Predictive Language Processing in the Complex Visual World in Children and Adults

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften der Fakultät HW

Bereich Empirische Humanwissenschaften

der Universität des Saarlandes

vorgelegt von

Linda Sommerfeld

aus Saarlouis

Saarbrücken, 2024

Dekan:
Prof. Dr. Peter Loos, Universität des Saarlandes

Berichterstatterinnen:
Prof. Dr. Jutta Kray, Universität des Saarlandes
Prof. Dr. Sabine Schäfer, Universität des Saarlandes

Tag der Disputation:
22.02.2024

Abstract

Given the sentence “On sunny days, Rose *rides* to work with her …”, it is likely that you predict the word “bicycle” before reading it. Notably, not only adults, but even children from an early age predict language, which is seen as one reason of why language comprehension is remarkably fast and accurate. In everyday-life language is usually received in visual contexts which can influence what a comprehender predicts. Imagine processing the above sentence while looking at the picture of a bicycle. This could make you even more likely to predict the noun “bicycle”. Thus, prediction research often applies the Visual World Paradigm. Here, participants listen to predictable sentences like the above while looking at visual scenes that show one visual prediction option that is (e.g., bicycle) and one distractor object that is not (e.g., cake) consistent with the predictive sentence context. When participants show an increase in fixations to the visual prediction option after the predictive cue was played (e.g., “ride”), but prior to the target noun, this indexes prediction.

Cognitive models argue that visually situated prediction involves two mechanisms. Predictive linguistic cues (e.g., the semantically constraining verb “ride”) cause the pre-activation of the mental representations of prediction options such as “bicycle” in long-term memory. If a visual context allows to commit to a prediction option, this option is pre-updated (i.e., pre-processed) in working memory. Given this, individual differences in verbal and cognitive abilities could influence visually situated prediction. That is, language experience could determine which long-term memory representations can be pre-activated, while working memory capacity could affect the ability to pre-update prediction options. Since children have smaller language experience and working memory capacity than adults, we used a developmental approach and compared children and adults in their prediction behavior in the visual world to test the above model assumptions.

First, we compared children and adults in their ability to make multiple predictions in parallel. With the Visual World Paradigm, adults have already been shown to rely on visual contexts to make multiple predictions: When hearing the sentence ("Rose *rides* to work with her …") while looking at multiple "ridable" objects, adults have been shown to predict up to four sentence continuations in parallel. We examined whether also children can follow a multiple predictions pattern, or whether their limited language experience and cognitive capacity prevent them from doing so. Besides, since working memory engages more mental resources when more stimuli are processed, we examined whether children and adults show an increase in cognitive load to pre-update multiple versus only single prediction options in working memory. We examined whether this effect is more prominent in children given their smaller cognitive capacity. We finally investigated whether processing load of a predictable target word (e.g., "bicycle") is smaller when that word was pre-updated alone or among multiple competitors.

In Chapter 1 we outline the theoretical background of this work. This is followed by an empirical section that addresses the above questions. We conducted two studies in which children and adults were presented with sentences with semantically constraining verbs and predictable target nouns (e.g., "The father *eats* the *waffle*") in visual scenes of four object pictures each. Across four conditions, the scenes varied in predictability: Either 0, 1, 3, or 4 visual objects were consistent with the verb constraints and thus viewed as visual prediction options. Chapter 2 shows a pretest of the sentences and the scenes with young children (4–6 years). Experiment 1 was an eye-tracking study in which children (5–6) and adults listened to the sentences while looking at the visual scenes. In Chapter 3, we used their anticipatory object fixations as an index of prediction behavior. Chapter 4 presents data collected in the same study. Here, the Index of Cognitive Activity (ICA) and pupil sizes were used as a measure of cognitive load engaged in sentence processing in the different visual conditions.

Chapter 5 presents Experiment 2, where literate children (8–12 years) and adults were presented with the same sentences and scenes in a self-paced reading task. They read the sentences word-by-word while inspecting the scenes. We relied on word processing times as an index of cognitive load.

Their anticipatory object fixations (Experiment 1) showed that children and adults followed a multiple predictions pattern. For children, this ability was positively related to their language experience, supporting the view that prediction involves the pre-activation of mental representations in long-term memory. We found no consistent evidence of whether children and adults engaged higher cognitive load to make multiple predictions. Both age groups' ICA and pupil size values did not (Experiment 1) but their word processing times did (Experiment 2) suggest additional processing costs for multiple predictions. The latter result is in line with the view that prediction involves the pre-updating of input in the cognitive system. Finally, both studies found children and adults to engage less processing load for target nouns that could be pre-updated alone versus among multiple competitors.

In sum, we provide indication that visual contexts can influence the ease of (predictive) language processing, which is discussed beyond cognitive perspectives of prediction in Chapter 6. Here, we also consider which questions about predictive language processing still remain open, in particular for children.

