% of all trials included an anomalous gaze cue, that is, gaze that was cueing a mismatching object (5 items, 3 fillers).

Latin square design was used to create 4 lists with one condition per item each. The trials were pseudo-randomized manually, and an additional four lists were created with the opposite trial order.[5] The experimental procedure was the same as in Experiment 1. The duration of the experiment was approximately 20 minutes.

4.2.2 Results

The same measures and analyses were conducted as in Experiment 4, except for the new inspections analysis, where only the gaze region of interest was considered, due to the differing experimental design of Experiment 5. Again, all independent variables were contrast coded for the statistical analyses, and all models included maximal converging random structure justified by the experimental design (Barr et al., 2013).

Proportion of Fixations

Fig. 4.6 shows the proportion of fixations to all presented objects during a trial. As previously, the plots are aligned to the onset of the gaze cue (marked with the solid line). In addition, the dashed line presents the onset of the article from the object noun phrase.

The verb (*spill)* shifts the focus of visual attention to one particular object (water). The mismatching object (sausage) is considered only upon being referred to linguistically (no-gaze condition), or earlier, at the point of the gaze cue (referent gaze condition). Thus, we observe the same pattern as in Experiment 4, namely, of the gaze cue shifting visual attention, on a par with the linguistic information.

New Inspections

Because of the unusual gaze cue (to sausage), which was, however, entirely congruent with the subsequent noun, we hypothesized that participants might get used to this over time and that this could change their perception and utilization of the cue. Hence, we include an

[5]This was done for both versions of the experiment.

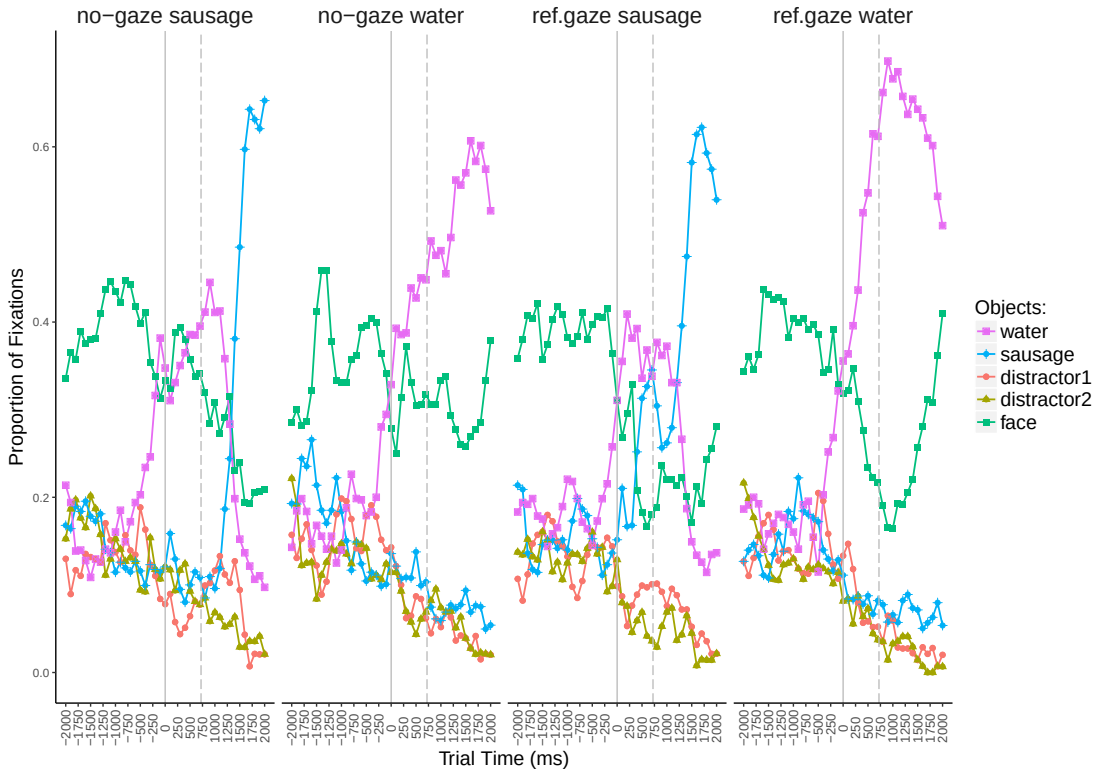


Figure 4.6 Exp. 5 – Proportion of fixations to presented objects in the four experimental conditions, aligned to the gaze cue onset (solid line). The dashed line presents article onset of the object noun phrase.

Table 4.3 Exp. 5 – Results of the main models fitted for the new inspections analysis (gaze region of interest.

	a) Water inspections				b) Sausage inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.617	0.098	-16.581	< 2e-16 ***	-2.434	0.126	-19.299	< 2e-16 ***
GAZE	0.129	0.156	0.826	0.409	0.555	0.224	2.476	0.013 *
FIT	-0.562	0.162	-3.467	0.001 ***	0.678	0.218	3.107	0.002 **
GAZE:FIT	-0.409	0.313	-1.308	0.191	1.199	0.436	2.749	0.006 **
HALF	0.138	0.155	0.890	0.373	-0.477	0.214	-2.226	0.026 *
HALF:GAZE	-0.005	0.310	-0.017	0.986	0.343	0.429	0.800	0.424

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) WaterInsp ~ Gaze * Fit + Half * Gaze + (1 + Fit || Subject) + (1 + Fit || Item), family = "binomial"
b) SausageInsp ~ Gaze * Fit + Half * Gaze + (1 + Gaze + Fit || Subject) + (1 + Fit || Item), family = "binomial"

additional variable in the analysis, namely, experiment Half, which includes the information of the experiment part in which a participant was presented with a given trial. Table 4.3 presents the main models and collected results. Detailed information about the results of the further comparisons is given in Table B.6 (Appendix B).

The model which included new inspections to water revealed only the main effect of Fit ($p = 0$ .001), suggesting that water was inspected more in the *spill water* condition ($M = 0.213$, $SD = 0.409$) than in *spill sausage* ($M = 0.133$, $SD = 0.34$). The lack of a main effect of Gaze ($p = 0.409$) and Gaze:Fit interaction ($p = 0.191$) suggests that water inspired more new inspections regardless of what, if anything, was gazed at.

The model with new inspections to sausage showed a Gaze:Fit interaction ($p = 0.006$), and main effects of Gaze ($p = 0.013$), Fit ($p = 0.002$), and Half ($p = 0.026$). Further comparisons show that in the *no gaze* subset, there was no effect of Fit ($p = 0.826$), while such an effect was present in the *referent gaze* subset ($\beta = 1.288$, $SE = 0.279$, $z = 4.618$, $p < .001$), suggesting that sausage was more readily inspected only when it was gazed at (ref. gaze: $M = 0.123$, $SD = 0.329$ vs. no-gaze: $M = 0.067$, $SD = 0.249$). Finally, the main effect of Half in the full model suggests that more new inspections of sausage were initiated in the first part of the experiment (1st part: $M = 0.113$, $SD = 0.317$ vs. 2nd part: $M = 0.082$, $SD = 0.274$).

In sum, new inspections of water were more likely both when there was no gaze cue, and when there was a referent gaze to water, while sausage was more likely to attract new inspections only when it was gazed at.[6] Considering the two halves of the experiment, the only observed difference lies in the new inspections of sausage; namely, there were more new inspections of sausage in the first part.[7]

The Index of Cognitive Activity

Table 4.4 presents the main models and collected results. Detailed information about the results of the further comparisons is given in Table B.7 (Appendix B).

We first considered the reference time-window and found a main effect of Fit on cognitive load ($p < 0.001$), suggesting that the anomalous *spill sausage* required more load ($M = 18.410$, $SD = 6.983$) than *spill water* ($M = 14.959$, $SD = 6.928$). Considering the effect of the gaze cue on the cost of the referent, a Half:Gaze interaction was observed ($p = 0.030$).

[6] Illustrated by Fig. B.8 in Appendix B.

[7] Fig. B.9 and Fig. B.10 in Appendix B show that while new inspections of water are patterned similarly in the two parts, new inspections to sausage are slightly reduced in the second half of the experiment. Thus, even though the gaze cue constantly led to a shift in visual attention, in the second half, the inspection of the fitting object, water, was not as inhibited after the gaze to sausage as was the case in the first part. That is, in the second half, upon a gaze cue to sausage a new inspection was equally likely to be directed to either sausage or water.

Table 4.4 Exp. 5 – Results of the main models fitted for the ICA analysis.

	a) Gaze time-window				b) Reference time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.773	0.038	72.90	<2e-16 ***	2.782	0.036	76.32	< 2e-16 ***
GAZE	0.062	0.034	1.86	0.063 .	-0.020	0.034	-0.58	0.562
FIT	0.037	0.034	1.09	0.275	0.222	0.045	4.94	7.76e-07 ***
GAZE:FIT	0.092	0.054	1.70	0.090 .	-0.022	0.050	-0.44	0.664
HALF	-0.040	0.039	-1.03	0.304	-0.052	0.035	-1.51	0.131
HALF:GAZE	0.069	0.059	1.16	0.246	-0.131	0.061	-2.17	0.030 *

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICA ~ Gaze * Fit + Half * Gaze + (1 + Gaze * Fit + Half * Gaze || Subject) + (1 + Fit | Item), family = Poisson (link = "log")
b) ICA ~ Gaze * Fit + Half * Gaze + (1 + Gaze * Fit + Half * Gaze || Subject) + (1 + Fit | Item), family = Poisson (link = "log")

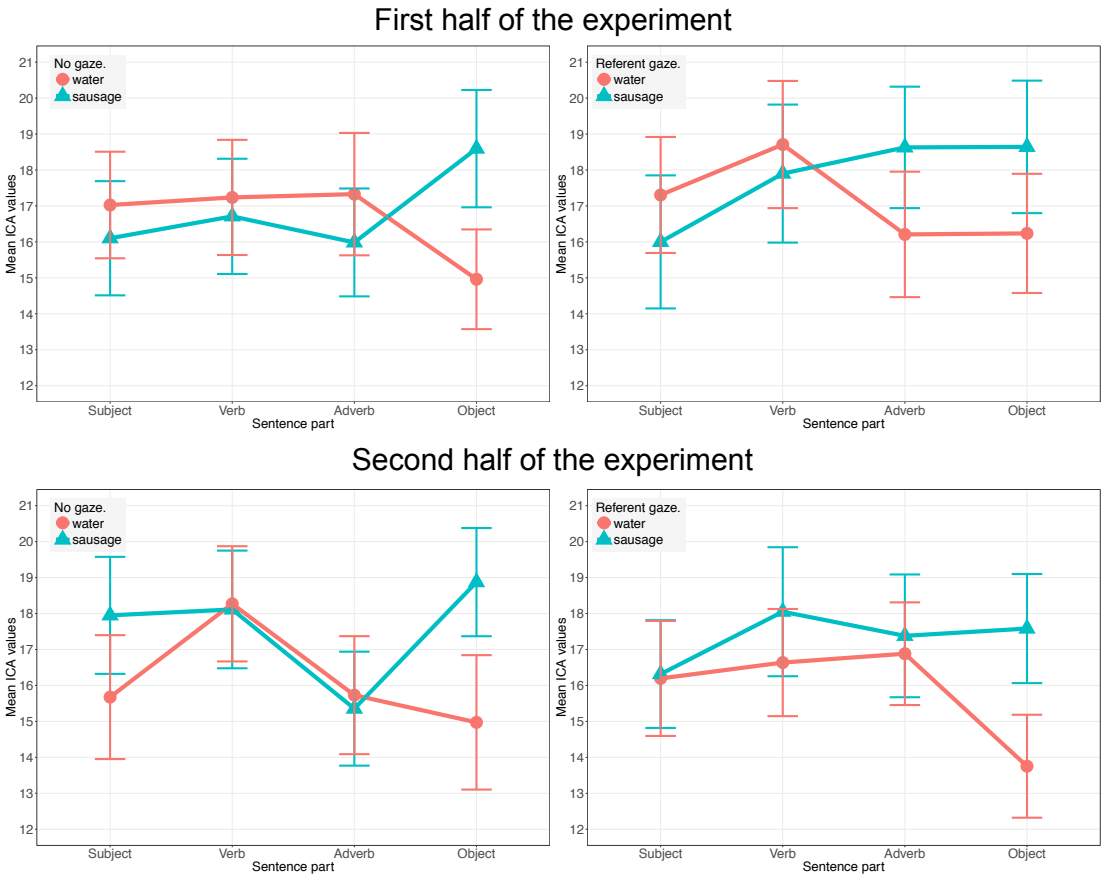


Figure 4.7 Exp. 5 – Mean ICA values in the four time-windows of a sentence in the first (above) and the second (below) half of the experiment, and the no-gaze (left) and referent gaze (right) conditions. Note that *Adverb* (Gaze window) and *Object* (Reference window) are the two time-windows relevant for the analysis (95% CI error bars).

Further analysis showed a marginal main effect of Gaze in the second half of the experiment (β = -0.091, SE = 0.047, z = -1.93, p = 0.054), suggesting that the referent gaze reduced cognitive load on the referent noun in both linguistic conditions (ref. gaze: M = 15.692, SD = 6.804 vs. no gaze: M = 16.987, SD = 7.631). No such effect was found in the first part (p = 0.392).

Since gaze affected referent processing in each experimental half differently, we also considered cognitive load on the cue itself (gaze window) for each experimental half separately. The first part of the experiment revealed a Gaze:Fit interaction (β = 0.179, SE = 0.084, z = 2.13, p = 0.033). Further comparisons required to interpret the interaction were not conducted due to power issues with too-small subsets. Nevertheless, Fig. 4.7 shows that the interaction is the result of the cognitive load induced by the anomalous gaze cue to *sausage* being higher than that resulting from the gaze cue to *water*, while such a difference between the two conditions is non-existent in the no-gaze condition. Finally, in the second half we found a main effect of Gaze (β = 0.099, SE = 0.045, z = 2.20, p = 0.028) suggesting that the referent gaze induced higher cognitive load (M = 17, SD = 6.971) than the no-gaze condition (M = 15.536, SD = 7.058).

Fig. 4.7 illustrates the results. The top two plots show the data in the first part of the experiment, while the lower two illustrate the second experiment half. The left-hand side shows the no-gaze condition and the right-hand side the referent gaze condition. As in Fig. 4.4, the Adverb point represents the gaze time-window, and the Object point the reference time-window. The pattern in the no-gaze condition did not change in the two experimental halves, while the referent gaze condition yielded significantly different effects. We see that in the first part the anomalous gaze cue was inducing more cognitive load than the congruent referent gaze. In the second half, though, the gaze cue became more costly in general. As for the cost of the referent noun, initially, the existence of the gaze cue did not reduce cognitive load either on the congruent or the incongruent noun. In the second half, however, we observe a marginal facilitation effect caused by the previous gaze cue for the processing of both congruent and incongruent references.

4.2.3 Discussion

The eye movement data show that the linguistic context inspired clear prediction of only one of the depicted objects (water). In addition, we see evidence of gaze following, even when the cued object was not only not predicted, but worse, did not fit the verb (*sausage*). In the case of mismatching gaze cue, the cued object was inspected the most (sausage); however, the fitting and preferred object (water) was nevertheless not discarded and still inspected more than the two distractor objects. This is particularly evident in the second part of the

experiment, where the mismatching gaze cue inspired equal inspection of both cued and predicted objects, suggesting a slight change in gaze cue following that occurred during the experiment.

Considering the cognitive load results, initially, the existence of gaze did not have an effect on the cost of processing the referent noun. No benefit of the gaze cue was found even in the fitting referent gaze condition, where the predicted object (*water*) was cued. Regarding the effort required on the gaze cue, the load induced by the mismatching cue itself was higher than that induced in both fitting gaze and no-gaze conditions. In the second half of the experiment, cognitive load on the referent noun was marginally reduced due to the gaze cue; while the cue itself (to both fitting and mismatching objects) now induced higher cognitive effort.

We understand that participants gradually started adapting to, and relying on, the surprising gaze cue. Initially, the mismatching gaze cue caused "concern" that anomalies were possible, not only at the point of the gaze, but potentially also later on the reference, resulting in no reduction of cognitive load on the referent noun. Over time, gaze proved to always be congruent with the referent noun. Simultaneously, the cognitive load on the gaze cue rose. Since gaze reliably gave away reference information, this led to more alertness on the cue (increasing the load on the cue in general) and inspired making use of its informativity (slightly lowering the load on the reference). The anomalous condition proved to be costly. However, in the presence of the gaze cue, we observed a tendency for facilitated processing of the subsequent anomalous reference.

Comparing the results obtained from eye movement and cognitive load analyses, we see that gaze was followed and induced a shift in visual attention throughout the experiment, even though the cognitive load results suggest that it was initially not exploited to the same extent. Higher cognitive effort induced by the anomalous gaze cue in the first experiment half was paired with new inspections of both the cued (*sausage*) and the anticipated object (*water*), where more inspections were directed to the cued object. At the same time, no increase in load on the fitting gaze cue was paired with new inspections of only that object. In the second half of the experiment, however, the cognitive load increased on both types of referent gaze, while the new inspections still patterned differently. The fitting gaze cue led to, again, new inspections to only the anticipated object (*water*), while the mismatching gaze cue equally inspired inspections to both the cued object (*sausage*) and the anticipated one (*water*).

4.3 Experiment 6

Manipulating the reliability of gaze and language while giving equal prominence to both cues, Macdonald and Tatler (2014) found that language was the most disruptive cue in their study, when least reliable. Thus, they concluded that language is the dominant cue, preferred over gaze. In addition, helpful gaze benefited performance by speeding up both first fixations to the target, and reaction times on the task. When language was 100% accurate, the gaze cue was superfluous, albeit still followed, and when incongruent with language it would slow down performance. In sum, gaze was found to be disruptive when incongruent to more informative language, but only when the cue was established to be reliable.

Again, measuring immediate cognitive load, we follow up on our previous findings and aim to answer the following research question:

2 b) Would a gaze cue, fitting with the previous linguistic context but incongruent with the following reference, affect the cognitive load on the (also fitting) referent noun?

As illustrated in Fig. 4.8, participants were presented with four objects, two of which were of **equally** good fit with the verb category (*tea* and *juice*). Thus, anticipation for either of those two objects (gazed at vs. referred to) was to be created prior to the gaze cue. The main manipulation was the incongruent gaze cue, that is, cuing one fitting object (*juice*) while the other one is subsequently referred to (*tea*).

Two developments could be expected from this experimental setup. The gaze cue could be automatically followed, as was the case in the previous two experiments, and thus, create expectation for the cued object. Subsequently, when the other fitting object is actually referred to, this would lead to an increase in cognitive load at the referent noun (otherwise fitting the context, and not surprising in itself). Alternatively, the incongruent gaze cue might not induce cost on the referent noun, since no clear expectation was created before the gaze cue (two equally probable target objects presented), and since the gaze cue is sometimes incongruent throughout the experiment (15% of trials), it could be equally acceptable to have either the cued object or the other fitting object complete the sentence.

Macdonald and Tatler (2014) found that attentional effects of gaze are affected by its reliability. Thus, we made sure that the gaze cue was reliable overall (85% congruent with the reference) in the context of the experiment.

4.3.1 Methods

One independent variable with three levels was created, namely Gaze (no-gaze vs. target gaze vs. competitor gaze). The linguistic reference was always plausible and fitting with the

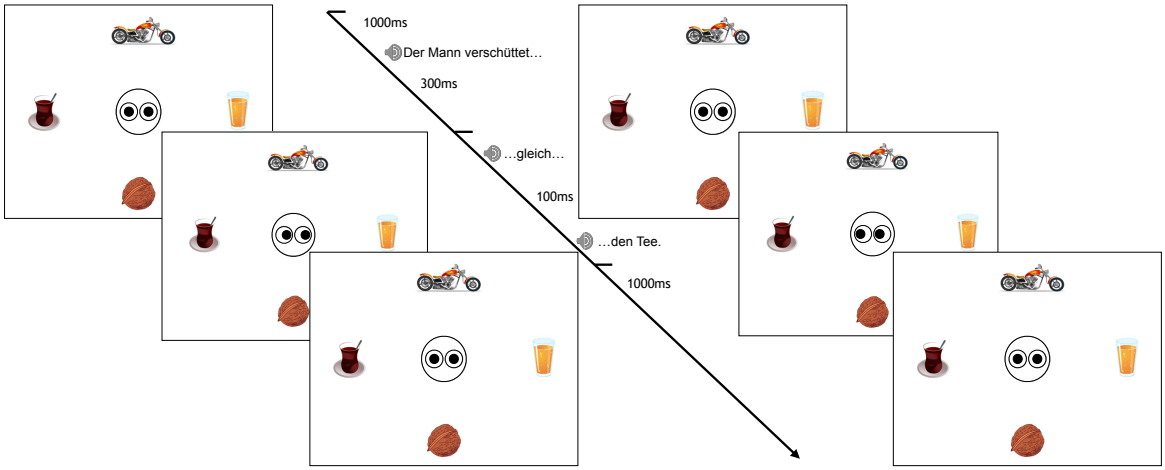


Figure 4.8 Exp. 6 – Trial timeline example: competitor gaze condition (left) and target gaze condition (right).

previous context; however, the previous gaze cue was not always cuing the target object, but was sometimes incongruent (i.e. cuing the competitor).

Participants

33 Saarland University students took part in this study (24 women) and were monetarily reimbursed for their participation. Their ages ranged from 18 to 39 years old (M = 23.52). Participants were all native speakers of the German language with normal or corrected-to-normal vision. None of the participants was familiar with the previously conducted experiments.

Items

21 item trials were created. They used the same sentence structure as was the case in the previous two studies. As in Experiment 5, only restrictive verbs were used (*spill*). The referent noun always fitted with the previous linguistic context (*spill*: *tea* or *juice*).[8]

The visual scenes were of the same structure as in the previous two experiments. Of the four presented objects, two did not fit the verb category (distractors: *motorbike*, *walnut*), while the other two were equally plausible in the context of the given verb (*spill*: *tea*, *juice*). Again, the position of target and competitor objects was rotated to all possible (individual and mutual) positions.

A pretest was conducted in which the displays were presented together with the beginning of the stimulus sentence: *The man spills now ...* . Written instructions stated that the sentence

[8]The list of all items is given in Appendix B, Table B.3 (for both experiment versions).

should fit the visual display, and the participants' task was to complete the sentence, by naming the object they would expect at the end of the sentence. Based on the results collected from 36 German native speakers, we chose 21 sentence and scene pairs where both potential target objects were selected equally often. In this way, we created 21 item trials that included visual scenes with two objects being equally plausible in the given linguistic and visual context.

As mentioned, the only experimental manipulation considered the gaze cue, which was either absent (no-gaze), or cued to one of the two fitting objects, that is, either the target (congruent gaze) or the competitor (incongruent gaze). The experiment was run in two versions. Version A used one sentence continuation (*spill tea*), while in version B the other object was referred to (*spill juice*).

A trial timeline is illustrated in Fig. 4.8. The left-hand side presents the competitor gaze condition, while the right-hand side shows the target gaze condition. The sentence used and the timeline were identical for different gaze conditions.[9]

Fillers

The experiment included 100 trials in total, 79 of which were fillers. The gaze cue was present in 2/3 of all trials (14 items, 53 fillers). Incongruent gaze made up only 10% of the overall number of trials, that is, 15% of the trials including a gaze cue (7 items, 3 fillers). Simple yes/no comprehension questions followed 57 of the filler trials. As was the case in previous experiments, filler trials differed from the experimental ones in terms of sentence structure and the number of objects that fit the verb selectional features. For instance, the sentence *Von allem Früchten mag der Bruder am liebsten Himbeeren* (literal translation: *Of all fruits likes the brother the most raspberries*) was presented with a display showing broccoli, a strawberry, a banana and a raspberry. Thus, three objects were potential targets in this example trial.

The experiment was preceded by a practice session that included 3 trials illustrating only no-gaze and target gaze conditions (no incongruent gaze cue).

Again, we used the Latin square design to create three lists where each item was presented in only one condition. The presentation of trials in a list was pseudo-randomized manually, and an additional three lists were created with opposite trial order. The experimental procedure was the same as in the previous two experiments. The duration of the experiment was approximately 30 minutes.

[9] An exhaustive list of visual displays used for this experiment is given in Appendix B, Table B.6.

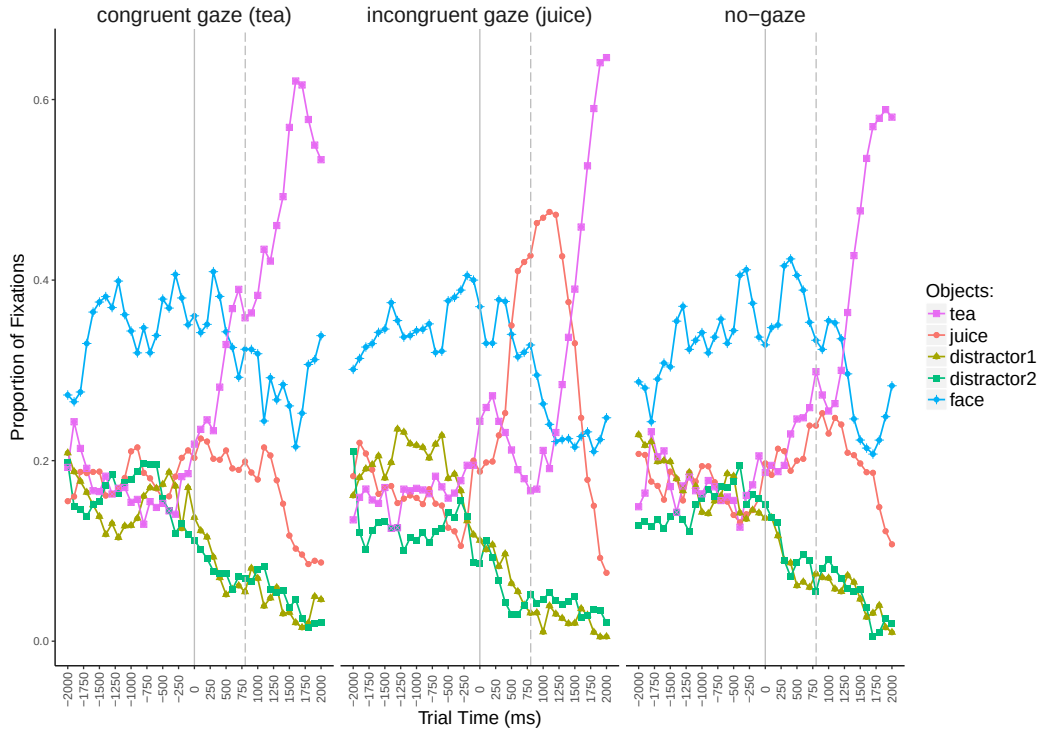


Figure 4.9 Exp. 6 – Proportion of fixations to presented objects in the three conditions: congruent (target), incongruent (competitor) gaze and no-gaze. Fixations aligned to the gaze cue onset (solid line). The dashed line presents article onset of the object noun phrase.

4.3.2 Results

As in the previous two studies, we consider the following three measurements: (a) the proportion of fixations throughout a whole trial, (b) the statistical analysis of new inspections in a region of interest, and (c) the analysis of the ICA events in the gaze time-window and the reference window. As done previously, all independent variables were contrast coded for the statistical analyses, and all models included maximal converging random structure justified by the experimental design (Barr et al., 2013).

Proportion of Fixations

Fig. 4.9 illustrates the proportion of fixations to the four presented objects (and the face) during a trial. We show the fixation patterns for the three gaze conditions, namely, congruent gaze (*tea*), incongruent gaze (*juice*) and the no-gaze condition. As previously, the plots are aligned to the onset of the gaze cue (or closing of the eyes in the no-gaze condition) which is marked with the solid line. The dashed line illustrates the onset of the article from the object noun phrase.