Zusammenfassung

Stellen Sie sich folgenden Satz vor: „Um an sonnigen Tagen zur Arbeit zu kommen, *fährt* Rosa mit ihrem ...“. Vermutlich haben Sie das Wort „Fahrrad“ antizipiert, ohne es gelesen zu haben. Dies wird prädiktive Sprachverarbeitung genannt und als ein Grund für die enorme Genauigkeit und Geschwindigkeit des Sprachverständnisses gesehen. Bemerkenswerterweise weisen nicht nur Erwachsene, sondern auch Kinder, die Fähigkeit zur sprachlichen Vorhersage auf. Im Alltag wird Sprache oft in visuellen Kontexten rezipiert, welche die Vorhersage beeinflussen. Stellen Sie sich vor, Sie hören obigen Satz, während Sie das Bild eines Fahrrades betrachten. Dies könnte die Wahrscheinlichkeit erhöhen, dass Sie das Wort „Fahrrad“ vorhersagen. Empirische Studien zur sprachlichen Vorhersage nutzen daher häufig das Visual World Paradigma. Hier hören Versuchspersonen vorhersagbare Sätze, wie den obigen, während sie visuelle Szenen betrachten. Diese zeigen typischerweise eine visuelle Vorhersageoption (z.B. das Bild eines Fahrrades) und ein weiteres Objekt, das inkonsistent mit dem prädiktiven Satzkontext ist (z.B. das Bild eines Kuchens). Dieses Paradigma weist sprachliche Vorhersage nach, wenn Versuchspersonen bereits nach dem prädiktive Hinweisreiz (z.B. „fahren“) und vor dem Zielwort (z.B. „Fahrrad“) einen Anstieg an Fixationen der visuellen Vorhersageoption im Vergleich zum inkonsistenten Objekt zeigen.

Kognitive Modelle postulieren, dass zwei Mechanismen an der Vorhersage im visuellen Kontext beteiligt sind. Prädiktive sprachliche Hinweisreize (z.B. das Verb „fahren“) erwirken die Voraktivierung von Vorhersageoptionen (z.B. Fortbewegungsmitteln) im Langzeitgedächtnis. Wenn zudem eine visuelle Vorhersageoption verfügbar ist (z.B. das Bild eines Fahrrades), wird diese Option im Arbeitsgedächtnis vorverarbeitet. Infolgedessen könnten verbale und kognitive Fähigkeiten die sprachliche Vorhersage im visuellen Kontext beeinflussen. So könnte die Spracherfahrung bestimmen, welche Informationen im Langzeitgedächtnis voraktiviert werden können. Die Arbeitsgedächtniskapazität hingegen

könnte die Fähigkeit zur Vorverarbeitung von Vorhersageoptionen beeinflussen. Da Kinder im Vergleich zu Erwachsenen über eine geringere Spracherfahrung sowie Kapazität des Arbeitsgedächtnisses verfügen, nutzte diese Arbeit einen entwicklungspsychologischen Ansatz, um obige Annahmen zur sprachlichen Vorhersage zu prüfen.

Zunächst wurden Kinder und Erwachsene in ihrer Fähigkeit verglichen, mehrere Vorhersagen gleichzeitig zu treffen. Mit dem Visual World Paradigma wurde bereits gezeigt, dass Erwachsene visuelle Kontexte nutzen, um mehrere Vorhersagen zu treffen: Erwachsene, die obigen Beispielsatz hören und gleichzeitig mehrere „fahrbare" Objekte betrachten, konnten nachweislich bis zu vier potentielle Zielwörter gleichzeitig vorhersagen. Diese Arbeit untersucht, ob auch Kinder mehrere Vorhersagen gleichzeig treffen oder ob ihre geringe Spracherfahrung und kognitive Kapazität ein solches Muster der Vorhersage einschränken. Weiterhin wird geprüft, ob Kinder und Erwachsene eine höhere kognitive Belastung zeigen, wenn sie mehrere, statt nur einer Vorhersageoption, vorverarbeiten. Dies wäre plausibel, da das Arbeitsgedächtnis in der Regel mehr mentale Ressourcen beansprucht, wenn es mehr Informationen verarbeitet. Zudem wird untersucht, ob dieser Effekt bei Kindern aufgrund ihrer geringen kognitiven Kapazität stärker ausgeprägt ist als bei Erwachsenen. Zuletzt wird ermittelt, ob mehr mentale Ressourcen zur Verarbeitung eines Zielwortes benötigt werden, wenn dieses Wort mit weiteren Vorhersageoptionen (statt als einzige Option) vorverarbeitet wurde.

Kapitel 1 präsentiert den theoretischen Hintergrund dieser Arbeit. Es folgt ein empirischer Teil, in dem obige Fragen adressiert werden. Dieser umfasst zwei Studien, in denen Kindern und Erwachsenen Sätze mit prädiktiven Verben und Zielwörtern gezeigt wurden (z.B. „Der Vater *isst* die *Waffel*"). Die Sätze wurden zusammen mit visuellen Szenen präsentiert, die jeweils vier Bilder von Objekten zeigten. Die Szenen variierten in ihrer Vorhersagbarkeit: Basierend auf dem prädiktiven Verb stellten 0, 1, 3 oder 4 der Objekte eine visuelle Vorhersageoption dar. Kapitel 2 zeigt eine Studie, in der die Sätze und Szenen mit Kindern (4–6 Jahre) normiert wurden. Experiment 1 war eine Eye-Tracking Studie, in der Kinder (5–6 Jahre) und Erwachsene die Szenen betrachteten, während ihnen die Sätze vorgespielt wurden. In Kapitel 3 wurden die Objektfixationen der Versuchspersonen als Index für das Vorhersageverhalten verwendet. Kapitel 4 präsentiert Daten, die in derselben Studie erhoben wurden. Hier wurde die Pupillengröße sowie der Index of Cognitive Activity (ICA) als Maß für die kognitive Belastung der Satzverarbeitung in den verschiedenen visuellen Konditionen verwendet. Kapitel 5 präsentiert Experiment 2. Hier wurden Kindern (8–12 Jahre) und Erwachsenen dieselben Sätze und Szenen präsentiert, jedoch wurden die Sätze auf dem Bildschirm innerhalb der Szenen gezeigt und Wort für Wort gelesen. Die Wortverarbeitungszeit wurde als Maß für die kognitive Belastung gewertet.

Anhand der Objektfixationen zeigte Experiment 1, dass beide Altersgruppen mehrere Vorhersagen gleichzeitig trafen. Bei Kindern stand diese Fähigkeit in positiver Relation zu ihrer Spracherfahrung. Wir fanden keine konsistente Evidenz, dass Kinder und Erwachsene eine höhere kognitive Belastung zeigen, wenn sie mehrere Vorhersagen gleichzeitig treffen. Dieser Effekt wurde durch die Wortverarbeitungszeiten beider Altersgruppen nachgewiesen (Experiment 2), nicht jedoch durch ihre Pupillengrößen und ICA-Daten (Experiment 1). In beiden Studien zeigten Kinder und Erwachsene eine höhere kognitive Belastung bei der Verarbeitung von Zielwörtern, die mit mehreren Vorhersageoptionen (statt als einzige Option) antizipiert wurden.

Insgesamt zeigen die Ergebnisse dieser Arbeit, dass visuelle Kontexte einen Einfluss auf die prädiktive Sprachverarbeitung und ihre Leichtigkeit haben können. Dies wird in Kapitel 6 vor dem Hintergrund kognitiver Modelle der Vorhersage diskutiert. Hier werden zudem offene Fragen zur sprachlichen Vorhersage, insbesondere bei Kindern, thematisiert.

Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Jutta Kray from whose professional support I could greatly benefit. She was very understanding and could encourage me to focus on the essentials and to trust in myself.

I would also like to thank Dr. Maria Staudte who supervised me at the beginning of this dissertation project and helped me with her competent and empathetic personality to have a positive start in this project.

Next, I would like to thank my research colleagues Katja Häuser, Marjolein van Os, and Yoana Vergilova who supported me with emotional and professional advice. I would also like to thank the research assistants Nora Graichen and Lena Sophie Oberkircher who helped me with data preparation.

A great thank you goes to Julia Meßmer. Julia, thank you for making this path more enjoyable and a lot easier. I am glad that you have supported me emotionally and technically since we both started studying psychology together in 2012.

Danke Mama, danke Papa, danke Robin, danke Nils. Ohne euch wäre ich nicht wer ich bin. Danke Nico, dass du mich bei dieser Arbeit stets unterstützt und an mich geglaubt hast. Zuletzt danke ich mir selbst, dass ich diese Arbeit abgegeben habe.

Table of Contents

Figures

Tables

Abbreviations

AIC	Akaike Information Criterion
ANOVA	Analysis of variance
EEG	Electroencephalogram
ERP	Event-related potential
fem	Feminine
fMRI	Functional magnetic resonance imaging
ICA	Index of Cognitive Activity
IPA	Index of Pupillary Activity
LC	Locus coeruleus
masc	Masculine
MRI	Magnetic resonance imaging
NE	Norepinephrine
Nr.	Number
RQ	Research question

1 Theoretical Background

1.1 The Role of Prediction in Language Comprehension

Language comprehension can be understood as the extraction of the meaning of spoken and written words or sentences (Kuperberg & Jaeger, 2016; Schrimpf et al., 2021) and is considered to be extremely rapid (e.g., Federmeier, 2007; Pickering & Gambi, 2018). In recent decades, a wide theoretical agreement has emerged that one reason for the speed of language comprehension is that comprehenders process language *incrementally*. This means that comprehenders continuously analyze each word of the linguistic input as soon as they encounter it (and not only when the whole utterance is completed) and integrate it with the previous linguistic context (e.g. with the parts of a sentence presented so far) in order to extract its meaning (e.g., Kamide, 2008; Kutas et al., 2011; Pickering & Gambi, 2018; van Petten & Luka, 2012). Incremental language processing has been evidenced empirically for adults and children by numerous studies (e.g., Rayner & Clifton, 2009; Snedeker & Trueswell, 2004; Swingley et al., 1999; Trueswell et al., 1999). Since the mid-20th century, incremental sentence processing is considered to be a precondition for comprehenders to not only process each word as they encounter it, but also to anticipate which word they will encounter next (e.g., Kamide, 2008; Pickering & Gambi, 2018). This way of language processing is named predictive language processing or *prediction* for short.[1]

1.2 Psycholinguistic Models of Language Prediction

The term prediction originates from the Latin words “pre” (meaning “before”) and “dicere” (meaning “to say”), well corresponding to the understanding of prediction in psycholinguistic research (van Petten & Luka, 2012). Here, prediction (which we also refer to as *anticipation*) is typically defined as the ability of comprehenders to actively rely on the

[1] Parts of this chapter were copied or adapted from a published manuscript (Sommerfeld et al., 2023).

linguistic context (e.g., the parts of a sentence presented so far) to anticipate upcoming words in the input rather than just passively receiving them (van Petten & Luka, 2012; Pickering & Gambi, 2018).

Notably, prediction can occur at multiple levels of representation. Thus, also the anticipation of only some aspects of words such as semantic, syntactic or form-related (morphosyntactic and perceptual) features is considered as prediction (e.g., Kamide, 2008; Pickering & Gambi, 2018). The sentence "The girl *eats* the …", for instance, does not provide a linguistic context specific enough to predict which exact word might reveal next. However, comprehenders could anticipate a word of a particular semantic category, namely an edible object. This is because the verb "to eat" is semantically constraining, thus only allows for a limited number of arguments (i.e., edible objects) to complete the sentence (e.g., Altmann & Kamide, 1999). Both types of prediction (i.e., prediction of exact words versus aspects of words) allow comprehenders to do some processing ahead of time to keep pace with the rapid and variable linguistic input. Prediction is therefore considered as one key mechanism of rapid and accurate language comprehension that frees up cognitive resources resulting in a more fluid communication between individuals (e.g., Dell & Chang, 2014; Huettig, 2015; Kamide, 2008; Kuperberg & Jaeger, 2016; Mani & Huettig, 2014; Pickering & Gambi, 2018; Reuter, 2020).