Table 4.5 Exp. 6 – Results of the two models fitted for the new inspections analysis (gaze region of interest).

	a) Target inspections				b) Competitor inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.798	0.105	-17.071	< 2e-16 ***	-1.773	0.090	-19.755	< 2e-16 ***
NTGAZE	-0.202	0.209	-0.966	0.334	0.127	0.214	0.591	0.554
TCGAZE	-0.733	0.226	-3.253	0.001 **	0.633	0.221	2.867	0.004 **

a) TargetInsp ~ NTgaze + TCgaze + (1 + NTgaze + TCgaze || Subject) + (1 | Item), family = "binomial"
b) CompetitorInsp ~ NCgaze + TCgaze + (1 + NCgaze + TCgaze || Subject) + (1 | Item), family = "binomial"

We see that the introduction of the gaze cue led to a shift in visual attention and the cued object became the most fixated one. However, at the same time, the other potential target was still considered, though to a lesser degree. In the no-gaze condition, where no object was cued, we see that both potential target objects were inspected more than the distractors, but with no preference for one of the two. Finally, upon the referent noun onset, the most inspected object was the named one.

New Inspections

We considered new inspections in the gaze region of interest, in order to see how the gaze cue influenced the immediate inspection of the presented objects.[10] We ran two different models, one for the target (tea) inspections and one for the inspections of the competitor (juice). The independent variable was contrast coded in such a way that NTgaze codes the comparison of no-gaze and target gaze conditions; NCgaze – no-gaze vs. competitor gaze, and TCgaze – target gaze vs. competitor gaze. Table 4.5 presents the fitted models and collected results.

The model run on the target inspections revealed no effect of NTgaze (p = 0.334), that is, no difference between the no-gaze and target gaze condition, but a main effect of TCgaze (p = 0.001), suggesting that there were more new inspections of the target (tea) when it was cued (M = 0.185, SD = 0.388) then when the competitor object (juice) was cued (M = 0.107, SD = 0.309).

The model run on the new inspections of the competitor showed, again, no difference between no-gaze and competitor gaze conditions (p = 0.554), but a main effect of TCgaze (p = 0.004), suggesting there was more looking at juice when that object was also cued (M = 0.201, SD = 0.401), than when water was cued (M = 0.112, SD = 0.316).

[10]The results are illustrated by Fig. B.11 in Appendix B.

Table 4.6 Exp. 6 – Results of the two models fitted for the ICA analysis. Note that the variable Gaze denotes different comparisons in the two models. In the gaze window, we compare the existence of the gaze cue: no-gaze vs. ref. gaze condition. In the reference window, the two conditions that behaved similarly are collapsed: congruent gaze vs. no-gaze & incongruent gaze.

	a) Gaze time-window				b) Reference time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.758	0.035	77.92	<2e-16 ***	2.723	0.042	64.27	< 2e-16 ***
GAZE	0.016	0.049	0.33	0.745	0.154	0.056	2.74	0.006 **
HALF	-0.063	0.034	-1.82	0.069 .	-0.082	0.035	-2.36	0.018 *
HALF:GAZE	-0.111	0.087	-1.28	0.201	-0.026	0.093	-0.28	0.778

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICA ~ Half * Gaze + (1 + Half * Gaze | Subject) + (1 | Item), family = Poisson (link = "log")
b) ICA ~ Half * Gaze + (1 + Half * Gaze | Subject) + (1 | Item), family = Poisson (link = "log")

The Index of Cognitive Activity

As Fig. 4.10 shows, there is a difference in cognitive load on the referent noun induced by target gaze, in comparison to both no-gaze and competitor gaze conditions. Since the incongruent and no-gaze conditions show no difference between each other, we treat them together in the statistical analysis of the reference window, and compare this to the congruent gaze condition. Table 4.6 presents the fitted models and collected results. A main effect of Gaze on cognitive load ($p = 0.006$) suggests that the congruent gaze cue led to the reduction of load on the subsequent reference (no gaze: $M = 16.307$, $SD = 7.678$; incongruent gaze: $M = 16.33$, $SD = 7.55$; congruent gaze: $M = 14.461$, $SD = 6.729$). In addition, a main effect of experimental Half was found ($p = 0.018$), but no interaction ($p = 0.778$), since cognitive load was overall lower in the second part of the experiment (1st part: $M = 16.548$, $SD = 7.374$ vs. 2nd part: $M = 14.925$, $SD = 7.297$).

In the Gaze window, we considered the two gaze conditions as one (target gaze and competitor gaze), since at the point of the gaze cue it is not known if the cue will end up being congruent or not; that is, we compared no-gaze with the gaze cue. No significant differences were found.

4.3.3 Discussion

Without the gaze cue, the eye movement data revealed that both potential target objects are considered equally. Gaze shifts the visual attention to the cued object. However, the other potential target is not immediately discarded. Thus, we see an interplay of linguistic and

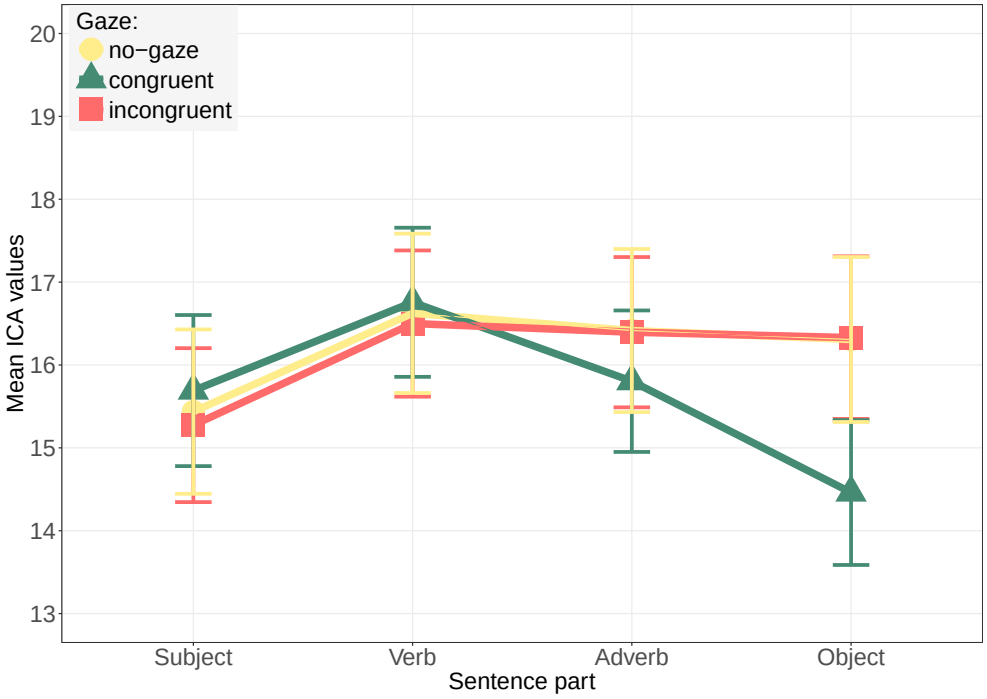


Figure 4.10 Exp. 6 – mean ICA values in the four time-windows of a sentence in no-gaze, congruent (target), and incongruent (competitor) gaze conditions. Points marked as *Adverb* (Gaze time-window) and *Object* (Reference time-window) are relevant for the analysis (95% CI error bars).

visual cues. The verb activates two visually presented objects as potential referents for the following noun phrase. Gaze further highlights one of those objects; however, this does not lead to disregarding the other fitting object – rather, this object continues being activated. It is only upon hearing the referent noun that the non-mentioned object is discarded.

Cognitive load results suggest that the incongruent gaze cue was not costly as such; rather, it induced the same effort as not having a cue. However, having a congruent (target) gaze cue has been shown to be beneficial for the effort required to process the target noun. Again, we found evidence that a fitting congruent gaze cue facilitates reference processing without inducing cost in itself.

Unlike the present study, previous explorations into gaze congruency speak of a cost of incongruency. Incongruent gaze led to longer reaction times in behavioral tasks, and is thus considered to be more costly than neutral or congruent gaze (Macdonald & Tatler, 2014; Staudte & Crocker, 2011). More recently, Jachmann et al. (2017) examined ERPs and similarly found a cost of an incongruent gaze cue. It is to be noted, however, that the design of the present experiment differs in important ways from the previous work.

Firstly, we used full sentences where linguistic constraints inspired anticipation of only certain depicted objects. Macdonald and Tatler (2014) used no linguistic context, gaze being

the only piece of information based on which a prediction for a specific object could have been made. When this cue was made reliable, it was followed and trusted, which led to disruption when it proved to be misleading. Similarly, Jachmann et al. (2017)'s gaze cue is the first piece of information based on which a prediction about the target can be made. Even though they use comparative sentences (*Compared to the car, the house is relatively small, I think*), the gaze cue was presented before the second noun is uttered, hence, also prior to the comparative adjective. In our study, the linguistic context included a restrictive verb based on which two illustrated objects could be expected as potential sentence continuations. The gaze cue only additionally directed attention to one of the two objects, but importantly, at that point both potential targets were already activated. Thus, when one of the two targets was mentioned, the comprehension followed smoothly. No disruption occurred when the object that was not cued was mentioned. In contrast, further activating an object with a gaze cue, and finally mentioning it, was shown to be facilitatory for the comprehension of the linguistic reference.

Staudte and Crocker (2011) found both an effect of congruent gaze facilitation, and a disruption caused by the incongruent gaze cue. Their study made use of sentences with comparatives which inspired anticipation of a certain object (*The cylinder is taller than the pyramid that is pink*). Incongruent gaze cue was directed to an object of the same type, but of different size (tall brown pyramid), rendering the gaze cue not only incongruent, but also mismatching with the previous linguistic context. We, on the other hand, used a linguistic context that activated two of the presented four objects, rendering them equally plausible to be mentioned next. Consequently, a gaze cue to one of those objects, even when incongruent with the upcoming noun, was not disruptive, since it was plausible given the context, highlighting an object already activated by the verbal constraint.

We conclude that the different results found in our experiment reflect the interplay of the linguistic and visual cues that are not present in the mentioned studies. On the one hand, when the gaze cue not only is incongruent with the subsequent target, but also does not match the previous linguistic context (as in Staudte & Crocker, 2011), it is not clear which of the two factors are driving the effect. On the other hand, when referential gaze cue is the only piece of information that inspires an anticipation for the target object, its incongruency turns out to be costly (Macdonald & Tatler, 2014; Jachmann et al., 2017). However, as we have shown, when the gaze cue is there as an additional cue to language, the linguistic information is not discarded once gaze is introduced; rather, they are considered together until the reference is resolved.

4.4 Chapter Discussion

Three studies were conducted in order to gain more insight into the immediate effect of referential gaze cue on language processing. We examined listeners' visual attention during simultaneous presentation of both linguistic and visual stimuli. In addition, we measured immediate cognitive effort required to perceive and process the information coming from both modalities, in the absence of a behavioral task. We summarize the findings of the presented studies by referring back to the questions raised at the beginning of the chapter. Moreover, Table 4.7 gives a short overview of the main gaze cue effects measured on the cognitive load.

First, we found evidence that the gaze cue influences the predictability of linguistic reference. All three conducted experiments indicate that gaze is followed, that is, it leads to a shift in visual attention, regardless of its fit to the linguistic context, or its slightly reduced reliability. This subsequently influences cognitive effort required for processing the reference.

Second, we found that the existence of the gaze cue facilitates the processing of the subsequent referent noun. Experiment 4 manipulated the existence of the (fitting and congruent) gaze cue and found that it is always followed, reducing the load on the reference, while maintaining the facilitation for the most plausible referent. Experiment 5 went a step further and examined a congruent, but mismatching cue. Again, gaze led to a (slight) reduction of cognitive load on the referent noun, even when they both did not match the linguistic context. Importantly, though, this effect was found only upon establishing trust in the gaze cue, that is, adapting to the sometimes anomalous gaze. Experiment 6 manipulated gaze congruency and found, again, a benefit of congruent gaze cues, but, interestingly, no cost of incongruent cues. In addition, eye movement analysis showed that even though gaze was followed and made one object more salient, the other plausible object was not dropped from attention until the reference was uttered.

Third, we wondered about the immediate cost of gaze cue perception and utilization. Experiments 4 and 6 give evidence that a gaze cue fitting the linguistic context does not induce higher cognitive effort. However, Experiment 5, which examined gaze cuing an object that does not fit the previous context, provided evidence that such anomalous gaze indeed induces immediate cost. This suggests that the cost is induced when the visual cue cannot be incorporated with the previously established linguistic context.

Fourth, inspired by the UID hypothesis, we had expected to detect a distribution of cognitive effort between the linguistic and visual cues. All three experiments showed that the existence of a fitting congruent gaze cue leads to the reduction of cognitive effort on the reference. However, this reduction was not preceded by an increase in cognitive effort

Table 4.7 Summary of the main cognitive load results.

		Gaze window	Reference window
Experiment 4:			
no gaze		n.s.	no gaze > congruent gaze
vs.			
congruent gaze			
Experiment 5:		**part 1**	
no gaze	*verb fitting*	gaze*fit	n.s.
vs.	vs.	**part 2**	
congruent gaze	*mismatched*	no gaze > congruent gaze	no gaze > congruent gaze
Experiment 6:			
no gaze			
vs.		n.s.	no gaze > congruent gaze
congruent gaze			no gaze = incongruent gaze
vs.			
incongruent gaze			

induced by perceiving the gaze cue itself. Interestingly, though, Experiment 5 indeed showed some evidence for such a distribution, where the surprising gaze cue induced higher cognitive load, which was followed by the reduction of effort on the referent noun (second half of the experiment). The same condition without the gaze cue showed that the referent noun bore all the effort required for processing the anomalous sentence. As it results from a nonsensical sentence, we are reluctant to understand this finding as a UID-like effect. The lack of a UID-like distribution of cognitive effort during a meaningful sentence is explainable in terms of prediction and surprisal. We argue that facilitated processing is a result of anticipation confirmation, while disconfirming anticipations is reflected in higher processing effort. A gaze cue inspires the anticipation for the target referent, resulting in a facilitated processing of the corresponding linguistic reference. The direction of the gaze cue, however, was not predictable in the context of our experiments, and thus, no effect was measured on the cue other than the effect of surprisal when the basic assumption was violated, namely, for the visual cue to be cooperative and to make sense in the current context. The creation of anticipations did not result in an increase in cognitive load, and thus, no UID-like effect was detected.

Finally, our findings suggest that the gaze cue is strategically used. It further highlights an object but does not inhibit previous activation of a different object. Thus, gaze is not disruptive when it is incongruent with the linguistic reference (given that the reference fits the context).

In sum, our assessment of anticipatory eye movements, visual attention shifts and immediate cognitive load allowed for gaining novel insight into the interplay between visual

perception and language processing. Our findings are in line with previous research showing that verb selectional preferences inspire activation of context-fitting visually presented objects, while gaze cue can further highlight the same, or lead to a shift in visual attention to an alternative object – such a shift happens even when the gaze is cuing an object not fitting the context. Also, as was previously assessed employing behavioral tasks, we found that the referential gaze cue indeed facilitates reference processing. Importantly, however, we showed this by measuring cognitive effort online, as it is induced, and without having to rely on a secondary task. Our findings suggest that the gaze cue reduces cognitive load required for processing the corresponding linguistic reference even when neither fit the previous context. In addition, we were able to further extend the current understanding of gaze cue perception by showing that the facilitatory effect of gaze on language processing is not preceded by an additional cost of perceiving the gaze cue. The perception of the gaze cue does not induce higher cognitive effort when it fits the context (i.e. when objects of different probability are cued); only mismatching, anomalous gaze was shown to be costly. Interestingly, we observed that the gaze cue always led to a shift in visual attention, which was, however, not always coupled with the same cognitive effort. This finding further suggests that visual attention is not indicative of the intensity of the induced processing effort.

Chapter 5

Quantitative Differences in Gaze Cuing

The three experiments presented in the previous chapter were carried out to examine the integration of the referential speaker gaze cue with the linguistic material. In addition to the shifts in visual attention, we assessed the immediate cognitive load induced at the points of perceiving the visual cue, and processing the linguistic reference. Shifts in visual attention towards the relevant object prior to its mention showed that anticipations about the target are created based on both the linguistic context and the information from the visual modality. The presentation of a referential visual cue proved facilitatory for subsequent reference processing, but no effects on the processing effort were detected on the gaze cue. Hence, the question of how the cue itself reflects and reduces the uncertainty about the target referent, and how this relates to the reduced processing on the referent noun, still remains open. The aim of the present chapter is thus to further examine the way in which the referential gaze cue is utilized as a visual means for reducing uncertainty about the target.

In the experiments from Chapter 4 the information contributed by the gaze cue (and the linguistic reference) was manipulated in a *qualitative* manner. We examined the effects of the probability of an object to be mentioned in a particular linguistic context. The results showed that the cognitive load on the referent noun was influenced by how certain the listener was about the upcoming reference. The gaze cue, as an additional piece of information, aided the creation of anticipations. The cognitive load induced on the visual cue was not influenced by cuing objects of different probability of mention given the linguistic context. Moreover, no difference was detected between perceiving the gaze cue, and having the eyes close without giving a hint about the target object. The qualitative difference was almost always equally difficult to process. The effect of target probability and listener (un)certainty was detected only on the noun.

In order to approach the question of gaze cue integration from a different perspective, the two experiments presented in this chapter use the same paradigm but manipulate the

information contributed by the gaze cue in a *quantitative* manner. When the gaze is cuing more than one object, its informativity (i.e., the amount of transmitted information) is not measured in terms of how predictable the cued object is given the linguistic context; but rather, the amount of transmitted information is relative to the sheer number of visually present objects that are highlighted and brought into focus by the visual cue. We follow up on the previously presented experiments by using the same paradigm, but keep the linguistic stimuli constant, while manipulating cue specificity.

We aimed to examine whether the referent identification during linguistic processing necessarily happens through the linguistic material, or if this information can be obtained from a visual cue. We manipulated cue specificity (visual cue in Exp. 7; linguistic cue in Exp. 8) and measured how it quantitatively affects the uncertainty about the target referent. In Experiment 7, we varied the degree to which the visual cue reduces the uncertainty about the upcoming reference. We assessed the consequences of such an effect on the processing load induced on the cue, and subsequently on the referent noun. In Experiment 8, we were interested in whether the gaze cue can take the role of reference resolution. Upon hearing an underspecified linguistic reference, the listeners were presented with a visual cue that disambiguated the target object. We assessed the processing effort required on the reference, as well as on the subsequent gaze cue.

Toward addressing the third overarching research question listed in Chapter 1 (p. 6) – *Is a referring expression resolved at the point of hearing the linguistic reference, or can a visual cue take the role of referent identification on a par with a noun?* – the two experiments presented in this chapter address the following specific questions:

1. Is the quantitative step-wise reduction of referential uncertainty, achieved by the combination of visual and linguistic cues, reflected in the immediately induced cognitive load (at the cues)?

2. If, upon hearing an underspecified referring expression, a referential gaze cue unambiguously identifies the target object, could gaze carry the same amount of surprisal as a referent noun does, and therefore elicit the same amount of effort?

    a) Does an underspecified linguistic reference induce a graded cognitive load effect at the referent noun, relative to the level of remaining referential ambiguity?

The present two experiments both introduce a stepwise reduction of referential uncertainty by the means of visual and linguisitc cues. The reduction of uncertainty is completed in two steps. We will refer to them as *step one* and *step two* for clarity and easier comparability between the two experiments. At step one the uncertainty about the target referent is reduced

to a certain degree, while it is at step two that the reference is resolved and the target is identified. In Experiment 7, the visual cue is introduced at step one, cuing a group of objects, thereby reducing the uncertainty about which is the target referent. It is the referent noun that disambiguates the target, at step two. In Experiment 8, we swapped the roles of the visual and linguistic cues. An underspecified referring expression is introduced at step one, which reduces the uncertainty, but does not disambiguate the target object. At step two, the visual cue is introduced that unambiguously points to the target referent and resolves the reference.

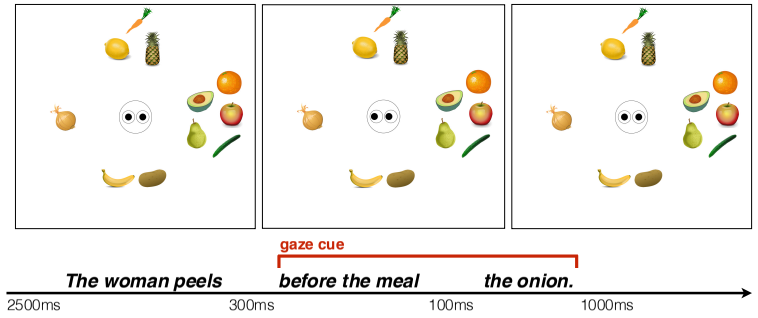
Both experiments were conducted in the VWP with a setting greatly resembling that of the previously presented studies. Simple SVAdvO sentences in the German language were paired with visual displays presenting the potential target objects and the schematic representation of the speaker gaze. Linguistic stimuli were kept constant, all manipulations stemming from the visual modality. As previously, we examined the anticipatory eye movements and measured immediate cognitive load at two points of interest – the linguistic reference and the visual cue. The crucial difference between the two experiments lies in the time point when the gaze cue was presented.

Fig. 5.1a illustrates a trial timeline from Experiment 7. We measured visual attention at the point of the gaze cue – examining whether it causes a shift in attention towards the cued (group of) object(s). Also, we measured immediately induced cognitive load, resulting from the reduction of referential uncertainty from eleven objects in the scene to the size of the cued group of objects. In addition, we considered the cognitive load induced by the referent noun finally naming the target, that is, (one of) the cued object(s). In Experiment 8, we also consider participants' visual attention and the cognitive load induced at the point of the introduction of the gaze cue, and the uttering of the referent noun. Importantly, as illustrated by Fig. 5.1b, here we introduce the gaze cue after the reference, that is, after the sentence has been uttered.

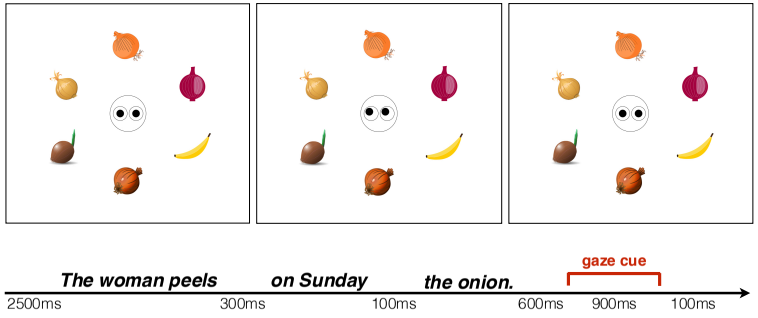
Experiment 7 shows graded effects of referential uncertainty on the referent noun, while Experiment 8 makes an attempt at swapping the roles of the gaze cue and the referent noun, and examines whether these effects are also reflected on the visual cue, when it takes over the role of reference resolution.

5.1 Experiment 7

This experiment manipulates gaze specificity (size of the cued group of objects), and thus simultaneously affects the uncertainty about which object will be mentioned. The aim of this experiment is to answer the first research question mentioned in the introduction of this chapter (see p. 86).



(a) Exp. 7 – Manipulating visual cue specificity and the reduction of referential uncertainty. Gaze cue onset approx. 800 ms prior to the reference onset.



(b) Exp. 8 – Manipulating noun specificity and the reduction of referential uncertainty. Gaze cue onset 600 ms after the reference offset.

Figure 5.1 Trial timeline illustrates the visual displays, the linguistic stimuli, and when and for how long the gaze cue was presented. Note the difference in gaze cue onset in the two experiments.

Figure 5.2 Exp. 7 – Three experimental conditions. From left to right: *GazeToOne*; *GazeToThree*; *GazeToFive*. Linguistic stimulus was kept constant within one item: *The woman peels, before the meal, the onion.*

We were interested in seeing whether the effect of gaze specificity on referential uncertainty affects cognitive load at the two steps where the uncertainty is reduced: (1) at the point of perceiving the gaze cue, and (2) at the point of hearing the referent noun. Participants were presented with visual displays of eleven objects (see Fig. 5.2), and were given no information, prior to the gaze cue, that could inspire anticipation of a particular target. Gaze cued only one position, occupied by either one object (*GazeToOne*); a group of three (*GazeToThree*); or five objects (*GazeToFive*). Please, see Fig. 5.1a again for an illustration of the trial timeline.

Our manipulation can be observed from two perspectives on estimating processing effort, namely, by considering the information-theoretic notions of entropy and surprisal. As already described in Chapter 2 (p. 13), both entropy and surprisal are used as metrics that predict the difficulty with which a linguistic item is perceived, hence typically linking language models with experimental measures of effort in language processing, such as, reading times. In the present work, however, the linguistic context is kept constant, and both surprisal and entropy are calculated by considering the immediate visual context which influences the probability of a referent to be mentioned (as done in Tourtouri et al., 2017; Ankener, Drenhaus, et al., 2018). The referential entropy is influenced by the competitors to the target object in the given visual context. If all competitors have an equal probability of mention, the more competitors there are, the higher the uncertainty about which is the target and thus, the higher the reduction in entropy once the target is mentioned. Similarly, the larger the number of competitors, the lower is the probability of each object to become the target, and thus, the higher the surprisal of the target referent.

Let us clarify the difference between entropy and surprisal by contrasting the two notions in a relevant example. Eleven objects are present in the visual scene, all equally likely to become the target given the linguistic context, since they all fit the verb selectional preferences equally well. The introduction of a referential gaze cue towards a group of five objects inspires a shift in visual attention towards the cued group, thereby reducing

the referential entropy. Subsequently, the referent noun is introduced, fully disambiguating the target object, and thereby completely reducing the entropy. Hence, the level of entropy reduction occurring at the gaze cue, and at the linguistic reference, both depend on the number of cued objects. The bigger the cued group, the less entropy is reduced at the gaze cue, but more at the reference, and vice versa.

Considering suprisal in this example, all eleven objects fitting equally well to the linguistic context, they are all equally likely to become the target. Thus, the level of surprisal at the gaze cue is not relative to the number of cued objects, since no number of objects was anticipated. Importantly, by highlighting them, the gaze cue increases the probability of some objects to be mentioned, and creates a basis for anticipation. The smaller the cued group, the higher is the probability for each cued object to be mentioned as the target. Consequently, at the point of the noun, the surprisal on the target noun will increase with the size of the cued group, because the bigger the group, the less likely each group member is to be mentioned as the target.

Table 5.1 Exp. 7 – Referential entropy (H) in the three experimental conditions at the start of the trial (H Start); upon the gaze cue (H Gaze); and after the referent noun has been uttered (H Ref).

Condition	H Start	H Gaze	H Ref.
GazeToOne	3.459	0	0
GazeToThree	3.459	1.585	0
GazeToFive	3.459	2.322	0

Table 5.2 Exp. 7 – Probability (P) and Surprisal (S) in the three experimental conditions at the start of the trial (Start); upon the gaze cue (Gaze); and after the referent noun has been uttered (Ref).

Condition	Start		Gaze		Ref.	
	P	S	P	S	P	S
GazeToOne	0.09	3.47	0.25	2	1	0
GazeToThree	0.09	3.47	0.25	2	0.33	1.59
GazeToFive	0.09	3.47	0.25	2	0.20	2.32

Consequently, the cognitive load induced on the gaze cue (step one) could follow two different hypotheses. If affected by referential entropy reduction, cuing more objects would lead to lower load, since the referential entropy is reduced by a lower degree (*GazeToFive* $<$

GazeToThree < *GazeToOne*). As shown in Table 5.1, referential entropy is either completely reduced by the gaze cue (*GazeToOne*: entropy reduction ΔH of 3.459 bits), or partially reduced by it (*GazeToThree*: ΔH of 1.874 bits; *GazeToFive*: (ΔH) of 1.137 bits). Cuing one object should induce high cognitive load, since the entropy of the scene is abruptly reduced by all 3.459 bits. The *GazeToThree* condition would induce less cognitive effort, due to its less abrupt reduction of 1.876 bits. Finally, *GazeToFive* reduces the entropy by 1.137 bits, which should induce the least effort. Alternatively, if affected by the levels of surprisal, we do not expect any change to be measurable at the gaze cue. As previously mentioned, all eleven objects were equally likely targets and no group size was more likely to be cued. Hence, there is no basis for creation of an anticipation for a gaze cue towards a certain group of objects. Table 5.2 shows that at the point of the gaze cue all four locations have the same probability of being cued (0.25), resulting in the same surprisal value (*GazeToOne* = *GazeToThree* = *GazeToFive*).