There is a large body of evidence showing that comprehenders continuously form predictions about upcoming linguistic input (for reviews, see Ferreira & Chantavarin, 2018; Huettig, 2015; Kamide, 2008; Kutas et al., 2011; Pickering & Gambi, 2018; Pickering & Garrod, 2013). This has been shown by empirical studies using neurophysiological (e.g., EEG), physiological (e.g., eye-tracking), and behavioral (e.g., reading times) methods.

ERP (event-related potential) studies measure changes in voltage on the scalp that are generated by the brain and time-locked to internal events (evoked by cognitive processes) or external events (evoked by the environment, e.g., Hagoort, 2003; Kutas & Federmeier, 2011). ERP studies have shown that adults who read semantically constraining sentences like "They wanted to make the hotel look like a tropical resort, so they planted …" ending with target words that are more (e.g., "palms") or less predictable (e.g., "tulips") in that context show a smaller N400 amplitude for predictable versus unpredictable target words (e.g., Federmeier & Kutas, 1999; Frank et al., 2015; Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012a, 2012b).[2] The N400 is a negative voltage deflection in the EEG signal with centro-parietal distributions that reaches its peak 400 ms after the onset of a critical stimulus (e.g., Brouwer et al., 2012; Delogu et al., 2019). The N400 is considered to index semantic processing or the semantic fit of a stimulus in a given context (e.g., Huettig, 2015; Kutas et al., 2011). Some authors even state that the N400 reflects the extent to which the cognitive system is engaged in retrieving information from semantic long-term memory (Kutas & Federmeier, 2000, 2011; van Berkum, 2009). A smaller N400 amplitude for predictable compared to unpredictable target words could therefore mean that comprehenders pre-activated the predictable input in semantic memory and therefore showed fewer activation processes (i.e. smaller N400 amplitudes) when the target word actually appeared (for reviews, see Kutas & Federmeier, 2011; van Petten & Luka, 2012).

Besides, as reflected in their reading times or fixation times of words, comprehenders typically spend less time to read predictable versus unpredictable words or even skip them during reading (e.g., Balota et al., 1985; Demberg & Keller, 2008; Ehrlich & Rayner, 1981; Frank & Thompson, 2012; Frisson et al., 2005; Haeuser & Kray, 2022; Kliegl, et al., 2006; Lowder et al., 2018; Monsalve et al., 2012; Rayner & Well, 1996; for a review, see Staub, 2015). Since the time spent on reading a word can index the difficulty of processing that

[2] Predictability of a word in a sentence context is typically extracted using cloze ratings which reflect the proportion of an independent group of participants who have completed the given sentence fragment with that word in an off-line task (e.g., Kutas & Hillyard, 1984; Rommers & Federmeier, 2018).

word (e.g., Frank & Thompson, 2012; McDonald & Shillcock, 2003), such results suggest that readers may have done some pre-processing for the predictable words.

However, the findings reviewed above are not unequivocally attributable to prediction. A smaller N400 amplitude and shorter reading times for predictable words could also be explained in terms of predictable words being easier to process because they are easier to integrate in the preceding context. According to this view, such findings rather demonstrate facilitatory effects of predictive contexts on language processing than supporting the conclusion that comprehenders predict language (for reviews, see Kamide, 2008; Kutas et al., 2011; Pickering & Gambi, 2018).

Nevertheless, it is possible to clearly demonstrate prediction, namely when effects of anticipation precede a target word (e.g., Huettig, 2015; Kutas et al., 2011; Pickering & Gambi, 2018). All studies from this point on cited as reference of prediction have, unless specifically mentioned, examined effects of prediction prior to target words. Delong et al. (2005), for instance, exploited the fact that in English the article forms “a” and “an” are used in relation to the initial phoneme of the following noun. They presented comprehenders with constraining sentences like “The day was breezy, so the boy went outside to fly …” followed by a more (“a kite”) or a less (“an airplane”) predictable continuation. Participants showed a smaller N400 for the article in “a kite” than for the article in “an airplane”. That this result revealed already on the article and thus prior to the predictable noun suggests that the N400 effect truly resulted from prediction and not from integration. This finding was replicated in some studies (e.g., Martin et al., 2013, 2018; Urbach et al., 2020), but failed to replicate in others (e.g., Ito et al., 2017; Nieuwland et al. 2018).

However, effects of prediction prior to target words have also been shown by a series of ERP studies in languages where articles and adjectives are morphosyntactically adapted to the grammatical gender of the subsequent noun (e.g., German: “Die$_{fem}$ neue$_{fem}$ Geldbörse$_{fem}$”, English: “The new wallet”). In those studies, participants experienced constraining sentences with articles and adjectives that were either consistent or inconsistent in gender with a predictable noun (e.g., “I opened my purse to put the money in the$_{fem/masc}$ new$_{fem/masc}$ wallet$_{fem}$”, e.g., Foucart et al., 2014; Nicenboim et al., 2020; Otten & van Berkum, 2008, 2009; Otten et al., 2007; Szewczyk & Schriefers, 2013; van Berkum et al., 2005; Wicha et al., 2004). Comprehenders showed larger ERP deflections on the prediction-inconsistent (e.g., “… the$_{masc}$ new$_{masc}$ wallet$_{fem}$”) versus the prediction-consistent (e.g., “… the$_{fem}$ new$_{fem}$ wallet$_{fem}$) determiners (i.e., larger N400 amplitudes and larger negative and positive deflections in the EEG signal that are also considered to index mismatch detections between predicted and received linguistic input). Notably, as those effects revealed already on the determiners, i.e., prior to the predictable nouns, this provides clear evidence that upcoming words were predicted by the sentence context, and that these predictions were specific enough to include information about the gender of the predicted word (for reviews, see Kochari & Flecken, 2019; Nicenboim et al., 2020).