Moreover, the cognitive load induced by the referent noun (step two) is expected to rise with the number of competitors in the target group (*GazeToFive* > *GazeToThree* > *GazeToOne*). This trend should follow both from the degree of entropy reduction, as well as from the surprisal hypothesis. Table 5.1 shows that in the *GazeToFive* condition, the reference reduces the entropy by 2.322 bits, and *GazeToThree* by 1.585 bits, while *GazeToOne* has no entropy. Similarly, table 5.2 shows that *GazeToFive* is the condition where the visual cue is least specific (four immediate competitors), and hence, each object has probability 0.2 of being mentioned, resulting in the highest surprisal value. In *GazeToThree* the target has fewer immediate competitors (two objects), and hence each object has probability 0.33 of being mentioned, resulting in lower surprisal. Finally, in *GazeToOne* the target object is already identified by the gaze, leaving no immediate competitors, and resulting in no surprisal. In conclusion, the fewer competitors in the relevant group of objects, the higher is the probability of the target object, and hence, the lower the surprisal on the referent noun.

5.1.1 Method

This study employed one independent variable, namely, Gaze Specificity. We manipulated the number of cued objects resulting in three experimental conditions: *GazeToOne* – gaze cuing one single object, *GazeToThree* – cuing a group of three objects, and *GazeToFive* – cuing a group of five objects.

Participants

40 students of Saarland University took part in our experiment and were monetarily reimbursed for their participation. All participants were native speakers of the German language with normal or corrected-to-normal vision.

Post-experimental briefings revealed that 10 participants ignored the gaze cue either due to mistrust, or belief that it was the representation of their own eye movements. Since our manipulation was fully dependent on the participants noticing and utilizing the gaze cue, we discarded the "ignorers" and continued data collection until reaching the necessary number of participants who engaged in gaze-following.[1] Thus, we present the analysis of 30 "followers" (25 female), with their age ranging from 18 to 32 years old (M = 23,3).

Items

The experiment included 30 item trials and 40 fillers. Each trial presented a visual display and the audio of one stimulus sentence. As illustrated by Fig. 5.2, the gaze cue, presented in the middle of the screen, was in the same format as the one used in the three experiments from Chapter 4. Visual displays included 11 objects organized in four groups around the center of the screen. The groups varied in size. Item trials always included groups of one, two, three and five objects. We counterbalanced both the position of the target (group of) object(s) as well as the position of different group sizes. As previously, audio consisted of natural language recorded by a German native speaker at a regular rate of speech.

All sentences used the same word order: Subject, Verb, Adverb, Object (example: *The woman peels, before the meal, the onion*; original: *Die Frau schält vor dem Essen die Zwiebel*). In order to reduce the variation in the adverbial phrase as much as possible in the item sentences, all 30 items included only 4 different adverbial phrases, namely, *vor dem Essen*, *nach dem Essen*, *vor der Arbeit* or *nach der Arbeit*.[2] The length of the audio presentation of the four adverbial phrases was kept constant (1350 ms) by including silent parts were necessary. Due to the complexity of the visual scene, we introduced a longer adverbial phrase in order to avoid an overlap in processing the gaze cue and the referent noun. Filler sentences contributed the needed variation in the meaning, as well as in position of the adverbial phrase.

The gaze cue was present in each trial (both item and filler) and was always reliable, cuing the target object, or the group of objects in which the target occurred. When a single object was cued in the *GazeToOne* condition, participants could be certain that this object

[1] The eye movement analyses of the 10 "ignorers" show clear evidence that the gaze cue was not followed. Naturally, those participants also showed no effect on cognitive load.

[2] An exhaustive list of all item sentences used in this experiment is given in Appendix C, Table C.1.

would be named. In the *GazeToThree* condition, one of the three objects in the cued group was the target. Finally, in the *GazeToFive* condition there were four competitors to the target. A pretest of the visual displays with German native speakers ensured that all eleven presented objects were fitting in the context of the verb. Hence, before the introduction of the gaze cue, there was no ground for creating anticipation as to what the target object might be.[3]

In order to provide enough time for the participants to familiarize themselves with the scene and its many objects, the visual display was presented for 2500 ms before the onset of the stimulus sentence (see Fig. 5.1a). The remainder of the trial resembles the timeline from previous experiments. In order to create clear and sufficient time-windows for the subsequent analyses, two short silent "breaks" were introduced during the sentence. A 300 ms silence was introduced between verb offset and the onset of the adverbial phrase. In addition, a 100 ms silence was inserted before the onset of the target noun phrase. Finally, an additional 1000 ms of silence was included upon sentence offset. Note that the onset of the adverb prepositional phrase also marked the gaze cue onset. The gaze cue was present until the end of the sentence. For the final silent second, the eyes were looking straight ahead.

Each trial was followed by a question regarding either the linguistic content of the sentence or the visual characteristics of the target object. For instance, the question after our example trial presented an image of an onion (visually similar but not identical to the target object) asking whether that was the exact onion just shown during the trial.

Fillers

40 filler trials introduced variation in terms of display organization and the linguistic structure of the sentence. The number of objects in the four groups varied in comparison to the items, and created an overall balance in the experiment, so that there was an equal representation of all group sizes. Fillers also introduced linguistic variation, using sentences of different length and structure. Since the item sentences included only four alternations of adverbial phrases, in order to mask this lack of variation, fillers used adverbial phrases of different length and at different positions in the sentence.

For instance, the sentence *Für die neue Küche braucht das Brautpaar nur noch eine Pfanne* (literal translation: *For the new kitchen needs the married couple only just a pan*) was presented with eleven images of different kitchen utensils, organized in three groups of three objects and one group of two objects. The pan was presented in a group together with a toaster and a martini glass.

The Latin square design was used to create 3 lists with 30 items and 40 fillers each. Hence, one list included each item in only one condition. The order of the trials was pseudo-

[3] An exhaustive list of all visual displays used in this experiment is given in Appendix C, Table C.5.

randomized manually, and an additional three lists were created with the opposite order of presentation.

Procedure

The experiment was prepared in the SR Experiment Builder software and run using the Eye-Link 1000+ eye-tracker (SR Research Ltd; Mississauga, Ont., Canada) tracking binocularly at a sampling rate of 500 Hz.[4]

Upon the calibration of the eye-tracker, participants were instructed to listen carefully to the sentences while freely moving their eyes to examine the visual displays. Also, they were informed that each trial would be followed by a question regarding the sentence or the visual characteristics of the target object. Before the start of the actual experimental session, participants first saw three practice trials and familiarized themselves with the experimental paradigm, as well as the difficulty of the post-trial questions. These were all yes/no questions that were answered with two keys on the keyboard. An additional key was used to advance the experiment after each question, and continue to the next trial. This empowered the participants to take their time, and create short breaks between trials, when needed. The experiment lasted for approximately 30 minutes.

Variable Coding and Data Analysis

Both the coding of the variables as well as the data analysis largely resemble those from the previous chapter, but will be described here, for the reader's convenience.

As mentioned earlier, we were interested in changes in visual attention that happen during a trial. For visualization, we looked into the proportion of fixations to the target object for the whole duration of a trial. Interested in the shift of attention occurring as a consequence of a visual cue, we conducted inferential statistics on the direction of new inspections upon the gaze cue. Consecutive fixations to the same interest area are considered as one inspection. A *new* inspection is thus the first inspection upon the relevant visual stimulus. This analysis compared the inspections to same-size groups in a condition when it was the one cued; compared to the conditions when another group was the target group. For the statistical analysis we treated new inspections as a binary variable, since trials with a new inspection to the area of interest (AoI) were given the value "1", and those without, a "0". For the analysis of new inspections upon the gaze cue, we considered the time-window starting from the

[4]This is the required setup for the subsequent use of the Workload Module software (EyeTracking.Inc) and the extraction of the ICA events from the pupil size data collected from this particular tracker.

onset of the gaze cue (onset of the adverbial phrase) up to the onset of the article from the referent noun phrase.[5]

Moreover, in order to assess the cognitive load induced at different points in the trial, we consider the analysis of the ICA events for the two relevant time-windows: at the point of the introduction of the gaze cue, and at the point of referring to the target object. For the analysis of the Gaze window, we take 600 ms from the onset of the visual cue.[6] For the analysis of the Reference window, however, we take 600 ms from the middle of the referent noun.[7] Data are collected from both eyes and combined by summing up the ICA events for the corresponding time-windows.

All independent variables were contrast coded for the statistical analyses. New inspections is a binary variable, and thus for the analysis we used generalized mixed effect models of binomial type. The mean ICA events for a time-window are treated as a count variable, and analyzed using generalized mixed effects models with Poisson distribution. All models included maximal converging random structure for both Item and Subject. Finally, the analyses were conducted in the R programming environment (R Core Team, 2017) using the *lme4* package (Bates et al., 2015).

5.1.2 Results

Before mentioning the results obtained from our measurements, we would like to mention that the accuracy of the participants' performance on the post-trial questions was 91.67%. This high percentage shows that even though the visual displays were rather complex, due to the large number of objects, by being alert and following the gaze cue the participants managed to attain very high accuracy on the task.

Proportion of Fixations

Fig. 5.3 illustrates the proportion of fixations to the target object during a trial, in each of the three experimental conditions. The solid vertical line represents the onset of the gaze cue, the first dashed line marks the onset of the article, and the second dashed line, the onset of the referent noun. We can clearly see that, upon the gaze cue, the target object is easily detected when it is the sole cued object (*GazeToOne*). The difference between target fixations in the

[5]The chosen window is relatively long; thus, we also conducted the analysis with a window of the same length as was used in Chapter 4, and found the same results.

[6]As is done throughout this thesis, we take 600 ms to be the appropriate size of the time-window for the analysis of the ICA events (Demberg & Sayeed, 2016; Tourtouri et al., 2017; Sekicki & Staudte, 2018; Ankener, Drenhaus, et al., 2018).

[7]The middle of the referent noun was calculated by taking the audio duration of a word, and taking its half as the starting point.

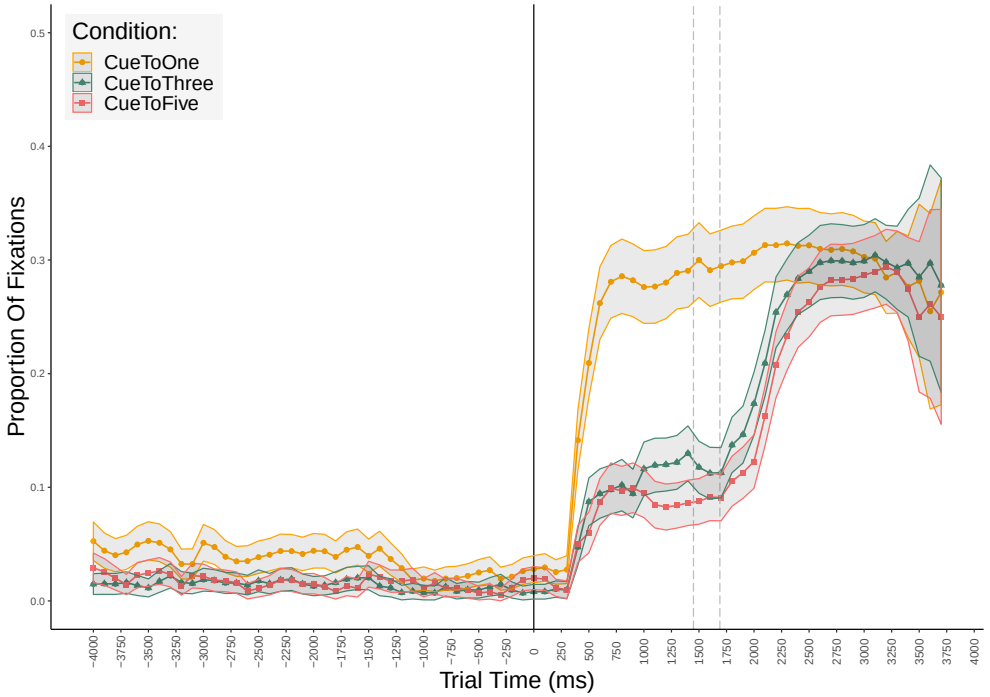


Figure 5.3 Exp. 7 – Proportion of fixations to the target object during the whole trial in the three conditions. The graph is centered on the gaze cue onset (solid vertical line). The two dashed lines show the onset of the referent article and the referent noun, respectively. (95% CI shading)

conditions *GazeToThree* and *GazeToFive* is less clear, however. We see a trend for more fixations of the target when it was presented in a group of three objects than when it was in a larger group of five objects.

New Inspections

In order to statistically evaluate the potential effect of gaze-following, we conduct the new inspections analysis. As described previously, all fixations that fell within one AoI and prior to a fixation outside of that AoI were grouped together and considered as one inspection. We compared the inspections of one (group of) object(s) in the condition when it was cued versus the other two conditions when it was not of importance. In other words, the inspections to the same-size groups were compared relative to whether the group was cued. Such a comparison allowed us to detect an effect of gaze-following. The results are presented in Table 5.3, while the illustration is to be found in Appendix C, Fig. C.6.

The model run for the new inspections to the single object compared the inspections in the *GazeToOne* condition versus the other two conditions where that object was not cued. These two conditions were considered together and labeled as *Other*. We found that the single object was looked at significantly more ($p < 0.001$) in the condition where it was cued

Table 5.3 Exp. 7 – Results of the main models fitted for the new inspections analysis (gaze region of interest).

	a) Single object inspections				b) Group of three inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-3.722	0.255	-14.625	<2e-16 ***	-2.755	0.088	-31.37	<2e-16 ***
G1VS.OTHER	-5.257	0.530	-9.914	<2e-16 ***	-	-	-	-
G3VS.OTHER	-	-	-	-	-4.003	0.197	-20.36	<2e-16 ***
	c) Group of five inspections							
INTERCEPT	-2.880	0.095	-30.45	<2e-16 ***				
G5VS.OTHER	-4.117	0.209	-19.72	<2e-16 ***				

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) New Insp. One ~ G1vs.Other + (1 + G1vs.Other || Subject) + (1 + G1vs.Other || Item), family = "binomial"
b) New Insp. Three ~ G3vs.Other + (1 + G3vs.Other || Subject) + (1| Item), family = "binomial"
c) New Insp. Five ~ G5vs.Other + (1 + G5vs.Other || Subject) + (1| Item), family = "binomial"

(*GazeToOne*: $M = 0.251$, $SD = 0.434$) than in the other two conditions (*GazeToThree*: $M = 0.013$, $SD = 0.115$; *GazeToFive*: $M = 0.01$, $SD = 0.101$).

The model ran for the new inspections to the group of three objects compared the *GazeToThree* condition with the other two conditions with the cue elsewhere. The group of three objects was inspected significantly more ($p < 0.001$) when the gaze cued that group (*GazeToThree*: $M = 0.319$, $SD = 0.467$; *GazeToOne*: $M = 0.027$, $SD = 0.163$; *GazeToFive*: $M = 0.019$, $SD = 0.138$).

The model fitting the new inspections to the group of five objects also showed that a group of five objects was inspected significantly more ($p < 0.001$) upon a gaze cue towards that group than in the other two conditions, where another (group of) object(s) was cued (*GazeToFive*: $M = 0.306$, $SD = 0.461$; *GazeToOne*: $M = 0.019$, $SD = 0.137$; *GazeToThree*: $M = 0.020$, $SD = 0.141$).

The Index of Cognitive Activity

Fig. 5.4 illustrates the mean ICA values in 600 ms time-windows at four points during a sentence. The analysis was relevant at two points where the entropy of the scene was reduced, namely the gaze time-window and the reference window. No difference is to be seen among the experimental conditions at the point of the gaze cue. At the point of the linguistic reference the three conditions differ gradually, relative to the number of objects in the cued group, *GazeToOne* inducing the least cognitive load, *GazeToFive* showing a higher level of load, and finally, the mean ICA value for *GazeToThree* falling between the other two conditions.

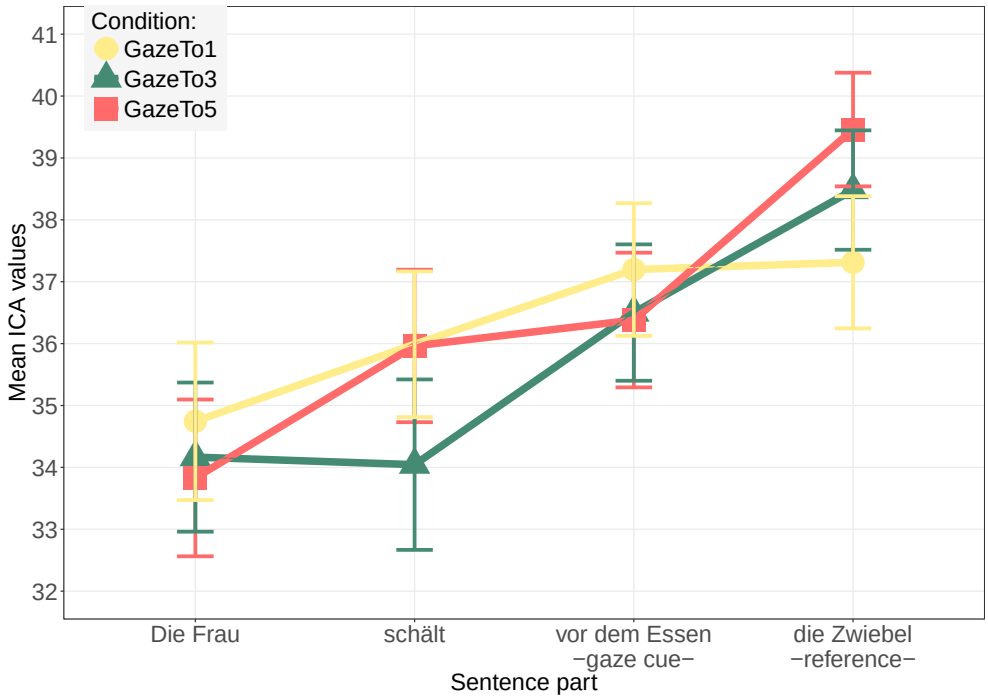


Figure 5.4 Exp. 7 – Mean ICA values at the four time-windows of a sentence. Points marked as "gaze cue" (Gaze time-winodw) and "reference" (Reference time-window) are relevant for the analysis. (95% CI error bars)

Table 5.4 presents the main models and collected results. Detailed information about the results of the further comparisons is given in Appendix C, Table C.2.

The analysis of the gaze window did not show any significant differences among the three conditions, either in the *GazeToOne* vs. *Other* comparison ($p = 0.101$), or the *GazeToThree* vs. *GazeToFive* comparison ($p = 0.760$). The full model showed only a significant interaction between the experiment Half and the *GazeToThree* vs. *GazeToFive* comparison. The following pairwise comparisons showed that the *GazeToThree* vs. *GazeToFive* comparison was not significant either in the first experiment half ($p = 0.215$), or in the second half ($p = 0.384$), but rather, the effect emerged due to the opposite direction of the trend in the two experimental halves. Thus, this effect is not of relevance for the interpretation of the results.

The model of the reference time-window showed a significant difference ($p = 0.004$) between *GazeToOne* and Other (*GazeToThree* and *GazeToFive* collapsed and taken together). In addition, the *GazeToThree* vs. *GazeToFive* comparison did not show a statistically significant difference. The findings show that conditions where more than one object was cued induced higher ICA values than the *GazeToOne* condition ($M = 37.314$, $SD = 9.248$), but were not significantly different from each other (*GazeToThree*: $M = 38.481$, $SD = 8.418$; *GazeToFive*: $M = 39.459$, $SD = 7.999$).

Table 5.4 Exp. 7 – Results of the main models fitted for the ICA analysis.

	a) Gaze time-window				b) Reference time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	3.609	0.021	172.03	<2e-16 ***	3.640	0.021	172.47	< 2e-16 ***
G1VS.OTHER	0.022	0.013	1.64	0.101	-0.046	0.016	-2.89	0.004 **
G3VS.G5	0.007	0.021	0.31	0.760	-0.024	0.019	-1.25	0.213
HALF	0.001	0.011	0.07	0.942	-0.016	0.022	-0.74	0.458
HALF : G1VS.OTHER	0.024	0.024	1.00	0.316	0.017	0.024	-0.74	0.459
HALF : G3VS.G5	0.067	0.028	2.39	0.017 *	0.001	0.027	0.03	0.974

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICAgaze ~ Half * G1vs.Other + Half * G3vs.G5 + (1 + G1vs.Other + G3vs.G5 || Subject) + (1 + G1vs.Other + G3vs.G5 || Item), family = Poisson (link = "log")
b) ICAref. ~ Half * G1vs.Other + Half * G3vs.G5 + (1 + G1vs.Other + G3vs.G5 + Half || Subject) + (1 + G1vs.Other + G3vs.G5 || Item), family = Poisson (link = "log")

5.1.3 Discussion

The results of the eye movements show a clear effect of gaze-following. Prior to the gaze cue, no anticipation or preference for a (group of) object(s) was created based on linguistic information. Upon the gaze cue, the target group was inspected more than another same-size group. Proportion of fixations show that the target object was considered more when it was part of a smaller group, and most of all when alone.

The cognitive load results show an effect of gaze-following only at the point of linguistic reference. We measured significantly lower values of cognitive load induced in the *GazeToOne* condition than in the other two conditions where more than one object was cued. Even though the difference between *GazeToThree* and *GazeToFive* was not statistically significant, we consider that this is not a "one vs. many" effect, but rather an indication of a graded effect that did not reach significance potentially due to a small difference in the number of cued objects between the neighboring conditions.

The more competitors there were in the target group, more ICA events were measured on the referent noun. Cognitive effort rose with the degree of entropy reduction, but only at step two (*GazeToOne* (ΔH = 0 bits) < *GazeToThree* (ΔH = 1.585 bits) < *GazeToFive* (ΔH = 2.322 bits)). Interestingly, no differences in cognitive load were measured at step one, on the gaze cue, where the *GazeToOne* condition introduced the overall highest reduction in entropy (ΔH = 3.459 bits). These findings are consistent with the results of Experiment 4 (Chapter 4), where the effect of (context fitting) referential gaze cue on the cognitive load was also detectable only on the referent noun.

Similarly, Tourtouri et al. (2017) also examined, among other things, the effect of referential entropy reduction. They presented scenes of six objects with a stimulus sentence like *Find the blue ball*. By changing the number of blue objects and the number of balls in the

visual scene, they manipulated the degree to which the adjective *blue* and the referent noun *ball* reduced the referential entropy. Using the same measurement of cognitive load, they found effects of entropy reduction on the noun, but not on the adjective. One could argue that this lack of effect could be related to the ICA measurement, which potentially reflects the cognitive load only at the point of reference resolution, not reflecting the stages of visual search that happen prior to processing the actual reference. However, in our Experiment 5, where gaze cued an object that did not match the previous linguistic context, we indeed measured an increase in cognitive load at the point of the gaze cue. This finding suggests that the cost of processing visual information when it does not fit the current sentence interpretation is reflected in the immediate cognitive effort, as measured by the ICA.

We argue that our results are accounted for by the notion of surprisal rather than entropy reduction. Prior to the appearance of the gaze cue, no concrete anticipation was created about the referent noun, due to the fact that all presented objects were equally likely to become the target. Hence, presenting a gaze cue that fits the previous linguistic context did not raise surprisal on the cue itself, and consequently, no cognitive effort was measured on it. Upon the gaze cue, however, more information was acquired about the potential target object, inspiring a more concrete anticipation, which was then measured on the cognitive load induced by the referent noun, reflecting the noun's surprisal given the linguistic and visual context. The fewer competitors an object had within the cued group, the higher the certainty that this object would be mentioned, and consequently, the lower the surprisal was on the referent noun. By inspiring a shift in visual attention, the gaze cue increased the probability of an object to become the next referent. Hence, the specificity of the gaze cue is reflected on the cognitive effort required for processing the linguistic reference.

Alternatively, the null effect at the gaze cue can be explained by the visual nature of the cue. A visual cue, stemming from a different modality, potentially does not reflect entropy reduction on a par with an explicit referential expression. In order to examine this argument, we conduct an additional experiment in which we swap the roles of the visual cue and the referent noun. The referential expression that completes the sentence is underspecified in the given visual context, and it is the visual cue, presented after the sentence, that highlights only one target object, thereby fully disambiguating the target referent. Such a design allows us to see the importance of the modality that the information comes from, and to disentangle the effects of surprisal and entropy reduction on cognitive load.

5.2 Experiment 8

This study aimed at shedding more light on the effect found in Experiment 7 by swapping the roles of the referent noun and the referential gaze cue in the step-wise reduction of referential uncertainty. Here, at step one, an underspecified linguistic reference highlights a group of objects, while it is the gaze cue that finally resolves the reference by identifying the target object at step two. In Experiment 7 we did not find any differences in cognitive load at step one, namely, on the gaze cue, even though it was gaze-following that allowed for the effect measured at the referent noun. Here, we are interested in seeing whether the gaze cue can take the role that the noun typically has, namely, of resolving the reference by identifying the target object. Simultaneously, the linguistic reference takes the role which the gaze cue had in Experiment 7, namely, of simply highlighting a group of objects without actually disambiguating the target. To this end, we changed the order of appearance of the two cues and presented the gaze cue after the referent noun. Even though this is not the typical ordering of gaze and referring expressions, it is nevertheless not uncommon. After having uttered an ambiguous reference, people tend to correct themselves by using a kind of linguistic repair, or a form of visual deixis. For example, one could first utter a sentence like *Could you pass me my cup?* and only later realize that the interlocutor cannot distinguish which is the intended one on the table with multiple cups. In this situation, one could either try to distinguish the cup by mentioning more detail *the dark blue one*, or cue to the right object by pointing with the hand, or, in case of small distance, with the gaze and head direction.

Experiment 8 was aimed at answering the following research questions:

2 If, upon hearing an underspecified referring expression, a referential gaze cue unambiguously identifies the target object, could gaze carry the same amount of surprisal as a referent noun does, and therefore elicit the same amount of effort?

a) Does an underspecified linguistic reference induce a graded cognitive load effect at the referent noun, relative to the level of remaining referential ambiguity?

could gaze carry the same amount of surprisal as a more/less predicted noun does and therefore elicit the same varying amounts of effort

We manipulated noun specificity and hypothesized a modulated effect on cognitive load, measured directly on the referent noun at step one, due to the number of objects that carry the same label. Even though no such effect was measured on the gaze cue in Experiment 7, we anticipated that a referring expression, as a point of linguistic reference resolution, would induce measurable differences in cognitive load, regardless of the fact that the given information was ambiguous. Moreover, at step two, we hypothesized the visual cue to show effects of reference resolution, resembling those measured on the noun in Experiment 7.

5.2.1 Method

This experiment made use of a 3×2 experimental design. We manipulated the number of named objects, resulting in three levels of **Noun Specificity**: *namingOne* – a single object in the scene selected by the referent noun, *namingThree* – three objects in the scene selected by the noun, and *namingFive* – five objects selected by the noun. Moreover, the existence of the gaze cue was manipulated, creating the **Gaze** independent variable (no gaze vs. referent gaze).