Similarly, some reading studies have revealed effects of prediction prior to target words. When reading constraining sentences with predictable target nouns, readers have shown accelerated reading times for articles and adjectives preceding the target noun when they were consistent (versus inconsistent) in grammatical gender with the predictable target noun (Cutter et al., 2023; McDonald & Shillcock, 2003; Staub & Clifton, 2006; van Berkum, 2009; van Berkum et al., 2005). This can be considered as evidence that readers may have done some processing in advance for the prediction consistent determiners, suggesting that they truly have used the sentence context to predict upcoming nouns.

Given the plethora of evidence for prediction in language comprehension, including but not limited to the above, prediction can be seen as an empirical robust phenomenon.

However, it remains interesting how prediction works on a functional level, that is, which cognitive processes are involved in prediction. We turn to this question in the next section.

1.3 Cognitive Perspectives on Language Prediction

According to cognitive models, which have emerged rather recently, prediction is considered as a top-down process that allows comprehenders to process the linguistic representations of an upcoming word before encountering that word (e.g., Huettig, 2015; Kamide, 2008; Kuperberg & Jaeger, 2012; Kutas et al., 2011; Pickering & Gambi, 2018; van Petten & Luka, 2012). It is assumed that this proceeds as follows. After a predictive cue (e.g., a semantically constraining verb) has been processed and its predictive power has been identified (e.g., its semantic constraints), two mechanisms of prediction operate (e.g., Koornneef, 2021; Lau et al., 2013; Ness & Meltzer-Asscher, 2018, 2021). First, the mental representations of potential sentence continuations are *pre-activated* in long-term memory, a storage of past (language) experience. An increase in the activation level of a predicted word (which we refer to as "prediction option") can result from spreading of activation or more controlled processes (Ness & Meltzer-Asscher, 2018, 2021).[3] Notably, since long-term memory can hold a seemingly infinite number of representations, numerous long-term representations can be pre-activated at the same time. The pre-activation of the mental representations of a word can result in facilitated retrieval of that word from long-term memory when it is finally presented which can reflect in smaller N400 amplitudes and shorter reading times for the pre-activated input (for empirical evidence, see above Chapter 1.2).

In some cases, a second mechanism of prediction can operate. Given a highly constraining linguistic signal (e.g., "Neil Armstrong was the first man on the …") which allows for a strong prediction of a particular sentence continuation (e.g., "moon"), the

[3] These "more controlled processes" are not defined to more detail in the presented cognitive model of prediction which is, however, not relevant for the present work.

cognitive system can make a commitment to that prediction option. As a result, the level of pre-activation for the related long-term representations reaches a certain threshold, leading to the mechanism of *pre-updating*. That is, the mental representations of the highly activated prediction option are transferred into working memory, a set of linked and interacting information processing components that maintain information in a short-term store for the purpose of the active manipulation (i.e., processing) of the stored items (Baddeley, 2003; Becker & Morris, 1999; Ecker et al., 2014). In working memory, the mental representations of a prediction option are then integrated with the sentence input received so far, generating an online model of the predicted sentence. This online model is then maintained and continuously integrated with incoming input until the prediction is cashed out (e.g., Ness & Meltzer-Asscher, 2018, 2021). Notably, the number of prediction options that can be pre-updated is limited because working memory is a capacity limited system that can only hold a limited number of information (e.g., Baddeley, 2000, 2003; Cowan, 2010; Green, 2017; Just & Carpenter, 1993; Miller, 1956; Seeber, 2011).

However, for those prediction options that are pre-updated, less cognitive processing is required when these words finally emerge. In turn, unpredicted input that was not pre-updated can result in additional processing load. This can be concluded from a series of ERP studies where adults encountered the same unpredictable target word either at the end of a constraining or an unconstraining sentence (Brothers et al., 2015; Federmeier, 2007; Kuperberg et al., 2020; for a review, see van Petten & Luka, 2012). Only in the constraining condition adults were found to elicit a late "frontal post-N400 positivity" (fPNP) for the target word. Since the fPNP is considered to reflect additional processing costs for prediction failure when comprehenders made a commitment to a specific prediction option (e.g., Federmeier, 2007; Ness & Meltzer-Asscher, 2018, 2021; for a review, see van Petten & Luka, 2012), this suggests the following: Encountering an unpredicted target word in the constraining condition disconfirmed adults' prediction. As a result, they involved additional processing load because the anticipated prediction option was already pre-updated and needed to be overridden to account for the incoming input (e.g., Ness & Meltzer-Asscher, 2018, 2021).

1.3.1 Summary and Implications

Comprehenders continuously predict language at different levels of representations (e.g., at the level of semantic information), which is considered to be one reason of why language comprehension is remarkably fast and accurate. Prediction has been evidenced by numerous studies with different methodologies (e.g., ERP, reading times). The focus was long time on a psycholinguistic view of prediction that aims to uncover which linguistic cues can be used to predict which kind of information. Recent approaches consider prediction from a cognitive perspective (i.e., at a functional level), thereby focusing on the cognitive mechanisms that are involved in prediction. Here, prediction is assumed to involve two cognitive mechanisms. When a predictive cue enters a cognitive system, its constraints allow for the pre-activation of the mental representations of prediction option(s) in long-term memory. This can be followed by the mechanism of pre-updating which refers to the integration of the pre-activated mental representations into working memory where an online model of the predicted sentence is formed and maintained. Notably, the level of pre-activation distinguishes whether a prediction option passes the mechanism of pre-updating.

In what follows from these cognitive model assumptions is that prediction should be influenced by individual differences in language experience and working memory capacity. First, it is likely that the mechanism of pre-activation depends on language experience. This is because language experience as the number of linguistic representations stored and linked in long-term memory (e.g., Mani & Huettig, 2012; Zhang et al., 2020) should determine which linguistic representations in long-term memory can be pre-activated by constraining linguistic signals. Second, since working memory is a capacity limited system that can only integrate and maintain a limited amount of information (e.g., Baddeley, 2000, 2003; Cowan, 2010; Green, 2017), working memory capacity is likely to influence whether or how thoroughly prediction options can be pre-updated in working memory.

Given this, in the current work, we use a developmental approach as a testing bed for the theoretical assumptions of the above presented cognitive model of prediction. This is because it is well known that language experience (e.g., Borovsky & Creel, 2014; Huettig, 2015; Rabagliati et al., 2016) and working memory capacity (e.g., Cowan et al., 2010; Johnson et al., 2014; Kharitonova et al., 2015) are strongly increasing from early childhood to adulthood. Thus, when young children encounter predictive input, they might not be able to pre-activate as many mental representations in long-term memory as adults. Besides, children with their limited working memory capacity could be in command of less cognitive resources required to pre-update a sentence's online model. Given that most developmental evidence for prediction derives from studies in which children received predictive linguistic signals together with visual information and since the cognitive mechanisms of pre-activation and pre-updating are thought to be influenced not only by the linguistic but also by the visual signal, we first focus on the effects of visual contexts on predictive processing before we provide a developmental perspective on prediction in Chapter 1.5.

1.4 Language Prediction in the Visual World

Predictions are not only formed by the constraints of linguistic input. Also extra-linguistic information such as the visual context in which language is presented can be constraining for upcoming input, thereby influence prediction (Ito et al., 2018a; Venhuizen et al., 2019). This is conceivable given that, in the real world, language is typically encountered in combination with visual information (Reuter et al., 2020). According to Pickering and Gambi (2018), for instance, comprehenders do not only track the linguistic signal of speakers but also the visual (and other extra-linguistic) context information shared with them. The combination of the linguistic and visual input enables comprehenders to infer a speaker's intention and run it through their own language production system, allowing for the generation of predictions (for similar accounts, see Dell & Chang, 2014; Federmeier, 2007; Huettig, 2015; Mani et al., 2016; Pickering & Garrod, 2013).

On the cognitive level, it has been proposed that the mechanisms of pre-activation and pre-updating may be influenced by visual contexts as follows (e.g., Huettig & Janse, 2016; Huettig et al., 2011a, 2011b; Magnuson, 2019; Özkan et al., 2022). Any information available during language processing passes the cognitive system and enters the separate modality-specific components of working memory (e.g., Baddeley, 2000, 2003; Just & Carpenter, 1993). The visual signal is temporarily stored and rehearsed (together with its spatial information) in visual components of working memory and the associated long-term memory representations are activated. The acoustic linguistic signal is temporarily stored and rehearsed in phonological/verbal components of working memory and the related long-term memory representations are activated. Besides, the constraints of the linguistic input are extracted (Özkan et al., 2022) and related long-term memory representations are activated. Then, all activated mental representations enter integrative components of working memory and an online model of the predicted sentence is generated and maintained.[4]

Collectively, this means that the same linguistic signal can — depending on the visual context — result in the pre-activation and pre-updating of different prediction options (e.g., Altmann & Mirković, 2009; Huettig & Janse, 2016; Huettig et al., 2011a, 2011b; Kamide, 2008; Magnuson, 2019; Özkan et al., 2022). For instance, when hearing the sentence fragment "The girl *eats* the …" without additional visual information, the constraining verb

[4] Few authors argue (Huettig et al., 2011a) or show (Özkan et al., 2022) that this integrative component could be the episodic buffer, a temporal storage that links input of different modalities and working memory components with each other and with long-term memory representations (Baddeley, 2003). This is not shown to more detail, since this work does not examine the role of different working memory components for prediction.

"eat" only allows for the pre-activation of the mental representations of a noun from the semantic category "edible". According to the view that pre-updating is only initiated when the activation level of a prediction option is sufficiently high, the constraining verb "eat" alone might not initiate the pre-updating of a particular prediction option. However, hearing the same sentence while looking at (the picture of) a cake allows for a more specific pre-activation, namely for the pre-activation of the mental representations of the noun "cake". This is because the linguistic and visual context cause the pre-activation of the mental representations of the semantic category "edible" but also of the object "cake". In turn, this allows for the integration of these mental representations into an online model of the predicted sentence, i.e., for the pre-updating of the particular prediction option "cake". This example shows how visual contexts can put additional constraints on linguistic signals and allow comprehenders to make a commitment to a particular prediction option. Thus, visual contexts play an important role for prediction. How comprehenders predict language in the visual world can be examined with eye-tracking studies applying the Visual World Paradigm (Altmann & Kamide, 1999; for a review, see Huettig et al., 2011b).