Participants

30 native speakers of the German language (23 female) with normal or corrected-to-normal vision took part in the study. We recruited Saarland University students and they were monetarily reimbursed for their participation. Participants' ages ranged from 19 to 39 years old (M = 23.7). None of the participants were familiar with the previous studies.

Items

Each participant was presented with 24 items. As previously, trials consisted of static visual displays and auditorily presented linguistic stimuli. Linguistic stimuli remained constant within one item. These were, again, simple German sentences with the SVAdvO structure. Sentence subjects were, as previously, kept relatively neutral and balanced within the items by using only German equivalents for *man* and *woman*.[8] Filler items introduced variation by using other subjects as well.

Visual displays included six objects in total, but only two object types. All objects presented in the scene fit the verb selectional preferences, and it was first upon hearing the noun that an object (type) could be selected as relevant. The results of two pretests, conducted with 20 German native speakers, showed that all presented objects were perceived to be equally likely to be mentioned as the target, as well as that all objects were labeled correctly, that is, recognized as depicted. Fig. 5.5 illustrates an example trial showing onions and bananas in three different ratios depending on the level of the Noun Specificity variable, namely, 1 : 5 (*namingOne*), 3 : 3 (*namingThree*) or 5 : 1 (*namingFive*). The position of the target object was counterbalanced in the whole experiment. The schematic eyes presented in the middle of the display created an additional manipulation. Sentence finally, they either cued the target object (referent gaze), or closed (no gaze). As was the case in previous experiments, in the no-gaze level the eyes closed in the relevant time-window, in order to

[8] An exhaustive list of all item sentences used for this experiment is given in Appendix C, Table C.3.

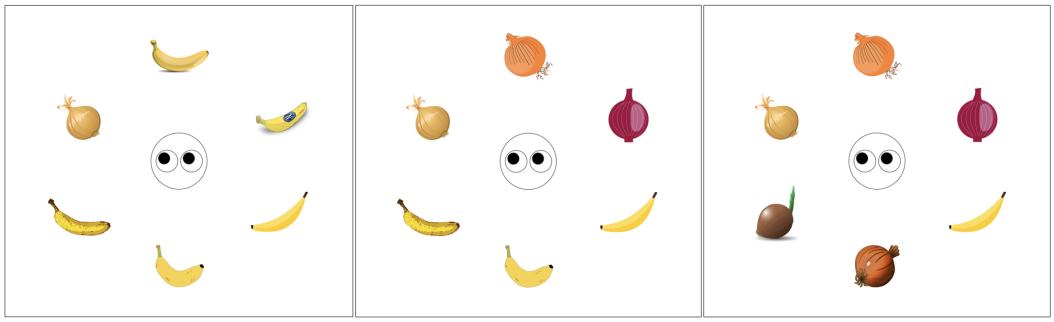


Figure 5.5 Exp. 8 – Three levels of Noun Specificity. From left to right: *namingOne*; *namingThree*; *namingFive*. Linguistic stimulus was kept constant within one item: *The woman peels on Sunday the onion.*

induce the same amount of change in the visual context without introducing any additional information. Half of the items (12) were presented in the no-gaze level of the Gaze variable.[9]

Please refer back to Fig. 5.1b for the illustration of a trial timeline. Initially, 2500 ms before sentence onset were allowed for the participants to familiarize themselves with the visual display. During the sentence, the schematic eyes were directed straight ahead, and it was only 600 ms after the final referent noun that the gaze cue appeared. The gaze cue was presented for 900 ms. Finally, the eyes went back to looking straight ahead for the last 100 ms.

After each trial a question was presented regarding the target object, or the sentence content. It was either a yes/no or a multiple-choice question, where the participant would choose one out of four given options.

Fillers

The experiment included 26 filler trials. Half of the filler trials were presented in the no-gaze level of the Gaze variable, maintaining the overall balance of the two Gaze levels during the experiment. Fillers presented the same pattern of the visual display, but additionally also introduced displays where two or four objects were named. Hence, balance was created within the experiment for each of the combinations of the number of named and unmentioned objects. Variation was introduced also in terms of the linguistic stimuli. For illustration: *In der Garage verleimt der Mann den Schrank* (literal translation: In the garage glues the man the wardrobe) is a filler sentence that was presented with four different wardrobes and two tables.

[9] An exhaustive list of all visual displays used for this experiment is given in Appendix C, Table C.10.

The Latin square design was used to create 6 lists with 24 items and 26 fillers each. Hence, one list included each item in only one condition. The order of the trials was pseudo-randomized manually, and an additional six lists were created with the opposite order of presentation.

The experimental procedure was the same as in Experiment 7 (see p. 94). The duration of the experiment was approximately 20 minutes.

Variable Coding and Data Analysis

The variables used in this experiment resemble those used previously (for a detailed description, see Experiment 7, p. 94). Here, we will mention only the differences in the current analyses.

The analysis of the new inspections included only the inspections of the target object and the representative distractor (the object that was present at each level of the Noun Specificity manipulation). We considered new inspections in two temporal regions of interest: a) the Reference region of interest: showing how the referent noun influenced visual attention, that is, which (group of) object(s) was anticipated to be the target after the uttering of the noun; and b) the Gaze region of interest: showing if the cue was followed, and if it changed the visual attention pattern.

5.2.2 Results

Before mentioning the results obtained from our measurements, we would like to mention that the accuracy of the participants' performance on the post-trial questions was 86.53% in total. Interestingly, the questions after the *no-gaze* trials were answered with 81.11% accuracy, while the accuracy was 91.94% after the trials with the *referent gaze*.

Proportion of Fixations

Fig. 5.6 illustrates the proportions of fixations to the target object during a trial. Three visual conditions are presented without the gaze cue (left-hand side) and in the referent gaze level of the Gaze variable (right-hand side). The plots are aligned to the referent noun onset, represented by the solid line. The dashed line presents the gaze cue onset. Note that the eyes would close at this point in the no-gaze manipulation.

It is evident that the referent noun guided participants' visual attention to the target object relative to the number of objects it selected (consider both Fig. 5.6a and b). Upon the reference, the target object was fixated the most in the *namingOne* manipulation. Also, there is a gradual difference between the other two conditions where the target is fixated more in

Table 5.5 Exp. 8 – Results of the main models fitted for the new inspections analysis for both reference and gaze regions of interest.

1. Reference region of interest

	a) Target inspections				b) Distractor inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.824	0.069	-26.470	< 2e-16 ***	-3.502	0.152	-23.092	<2e-16 ***
N1VS.OTHER	1.025	0.119	8.619	< 2e-16 ***	0.228	0.260	0.877	0.381
N3VS.N5	0.504	0.167	3.013	0.003 **	0.340	0.304	1.283	0.200

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1 a), b) TargetInsp ~ N1vs.Other + N3vs.N5 + (1 + N1vs.Other + N3vs.N5 || Subject) + (1 + N1vs.Other || Item), family = "binomial"

2. Gaze region of interest

	a) Target inspections				b) Distractor inspections			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.020	0.066	-15.496	< 2e-16 ***	-4.463	0.352	-12.687	< 2e-16 ***
N1VS.OTHER	0.324	0.121	2.680	0.007 **	0.660	0.340	1.650	0.099
N3VS.N5	0.406	0.149	2.732	0.006 **	-0.394	0.574	-0.686	0.493
GAZE	-0.950	0.118	-8.062	7.53e-16 ***	1.121	0.426	2.631	0.009 **
GAZE : N1VS.OTHER	1.095	0.242	4.531	5.86e-06 ***	-1.099	0.780	-1.374	0.169
GAZE : N3VS.N5	0.619	0.297	2.083	0.03728 *	-0.814	1.148	-0.709	0.479

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

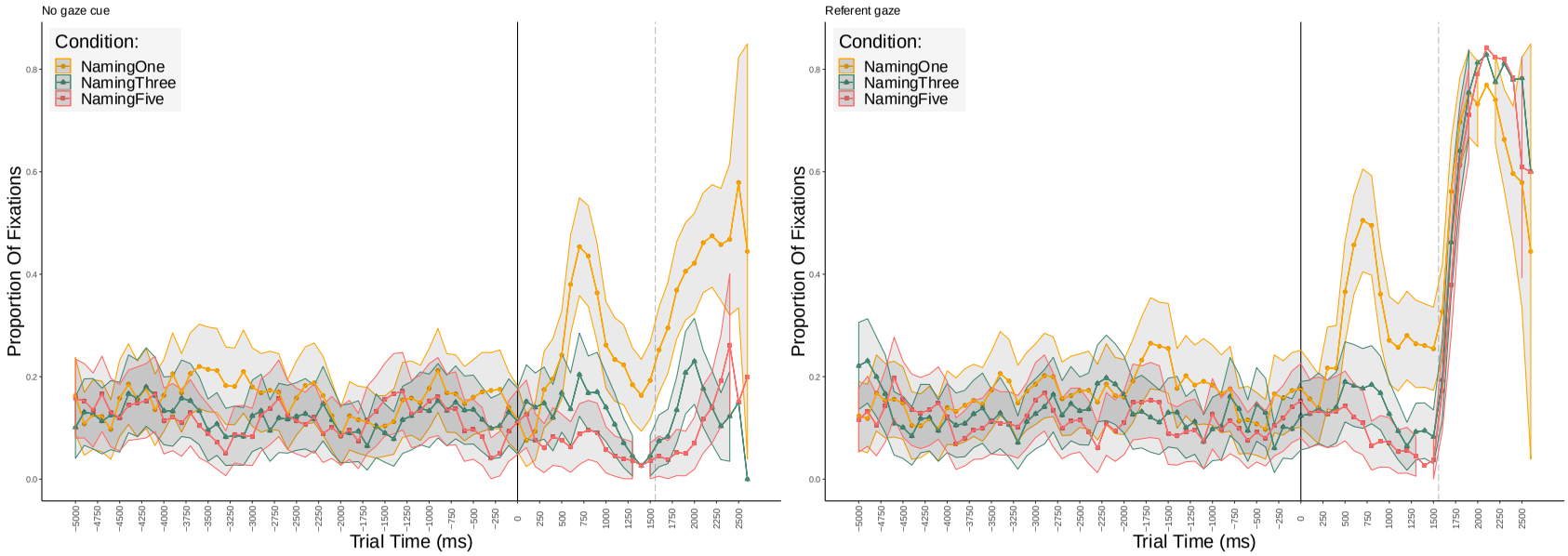
2 a) TargetInsp ~ Gaze * N1vs.Other + Gaze * N3vs.N5 + (1 + N1vs.Other + N3vs.N5 + Gaze || Subject) + (1 + N1vs.Other + N3vs.N5 || Item), family = "binomial"
2 b) DistractInsp ~ Gaze * N1vs.Other + Gaze * N3vs.N5 + (1 + N1vs.Other || Subject) + (1 | Item), family = "binomial"

namingThree than in *namingFive*. Moreover, Fig. 5.6b shows a clear effect of gaze-following, since, upon the gaze cue, the target object is equally fixated in all three levels of the Noun Specificity manipulation.

New Inspections

We considered the probability of an inspection to fall in an AoI during a temporal region of interest (from noun onset; from gaze cue onset). First, we conducted statistical analysis of new inspections of the target and the representative distractor in the reference region of interest (200 ms from noun onset to gaze onset). Second, we considered the gaze region of interest (during the presentation of the gaze cue). The main models and collected results are given in Table 5.5. Detailed information about the results of the further comparisons is given in Appendix C, Table C.4, and illustrated by Fig. C.11 and Fig. C.12.

Considering the reference region of interest, the analysis of target inspections showed significantly more inspections of the target in *namingOne* ($M = 0.243$, $SD = 0.429$) than in the other two conditions ($p < 0.001$). In addition, *namingThree* ($M = 0.129$, $SD = 0.336$)



(a) Three visual conditions in the no-gaze manipulation

(b) Three visual conditions in the referent gaze manipulation

Figure 5.6 Exp. 8 – Proportion of fixations to the target object aligned to the referent noun onset (solid line). The dashed line represents the onset of the gaze cue.

inspired more fixations of the target ($p = 0.003$) than the *namingFive* condition ($M = 0.083$, $SD = 0.275$).

Distractor inspections showed no such effects, suggesting that the representative distractor object was looked at with no significant difference within the Noun Specificity manipulation.

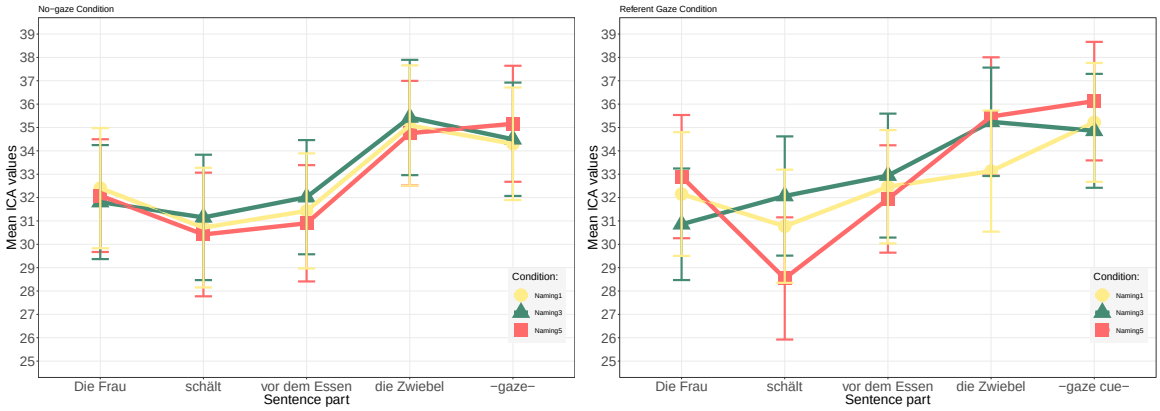
Furthermore, we considered the gaze temporal region of interest which shows a shift in attention inspired by the gaze cue. The model fitted for target inspections shows a main effect of Gaze ($p < 0.001$) proving that the target was inspected more when the referent gaze was presented (no-gaze: $M = 0.190$, $SD = 0.392$; ref. gaze: $M = 0.367$, $SD = 0.482$). Moreover, we find significant interactions between Gaze and the *namingOne* vs. Other comparison ($p < 0.001$), as well as Gaze and the *namingThree* vs. *namingFive* comparison ($p = 0.037$). Further comparisons reveal that without the gaze cue, *namingOne* ($M = 0.286$, $SD = 0.453$) inspired significantly more new inspections to the target than the other two conditions ($p < 0.001$). Also, *namingThree* ($M = 0.193$, $SD = 0.396$) showed significantly more new inspections of the target ($p = 0.004$) than *namingFive* ($M = 0.105$, $SD = 0.307$). Importantly, in the subset of referent gaze, none of the two comparisons showed significant differences (*namingOne*: $M = 0.332$, $SD = 0.472$; *namingThree*: $M = 0.394$, $SD = 0.489$; *namingFive*: $M = 0.371$, $SD = 0.484$).

The model fitting new inspections of the distractor reveals only a main effect of Gaze ($p = 0.009$), suggesting that the distractor was inspected more when the gaze cue was not presented (no-gaze: $M = 0.031$, $SD = 0.173$; ref. gaze: $M = 0.011$, $SD = 0.106$).

The Index of Cognitive Activity

Fig. 5.7 illustrates the mean ICA values in 600 ms time-windows at five relevant points during a trial. We show two graphs for the two levels of the Gaze variable. The analysis was relevant at two points where the entropy of the scene was reduced, namely, the gaze time-window and the reference window. From the two graphs, it is evident that no differences among the conditions are to be expected either in the reference or in the gaze time-window.

Presented in Table 5.6, the full model fitted for the mean ICA events in the reference time-window did not show any significant differences among the relevant conditions (*namingOne*: $M = 34.276$, $SD = 14.289$; *namingThree*: $M = 35.471$, $SD = 13.123$; *namingFive*: $M = 35.328$, $SD = 13.106$). The only significant effect was the interaction between experiment Half and the comparison *namingThree* vs. *namingFive* ($p = 0.027$). Further comparisons, however, showed that *namingThree* vs. *namingFive* was not significant either in the first experiment half ($p = 0.283$), or in the second half ($p = 0.283$), but that the interaction is carried by the trends in the opposite direction in the two halves. Hence, we will not consider this finding as relevant for the interpretation of the current results. Interestingly, the subset of



(a) No-gaze condition.

(b) Referent gaze condition.

Figure 5.7 Exp. 8 – Mean ICA values during the four time-windows of a sentence and the additional gaze time-window. Note that the points labeled as *die Zwiebel* (Reference time-window) and *-gaze-* (Gaze time-window) represent the time-windows relevant for the analysis (95% CI error bars).

the first experiment half revealed a significant *namingOne* vs. Other comparison ($p = 0.046$). However, due to the very small size of this effect, we will also not consider it as relevant for the interpretation.[10]

Considering the gaze time-window, the full model revealed only a main effect of experiment Half ($p = 0.014$), showing that the overall cognitive effort was higher in the second half of the experiment (1st half: $M = 34.479$, $SD = 13.411$; vs. 2nd half: $M = 35.490$, $SD = 13.699$).

5.2.3 Discussion

The results of the participants' eye movements show that, as expected, hearing the referent noun drew the visual attention to the mentioned objects. When the reference was underspecified, and thus visually ambiguous, all potential targets were considered. Hence, more attention was focused on the target object when there were fewer competitors. Upon perceiving the gaze cue, however, the three levels of Noun Specificity did not differ any more, proving that the visual cue inspired gaze-following, which focused the visual attention on the specific target object. In addition, the representative distractor object, present in all conditions, was inspected more when the gaze cue was not present.

The analysis of the ICA events showed no relevant significant differences, either at step one or step two of uncertainty reduction. We understand the null effect at step one as a

[10]For the results of the further comparisons see Appendix C, Table C.5.

Table 5.6 Exp. 8 – Results of the main models fitted for the ICA analysis.

	a) Reference time-window				b) Gaze time-window			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	3.498	0.064	54.50	< 2e-16 ***	3.487	0.071	49.23	< 2e-16 ***
N1VS.OTHER	-0.040	0.027	-1.50	0.133	-0.008	0.018	-0.44	0.657
N3VS.N5	0.004	0.025	-0.14	0.888	-0.030	0.016	-1.89	0.059
HALF	0.016	0.015	1.07	0.284	-0.322	0.013	-2.46	0.014 *
HALF : N1VS.OTHER	0.062	0.037	1.69	0.092	-0.007	0.028	-0.25	0.805
HALF : N3VS.N5	-0.071	0.032	-2.22	0.027 *	-0.027	0.032	-0.86	0.391
GAZE	-	-	-	-	-0.016	0.017	-0.95	0.342
GAZE : N3VS.N5	-	-	-	-	-0.013	0.027	-0.47	0.640
GAZE : N3VS.N5	-	-	-	-	-0.012	0.031	0.39	0.698

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) ICAref. ~ Half * N1vs.Other + Half * N3vs.N5 + (1 + Half * N1vs.Other + Half * N3vs.N5 || Subject) + (1 + N1vs.Other + N3vs.N5 || Item), family = Poisson (link = "log")
b) ICAgaze ~ Half * N1vs.Other + Half * N3vs.N5 + Gaze * N1vs.Other + Gaze * N3vs.N5 + (1 + N1vs.Other + N3vs.N5 + Half + Gaze || Subject) + (1 + N1vs.Other + N3vs.N5 || Item), family = Poisson (link = "log")

replication of the result from Experiment 7. The null effect at step two can be explained by the visual nature of the cue. We will thus discuss the results obtained from this experiment in the following section, considered together with the previous experiment.

There are two limitations of the design of this experiment that are worth noting before interpreting the results. We argue, however, that the choices made regarding the experimental design were unavoidable in the present context, that is, necessary to create a well-controlled experiment that allows for a comparison with Experiment 7.

First, since this experiment was the first in which we presented the gaze cue sentence finally, in order to have a baseline for the potential effects of the gaze cue, a no-gaze condition was necessary. As mentioned previously, the manipulation of gaze existence was balanced, resulting in half of the trials not having a gaze cue. Consequently, even though the referent noun made some objects more probable by naming them, one could argue that having the eyes close, rather than cue the target object, made it equally probable that none of the objects would be revealed as the target. For instance, in the *namingOne* condition, the probability of the mentioned object to be cued was not 1, but due to the no-gaze appearing as often as referent gaze, it was actually only 0.5. Furthermore, if we consider the probability of an objects in a named group of three, it was not 0.33, but due to the 0.5 probability of the target not being identified at all, each object had only a 0.17 probability of being identified as the target. Finally, the probability of an object from a mentioned group of five to be cued was only 0.1. Hence, it is reasonable to assume that in the context of Experiment 8 it was too costly to create predictions about which object would be cued.

Second, in order to make sure that participants do not give up in the potentially overwhelming visual context, but pay attention to both linguistic and visual information, each trial was followed by a question. The questions concerned the visual characteristics of the presented target object, other objects in the highlighted group, or the content of the sentence. The questions always referred only to the information available to the participants. Hence, since the target referent was not always disambiguated, the question about the target object would never follow such a trial. Even though the trials with an identified target would not always be followed by a question about that object, participants could realize that they would never be asked about something that they do not know about. Hence, step two (of uncertainty reduction) in this experiment, rather than only concluding the processing of the sentence, also becomes a prelude to the upcoming task. Assuming that the participants' motivation during the trial was, rather than to correctly comprehend the message, more likely to correctly answer the subsequent question, this might have influenced the perception of the gaze cue. Consequently, receiving the visual information about the target potentially became less important, since, if it was lacking, it could be expected not to be relevant.

Finally, we would like to reiterate that it was important to include a no-gaze condition, since without a baseline, the potential effects on the gaze cue sentence finally could not be reliably interpreted. Moreover, the post-trial questions were essential for motivating the participants to engage not only with the linguistic, but also the visual information from a rich visual display. Avoiding the mentioned drawbacks of the current design would mean creating an experiment that was very distinct from Experiment 7. For reasons of comparability with the previous experiment, we could not change the design to the extent required to assure controlled conditions and sufficiently strong manipulation. In the following section, we show that the obtained results are interpretable together with those obtained in the previous experiment.

On a minor note, the cognitive load differences seemingly show a trend at the point of the verb, which may mislead the reader to conclude that there was an effect of visual priming. Seeing that *namingFive* induced less cognitive load than other conditions, one might assume that it is the fact that there are five onions in the scene that primes the onion as the potential target. We would like to clarify this point against that interpretation. First, none of the apparent differences in the verb window are statistically significant. Second, *namingFive* and *namingOne* are, at the point of the verb, identical, in that there are five objects of the same type (onions or bananas, respectively). Hence, we should expect, if anything, that these two levels cluster together and differ from *namingThree*, which is a scene with three onions and three bananas. This trend is not there; rather, *namingOne* and *namingThree* cluster together, which is not meaningful in this setting.

In sum, even though we saw clear effects of both linguistic and visual cues directing participants' visual attention accordingly, this was not reflected in the measure of simultaneously induced cognitive load. We find these results to be consistent with our interpretation of the findings from Experiment 7, and we will hence discuss them together in the following section.

5.3 Chapter Discussion

The two experiments conducted in this chapter examined situated sentence processing in a context where a combination of visual and linguistic cues allowed for a step-wise reduction of referential uncertainty. The reduction of uncertainty was quantitative, since, rather than having different probabilities of mention in the given context, all presented competitors were equally probable, while the number of relevant objects was gradually reduced. The referential uncertainty was reduced in two steps: visually, by the means of cuing objects, and linguistically, by naming them. The two experiments altered the position of a visual cue and a referent noun at the two steps. As previously, we examined the visual attention and cognitive load induced during sentence processing.

At step one, we saw immediate shifts of visual attention towards the relevant objects, due to the gaze cue or the referent noun. The target object was inspected the most when being the only highlighted object, and gradually less, relative to the number of immediate competitors in the highlighted group. Hence, having a certain (group of) object(s) highlighted as relevant, either by linguistic or visual cues, draws one's visual attention to the object(s). This finding is not surprising in itself, but it is relevant for the assessment of the way in which this visual attention is reflected on the immediately induced cognitive load.

Moreover, visually or linguistically cuing a number of objects, and thereby subsetting the group of potential target objects, did not reflect on the cognitive load. Acquiring new information motivated a shift in one's visual attention to the relevant group of objects. However, considering more or fewer newly relevant objects, and thereby being able to exclude the others, did not induce different degrees of effort. We argue that this finding can be explained in terms of surprisal. No effects found either on the visual or linguistic cue at step one of uncertainty reduction suggest that at that point no anticipation was created for a specific target object. The given context included a number of presented objects that all fit the verb selectional preferences equally well. They all had the same probability of being mentioned as the target, thus not creating a basis for an informed anticipation regarding the target object. Hence, either seeing a visual cue towards a group of objects, or hearing a noun referring to them, did not show any gradual differences relative to the size of the

selected group. The lack of a difference between the referent noun naming (or gaze cueing) the target object versus more potential targets gives even stronger evidence for this claim, by showing that it is not only that different levels of ambiguity were not detected, but not even a difference between a clear unambiguous message and an underspecified ambiguous referring expression. These results are consistent with the findings from the previous chapter. There we found an effect on the gaze cue only when the cued object did not fit the previous linguistic context, that is, when it violated the prediction, or basic assumption of the visual and linguistic context contributing together to the meaning of the message.

At step two, where the reference was resolved, we found a graded effect relative to the number of immediate competitors measured on the cognitive load on the referent noun. Referring to the target when it was the sole visually highlighted object induced the least cognitive effort. This was the condition in which the target object had the highest probability of mention, due to the previous visual cue. The more objects there were in the cued group, the lower the probability of mention was for each of them. Consequently, the cognitive load measured on the noun rose with the size of the highlighted group. In contrast, when it was the visual cue that identified the target at step two, even though it was followed, it did not yield an effect on cognitive load. Not even an effect of gaze existence was detected.

We argue that in Experiment 8, at step one, the noun had already linguistically identified the target (although not visually), which led to the processing of the target label. Consequently, at step two, what was left was to visually identify the target, that is, to map the cued object to the already known label. Apparently, this visual mapping without linguistic processing does not undergo a modulation due to the different level of ambiguity in the visual context. In Experiment 7, reference resolution happened at the point of the noun where the target object was linguistically labeled, and hence simultaneously visually identified. In Experiment 8, however, reference resolution actually happened in a step-wise manner. First, the linguistic label for the target was given prior to the visual disambiguation of the object. Subsequently, the target object was visually identified through a visual cue. The processing of the label had already happened earlier, and what was done at the point of cue perception was just the step of mapping the label to one of the previously identified potential target objects.

Recently, comparable findings have been obtained by Ankener, Drenhaus, et al. (2018). They have examined the effects of multimodal surprisal in the visual world by employing and comparing both the N400 ERP component and the ICA measurement. While keeping the linguistic material constant, the authors manipulated the number of objects presented on the display that fit the selectional preferences of the verb, and were thus to be seen as potential targets. The verb was the first piece of information that allowed for the creation of anticipations for the upcoming target. Their VWP experiment showed that hearing the verb

led to a shift in listeners' visual attention towards the object(s) that fit the verb's selectional preferences, and could thus be anticipated to be mentioned. The effect of target probability was found on the cognitive load induced on the referent noun, namely, the fewer objects were present in the scene that fit the verb selectional preferences, the lower the cognitive effort was on the noun. However, even though the verb created an immediate shift in visual attention towards the fitting object(s), no differences in the induced cognitive load relative to the number of relevant objects were found on the verb. The results of their subsequent EEG experiment support these findings. These findings are in line with our null effect at step one in both experiments. No anticipation being created for the verb, no surprisal effects were measured on it. The fact that the verb allowed for subsequent creation of anticipations, due to raising the probabilities of some objects or one object to become the target, is not measured immediately on the verb.

In sum, as in the previous chapter, the results of the two experiments conducted here show effects of surprisal on cognitive load. Our findings indicate that the visual cues are incrementally included and actively influence sentence processing. Interestingly, when a visual cue was the piece of information that disambiguated the target referent, this was not reflected on the cognitive load. Not including the processing of labels, their integration into the sentence interpretation cannot be compared with that of a linguistic cue, such as a noun.

Chapter 6

General Discussion

In this chapter we conclude the thesis by firstly summarizing the results of the presented experimental work (Section 6.1). Secondly, we put these findings in the wider context of the current literature (Section 6.2). Thirdly, we mention the questions that remain open after our experimental investigations and present ideas for future work (Section 6.3). Finally, we give our concluding remarks for the presented work (Section 6.4).

6.1 Summary of Results

In Chapter 3 we presented our initial experimental work which examined the effects of multimodal predictive context on reference processing. **Exp. 1** and **Exp. 2** showed that, in comparison to a non-predictive context, our predictive linguistic context was too weak to create a facilitation for reference processing. In **Exp. 3**, however, once the relevant visual context was additionally introduced, we indeed detected such a difference. The restrictive verb changed the probabilities of the visually presented objects, only one of them becoming the most likely referent. Participants' eye movements showed that this object was anticipated upon presentation of the verb. Consequently, we measured lower cognitive load on the processing of the corresponding referent noun, and higher load on processing the other, unanticipated referent. Processing loads for the same two referent nouns in the non-restrictive context were not different from one another, and were positioned between the other two mentioned values. In addition, it is noteworthy that the cognitive load measured in Exp. 3, when the experimental sentences were presented together with the visual stimuli, was higher overall than in Exp. 2.

We found that the visual context has an immediately measurable effect on the cognitive load measured on the linguistic reference. First, the simultaneous presentation of linguistic and visual material led to an increase in cognitive load, due to the additional information

introduced through the visual modality. Second, the illustration of potential referents led to the anticipation of the target referent, which consequently facilitated the processing of its linguistic reference.

In **Exp. 4**, we introduced the speaker referential gaze cue to the same linguistic and visual stimuli. The cue was presented after the predictive information was introduced by the verb, and before the actual mentioning of the target referent. The gaze cue was always congruent, cuing the object that was about to be referred to. The participants' eye movements patterned differently relative to the existence of the gaze cue. The gaze cue was followed, leading to a shift in visual attention. Our results revealed that the existence of the visual cue in addition to the visual context further facilitates the processing of the linguistic reference. The cognitive load induced on the referent noun was overall lower when the gaze cue was presented, than when no visual cue was shown. Moreover, the one condition in which anticipation for a specific referent was created after the verb was the condition in which the load on the noun was even lower. Importantly, perceiving the gaze cue and considering the cued object as the potential target referent did not induce additional cognitive load. We initially hypothesized that if the cognitive effort required for the processing of the linguistic reference is reduced due to the existence of the gaze cue, this reduction of load should be compensated for at an earlier stage. If processing the referent noun becomes easier, this suggests that some of the work was already completed, presumably at the point when the additional information was introduced – at the gaze cue. Hence, we expected to see an increase in cognitive load on the gaze cue highlighting a certain object, followed by a reduction of effort required for processing the linguistic reference to that object. Interestingly, the observed reduction of load on the referent noun was not preceded by an increase in load on gaze. Hence, we did not find support for the hypothesis that visual cues facilitate the creation of a measurable uniform information density throughout a sentence. Nevertheless, referential visual cues have proven facilitatory for the creation of anticipations and the subsequent processing of linguistic reference.

Our subsequent two experiments were designed to further examine the perception of the gaze cue by manipulating its fit with the linguistic context, and its congruency with the following linguistic reference. In **Exp. 5**, presenting a gaze cue that does not fit the previous linguistic context, but gives a hint that the subsequent referent noun will not fit either, has revealed interesting results. First, even though the gaze cue was followed, and the cued object was considered, the preferred referent, that is, the one fitting the context, was continuously fixated in addition to the cued object. Second, we observed different cognitive load results in the two halves of the experiment. Initially, we measured an immediate increase in cognitive load on the mismatching cue itself. Interestingly, in this phase of the experiment,

the previously observed facilitation for the referent noun processing (Exp. 4) was not found, either for the fitting, or for the mismatching gaze cue. Later on, in the second half of the experiment, the gaze cue as such became costly, regardless of its fit with the linguistic context. This effect was followed by the facilitation of referent noun processing, whether or not the noun fit the linguistic context. As expected, we measured increased cognitive load on the mismatching referent noun throughout the experiment, and regardless of the gaze condition.

We argue that the anomalous condition with nonsensical sentences led to an accommodation effect, reflected in the difference between the first and the second half of the experiment. In the first half, the gaze cue towards a nonsensical (mismatching) object was very surprising, inducing high cognitive load. Since it was sometimes mismatching, the gaze cue was not trusted; rather, the participants awaited the referent noun to resolve the reference. Consequently, we measured no facilitation of the referent noun processing due to the existence of the gaze cue. In the second half, however, participants got accustomed to the occasional anomalous trials, and realized that the gaze cue, however surprising, was a reliable indicator of which referent was about to be mentioned. We argue that the reliability gave the gaze cue higher salience, which is reflected in the higher load on the cue overall, regardless of the fit of the cued object. Moreover, the established trust in the gaze cue in the second half of the experiment led to the facilitation of the referent noun processing, regardless of its fit to the sentence, replicating the facilitation effect from Exp. 4.

In **Exp. 6**, we manipulated the congruency of the gaze cue with the following linguistic reference. We presented four objects, two of which were equally likely potential target referents. Gaze was either congruent, cuing the target object, or incongruent, cuing the other fitting object which was not subsequently mentioned. The results showed that even though only one object was cued, the other fitting referent was also being fixated until the reference was made. We measured facilitation for the referent noun processing following the congruent gaze cue. Interestingly, the condition where the referent noun and the previous gaze cue were not congruent did not induce any disruption. In other words, the cognitive load induced by processing the reference was the same when no gaze cue was presented, and when gaze cued the other fitting object that was not mentioned. We argue that the gaze cue simply highlighted one potential referent, thus making it more salient. Since it was not addressed by the cue, but was equally fitting with the previous linguistic context, the other referent was thereby not actively excluded. If the referent noun fits the linguistic context, an additional visual cue elsewhere does not disrupt the processing of that fit.

The examination of the referential gaze cue conducted in Exp. 4, Exp. 5 and Exp. 6 manipulated the introduced information in a *qualitative* way. Only one referent was ever mentioned, but the probability of its mention varied. First, we looked into the effect of

the existence of the gaze cue. Second, we examined the effect of cue towards one object of various degrees of probability of mention given the previous context. We found that such *qualitative* differences in the transmitted amount of information are reflected on the cognitive load induced on the linguistic reference. However, we found no proof that this reduction of processing load was due to more load being exerted previously. The gaze cue towards referents of different probabilities did not show different processing requirements. We subsequently conducted two experiments that manipulated the information conveyed by the gaze cue and the referent noun in a *quantitative* manner. Rather than manipulating the probability of the target referent based on its fit to the context, we employed the gaze cue which highlighted the relevant object(s) from a larger group of equally likely potential referents.

In **Exp. 7**, gaze reduced the set of visually presented equally probable potential target objects to a numerically smaller group. Subsequently, the referent noun named the single target referent. The gaze cue led to a shift in visual attention towards the cued group of objects, and, depending on the size of the cued group, we measured differences in the cognitive load induced on the referent noun. Hence, once again we measured an effect of referential visual cue on the processing of the subsequent linguistic reference, but no evidence of the previous compensation for that facilitation. Subsequently, in **Exp. 8** we examine if the measured effect had to do with the modality that the relevant information came from. In this experiment, the referent noun was ambiguous in referring to a different number of objects (of the same type). The subsequent gaze cue was the disambiguating piece of information, always cuing only the single target referent. Hence, the roles of the visual referential cue and the linguistic reference were swapped from those assigned to them in Exp. 7. The referent noun, like the gaze cue in Exp. 7, inspired a shift in visual attention to the mentioned potential target objects. Interestingly, no differences in cognitive load were measured on the noun, even though it selected one or more objects; that is, the reference was either clear or ambiguous as to which was the target. Subsequently, the gaze cue was followed and shifted the attention to the specific target object. However, this reference resolution did not carry any differences in the induced cognitive load, regardless of the degree of previous ambiguity that was being resolved.

We argue that the null effect on the referent noun in Exp. 8 is comparable with the null effect on the referential gaze cue in Exp. 7. Based on the previous context, no specific anticipation was created for the piece of information contributed by the gaze cue, or by the referent noun in Exp. 8, beyond general context fit. The only cognitive load effect on the gaze cue we measured was when the cue did not match the current context. This finding suggests that even though there was an integration of the gaze cue with the current

situation model, there was no specific anticipation for the information it should introduce. In contrast, at the point of the referent noun in Exp. 4–7, an anticipation was created for the target referent, which was then reflected on the cognitive load induced by its processing. We have clearly seen that this anticipation is built based on the information gathered from both modalities. The relevant information from the linguistic context was enriched by the referential gaze cue highlighting specific referents, which resulted in facilitation of referent noun processing. Exp. 8, however, did not replicate this anticipatory effect on the piece of information resolving the reference. We argue that this is due to the referent noun explicitly introducing the linguistic label, when selecting a group of objects as potential target referents. Hence, even though it was not possible to identify the exact target referent, its label was already processed at the point of the noun. When the gaze cue subsequently resolved the referential ambiguity, the referent's label was already processed, and it was only mapping with its visual representation that took place. Hence, the gaze cue in Exp. 8 did not contribute any additional semantic information, and thus its contribution was qualitatively different than that made by the referent noun. We argue that the visual identification of the referent and the processing of its label are two different processes, where the visual identification does not carry additional load beyond the load of labeling the object.

6.2 Interpretations

Visual Attention

In Chapters 3 and 4 we have seen that the verb selectional preferences motivate inspection of the fitting objects before any of them is referred to, thereby replicating Altmann and Kamide (1999)'s seminal finding of anticipatory eye movements. Moreover, the work from Chapters 4 and 5 showed that the referential visual cue inspires an immediate shift in visual attention towards the cued object, allowing listeners to anticipate the upcoming linguistic material. This is a replication of a well-known effect of gaze-following (e.g., Hanna & Brennan, 2007; Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013; Staudte & Crocker, 2011; Staudte et al., 2014).

Moreover, in Chapter 4 we found that the shift in visual attention inspired by a visual cue does not override the anticipatory inspections of another object which was seen as a potential referent in the given context. Also, Exp. 5 gives an interesting insight, since the two halves of the experiment differed in terms of cognitive load effects, but not in terms of eye movement patterns. Hence, we conclude that the anticipatory considerations, motivated by linguistic or visual material, are not accompanied with additional cognitive effort. Also, it is possible for

the pattern of visual attention to remain the same, while the underlying integration of the perceived material changes.

In addition, the present work provides an insight into the cognitive load induced simultaneously with the mentioned eye movements. We have contrasted visual attention shifts which are a) motivated by linguistic information vs. a visual cue; b) towards more vs. fewer objects; c) towards a highly probable object vs. a possible but not previously considered object, and found that none of these comparisons resulted in a difference in the immediately induced cognitive load. A shift in visual attention was accompanied with immediate processing cost only when gaze cued an object which did not fit the established context.

Exploiting speaker referential gaze and shifting one's attention to consider (a group of) object(s) did not prove costly, but since it allowed the formation of expectations for the upcoming reference, it resulted in facilitatory processing of the upcoming linguistic material. Thus, by providing more direct evidence of processing effort than post-trial task performance, we have not only replicated, but further enriched the existing literature on the facilitatory effect of gaze on sentence processing (see Staudte & Crocker, 2011; Staudte et al., 2014; Knoeferle & Kreysa, 2012).

A Multimodal Account of Information-Theoretic Notions

If we approach the examination of language processing from the perspective of information processing, we must consider that the relevant information is available from different modalities. In situated communication, both linguistic and (extra-linguistic) visual information are combined to create a situation model. Our present findings show that the information from both modalities is incrementally combined in order to create anticipations for the upcoming linguistic material. Creating such anticipations allows for prompt and effortless processing of the upcoming material. If the anticipations are not confirmed, though, this can result in a higher processing cost than not having expected anything (Exp. 4). Hence, optimal listeners must always weigh the cost-benefit ratio of committing to an anticipation. As we have seen in the present work, listeners do not fully commit to the cued object when another object is objectively likely to be mentioned (Exp. 4, 5, 6).

After having examined the integration of multimodal information in light of the information-theoretic notions of surprisal, entropy reduction and uniform information density, we argue that the cognitive load results measured in all our experiments are essentially surprisal effects. Before explaining this claim, we will first show how the other two notions, namely, the uniform information density hypothesis and entropy reduction, are not compatible with our data.

Firstly, in all the presented studies, we found a null-effect on the first piece of information that allows for the subsetting of the visual context, and for creating an anticipation for the upcoming reference. However, we indeed find a reduction of processing cost on the subsequent referent noun. This finding speaks against the canonical understanding of the UID hypothesis in terms of information being distributed over the sentence, and thus, also its processing cost. In other words, the UID expects that the processing facilitation of one linguistic unit would be preceded by a higher processing effort made at a previous stage, which allows for that facilitation. Our results from the visual world paradigm do not provide evidence for this hypothesis. Rather, we find facilitation effects due to more information being provided either verbally or visually prior to the referent noun; however, the integration of this information did not prove costly in any way. Hence, we find no distribution of processing effort throughout the sentence, but only an effect on the point of reference resolution.

We argue that the cognitive load as measured by the ICA reflects the processing effort required for the integration of the given piece of information in the current situation model. Since the context is gradually built, earlier in the sentence it is only loosely defined. Hence, subtle effects of integration difficulty or facilitation are to be measured only in the later stages of a more defined context. We argue that at the point of the restrictive verb, the previous context is too unspecific to result in an effect of integrating such a verb as opposed to an unconstraining one.

Regarding the gaze cue, in the context of the current experiments, and typically in situated communication, the referential visual cues are tightly coupled with language. The gaze cue is thus not seen as a substitute for the linguistic reference, but rather as an accompanying visual cue. We argue that listeners did not integrate the piece of information contributed by gaze as final and definite, but rather as only relevant and valuable. Consequently, as shown by the eye movement data, when perceiving a gaze cue towards an object they did not anticipate, listeners broadened their anticipation to include that object, but never dropped their already created anticipation for another possible referent.

In sum, we measured effects of the cost of integration in the situation model. As the sentence continued and the context became more restrictive, such effects became large enough to be detected. The visual cue not being seen as a substitute for an explicit and definite mention, the cost of processing a piece of information may be lowered due to a previous visual cue, but this is not necessarily measurable on the cost that the cue integration induces.

Secondly, in Chapter 5, we framed the experimental manipulations in such a way that they allowed for effects of entropy reduction. The entropy about the referent noun was gradually reduced in two steps, in an abrupt or more gradual manner. We found no differences in

immediate cognitive load due to the step-wise entropy reduction, but rather only differences on the referent noun occurring as a result. The obtained results cannot be accounted for in terms of entropy reduction. The null-effect on the first step of entropy reduction was found when it was either a visual cue or a linguistic cue that allowed for the reduction of entropy. This excludes the possibility that the effect was due to the modality from which the relevant information was perceived.

A similar lack of entropy reduction effect was also found by Ankener, Drenhaus, et al. (2018). In contrast to our visual cue, the authors examined entropy reduction by using linguistic cues (verb selectional preferences) in a visual context. Comparably to our findings, they also saw cognitive load effects on the referent noun, but not on the piece of information that allowed for the reduction of entropy, namely, the verb. Moreover, Tourtouri et al. (2017) also examined the gradual reduction of referential entropy in the visual world paradigm, and found comparable effects only on the final point of disambiguating the referent.

Even though the concept of entropy reduction fits well with the fact that it is the reduction of uncertainty about the referent that is essentially happening in our last two experiments, it should be noted that our potential referents are always present in the visual context. Hence, the target object is, in a sense, already revealed at trial onset – one just needs to pinpoint it, while the other objects still remain visually present.

Finally, we claim that the obtained differences in cognitive load are measured as a consequence of the previously created anticipations for the target referent. The confirmation of a specific anticipation led to facilitated processing of the referent noun, while its contradiction led to higher processing load. In other words, the cost of incorporating a referent noun in the situation model is relative to its probability of mention given the previous context. The creation of anticipations was based on both the linguistic and visual context. Verb selectional preferences made some of the visually present objects more likely to be mentioned. Similarly, the visual cue highlighting one or more objects made those particular objects more salient, and thus considered more likely to be mentioned as the target. We found evidence that the immediate linguistic processing measured on the referent noun is affected by the anticipations created based on a) the verb selectional preferences and the visual context (Exp. 3 and 4), b) the information contributed by the visual cue in combination with the linguistic and visual context (Exp. 4, 5 and 6), and finally c) the visual context and visual cue (Exp. 7).

Our findings are in accordance with the previous literature on surprisal effects in language processing (e.g., Demberg & Keller, 2008; Fernandez Monsalve et al., 2012; Fossum & Levy, 2012; Smith & Levy, 2013; S. Frank et al., 2015) Moreover, they further contribute to this literature by providing insight into visually motivated surprisal effects. Also, Ankener,

Drenhaus, et al. (2018) found comparable effects of multimodal surprisal, additionally comparing the presently used measure, the ICA, with the ERP components.

Linguistic vs. Visual Cues

In the introduction of the dissertation we hypothesized that the integration of referential visual cues during sentence processing could potentially be compared to that of nouns, since they similarly typically refer to only one referent. Entertaining this idea, we have created experimental conditions which allow for a comparison of the referential gaze cue to both noun and verb processing.

In Chapter 4, we described three studies where the referential gaze cue was highlighting only one object, and in that sense it resembled the referent noun that would name that object. We measured elevated processing cost immediately on the gaze cue **only** when it did not fit the previous linguistic context (Exp. 5). This suggests that the information contributed by the visual cue is integrated in the current situation model, since we measured a disruption when the listener failed to do so. However, when the cued object, that is, the information contributed by the gaze cue, fit the linguistic context, we did not measure any differences as a result of the different probabilities of mention of the cued objects. Such differences were reliably measured on the subsequent referent noun; that is, the differences were relative to the probability of mention of the target referent. Hence, the processing of the information regarding the referent was not moved to the point of the gaze, but was still integrated on the noun. In addition, we saw no disruption of referent noun processing upon a gaze cue that had cued a different, but equally probable object (Exp. 6). Moreover, the anticipation created based on the previous context was not abandoned due to the gaze cue on a different object (Exp. 4–6). These findings suggest that the information contributed by visual cues is not integrated in the listener's situational model in a definite manner, rather, this information is made more salient, but not considered explicit until referred to linguistically.

If we contrast Exp. 3 and Exp. 4, the gaze cue in Exp. 4 is essentially introducing the same piece of information as the noun in Exp. 3. One could anticipate that the same effect found in Exp. 3 would now be detected on the gaze cue in Exp. 4. Such an effect was not found. We argue that this is so while the gaze cue is seen as accompanying language, and, at least in the context of these two experiments, it was always followed by a linguistic reference. Knowing that a linguistic reference will follow and will definitely disambiguate the target referent, the information contributed by the gaze cue was, similarly to the verb, used to further define the situation model in combination with the other visual and linguistic information. Hence, the gaze cue is accompanying language, but not substituting it. This claim holds in the context of our experiments, but, we argue, also for the most typical situated communication situations.

The information contributed by the gaze cue is integrated in the situation model, as confirmed by the higher cognitive load measured on the cue when it did not fit the context (Exp. 5). As that experiment continued, and listeners noticed that the gaze cue was always a reliable indicator of the final referent noun, the gaze cue as such became more costly than not having a gaze cue, and not only when it was anomalous. This finding we explained in terms of saliency and attention. The gaze cue as such became more salient, and hence, the listeners were more alert to the information it was contributing, which induced higher immediate load.

When the gaze cue added new insight it was considered, but not costly. So, when it was misleading, it was the same as if it had been absent. When providing correct information, however, it facilitated the subsequent noun processing. Hence, the semantic information contributed by the gaze cue is processed and incorporated as relevant for the situation, but not determinant. The final say was expected from the noun. As we could see from the eye movement patterns, the existing predictions about fitting objects were not falsified on the gaze cue. They were broadened to include the cued object, but the previously considered object was never dropped due to the gaze information. This shows that the gaze was not the place of falsifying one's anticipations; rather, that was the noun. We argue that this is not due to listeners' inability to process a visual cue as determinant information, but rather that the typical situational context, and use of the speaker referential gaze cue, creates such an interdependent situation in which gaze hints at what will be referred to, without making the explicit verbal reference redundant. Hence, even though the gaze cue contributed to the facilitated processing of the corresponding referent noun, the processing of the two cues is obviously not comparable.

Tentatively, we can compare the visual cue to the information contributed by the verb, in that it highlights the cued material, without demanding a commitment to it, but rather motivating anticipations for the upcoming explicit reference. We have seen that verb selectional preferences create anticipations for fitting objects to be mentioned, the violation of which creates a disruption (Exp. 5). However, when we violated the context created by the verb and visual cue (Exp. 6), such a disruptive effect was not found, since the referent object still fit the linguistic context. This suggests that the visual and linguistic cues are nevertheless considered with a different importance level, visual cues being seen as an addition to linguistic ones. As mentioned previously, it is our understanding that this is due to the relationship between referential visual cues and language, in that they accompany, rather than substitute for each other.

In addition, Exp. 7 introduced gaze that cued more than one object, thereby subsetting the group of potential target referents from 11 to, for instance, 5 objects. The function of the gaze cue was thus comparable to that of the constraining verb, since rather than disambiguating

the target object, it reduced the ambiguity about the target to a certain degree. The null effect measured both on the constraining verb and on the gaze cue is comparable in the context of our present work, since they were both integrated in an early stage of sentence processing, when the situation model was not constrained enough for strong integration effects. Rather, they both contributed to the constraining of the multimodal context, and the creation of more defined anticipations, that influenced the integration of the subsequent linguistic reference.

Moreover, we would like to mention a recent study conducted in our lab that used the same paradigm, but created a multimodal context in which the noun and the presented scenes allowed for the anticipation of the verb. Muljadi (2018) depicted actions, rather than objects, and, due to sentence restructuring, nouns came before verbs. Hence, some of the depicted actions could be discarded since they did not include the mentioned noun, thereby motivating the anticipation for a verb. Nouns reduced the uncertainty about which action will be mentioned by the verb, making the verb less surprising. The results revealed cognitive load differences, as measured by the ICA, only at the point of the verb, and not on the noun. Consistent with our interpretation, in the context of this study, it was the nouns that together with the illustrated actions created a defined model that the verb was to be integrated into.

Additionally, we found that after the noun processing was completed, there was no effect on the subsequent disambiguating referential gaze cue (Exp. 8). When interchanging the referent noun with the referential gaze cue, we observe differences in processing cost. Even when resolving the reference ambiguity, the cue proved not to be costly, since the linguistic label was processed previously, on the noun. Hence, the information contributed by the gaze cue was not integrated with a cost, since the semantic information was already processed at the noun. Visual mapping of a label to a specific object seems not to recruit the same processes as the integration of the label. Hence, the gaze cue, when occurring post-noun, does not activate a specific label for an object when shifting the visual attention, and hence does not induce cognitive load, as initially expected.

There is evidence that the semantic information from pictures and words is integrated in the sentence context in similar ways in the brain. Willems, Özyürek, and Hagoort (2008) have contrasted nouns with illustrations of the same object, by presenting the image of an object during the auditory presentation of a sentence. The semantic information was found to be incorporated with the same neural time course (ERP evidence) and by recruiting overlapping brain areas (fMRI evidence). The authors conclude that the language comprehension system does not restrict itself to one source of information. In addition, they note that pictures are different from co-speech gestures, in that the latter are bound to language, not clearly representing their meaning without the language (McNeill, 1992; Krauss, Morrel-Samuels, & Colasante, 1991). Our results are compatible with such interpretation of the referential visual

cues. The gaze cue is also not clearly representing its meaning in itself – it suggests what is relevant, but does not explicitly determine the referent. We argue that this is due to the typical way of coupling speaker gaze cues with the uttered linguistic material. They accompany each other, rather than exclude and substitute. Moreover, we argue that due to its more explicit nature, language is relied on for the acquisition of definite information. Arguably, it is only when the language fails to make sense that the visual cues are relied on for the creation of a definite model of the intended meaning. The information contributed by the gaze cue is being considered as relevant, but not as final. Such a commitment to gaze would be too risky, since it could often prove wrong. The situational model is defined gradually and incrementally, and there are no costs on the processing effort of considering relevant information prior to the creation of a relatively defined situation model. Since it is coupled with language, the gaze cue is seen as accompanying it, but not substituting for it.

In sum, we have found no evidence that the processing of the referential gaze cue is comparable to that of a referent noun. We find reliable evidence that the referential visual cues are contributing incrementally to the creation of a situation model, strongly impacting the noun processing, without being interchangeable with the corresponding linguistic reference.

6.3 Future Work

The burning question that remains open, and the main idea for continuing this line of work, is whether the processing load effects measured throughout the present work could also be measured on the visual cue. A way to test this would be to create such a context in which it is the gaze cue that introduces the target referent, accompanied by language, as is typically the case, but in such a way that the linguistic cue does not contribute semantic information about the target. We are currently running such a study, in which the gaze cue occurs together with the referring expression *sowas* (literal translation: *such a thing*). The linguistic and visual information allows for an unambiguous sentence, but the gaze is contributing not only the visual (as in Exp. 8), but also the semantic information of the target referent. In this manner, we "force" the gaze cue into substituting for the referent noun, and expect to measure surprisal effects on the cue itself, similar to those measured on the referent noun in the context of this work.

In addition, it would be of relevance to compare the examinations presented here with similar investigations using a direct measurement of brain activity, such as the ERP components. Such an examination would allow for a better understanding of the processes that lie behind the measured cognitive load. The only current example of such a study is that conducted by Jachmann et al. (2017). The authors found an increased N400 at the referent noun when there

was no gaze cue previously, or when the cue and the noun were not congruent. In the latter case of the incongruent gaze cue, they also found a late (sustained) positivity. The authors conclude that speaker gaze influences the lexical retrieval as well as the integration processes. It remains unaddressed what could be measured on the gaze cue had it been introduced into an already-constrained, predictable context.

Moreover, previous literature has named P3 amplitude as a measure of LC-NE activity, similarly to pupil dilation (Murphy, Robertson, Balsters, & O'Connell, 2011; Nieuwenhuis, Aston-Jones, & Cohen, 2005). However, recent findings have shown that the two are not interchangeable, suggesting that they reflect at least partially different mechanisms (Kamp & Donchin, 2015). As noted by Eckstein, Guerra-Carrillo, Singley, and Bunge (2017), this is not surprising, pupillometry being a more indirect global measure of brain function. In the context of the linguistic processing literature, however, the present effects could potentially be correlated with the N400 component, found to be sensitive to an item's cloze probability (Kutas & Hillyard, 1980, 1984). As Brouwer, Crocker, Venhuizen, and Hoeks (2017) recently suggested in their Retrieval-Integration account, the N400 component indexes the retrieval of word meaning from semantic memory, while P600 reflects the integration of this meaning into the unfolding utterance interpretation. Hence, we expect that the cognitive load results as obtained by the ICA measurement are not directly translatable to one ERP component, but that it rather reflects the load that can be backed up by a combination of N400 and P600 effects.

Finally, since the current series of seven experiments conducted in the visual world paradigm has successfully established the ICA as a reliable measure of cognitive load that can be utilized together with the examination of eye movements, one could follow up from here and venture to use more elaborate experimental designs. More complex visual environments could be used, as well as even more complex communicative situations, where the processing effort during both comprehension and production could be assessed.

6.4 Conclusion

The present thesis aimed at reaching a better understanding of how multimodal information is integrated in the mental representation that the listener creates about the intended meaning of a sentence. In addition to the visual context, we have introduced an additional referential cue in the visual modality – speaker gaze – and evaluated when and how the information contributed by such a visual cue is considered in the sentence representation. By assessing the visual attention together with the simultaneously induced cognitive load, we were able to see when each piece of information was considered, and how the two modalities interrelate.