1.4.1 The Visual World Paradigm

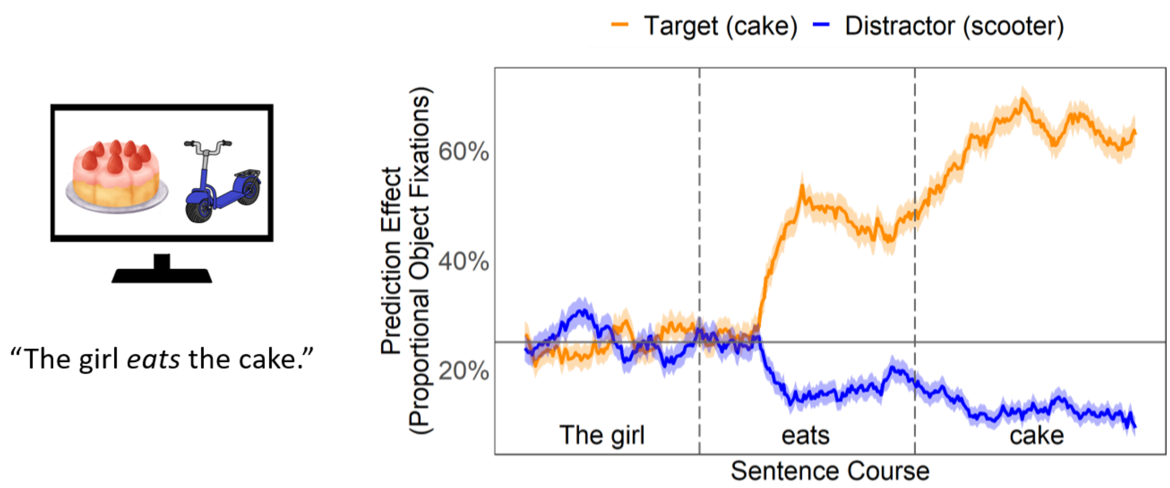
Recently, the Visual World Paradigm was one of the most frequently used paradigms to examine whether and based on which linguistic cues comprehenders of different populations anticipate language. It is typically applied in combination with eye-tracking. Here, participants' eye-movements are recorded while they look at visual scenes and listen to language (Allopenna et al., 1998; Altmann & Kamide, 1999; Tanenhaus et al., 1995; for a review, see Huettig et al., 2011b). Since comprehenders typically guide their eyes to that object in a visual scene that refers to a word in the input closely time-locked to the moment when encountering the word (Allopenna et al., 1998; Cooper, 1974), eye-movements provide a continuous insight into online sentence processing with high temporal resolution (Borys &

Plechawska-Wójcik, 2017; Duchowski, 2017; Tanenhaus et al., 2000). In prediction research, comprehenders are typically presented with simple scenes of two visual objects (see Figure 1) while they listen to predictive sentences, for instance, such with semantically constraining verbs (e.g., “The girl *eats* the cake”). Typically, one of the objects onscreen, the target object (e.g., cake), is consistent with the verb constraints (e.g., is edible) while the other object is not (e.g., scooter), and thus considered as distractor.

As shown in Figure 1, prediction is then indexed by *anticipatory eye-movements* to the target object, i.e., when comprehenders anticipatorily fixate the target object (cake) after hearing the constraining verb (“eat”), but, importantly, prior to the noun (e.g., Altmann & Kamide, 1999; Mani & Huettig, 2012, 2014). Anticipatory eye-movements are thought to derive from the cognitive system forming an online model of the predicted sentence and integrating it with the spatial information (i.e., the position onscreen) of the prediction option in working memory (e.g., Huettig et al., 2011a, 2011b; Huettig & Janse, 2016).

Figure 1

Example of a Visual-World Eye-Tracking Study on Prediction



Note. Left side: Example stimulus of an eye-tracking study applying the Visual World Paradigm. While looking at visual scenes of one target (cake) and one distractor (scooter) object, participants listen to predictive sentences such as "The girl *eats* the cake". Right side: Fictional graph of participants' proportion of fixations to the target relative to the distractor object across the sentence course (unpredictive baseline, verb, noun). Upon hearing the semantically constraining verb "eat" fixations increase for the edible and decrease for the inedible object. As this effect reveals prior to the noun, this evidences prediction.

1.4.2 Adults Predict Language in the Visual World

Numerous eye-tracking studies applying the Visual World Paradigm provide evidence that adults predict sentence input in the visual world (for reviews, see Huettig, 2011; Pickering & Gambi, 2018). Adults have shown anticipatory fixations of target objects based on a variety of predictive linguistic cues, including but not limited to *semantic* cues such as constraining verbs (e.g., Altmann & Kamide, 1999; Andreu et al., 2013; Brouwer et al., 2018; Gambi et al., 2016; Hintz et al., 2017, 2020; Kamide, 2003; Lee et al., 2022; Mani et al., 2016; Reuter et al., 2020), constraining adjectives (Tribushinina & Mak, 2016), and semantic roles of agents or speakers (Borovsky & Creel, 2014; Borovsky et al., 2012). In the visual world, adults have also been shown to use their general semantical knowledge (e.g., that Neil Armstrong was the first man on the moon) to predict not only semantic but also phonological and perceptual features of upcoming nouns (Ito et al., 2018b; Kukona, 2020; Rommers et al., 2013). Besides, they rely on *morphosyntactic* cues like gender or number marked articles (Brouwer et al., 2017a; Huettig & Brouwer, 2015; Huettig & Janse, 2016; Hopp, 2016; Hopp & Lemmerth, 2018; Stone et al., 2021) and adjectives (Aumeistere et al., 2022; Garrido-Pozú, 2022; Hopp & Lemmerth, 2018; Mishra et al., 2012; Sekerina, 2015) as well as pronoun (Stone et al., 2021) and verb morphology (Altmann & Kamide, 2007; Koch et al., 2021; Lukyaneko & Fisher, 2016) to anticipatorily guide their eyes to predictable target objects.

1.4.3 ***Summary and Implications***

A large body of research provides insights into adults' use of prediction in language comprehension. Adults rely on the constraints of various linguistic cues to form predictions which can result in anticipatory fixations of visual prediction options. Anticipatory fixations are thought to be initiated when the online model of a predicted sentence is integrated with the spatial information of a visual prediction option that is temporarily stored in working memory. Thus, when predicting language in the visual world, adults seem to rely on predictive linguistic cues to not only pre-activate but also to pre-update prediction options.