Moreover, we examined different information-theoretic notions in order to estimate the amount of information contributed by each item, and relate it with the measured cognitive load.

In conclusion, we would like to reiterate our findings in light of the initially posed research questions. First, we found that the presence of the visual context indeed has a general impact on cognitive load induced during sentence processing. We argue that the required cognitive load was increased with the visual context, due to the constant availability of the information from the visual modality. Together with the information introduced from the linguistic stream, the visual information was continuously being considered for incorporation in the sentence interpretation. Moreover, the visual context affected the processing of the linguistic reference. The set of potential referents was reduced to the visually presented objects, thereby motivating anticipations for the upcoming referent, which helped facilitate the processing of the corresponding referent noun.

Second, we examined a referential visual cue – speaker gaze, and established that it facilitates the processing of the corresponding linguistic reference. We found evidence that the processing of the visual cue itself, that is, the referent it selects, does not induce higher processing effort, unless it is mismatching the previous context. Hence, we argued that the specific information conveyed by the gaze cue was not anticipated, and thus, its processing did not induce a different processing cost. This information is, however, incorporated in the sentence interpretation and influences the processing load at the linguistic reference. When the gaze cue was helpful for the creation of anticipation for the upcoming referent, the processing of the linguistic reference was facilitated. However, when the gaze cue was misleading, this did not lead to a disruption of reference processing. This finding suggests, we argue, that the gaze cue is not perceived on a par with other linguistic information, but is rather taken as accompanying relevant information which is used to broaden or further specify an anticipation, without committing to it as an explicit reference that is able to substitute for the linguistic material. Throughout this work the cognitive load effects we measured were the effects of incorporating a piece of information into a current situation model, and updating the model if necessary. The effects were not measured prior to the noun, we argue, because the situation model is gradually and incrementally built, and it is only when it becomes defined enough that slight differences in the probability of the introduced information are detected.

Third, we have seen that besides the linguistic reference, a referential gaze cue can also fully resolve the referential ambiguity and identify the target referent. However, we found that once the referent is named, even though it remains ambiguous, the subsequent gaze cue contributes only to its visual mapping. Processing this information has proven not to be as

costly as processing the referent noun that distinguishes the referent both semantically and visually. We argue that following a visual cue does not necessarily include the activation of an exact label for the cued referent, and that such visual mapping does not require the same processing load as does the mapping of a linguistic label to the visually presented referent.

Linguistic and visual information are both part of the message that is to be understood. We see that restrictive verbs are used to make sense of the visual context, while gaze is used to anticipate the noun. Both modalities are incrementally and interchangeably utilized. We measured only effects of surprisal, that is, integrating a piece of information in a situation model. The more defined the model is, the larger are the effects. The language, however, seems to be perceived as a more definite and explicit information source than the cues from the visual modality. We argue that this is due to the strong coupling of the referential visual cues such as gaze with language; the referential gaze is not seen as a substitute for the referent noun, but rather as an additional contribution for sentence processing.

To conclude, the present series of eight experiments has contributed to a) shedding light on the processing effort that accompanies anticipatory eye movements; b) establishing a more detailed account of the facilitatory effect of referential visual cues on language processing; c) an account of multimodal understanding of the notion of surprisal; and d) establishing the ICA as a reliable and robust measure of cognitive load in the visual world paradigm, allowing for the simultaneous inspection of unconstrained eye movements.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*(4412), 1289–1292.

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye-movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518.

Altmann, G., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609.

Ankener, C., Drenhaus, H., Crocker, M. W., & Staudte, M. (2018). Multimodal surprisal in the n400 and the index of cognitive activity. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 94–99). Madison, WI.

Ankener, C., Sekicki, M., & Staudte, M. (2018). The influence of visual uncertainty on word predictability and processing effort. *Frontiers in Psychology*, *9*, 2387. doi: 10.3389/fpsyg.2018.02387

Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge, UK: Cambridge University Press.

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 289–304.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403–450.

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, *46*(9), 1309–1320.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–9.

Baron-Cohen, S., Baldwin, D., & Crowson, M. (1995). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, *68*, 48–57.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, *4*(3), 311–331.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review*, *13*(6), 1061–1066.

Beatty, J. (1982). Task-evoked pupilary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371–372.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 142–162). Cambridge, UK: Cambridge University Press.

Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Pupil constrictions to photographs of the sun. *Journal of Vision*, *13*(6), 8.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition*, *120*(2), 292–298.

Bradshaw, J. L. (1968). Load and pupillary changes in continuous processing tasks. *British Journal of Psychology*, *59*(3), 265–271.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the n400 and the p600 in language processing. *Cognitive science*, *41*, 1318–1352.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Dahan, D., & Tanenhaus, M. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459.

DeLong, K., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162.

DeLong, K., Urbach, T., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210.

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS ONE*, *11*, e0146194.

DeReWo. (2012). *Korpusbasierte Wortlisten DeReWo v-ww-bll-320000g-2012-12-31-1.0.* Available at: `http://www.ids-mannheim.de/derewo`. Accessed November 26, 2016.

Dodd, M. D., Weiss, N., McDonnell, G. P., Sarwal, A., & Kingstone, A. (2012). Gaze cues influence memory? But not for long. *Acta Psychologica*, *141*(2), 270–275.

Dovidio, J. F., & Ellyson, S. L. (1982). Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening. *Social Psychology Quarterly*, 106–113.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540.

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., ... Giannopoulos, I. (2018). The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 282).

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, *23*(2), 283.

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69–91.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioural Reviews*, *24*, 581–604.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, *63*(4), 639–645.

Federmeier, K., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Fernandez Monsalve, I., Frank, S., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. *EACL '12 Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398–408.

Flom, R., Lee, K., & Muir, D. (2007). *Gaze-following: Its development and significance*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Fossum, V., & Levy, R. (2012). Sequential vs. hierarchical syntactic models of human incremental sentence processing. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 61–69).

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society.*

Frank, S. (2010). Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics* (pp. 81–89).

Frank, S. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494.

Frank, S., Otten, L., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts n400 amplitude during reading. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 878–883). Sofia, Bulgaria: Association for Computational Linguistics.

Frank, S., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11.

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology*, *66*(12), 2289–2294.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490 - 495.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–329.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 199–206).

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269.

Greenbaum, P. E. (1985). Nonverbal differences in communication style between American

Indian and Anglo elementary classrooms. *American Educational Research Journal*, *22*(1), 101–115.
Gregory, S. E., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 237.
Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14.
Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274-279.
Hakerem, G. (1967). Pupillography. *A manual of Psychophysiological Methods*, 335–349.
Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies, NAACL '01* (pp. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics.
Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101–123.
Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.
Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10*(9), 397–412.
Hanna, J., & Brennan, S. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*, 596-615.
Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*(2), 510–527.
Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350.
Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192.
Huettig, F., & Altmann, G. (2004). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*, B23–B32.
Huettig, F., & Altmann, G. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985-1018.
Huettig, F., & Altmann, G. T. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology*, *64*(1), 122–145.
Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.
Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151–171.
Jachmann, T., Drenhaus, H., Staudte, M., & Crocker, M. (2017). The influence of speaker's gaze on sentence comprehension: An ERP investigation. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society.* London, UK: Cognitive Science Society.
Jackendoff, R. (2002). *Foundations of Language*. New York: University Press.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* (Unpublished doctoral dissertation). Stanford University Stanford, CA.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 310.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.

Kahnemann, D., & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, *2*(3), 101–105.

Kamide, Y., Altmann, G. T., & Haywood, S. (2003). Prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Kamp, S.-M., & Donchin, E. (2015). ERP and pupil responses to deviance in an oddball paradigm. *Psychophysiology*, *52*(4), 460–471.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22-63.

Knoeferle, P., Crocker, M., Scheepers, C., & Pickering, M. (2005). The influence of immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529.

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assigment during spoken sentence comprehension? *Frontiers in Psychology*, *3*, 538.

Koss, M. C. (1986). Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *Journal of Pharmacological Methods*, *15*(1), 1–19.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, *61*(5), 743.

Kuhn, G., & Benson, V. (2007). The influence of eye-gaze and arrow pointing distractor cues on voluntary eye movements. *Perception & Psychophysics*, *69*(6), 966–971.

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics*, *71*(2), 314–327.

Kuipers, J.-R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex*, *49*(10), 2853–2860.

Kuperberg, G. R., & Jaeger, T. (2016). An integrative theory of locus coeruleus norepinephrine function: Adaptive gain and optimal performance. *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *4427*(207), 203–205.

Kutas, M., & Hillyard, S. A. (1984, jan). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161.

Laeng, B., & Endestad, T. (2012). Bright illusions reduce the eye's pupil. *Proceedings of the National Academy of Sciences*, *109*(6), 2162–2167.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27.

LaFrance, M., & Mayo, C. (1976). Racial differences in gaze behavior during conversations: Two systematic observational studies. *Journal of Personality and Social Psychology*, *33*(5), 547.

Langton, S. R., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, *6*(5), 541–567.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems* (pp. 849–856).

Linzen, T., & Jaeger, T. F. (2014). Investigating the role of entropy in sentence processing. *Proceedings of the 2014 Workshop on Cognitive Modelling and Computational Linguistics (CMCL). Association for Computational Linguistics*.

Loewenfeld, I., & Lowenstein, O. (1993). The light reflex. *The Pupil: Anatomy, Physiology and Clinical Applications*, 189–193.

Lowenstein, O., & Loewenfeld, I. (1962). The pupil. In D. H (Ed.), *The Eye. Muscular Mechanisms, vol. 3.* New York: Academic Press.

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 1–12.

Macdonald, R. G., & Tatler, B. W. (2014). Eye can't ignore what you're saying: Varying the reliability of gaze and language. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 910–915). Austin, TX: Cognitive Science Society.

MacLachlan, C., & Howland, H. C. (2002). Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. *Ophthalmic and Physiological Optics*, *22*(3), 175–182.

Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction signatures in the brain: Semantic pre-activation during language comprehension. *Frontiers in Human Neuroscience*, *10*, 1–11.

Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. *US Patent*, *6,090,051*.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the 7th Conference on Human Factors and Power Plants* (pp. 7-5–7-9). IEEE.

Marshall, S. P. (2007). Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine*, *78*, B165–B175.

Martin, C., Thierry, G., Kuipers, J., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588.

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary responses to words that convey a sense of brightness or darkness. *Psychological Science*, *28*(8), 1116-1124. doi: 10.1177/0956797617702699

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. (2010). "social" robots are psychological agents for infants: A test of gaze following. *Neural Networks*, *23*(8-9), 966–972.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25–B33.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability – Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1).

Muljadi, C. (2018). *Predicting verbs: How visual context affects verb surprisal and processing effort* (Unpublished master's thesis). Saarland University, DE.

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with bold activity in human locus coeruleus. *Human brain mapping*, *35*(8), 4140–4154.

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543.

Naber, M., & Nakayama, K. (2013). Pupil responses to high-level image content. *Journal of Vision*, *13*(6), 7.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, *15*(7), 1040.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin*, *131*(4), 510.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... others (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, e33468.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rayner, K., Juhasz, B. J., Warren, T., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology*, *30*(6), 1290–1301.

Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504-509.

Research, S. (2008). Eyelink user manual [Computer software manual]. 5516 Main St., Osgoode, ON, Canada. Retrieved from `Version1.4.0`

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport*, *13*(17), 2259–2264.

Ristic, J., Wright, A., & Kingstone, A. (2007). Attentional control and reflexive orienting to gaze and arrow cues. *Psychonomic Bulletin & Review*, *14*(5), 964–969.

Rommers, J., Meyer, A., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*, 437-447.

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part i: Principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235–253.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and

arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. *The On-line Study of Sentence Comprehension: Eyetracking, ERP, and Beyond*, 167–185.

Schilling, H., Rayner, K., & Chumbley, J. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition*, *26*(6), 1270–1281.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, *5*.

Schmidtke, J. (2017). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, 1–21.

Schofield, T. J., Parke, R. D., Castañeda, E. K., & Coltrane, S. (2008). Patterns of gaze between parents and children in European American and Mexican American families. *Journal of Nonverbal Behavior*, *32*(3), 171–186.

Schwalm, M., Keinath, A., & Zimmer, H. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. In D. de Waard, F. Flemisch, B. Lorenz, H. Oberheid, & K. Brookhuis (Eds.), *Human Factors for Assistance and Automation.* Maastricht: Shaker.

Sekicki, M., & Staudte, M. (2017). The facilitatory effect of referent gaze on cognitive load in language processing. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 3107–3112). London, UK.

Sekicki, M., & Staudte, M. (2018). Eye'll help you out! How the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*, *42*(8), 2418-2458.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society.*

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, *120*, 268–291.

Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, *133*, 317–328.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*(30), 415–433.

Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception & Psychophysics*, *70*(1), 77–87.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.

Tourtouri, E., Delogu, F., & Crocker, M. (2017). Specificity and entropy reduction in situated referential processing. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 3356–3361). London, UK.

Van Berkum, J., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235–1249.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32.

Wyatt, H. J. (1995). The form of the human pupil. *Vision Research*, *35*(14), 2021–2036.

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, *43*(2), 303–311.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–284.

Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*(3), 133–141.

Appendix A

Appendix to Chapter 3

Table A.1 Exp. 1 – Linguistic Stimuli. Constraint was manipulated by verb restrictiveness, and Plausibility by noun fit with the restrictive verb.

Item	Verb	Object	Sentence
1	1	1	Die Mutter löffelt gleich die Suppe in der Küche.
1	1	2	Die Mutter löffelt gleich den Kaffee in der Küche.
1	1	3	Die Mutter löffelt gleich das Kleid in der Küche.
1	2	1	Die Mutter bewertet gleich die Suppe in der Küche.
1	2	2	Die Mutter bewertet gleich den Kaffee in der Küche.
1	2	3	Die Mutter bewertet gleich das Kleid in der Küche.
2	1	1	Der Großvater verschüttet gleich das Wasser während des Restaurantbesuchs.
2	1	2	Der Großvater verschüttet gleich das Eis während des Restaurantbesuchs.
2	1	3	Der Großvater verschüttet gleich das Buch während des Restaurantbesuchs.
2	2	1	Der Großvater bestellt gleich das Wasser während des Restaurantbesuchs.
2	2	2	Der Großvater bestellt gleich das Eis während des Restaurantbesuchs.
2	2	3	Der Großvater bestellt gleich das Buch während des Restaurantbesuchs.
3	1	1	Die Schwester schmilzt gleich die Butter für die Soße.
3	1	2	Die Schwester schmilzt gleich den Honig für die Soße.
3	1	3	Die Schwester schmilzt gleich das Brot für die Soße.
3	2	1	Die Schwester kontrolliert gleich die Butter für die Soße.
3	2	2	Die Schwester kontrolliert gleich den Honig für die Soße.
3	2	3	Die Schwester kontrolliert gleich das Brot für die Soße.
4	1	1	Die Cousine montiert gleich die Antenne hinter dem Haus.
4	1	2	Die Cousine montiert gleich das Motorrad hinter dem Haus.
4	1	3	Die Cousine montiert gleich den Ball hinter dem Haus.
4	2	1	Die Cousine ersetzt gleich die Antenne hinter dem Haus.
4	2	2	Die Cousine ersetzt gleich das Motorrad hinter dem Haus.
4	2	3	Die Cousine ersetzt gleich den Ball hinter dem Haus.
5	1	1	Der Großvater kocht gleich die Kartoffel für die Dinnerparty.
5	1	2	Der Großvater kocht gleich die Banane für die Dinnerparty.
5	1	3	Der Großvater kocht gleich das Klavier für die Dinnerparty.
5	2	1	Der Großvater verpackt gleich die Kartoffel für die Dinnerparty.
5	2	2	Der Großvater verpackt gleich die Banane für die Dinnerparty.
5	2	3	Der Großvater verpackt gleich das Klavier für die Dinnerparty.
6	1	1	Der Großmutter trinkt gleich den Kaffee bei dem Kaffeekränzchen.
6	1	2	Der Großmutter trinkt gleich den Joghurt bei dem Kaffeekränzchen.
6	1	3	Der Großmutter trinkt gleich den Apfel bei dem Kaffeekränzchen.
6	2	1	Der Großmutter testet gleich den Kaffee bei dem Kaffeekränzchen.
6	2	2	Der Großmutter testet gleich den Joghurt bei dem Kaffeekränzchen.
6	2	3	Der Großmutter testet gleich den Apfel bei dem Kaffeekränzchen.
7	1	1	Der Großvater serviert gleich das Eis für die Enkel.
7	1	2	Der Großvater serviert gleich die Zitrone für die Enkel.
7	1	3	Der Großvater serviert gleich den Luftballon für die Enkel.
7	2	1	Der Großvater fotografiert gleich das Eis für die Enkel.
7	2	2	Der Großvater fotografiert gleich die Zitrone für die Enkel.
7	2	3	Der Großvater fotografiert gleich den Luftballon für die Enkel.
8	1	1	Der Großvater isst gleich das Brot aus dem Schrank.
8	1	2	Der Großvater isst gleich den Pfeffer aus dem Schrank.
8	1	3	Der Großvater isst gleich die Kassette aus dem Schrank.
8	2	1	Der Großvater nimmt gleich das Brot aus dem Schrank.
8	2	2	Der Großvater nimmt gleich den Pfeffer aus dem Schrank.
8	2	3	Der Großvater nimmt gleich die Kassette aus dem Schrank.
9	1	1	Der Cousin poliert gleich das Auto auf dem Nachhauseweg.
9	1	2	Der Cousin poliert gleich den Zug auf dem Nachhauseweg.
9	1	3	Der Cousin poliert gleich die Milch auf dem Nachhauseweg.
9	2	1	Der Cousin erblickt gleich das Auto auf dem Nachhauseweg.
9	2	2	Der Cousin erblickt gleich den Zug auf dem Nachhauseweg.
9	3	3	Der Cousin erblickt gleich die Milch auf dem Nachhauseweg.
10	1	1	Der Vater kühlt gleich den Wein für die Grillparty.
10	1	2	Der Vater kühlt gleich die Waffel für die Grillparty.
10	1	3	Der Vater kühlt gleich das Sofa für die Grillparty.
10	2	1	Der Vater reklamiert gleich den Wein für die Grillparty.
10	2	2	Der Vater reklamiert gleich die Waffel für die Grillparty.
10	2	3	Der Vater reklamiert gleich das Sofa für die Grillparty.

Item	Verb	Object	Sentence
11	1	1	Die Schwester bestickt gleich das Kissen für die Freundin.
11	1	2	Die Schwester bestickt gleich den Stiefel für die Freundin.
11	1	3	Die Schwester bestickt gleich den Wecker für die Freundin.
11	2	1	Die Schwester behalt gleich das Kissen für die Freundin.
11	2	2	Die Schwester behalt gleich den Stiefel für die Freundin.
11	2	3	Die Schwester behalt gleich den Wecker für die Freundin.
12	1	1	Der Großmutter zuckert gleich den Tee für den Festabend.
12	1	2	Der Großmutter zuckert gleich die Zitrone für den Festabend.
12	1	3	Der Großmutter zuckert gleich das Hemd für den Festabend.
12	2	1	Der Großmutter prüft gleich den Tee für den Festabend.
12	2	2	Der Großmutter prüft gleich die Zitrone für den Festabend.
12	2	3	Der Großmutter prüft gleich das Hemd für den Festabend.
13	1	1	Der Mann fährt gleich das Auto während des Ausflugs.
13	1	2	Der Mann fährt gleich das Schiff während des Ausflugs.
13	1	3	Der Mann fährt gleich die Tasse während des Ausflugs.
13	2	1	Der Mann sieht gleich das Auto während des Ausflugs.
13	2	2	Der Mann sieht gleich das Schiff während des Ausflugs.
13	2	3	Der Mann sieht gleisch die Tasse während des Ausflugs.
14	1	1	Der Cousin näht gleich die Jacke von der Freundin.
14	1	2	Der Cousin näht gleich das Sofa von der Freundin.
14	1	3	Der Cousin näht gleich den Herd von der Freundin.
14	2	1	Der Cousin berührt gleich die Jacke von der Freundin.
14	2	2	Der Cousin berührt gleich das Sofa von der Freundin.
14	2	3	Der Cousin berührt gleich den Herd von der Freundin.
15	1	1	Die Frau schneidet gleich das Brot für die Gäste.
15	1	2	Die Frau schneidet gleich die Pommes für die Gäste.
15	1	3	Die Frau schneidet gleich den Cocktail für die Gäste.
15	2	1	Die Frau holt gleich das Brot für die Gäste.
15	2	2	Die Frau holt gleich die Pommes für die Gäste.
15	2	3	Die Frau holt gleich den Cocktail für die Gäste.
16	1	1	Die Mutter flickt gleich die Jeans bei der Tante.
16	1	2	Die Mutter flickt gleich den Besen bei der Tante.
16	1	3	Die Mutter flickt gleich das Brötchen bei der Tante.
16	2	1	Die Mutter vergisst gleich die Jeans bei der Tante.
16	2	2	Die Mutter vergisst gleich den Besen bei der Tante.
16	2	3	Die Mutter vergisst gleich das Brötchen bei der Tante.
17	1	1	Die Frau bügelt gleich das T-Shirt in der Washküche.
17	1	2	Die Frau bügelt gleich die Socke in der Washküche.
17	1	3	Die Frau bügelt gleich den Sessel in der Washküche.
17	2	1	Die Frau beschreibt gleich das T-Shirt in der Washküche.
17	2	2	Die Frau beschreibt gleich die Socke in der Washküche.
17	2	3	Die Frau beschreibt gleich den Sessel in der Washküche.
18	1	1	Die Mutter strickt gleich den Schal nach der Arbeit.
18	1	2	Die Mutter strickt gleich die Decke nach der Arbeit.
18	1	3	Die Mutter strickt gleich die Wurst nach der Arbeit.
18	2	1	Die Mutter bekommt gleich den Schal nach der Arbeit.
18	2	2	Die Mutter bekommt gleich die Decke nach der Arbeit.
18	2	3	Die Mutter bekommt gleich die Wurst nach der Arbeit.
19	1	1	Die Frau erntet gleich den Apfel für den Nachtisch.
19	1	2	Die Frau erntet gleich den Reis für den Nachtisch.
19	1	3	Die Frau erntet gleich die Pfanne für den Nachtisch.
19	2	1	Die Frau wäscht gleich den Apfel für den Nachtisch.
19	2	2	Die Frau wäscht gleich den Reis für den Nachtisch.
19	2	3	Die Frau wäscht gleich die Pfanne für den Nachtisch.
20	1	1	Der Bruder repariert gleich den Laptop nach der Schule.
20	1	2	Der Bruder repariert gleich das Raumschiff nach der Schule.
20	1	3	Der Bruder repariert gleich die Nuss nach der Schule.
20	2	1	Der Bruder zeichnet gleich den Laptop nach der Schule.
20	2	2	Der Bruder zeichnet gleich das Raumschiff nach der Schule.
20	2	3	Der Bruder zeichnet gleich die Nuss nach der Schule.
21	1	1	Die Cousine füllt gleich die Tasse in ihrer Pause.
21	1	2	Die Cousine füllt gleich den Ordner in ihrer Pause.
21	1	3	Die Cousine füllt gleich den Nagel in ihrer Pause.
21	2	1	Die Cousine sucht gleich die Tasse in ihrer Pause.
21	2	2	Die Cousine sucht gleich den Ordner in ihrer Pause.
21	2	3	Die Cousine sucht gleich den Nagel in ihrer Pause.

Item	Verb	Object	Sentence
22	1	1	Der Mann entzündet gleich die Kerze auf dem Regal.
22	1	2	Der Mann entzündet gleich die Laterne auf dem Regal.
22	1	3	Der Mann entzündet gleich die Trommel auf dem Regal.
22	2	1	Der Mann verstaut gleich die Kerze auf dem Regal.
22	2	2	Der Mann verstaut gleich die Laterne auf dem Regal.
22	2	3	Der Mann verstaut gleich die Trommel auf dem Regal.
23	1	1	Die Schwester schält gleich die Banane aus der Küche.
23	1	2	Die Schwester schält gleich den Pilz aus der Küche.
23	1	3	Die Schwester schält gleich das Saxophon aus der Küche.
23	2	1	Die Schwester bringt gleich die Banane aus der Küche.
23	2	2	Die Schwester bringt gleich den Pilz aus der Küche.
23	2	3	Die Schwester bringt gleich das Saxophon aus der Küche.
24	1	1	Der Cousin zerbricht gleich die Tasse aus dem Wintergarten.
24	1	2	Der Cousin zerbricht gleich die Geige aus dem Wintergarten.
24	1	3	Der Cousin zerbricht gleich den Baum aus dem Wintergarten.
24	2	1	Der Cousin malt gleich die Tasse aus dem Wintergarten.
24	2	2	Der Cousin malt gleich die Geige aus dem Wintergarten.
24	2	3	Der Cousin malt gleich den Baum aus dem Wintergarten.
25	1	1	Der Vater betankt gleich das Motorrad in seiner Freizeit.
25	1	2	Der Vater betankt gleich das Boot in seiner Freizeit.
25	1	3	Der Vater betankt gleich das Fass in seiner Freizeit.
25	2	1	Der Vater tauscht gleich das Motorrad in seiner Freizeit.
25	2	2	Der Vater tauscht gleich das Boot in seiner Freizeit.
25	2	3	Der Vater tauscht gleich das Fass in seiner Freizeit.
26	1	1	Die Cousine besichtigt gleich das Haus in der Nachbarschaft.
26	1	2	Die Cousine besichtigt gleich das Auto in der Nachbarschaft.
26	1	3	Die Cousine besichtigt gleich die Schokolade in der Nachbarschaft.
26	2	1	Die Cousine kauft gleich das Haus in der Nachbarschaft.
26	2	2	Die Cousine kauft gleich das Auto in der Nachbarschaft.
26	2	3	Die Cousine kauft gleich die Schokolade in der Nachbarschaft.
27	1	1	Der Vater liest gleich die Zeitung auf der Couch.
27	1	2	Der Vater liest gleich das Lexikon auf der Couch.
27	1	3	Der Vater liest gleich den Roboter auf der Couch.
27	2	1	Der Vater findet gleich die Zeitung auf der Couch.
27	2	2	Der Vater findet gleich das Lexikon auf der Couch.
27	2	3	Der Vater findet gleich den Roboter auf der Couch.
28	1	1	Der Mann steuert gleich das Motorrad aus der Werkstatt.
28	1	2	Der Mann steuert gleich den Panzer aus der Werkstatt.
28	1	3	Der Mann steuert gleich die Socke aus der Werkstatt.
28	2	1	Der Mann stiehlt gleich das Motorrad aus der Werkstatt.
28	2	2	Der Mann stiehlt gleich den Panzer aus der Werkstatt.
28	2	3	Der Mann stiehlt gleich die Socke aus der Werkstatt.
29	1	1	Der Mann öffnet gleich das Fenster nach dem Hagelschauer.
29	1	2	Der Mann öffnet gleich die Limousine nach dem Hagelschauer.
29	1	3	Der Mann öffnet gleich den Besen nach dem Hagelschauer.
29	2	1	Der Mann begutachtet gleich das Fenster nach dem Hagelschauer.
29	2	2	Der Mann begutachtet gleich die Limousine nach dem Hagelschauer.
29	2	3	Der Mann begutachtet gleich den Besen nach dem Hagelschauer.
30	1	1	Der Cousin startet gleich den Rechner nach der Pause.
30	1	2	Der Cousin startet gleich den Hubschrauber nach der Pause.
30	1	3	Der Cousin startet gleich die Gabel nach der Pause.
30	2	1	Der Cousin verkauft gleich den Rechner nach der Pause.
30	2	2	Der Cousin verkauft gleich den Hubschrauber nach der Pause.
30	2	3	Der Cousin verkauft gleich die Gabel nach der Pause.
31	1	1	Der Mann baut gleich den Kamin für seine Frau.
31	1	2	Der Mann baut gleich die Mauer für seine Frau.
31	1	3	Der Mann baut gleich die Traube für seine Frau.
31	2	1	Der Mann inspiziert gleich den Kamin für seine Frau.
31	2	2	Der Mann inspiziert gleich die Mauer für seine Frau.
31	2	3	Der Mann inspiziert gleich die Traube für seine Frau.
32	1	1	Die Frau giesst gleich die Blume hinter dem Gartentisch.
32	1	2	Die Frau giesst gleich den Baum hinter dem Gartentisch.
32	1	3	Die Frau giesst gleich das Brötchen hinter dem Gartentisch.
32	2	1	Die Frau entdeckt gleich die Blume hinter dem Gartentisch.
32	2	2	Die Frau entdeckt gleich den Baum hinter dem Gartentisch.
32	2	3	Die Frau entdeckt gleich das Brötchen hinter dem Gartentisch.

Item	Verb	Object	Sentence
33	1	1	Der Junge zerreisst gleich das Buch während des Besuchs.
33	1	2	Der Junge zerreisst gleich die Jeans während des Besuchs.
33	1	3	Der Junge zerreisst gleich den Wein während des Besuchs.
33	2	1	Der Junge verschenkt gleich das Buch während des Besuchs.
33	2	2	Der Junge verschenkt gleich die Jeans während des Besuchs.
33	2	3	Der Junge verschenkt gleich den Wein während des Besuchs.
34	1	1	Der Junge bedient gleich das Radio von seinem Vater.
34	1	2	Der Junge bedient gleich den Bagger von seinem Vater.
34	1	3	Der Junge bedient gleich den Schuh von seinem Vater.
34	2	1	Der Junge übersieht gleich das Radio von seinem Vater.
34	2	2	Der Junge übersieht gleich den Bagger von seinem Vater.
34	2	3	Der Junge übersieht gleich den Schuh von seinem Vater.
35	1	1	Der Bruder restauriert gleich die Statue nach der Arbeit.
35	1	2	Der Bruder restauriert gleich das Boot nach der Arbeit.
35	1	3	Der Bruder restauriert gleich die Socke nach der Arbeit.
35	2	1	Der Bruder verleiht gleich die Statue nach der Arbeit.
35	2	2	Der Bruder verleiht gleich das Boot nach der Arbeit.
35	2	3	Der Bruder verleiht gleich die Socke nach der Arbeit.
36	1	1	Der Vater schärft gleich die Schere aus dem Keller.
36	1	2	Der Vater schärft gleich das Schwert aus dem Keller.
36	1	3	Der Vater schärft gleich die Treppe aus dem Keller.
36	1	1	Der Vater benutzt gleich die Schere aus dem Keller.
36	1	2	Der Vater benutzt gleich das Schwert aus dem Keller.
36	1	3	Der Vater benutzt gleich die Treppe aus dem Keller.

Table A.2 Exp. 2, Exp. 3, Exp. 4,– Linguistic Stimuli. Constraint was manipulated by verb restrictiveness, and Plausibility by noun fit with the restrictive verb.

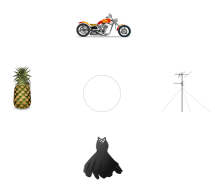
Item	Verb	Object	Sentence
1	1	1	Die Mutter löffelt gleich die Suppe.
1	1	2	Die Mutter löffelt gleich den Kaffee.
1	2	1	Die Mutter bewertet gleich die Suppe.
1	2	2	Die Mutter bewertet gleich den Kaffee.
2	1	1	Der Großvater verschüttet gleich das Wasser.
2	1	2	Der Großvater verschüttet gleich das Eis.
2	2	1	Der Großvater bestellt gleich das Wasser.
2	2	2	Der Großvater bestellt gleich das Eis.
3	1	1	Die Schwester schmilzt gleich die Butter.
3	1	2	Die Schwester schmilzt gleich den Honig.
3	2	1	Die Schwester kontrolliert gleich die Butter.
3	2	2	Die Schwester kontrolliert gleich den Honig.
4	1	1	Die Cousine montiert gleich die Antenne.
4	1	2	Die Cousine montiert gleich das Motorrad.
4	2	1	Die Cousine ersetzt gleich die Antenne.
4	2	2	Die Cousine ersetzt gleich das Motorrad.
5	1	1	Der Großvater kocht gleich die Kartoffel.
5	1	2	Der Großvater kocht gleich die Banane.
5	2	1	Der Großvater verpackt gleich die Kartoffel.
5	2	2	Der Großvater verpackt gleich die Banane.
6	1	1	Der Großmutter trinkt gleich den Kaffee.
6	1	2	Der Großmutter trinkt gleich den Joghurt.
6	2	1	Der Großmutter testet gleich den Kaffee.
6	2	2	Der Großmutter testet gleich den Joghurt.
7	1	1	Der Großvater serviert gleich das Eis.
7	1	2	Der Großvater serviert gleich die Zitrone.
7	2	1	Der Großvater fotografiert gleich das Eis.
7	2	2	Der Großvater fotografiert gleich die Zitrone.
8	1	1	Der Großvater isst gleich das Brot.
8	1	2	Der Großvater isst gleich den Pfeffer.
8	2	1	Der Großvater nimmt gleich das Brot.
8	2	2	Der Großvater nimmt gleich den Pfeffer.
9	1	1	Der Cousin poliert gleich das Auto.
9	1	2	Der Cousin poliert gleich den Zug.
9	2	1	Der Cousin erblickt gleich das Auto.
9	2	2	Der Cousin erblickt gleich den Zug.
10	1	1	Der Vater kühlt gleich den Wein.
10	1	2	Der Vater kühlt gleich die Waffel.
10	2	1	Der Vater reklamiert gleich den Wein.
10	2	2	Der Vater reklamiert gleich die Waffel.
11	1	1	Die Schwester bestickt gleich das Kissen.
11	1	2	Die Schwester bestickt gleich den Stiefel.
11	2	1	Die Schwester behalt gleich das Kissen.
11	2	2	Die Schwester behalt gleich den Stiefel.
12	1	1	Der Großmutter zuckert gleich den Tee.
12	1	2	Der Großmutter zuckert gleich die Zitrone.
12	2	1	Der Großmutter pruft gleich den Tee.
12	2	2	Der Großmutter pruft gleich die Zitrone.

Item	Verb	Object	Sentence
13	1	1	Der Mann fährt gleich das Auto.
13	1	2	Der Mann fährt gleich das Schiff.
13	2	1	Der Mann sieht gleich das Auto.
13	2	2	Der Mann sieht gleich das Schiff.
14	1	1	Der Cousin näht gleich die Jacke.
14	1	2	Der Cousin näht gleich das Sofa.
14	2	1	Der Cousin berührt gleich die Jacke.
14	2	2	Der Cousin berührt gleich das Sofa.
15	1	1	Die Frau schneidet gleich das Brot.
15	1	2	Die Frau schneidet gleich die Pommes.
15	2	1	Die Frau holt gleich das Brot.
15	2	2	Die Frau holt gleich die Pommes.
16	1	1	Die Mutter flickt gleich die Jeans.
16	1	2	Die Mutter flickt gleich den Besen.
16	2	1	Die Mutter vergisst gleich die Jeans.
16	2	2	Die Mutter vergisst gleich den Besen.
17	1	1	Die Frau bügelt gleich das T-Shirt.
17	1	2	Die Frau bügelt gleich die Socke.
17	2	1	Die Frau beschreibt gleich das T-Shirt.
17	2	2	Die Frau beschreibt gleich die Socke.
18	1	1	Die Mutter strickt gleich den Schal.
18	1	2	Die Mutter strickt gleich die Decke.
18	2	1	Die Mutter bekommt gleich den Schal.
18	2	2	Die Mutter bekommt gleich die Decke.
19	1	1	Die Frau erntet gleich den Apfel.
19	1	2	Die Frau erntet gleich den Reis.
19	2	1	Die Frau wäscht gleich den Apfel.
19	2	2	Die Frau wäscht gleich den Reis.
20	1	1	Der Bruder repariert gleich den Laptop.
20	1	2	Der Bruder repariert gleich das Raumschiff.
20	2	1	Der Bruder zeichnet gleich den Laptop.
20	2	2	Der Bruder zeichnet gleich das Raumschiff.

(a) Item 1

(b) Item 2

(c) Item 3



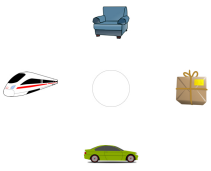
(d) Item 4

(e) Item 5

(f) Item 6

(g) Item 7

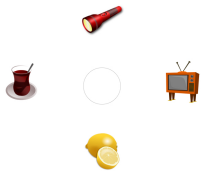
(h) Item 8



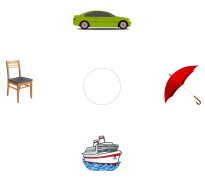
(i) Item 9

(j) Item 10

(k) Item 11



(l) Item 12



(m) Item 13

(n) Item 14

(o) Item 15

(a) Item 16 (b) Item 17 (c) Item 18

(d) Item 19 (e) Item 20

Figure A.2 Exp. 3 – Visual stimuli given in the state prior to the gaze cue.

Appendix B

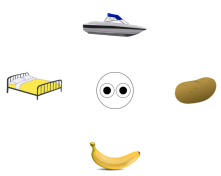
Appendix to Chapter 4

(a) Item 1

(b) Item 2

(c) Item 3

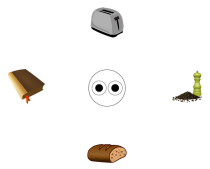
(d) Item 4



(e) Item 5

(f) Item 6

(g) Item 7

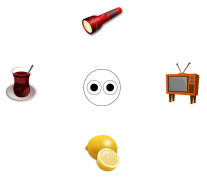


(h) Item 8

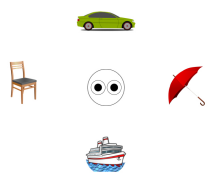
(i) Item 9

(j) Item 10

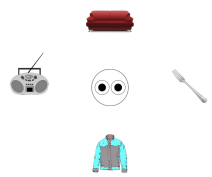
(k) Item 11



(l) Item 12



(a) Item 13



(b) Item 14

(c) Item 15

(d) Item 16

(e) Item 17

(f) Item 18

(g) Item 19

(h) Item 20

Figure B.2 Exp. 4 – Visual stimuli given in the state prior to the gaze cue.

Table B.1 Exp. 5 – Linguistic Stimuli (version A). Fit was manipulated by whether the referent noun fits the verb.

Item	Object	Sentence
1	1	Der Mann löffelt gleich die Suppe.
1	2	Der Mann löffelt gleich den Schuh.
2	1	Der Mann verschüttet gleich das Wasser.
2	2	Der Mann verschüttet gleich die Wurst.
3	1	Der Mann schmilzt gleich die Butter.
3	2	Der Mann schmilzt gleich den Schrank.
4	1	Der Mann montiert gleich die Antenne.
4	2	Der Mann montiert gleich die Rose.
5	1	Die Frau kocht gleich die Kartoffel.
5	2	Die Frau kocht gleich den Stuhl.
6	1	Die Frau trinkt gleich den Kaffee.
6	2	Die Frau trinkt gleich die Zwiebel.
7	1	Die Frau serviert gleich das Eis.
7	2	Die Frau serviert gleich die Büroklammer.
8	1	Der Mann isst gleich die Waffel.
8	2	Der Mann isst gleich die Zeitung.
9	1	Der Mann poliert gleich das Auto.
9	2	Der Mann poliert gleich den Salat.
10	1	Die Frau kühlt gleich den Wein.
10	2	Die Frau kühlt gleich den Buntstift.
11	1	Die Frau bestickt gleich das Kissen.
11	2	Die Frau bestickt gleich den Topf.
12	1	Der Mann zuckert gleich den Tee.
12	2	Der Mann zuckert gleich den Hubschrauber.
13	1	Die Frau fährt gleich das Auto.
13	2	Die Frau fährt gleich das Klavier.
14	1	Die Frau näht gleich die Jacke.
14	2	Die Frau näht gleich die Nuss.
15	1	Der Mann filetiert gleich den Fisch.
15	2	Der Mann filetiert gleich die Geige.
16	1	Die Frau flickt gleich die Jeans.
16	2	Die Frau flickt gleich die Blume.
17	1	Die Frau bügelt gleich das T-Shirt.
17	2	Die Frau bügelt gleich den Keks.
18	1	Die Frau strickt gleich den Schal.
18	2	Die Frau strickt gleich die Orange.
19	1	Der Mann mauert gleich die Wand.
19	2	Der Mann mauert gleich den Koffer.
20	1	Der Mann repariert gleich den Laptop.
20	2	Der Mann repariert gleich die Zigarre.

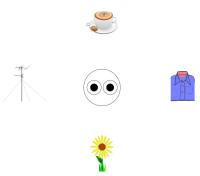
Table B.2 Exp. 5 – Linguistic Stimuli (version B). Fit was manipulated by whether the referent noun fits the verb.

Item	Object	Sentence
1	1	Der Mann bindet gleich den Schuh.
1	2	Der Mann bindet gleich die Suppe.
2	1	Der Mann grillt gleich die Wurst.
2	2	Der Mann grillt gleich das Wasser.
3	1	Der Mann verleimt gleich den Schrank.
3	2	Der Mann verleimt gleich die Butter.
4	1	Der Mann pflanzt gleich die Rose.
4	2	Der Mann pflanzt gleich die Antenne.
5	1	Die Frau streicht gleich den Stuhl.
5	2	Die Frau streicht gleich die Kartoffel.
6	1	Die Frau schält gleich die Zwiebel.
6	2	Die Frau schält gleich den Kaffee.
7	1	Die Frau verbiegt gleich die Büroklammer.
7	2	Die Frau verbiegt gleich das Eis.
8	1	Der Mann ließt gleich die Zeitung.
8	2	Der Mann ließt gleich die Waffel.
9	1	Der Mann würzt gleich den Salat.
9	2	Der Mann würzt gleich das Auto.
10	1	Die Frau spitzt gleich den Buntstift.
10	2	Die Frau spitzt gleich den Wein.
11	1	Die Frau spült gleich den Topf.
11	2	Die Frau spült gleich das Kissen.
12	1	Der Mann lenkt gleich den Hubschrauber.
12	2	Der Mann lenkt gleich den Tee.
13	1	Die Frau stimmt gleich das Klavier.
13	2	Die Frau stimmt gleich das Auto.
14	1	Die Frau knackt gleich die Nuss.
14	2	Die Frau knackt gleich die Jacke.
15	1	Der Mann spielt gleich die Geige.
15	2	Der Mann spielt gleich den Fisch.
16	1	Die Frau gießt gleich die Blume.
16	2	Die Frau gießt gleich die Jeans.
17	1	Die Frau bäckt gleich den Keks.
17	2	Die Frau bäckt gleich das T-Shirt.
18	1	Die Frau entsaftet gleich die Orange.
18	2	Die Frau entsaftet gleich den Schal.
19	1	Der Mann packt gleich den Koffer.
19	2	Der Mann packt gleich die Wand.
20	1	Der Mann raucht gleich die Zigarre.
20	2	Der Mann raucht gleich den Laptop.

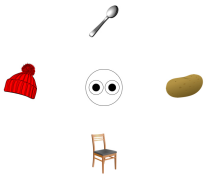
(a) Item 1

(b) Item 2

(c) Item 3

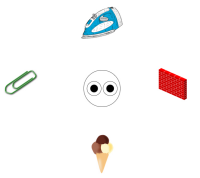


(d) Item 4

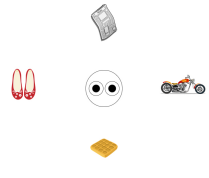


(e) Item 5

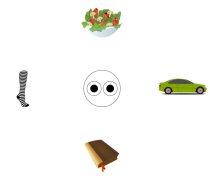
(f) Item 6



(g) Item 7



(h) Item 8

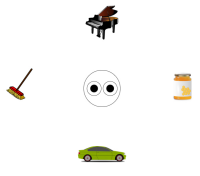


(i) Item 9

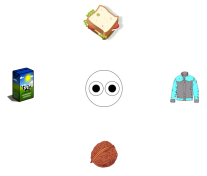
(j) Item 10

(k) Item 11

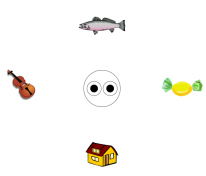
(l) Item 12



(m) Item 13



(n) Item 14



(o) Item 15

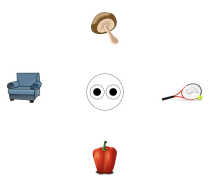
(a) Item 16 (b) Item 17 (c) Item 18

(d) Item 19 (e) Item 20

Figure B.4 Exp. 5 – Visual stimuli given in the state prior to the gaze cue.

Table B.3 Exp. 6 – Linguistic Stimuli (versions A and B).

Item	Version	Sentence
1	A	Die Frau dünstet gleich den Pilz.
1	B	Die Frau dünstet gleich die Paprika.
2	A	Der Mann sät gleich den Weizen.
2	B	Der Mann sät gleich den Mais.
3	A	Der Mann versichert gleich das Handy.
3	B	Der Mann versichert gleich das Fahrrad.
4	A	Die Frau etikettiert gleich die Marmelade.
4	B	Die Frau etikettiert gleich die Flasche.
5	A	Die Frau trocknet gleich die Gabel.
5	B	Die Frau trocknet gleich den Löffel.
6	A	Die Frau frühstückt gleich den Pfannkuchen.
6	B	Die Frau frühstückt gleich das Obst.
7	A	Der Mann nascht gleich das Gummibärchen.
7	B	Der Mann nascht gleich das Bonbon.
8	A	Der Mann schneidet gleich den Kuchen.
8	B	Der Mann schneidet gleich die Pizza.
9	A	Die Frau grillt gleich die Wurst.
9	B	Die Frau grillt gleich das Steak.
10	A	Die Frau kocht gleich die Zucchini.
10	B	Die Frau kocht gleich die Aubergine.
11	A	Die Frau bäckt gleich den Keks.
11	B	Die Frau bäckt gleich den Muffin.
12	A	Der Mann spielt gleich das Akkordeon.
12	B	Der Mann spielt gleich das Saxofon.
13	A	Die Frau spült gleich den Topf.
13	B	Die Frau spült gleich den Teller.
14	A	Die Frau schält gleich die Kartoffel.
14	B	Die Frau schält gleich die Zwiebel.
15	A	Der Mann bindet gleich den Schuh.
15	B	Der Mann bindet gleich die Krawatte.
16	A	Der Mann würzt gleich den Salat.
16	B	Der Mann würzt gleich die Suppe.
17	A	Der Mann isst gleich die Waffel.
17	B	Der Mann isst gleich das Croissant.
18	A	Der Mann kühlt gleich den Wein.
18	B	Der Mann kühlt gleich das Bier.
19	A	Die Frau serviert gleich das Eis.
19	B	Die Frau serviert gleich die Torte.
20	A	Der Mann verschüttet gleich den Tee.
20	B	Der Mann verschüttet gleich den Saft.
21	A	Der Mann zuckert gleich den Tee.
21	B	Der Mann zuckert gleich den Kaffee.



(a) Item 1

(b) Item 2

(c) Item 3

(d) Item 4

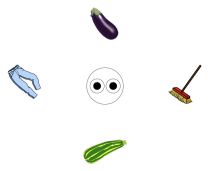
(e) Item 5

(f) Item 6

(g) Item 7

(h) Item 8

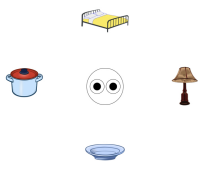
(i) Item 9



(j) Item 10

(k) Item 11

(l) Item 12



(m) Item 13

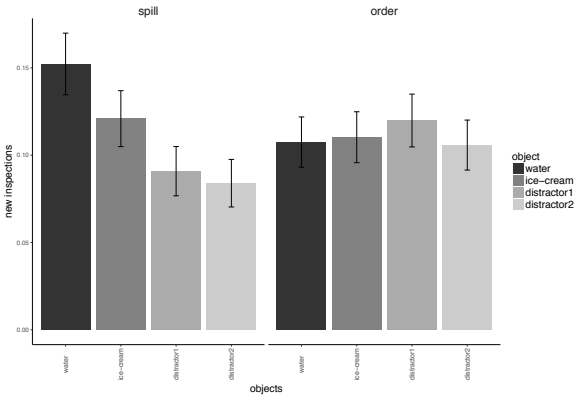
(n) Item 14

(o) Item 15

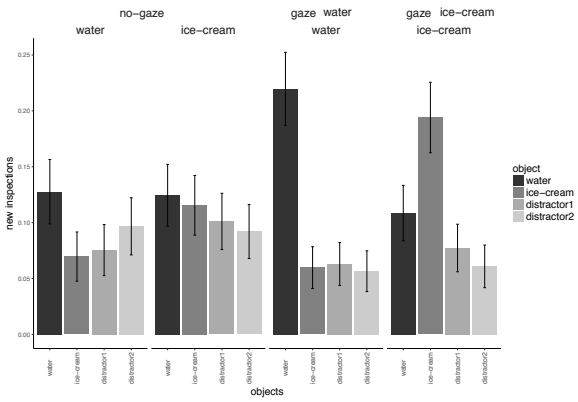
(a) Item 16 (b) Item 17 (c) Item 18

(d) Item 19 (e) Item 20 (f) Item 21

Figure B.6 Exp. 6 – Visual stimuli given in the state prior to the gaze cue.



(a) Verb region of interest: *spill* (left) and *order* (right).



(b) Gaze region of interest: gaze cuing *water* or *ice cream*: no-gaze (left) and referent gaze (right) condition.

Figure B.7 Exp. 4 – New inspections of *water*, *ice cream* and two distractor objects (95% CI error bars).

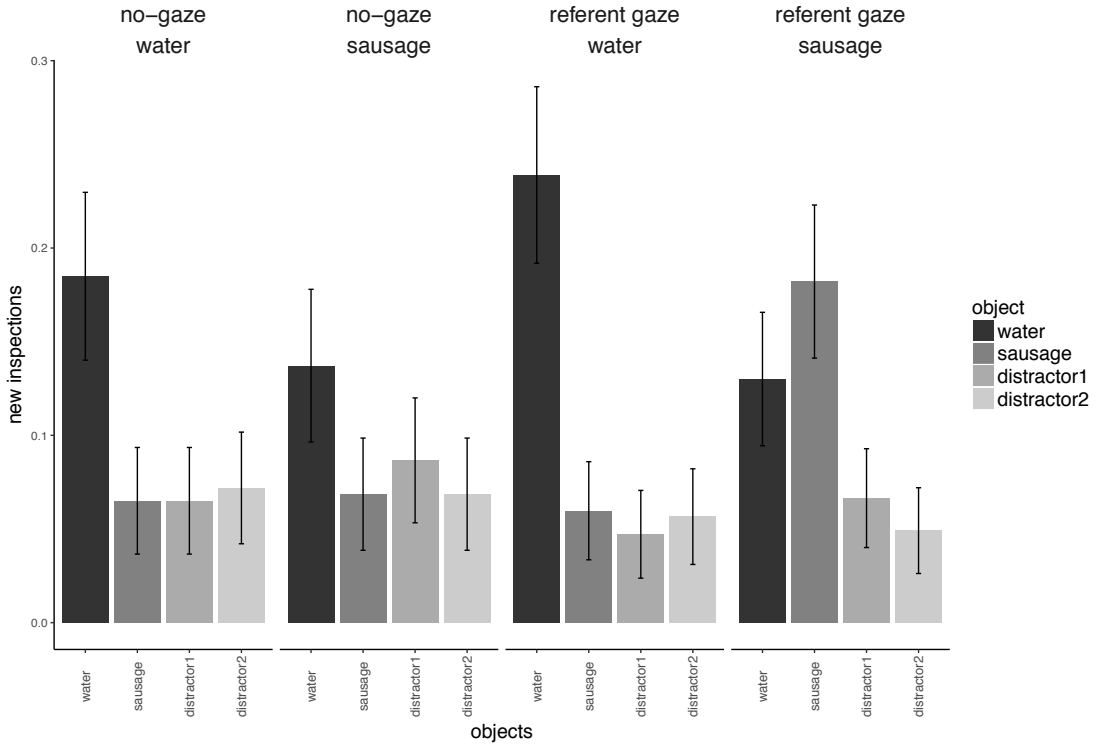
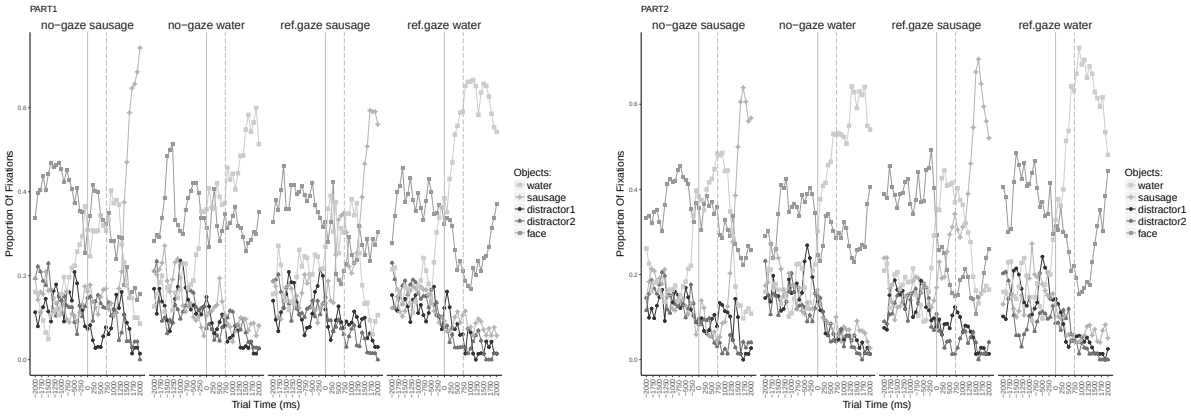
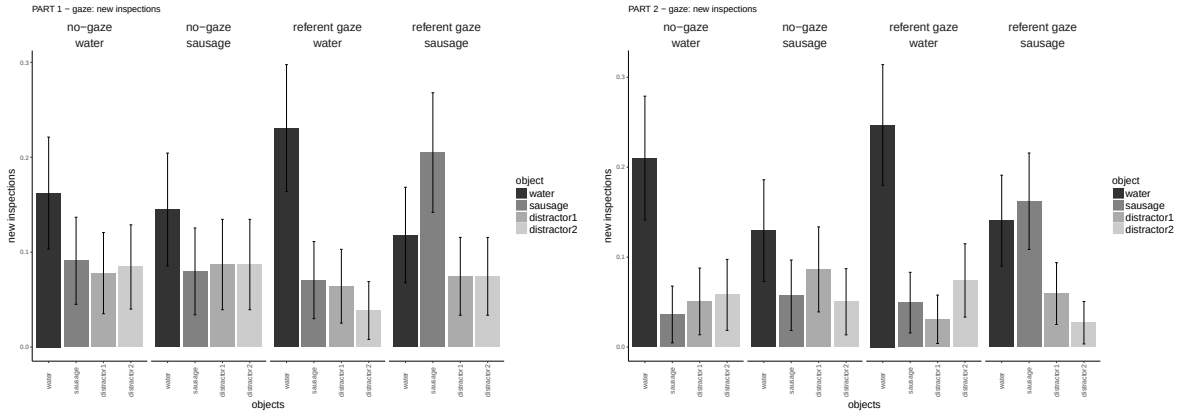


Figure B.8 Exp. 5 – New inspections of *water*, *sausage* and distractors when *sausage* or *water* are gazed at (right) vs. no-gaze condition (95% CI error bars).



(a) 1st half of the experiment (b) 2nd half of the experiment

Figure B.9 Exp. 5 – Proportion of fixations in the two halves of the experiment.



(a) 1st half of the experiment (b) 2nd half of the experiment

Figure B.10 Exp. 5 – New inspections of target and competitor, in the gaze region of interest (95% CI error bars).

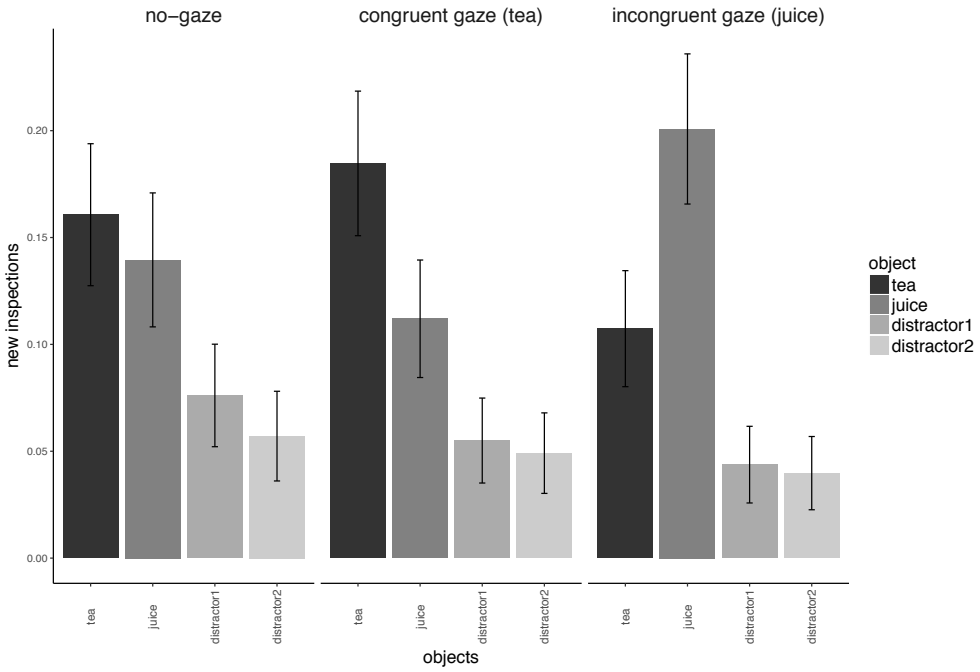


Figure B.11 Exp. 6 – New inspections of the four presented objects in the gaze region of interest (95% CI error bars).

Table B.4 Exp. 4 – Further comparisons for new inspections (gaze region of interest).

	1. Target inspections							
	a) subset *water*				b) subset *ice cream*			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.597	0.089	-17.959	< 2e-16 ***	-2.032	0.092	-22.135	<2e-16 ***
GAZE	0.648	0.178	3.641	0.0003 ***	-0.150	0.184	-0.815	0.415
	2. Competitor inspections							
	c) subset *water*				d) subset *ice cream*			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-2.788	0.175	-15.957	<2e-16 ***	-1.752	0.098	-17.877	< 2e-16 ***
GAZE	-0.204	0.296	-0.688	0.491	0.721	0.187	3.853	0.0001 ***

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a), b) Target ~ Gaze + (1 + Gaze | Subject) + (1 | Item), family = "binomial")
c) Competitor ~ Gaze + (1 + Gaze || Subject) + (1 | Item), family = "binomial")
d) Competitor ~ Gaze + (1 + Gaze | Subject) + (1 | Item), family = "binomial")

Table B.5 Exp. 4 – Further comparisons for the ICA in the subsets of the two verbs.

	1. Gaze window							
	a) subset *spill*				b) subset *order*			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.877	0.030	96.67	<2e-16 ***	2.880	0.0314	91.80	<2e-16 ***
PLAUSIBILITY	0.051	0.031	1.62	0.104	-0.013	0.027	-0.49	0.624
GAZE	-0.003	0.055	-0.05	0.957	0.012	0.060	0.20	0.842
HALF	-0.018	0.019	-0.96	0.338	-0.044	0.019	-2.30	0.022 *
	2. Reference window							
	a) subset *spill*				b) subset *order*			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.897	0.030	97.03	< 2e-16 ***	2.937	0.028	103.54	<2e-16 ***
PLAUSIBILITY	0.152	0.035	4.37	1.23e-05 ***	-0.043	0.037	-1.15	0.249
GAZE	-0.148	0.056	-2.63	0.009 **	-0.103	0.054	-1.93	0.054 .
HALF	-0.049	0.019	-2.59	0.010 **	-0.044	0.019	-2.39	0.017 *

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a), b) ICA ~ Plausibility + Gaze + Half + (1 + Plausibility | Subject) + (1 + Plausibility | Item), family = Poisson (link = "log")

Table B.6 Exp. 5 – Further comparisons for new inspections of *sausage*.

	sausage inspections							
	a) subset no-gaze				b) subset ref. gaze			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-2.699	0.201	-13.445	<2e-16 ***	-2.186	0.168	-12.993	< 2e-16 ***
FIT	0.077	0.351	0.220	0.826	1.288	0.279	4.618	3.87e-06 ***
HALF	-0.645	0.353	-1.825	0.068 .	-0.312	0.254	-1.227	0.22

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a) Competitor ~ Fit + Half (1 + Fit + Half || Subject) + (1 + Fit | Item), family = "binomial")
b) Competitor ~ Fit + Half (1 + Fit + Half || Subject) + (1 + Fit || Item), family = "binomial")

Table B.7 Exp. 5 – Further comparisons for the ICA in the subsets of the two halves of the experiment for both the gaze time-window and the reference time-window.

	1. Gaze time-window							
	a) subset 1st half				b) subset 2nd half			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.783	0.044	63.28	<2e-16 ***	2.751	0.047	58.76	<2e-16 ***
GAZE	0.020	0.048	0.43	0.6670	0.099	0.045	2.20	0.028 *
FIT	0.051	0.066	0.77	0.440	0.013	0.048	0.27	0.784
GAZE:FIT	0.180	0.084	2.13	0.033 *	-0.008	0.090	-0.09	0.929

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1 a), b) ICA ~ Gaze * Fit + (1 + Gaze * Fit || Subject) + (1 + Fit | Item), family = Poisson (link = "log")

	2. Reference time-window							
	a) subset 1st half				b) subset 2nd half			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	2.803	0.039	71.72	< 2e-16 ***	2.745	0.044	61.95	< 2e-16 ***
GAZE	0.043	0.051	0.86	0.392	-0.091	0.047	-1.93	0.054 .
FIT	0.199	0.064	3.11	0.002 **	0.250	0.067	3.76	0.0002 ***
GAZE:FIT	-0.059	0.092	-0.64	0.520	-0.026	0.073	-0.35	0.726

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2 a), b) ICA ~ Gaze * Fit + (1 + Gaze * Fit | Subject) + (1 + Fit | Item), family = Poisson (link = "log")

Appendix C

Appendix to Chapter 5

Table C.1 Exp. 7 – Linguistic Stimuli.

Item	Sentence
1	Die Frau schält vor dem Essen die Zwiebel.
2	Die Frau frühstückt vor der Arbeit den Pfannkuchen.
3	Der Mann nascht nach dem Essen das Gummibärchen.
4	Der Mann schneidet vor dem Essen den Käse.
5	Die Frau grillt nach der Arbeit die Wurst.
6	Die Frau bäckt nach der Arbeit den Keks.
7	Die Frau spült nach dem Essen den Topf.
8	Der Mann verschüttet vor dem Essen den Saft.
9	Der Mann repariert vor der Arbeit den Laptop.
10	Der Mann fährt nach der Arbeit das Motorrad.
11	Der Mann spielt nach dem Essen das Akkordeon.
12	Der Mann bestellt vor dem Essen die Suppe.
13	Die Frau isst vor der Arbeit das Croissant.
14	Der Mann trinkt nach dem Essen den Kaffee.
15	Die Frau wäscht vor dem Essen die Birne.
16	Die Frau erntet vor dem Essen die Tomate.
17	Die Frau kocht nach der Arbeit den Reis.
18	Die Frau serviert nach dem Essen das Eis.
19	Der Mann würzt vor dem Essen den Salat.
20	Die Frau dünstet nach der Arbeit den Pilz.
21	Die Frau gießt vor der Arbeit die Rose.
22	Die Frau trocknet nach der Arbeit das Kleid.
23	Der Mann verpackt nach der Arbeit den Wein.
24	Der Mann kühlt vor dem Essen das Wasser.
25	Der Mann schokoliert nach dem Essen die Erdbeere.
26	Die Frau bügelt vor der Arbeit das T-Shirt.
27	Die Frau versichert nach der Arbeit das Haus.
28	Der Mann zuckert nach dem Essen den Tee.
29	Der Mann parkt vor der Arbeit den Fahrrad.
30	Der Mann mietet vor der Arbeit das Auto.

Table C.2 Exp. 6 – Further comparisons for the ICA in the subsets of the two halves of the experiment in the gaze time-window.

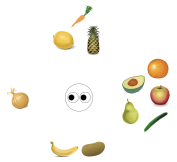
	a) subset 1st half				b) subset 2nd half			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	3.609	0.021	173.35	<2e-16 ***	3.606	0.024	152.19	<2e-16 ***
G1VS.OTHER	0.035	0.024	1.48	0.138	0.010	0.019	0.50	0.615
G3VS.G5	0.038	0.030	1.24	0.215	-0.029	0.033	-0.87	0.384

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1 a), b) ICA ~ G1vs.Other + G3vs.G5 + (1 + G1vs.Other + G3vs.G5 || Subject) + (1 + G1vs.Other + G3vs.G5 || Item), family = Poisson (link = "log")

Table C.3 Exp. 8 – Linguistic Stimuli.

Item	Sentence
1	Die Frau schält für die Suppe am Sonntag die Zwiebel.
2	Die Frau schneidet für die Vorspeise am Samstag den Käse.
3	Der Mann verschüttet in der Disco am Abend den Cocktail.
4	Der Mann fährt zum Auspowern am Morgen das Fahrrad.
5	Der Mann bestellt als Vorspeise am Mittag die Suppe.
6	Die Frau serviert für die Kinder am Mittag das Eis.
7	Der Mann kühlt für das Kind am Mittag den Saft.
8	Der Mann repariert vor dem Rennen am Sonntag das Auto.
9	Der Mann trinkt zum Frühstück am Sonntag den Kaffee.
10	Der Mann spielt in dem Blasorchester am Abend das Saxofon.
11	Die Frau bäckt für das Abendessen am Sonntag die Pizza.
12	Die Frau erntet für den Kuchen am Mittag die Kirschen.
13	Die Frau wäscht für die Gemüsesuppe am Samstag die Karotte.
14	Die Frau isst zum Frühstück am Sonntag das Croissant.
15	Die Frau sammelt am Strand am Sonntag die Muschel.
16	Die Frau spült nach dem Kochen am Abend den Topf.
17	Der Mann versichert vor dem Einzug am Morgen das Haus.
18	Der Mann mietet am See am Morgen das Boot.
19	Der Mann zuckert vor dem Trinken am Morgen den Kaffee.
20	Der Mann graviert für die Hochzeit am Sonntag den Ring.
21	Die Frau trägt zum Joggen am Samstag die Turnschuhe.
22	Die Frau kauft für das Kind am Samstag den Teddy.
23	Der Mann restauriert für das Wohnzimmer am Samstag das Sofa.
24	Die Frau garniert für die Nachspeise am Mittag die Torte.



(a) Item 1, *GazeToOne*

(b) Item 1, *GazeToThree*

(c) Item 1, *GazeToFive*

(d) Item 2, *GazeToOne*

(e) Item 2, *GazeToThree*

(f) Item 2, *GazeToFive*

(g) Item 3, *GazeToOne*

(h) Item 3, *GazeToThree*

(i) Item 3, *GazeToFive*

(j) Item 4, *GazeToOne*

(k) Item 4, *GazeToThree*

(l) Item 4, *GazeToFive*

(m) Item 5, *GazeToOne*

(n) Item 5, *GazeToThree*

(o) Item 5, *GazeToFive*

(p) Item 6, *GazeToOne*

(q) Item 6, *GazeToThree*

(r) Item 6, *GazeToFive*

(a) Item 7, *GazeToOne*

(b) Item 7, *GazeToThree*

(c) Item 7, *GazeToFive*

(d) Item 8, *GazeToOne*

(e) Item 8, *GazeToThree*

(f) Item 8, *GazeToFive*

(g) Item 9, *GazeToOne*

(h) Item 9, *GazeToThree*

(i) Item 9, *GazeToFive*

(j) Item 10, *GazeToOne*

(k) Item 10, *GazeToThree*

(l) Item 10, *GazeToFive*

(m) Item 11, *GazeToOne*

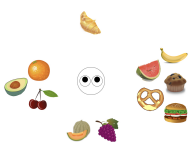
(n) Item 11, *GazeToThree*

(o) Item 11, *GazeToFive*

(p) Item 12, *GazeToOne*

(q) Item 12, *GazeToThree*

(r) Item 12, *GazeToFive*



(s) Item 13, *GazeToOne*

(t) Item 13, *GazeToThree*

(u) Item 13, *GazeToFive*

(a) Item 14, *GazeToOne*

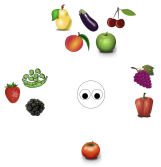
(b) Item 14, *GazeToThree*

(c) Item 14, *GazeToFive*

(d) Item 15, *GazeToOne*

(e) Item 15, *GazeToThree*

(f) Item 15, *GazeToFive*



(g) Item 16, *GazeToOne*

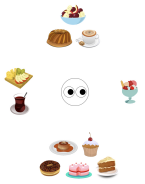
(h) Item 16, *GazeToThree*

(i) Item 16, *GazeToFive*

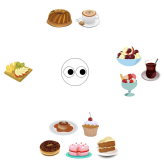
(j) Item 17, *GazeToOne*

(k) Item 17, *GazeToThree*

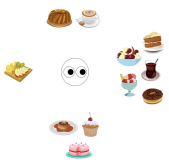
(l) Item 17, *GazeToFive*



(m) Item 18, *GazeToOne*



(n) Item 18, *GazeToThree*



(o) Item 18, *GazeToFive*

(p) Item 19, *GazeToOne*

(q) Item 19, *GazeToThree*

(r) Item 19, *GazeToFive*

(a) Item 20, *GazeToOne*

(b) Item 20, *GazeToThree*

(c) Item 20, *GazeToFive*

(d) Item 21, *GazeToOne*

(e) Item 21, *GazeToThree*

(f) Item 21, *GazeToFive*

(g) Item 22, *GazeToOne*

(h) Item 22, *GazeToThree*

(i) Item 22, *GazeToFive*

(j) Item 23, *GazeToOne*

(k) Item 23, *GazeToThree*

(l) Item 23, *GazeToFive*

(m) Item 24, *GazeToOne*

(n) Item 24, *GazeToThree*

(o) Item 24, *GazeToFive*

(p) Item 25, *GazeToOne*

(q) Item 25, *GazeToThree*

(r) Item 25, *GazeToFive*

(a) Item 26, *GazeToOne*

(b) Item 26, *GazeToThree*

(c) Item 26, *GazeToFive*

(d) Item 27, *GazeToOne*

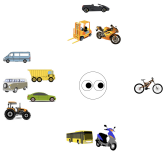
(e) Item 27, *GazeToThree*

(f) Item 27, *GazeToFive*

(g) Item 28, *GazeToOne*

(h) Item 28, *GazeToThree*

(i) Item 28, *GazeToFive*



(j) Item 29, *GazeToOne*

(k) Item 29, *GazeToThree*

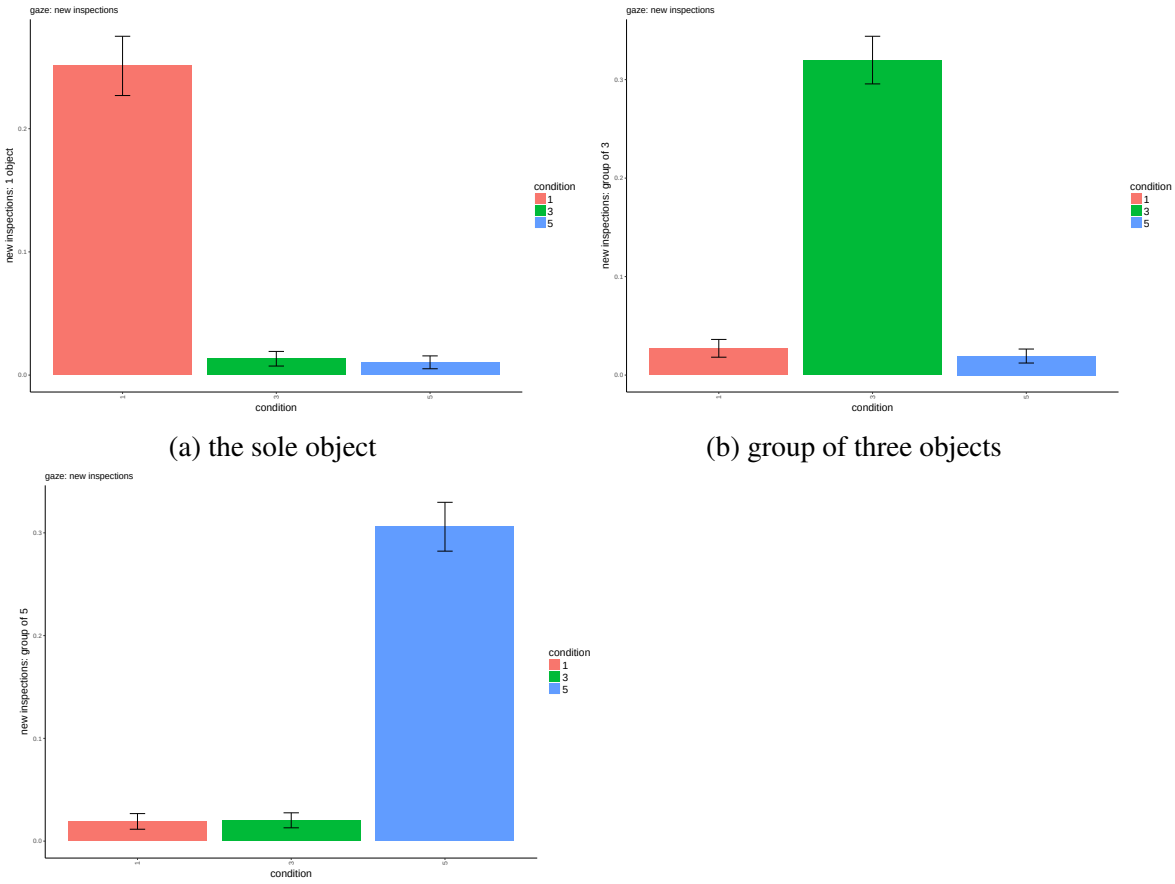
(l) Item 29, *GazeToFive*

(m) Item 30, *GazeToOne*

(n) Item 30, *GazeToThree*

(o) Item 30, *GazeToFive*

Figure C.5 Exp. 7 – Visual stimuli given in the state when the gaze cue was presented. Each item is presented in the three experimental conditions.



(a) the sole object

(b) group of three objects

(c) group of five objects

Figure C.6 Exp. 7 – New inspections of the target group of objects during the gaze region of interest (95% CI error bars).

Table C.4 Exp. 8 – Further comparisons for new inspections of the target in the gaze region of interest.

	Target inspections							
	a) subset no-gaze				b) subset ref. gaze			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	-1.498	0.094	-15.876	<2e-16 ***	-0.553	0.070	-7.893	< 2e-16 ***
N1VS.OTHER	0.087	0.191	4.561	5.1e-06 ***	-0.219	0.153	-1.430	0.153
N3VS.N5	0.717	0.246	2.918	0.004 **	0.094	0.166	0.568	0.570

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a)b) TargetInsp ~ Gaze * N1vs.Other+ Gaze * N3vs.N5 + (1 + N1vs.Other+ N3vs.N5 + Gaze || Subject) + (1 + N1vs.Other+ N3vs.N5 || Item), family = "binomial")

Table C.5 Exp. 8 – Further comparisons for the ICA in the subsets of the two halves of the experiment in the reference time-window.

	a) subset 1st half				b) subset 2nd half			
Predictor	β	SE	z	p	β	SE	z	p
INTERCEPT	3.491	0.064	54.52	<2e-16 ***	3.507	0.065	53.59	<2e-16 ***
N1VS.OTHER	-0.068	0.034	-2.00	0.046 *	-0.009	0.036	-0.24	0.812
N3VS.N5	0.040	0.037	1.07	0.283	-0.034	0.032	-1.07	0.283

. $p < 0.1$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1 a), b) ICA ~ N1vs.Other + N3vs.N5 + (1 + N1vs.Other + N3vs.N5 || Subject) + (1 + N1vs.Other + N3vs.N5 || Item), family = Poisson (link = "log")

(a) Item 1, *namingOne*

(b) Item 1, *namingThree*

(c) Item 1, *namingFive*

(d) Item 2, *namingOne*

(e) Item 2, *namingThree*

(f) Item 2, *namingFive*

(g) Item 3, *namingOne*

(h) Item 3, *namingThree*

(i) Item 3, *namingFive*

(j) Item 4, *namingOne*

(k) Item 4, *namingThree*

(l) Item 4, *namingFive*

(m) Item 5, *namingOne*

(n) Item 5, *namingThree*

(o) Item 5, *namingFive*

(p) Item 6, *namingOne*

(q) Item 6, *namingThree*

(r) Item 6, *namingFive*

(a) Item 7, *namingOne*

(b) Item 7, *namingThree*

(c) Item 7, *namingFive*

(d) Item 8, *namingOne*

(e) Item 8, *namingThree*

(f) Item 8, *namingFive*

(g) Item 9, *namingOne*

(h) Item 9, *namingThree*

(i) Item 9, *namingFive*

(j) Item 10, *namingOne*

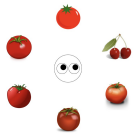
(k) Item 10, *namingThree*

(l) Item 10, *namingFive*

(m) Item 11, *namingOne*

(n) Item 11, *namingThree*

(o) Item 11, *namingFive*

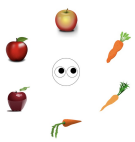


(p) Item 12, *namingOne*

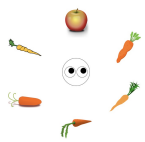
(q) Item 12, *namingThree*

(r) Item 12, *namingFive*

(s) Item 13, *namingOne*



(t) Item 13, *namingThree*



(u) Item 13, *namingFive*

(a) Item 14, *namingOne*

(b) Item 14, *namingThree*

(c) Item 14, *namingFive*

(d) Item 15, *namingOne*

(e) Item 15, *namingThree*

(f) Item 15, *namingFive*

(g) Item 16, *namingOne*

(h) Item 16, *namingThree*

(i) Item 16, *namingFive*

(j) Item 17, *namingOne*

(k) Item 17, *namingThree*

(l) Item 17, *namingFive*

(m) Item 18, *namingOne*

(n) Item 18, *namingThree*

(o) Item 18, *namingFive*

(p) Item 19, *namingOne*

(q) Item 19, *namingThree*

(r) Item 19, *namingFive*

(a) Item 20, *namingOne*

(b) Item 20, *namingThree*

(c) Item 20, *namingFive*

(d) Item 21, *namingOne*

(e) Item 21, *namingThree*

(f) Item 21, *namingFive*

(g) Item 22, *namingOne*

(h) Item 22, *namingThree*

(i) Item 22, *namingFive*

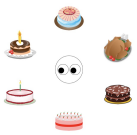
(j) Item 23, *namingOne*

(k) Item 23, *namingThree*

(l) Item 23, *namingFive*

(m) Item 24, *namingOne*

(n) Item 24, *namingThree*



(o) Item 24, *namingFive*

Figure C.10 Exp. 8 – Visual stimuli given in the state when the gaze cue was presented. Each item is presented in the three experimental conditions.

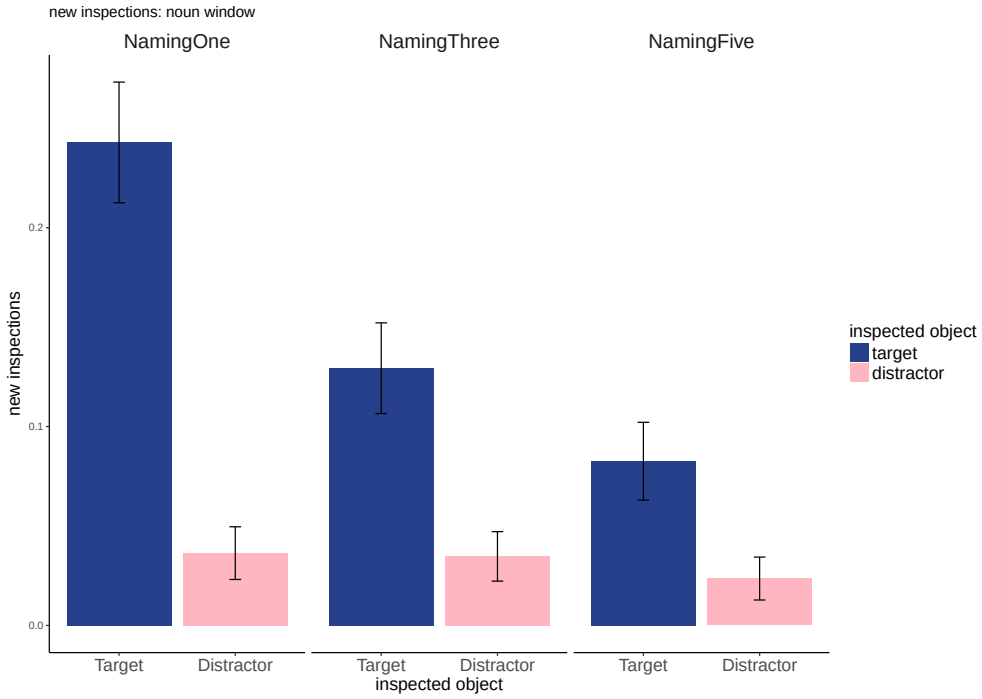
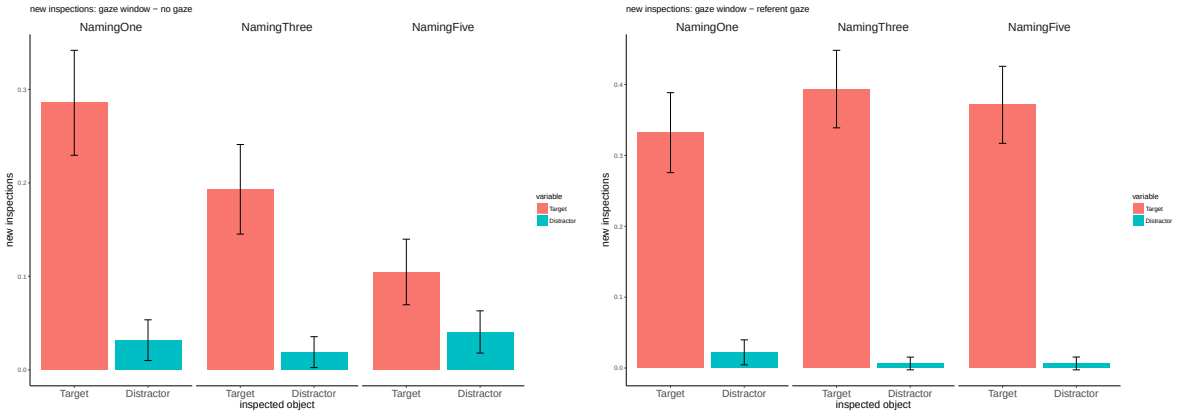


Figure C.11 Exp. 8 – New inspections of the target and the distractor object in the reference region of interest (95% CI error bars).



(a) no gaze

(b) referent gaze

Figure C.12 Exp. 8 – New inspections of target and distractor, in the gaze region of interest (95% CI error bars).