While evidence on (visually situated) language prediction is strong for adults, a recent interest has emerged regarding the question of how individual factors such as language experience and working memory may influence prediction. It is therefore interesting to consider prediction from a developmental perspective, that is, in populations of young children for whom it is known that their language experience and cognitive capacity is still developing. We turn to this point in the next chapter.

1.5 Language Prediction in Children

Notably, not only adults, but even children as young as two years process language extremely rapidly and accurately (Brouwer et al., 2018; Mani & Huettig, 2014; Mani et al., 2016). Recent research attributes children's proficient real-time language comprehension to the fact that they also generate predictions about upcoming linguistic input (e.g., Brouwer et al., 2018; Mani & Huettig, 2014; Mani et al., 2016; Rabagliati et al., 2016; van Alphen et al., 2021). Children from the age of five years show smaller N400 amplitudes (Holcomb et al., 1992; Vergilova et al., 2022) and shorter reading times (Connor et al., 2015; van der Schoot et al., 2012; Wassenburg et al., 2015; Zabrucky & Ratner, 1986; Zargar et al., 2020) for input that was (versus such that was not) predictable by a linguistic context. In other studies, children (6–13 years) have revealed a broader, more unspecific negative ERP signal for unpredictable versus predicable words which was related to the well-known N400 effect (Clahsen et al., 2007; Hahne et al., 2004). These findings suggest that also children have a processing advantage for predictable input which could result from children having pre-processed (i.e., predicted) that input. However, as critically reviewed above, these results could also mean that the predictable input was easier to integrate in the preceding context (for reviews, see Kamide, 2008; Kutas et al., 2011; Pickering & Gambi, 2018).

Nevertheless, numerous eye-tracking studies in combination with the Visual World Paradigm provide clear evidence for prediction in children. Such studies have shown that (young) children rely on a variety of linguistic cues to predict sentence input in the visual world (for reviews, see Huettig & Mani, 2016; Pickering & Gambi, 2018; Rabagliati et al., 2016). First, children use *semantic* cues to predict. When listening to sentences with semantically constraining verbs (e.g., "The girl *eats* the cake") they anticipatorily fixate a target object that is consistent with the semantic verb constraints (e.g., cake) more than an inconsistent distractor object (e.g., scooter). This has been shown for children from two to eleven years of age in a variety of languages like Dutch (Brouwer et al., 2017b, 2018; van Alphen et al., 2021), English (Gambi et al., 2016, 2021; Lee et al., 2022; Nation et al., 2003; Prescott et al., 2022; Reuter et al., 2020), German (Mani & Huettig, 2012, 2014; Mani et al., 2016), Mandarin (Zhou et al., 2019), and Spanish (Andreu et al., 2013; Arias-Trejo et al., 2019). Interestingly, in line with the view that language experience could influence prediction, Theimann et al. (2021) have shown that Norwegian-English bilinguals (2.5–3 years) anticipatorily fixate target objects in both of their languages based on constraining verbs, but they do so more quickly in their dominant language (i.e., in the language they have more experience with). Besides, in languages where children are not familiar with sentences in which constraining verbs precede referential nouns (Turkish), children (4 years) do not rely on verb semantics to anticipatorily fixate target nouns (Brouwer et al., 2018).

In addition, children have been shown to use their semantic knowledge of adjectives and nouns to anticipate input. When hearing sentences with constraining adjectives (e.g., "There is a *soft* pillow"), Dutch speaking 3-year-olds anticipatorily fixate an object matching the constraints of the adjective (e.g., pillow) more than a distractor object (e.g., book, Tribushinina & Mak, 2016). When listening to sentences like "The boy gets his *bucket* and his shovel" English speaking children (30 months) anticipatorily fixate a shovel more than a distractor object (glasses) upon hearing the semantic cue "bucket". Notably, they also anticipatorily fixate a fork (an object very similar in shape to a shovel), indicating that they can predict not only linguistic but also perceptual representations of words (Bobb et al., 2016). Besides, English speaking children (3–10 years) predict based on their long-term knowledge of semantic roles of agents (Borovsky et al., 2012) and speakers (Borovsky & Creel, 2014). They show more anticipatory fixations of a sword than of a magic wand when hearing the sentence "The *pirate* holds the …" or when hearing the sentence "I want to hold the …" in a voice that was introduced as belonging to a pirate character.

Although the focus of the present work is on semantic prediction, it is remarkable that children also predict based on higher-level information such as *morphosyntactic* cues. This has been shown in studies with languages in which verbs, articles, and adjectives must be adapted morphosyntactically to subsequent nouns in terms of grammatical gender and number (e.g., German: “Da ist$_{sing}$ ein$_{sing\text{-}masc}$ großer$_{sing\text{-}masc}$ Kuchen$_{sing\text{-}masc}$”). It has been shown that children (2–11 years) rely on the grammatical gender of articles and adjectives to form predictions about the genus of subsequent nouns (Arias-Trejo et al., 2013; Aumeistere et al, 2022; Bosch et al., 2022; Brouwer et al., 2017a; Cholewa et a., 2019; Lemmerth & Hopp, 2019). In some studies, this effect was weaker in children versus adults (Aumeistere et al., 2022) or did not reveal in younger (24 months) versus older children (30–36 months, Arias-Trejo et al., 2013). Besides, children (2–11 years) have been shown to use number marked verbs and adjectives to predict the numerus of upcoming nouns (Bosch et al., 2022; Deevy et al., 2017; Deevy & Leonard, 2018; Kouider et al., 2006; Lukyaneko & Fisher, 2016; Sekerina, 2011, 2015; Smolík & Bláhová, 2019, 2022). This effect has once been shown to emerge only in older (3 years) but not in younger (2.5 years) children (Lukyaneko & Fisher, 2016). Finally, Gambi et al. (2016) have shown that children (1–5 years) tough rely on simple numerical cues (e.g., “Can you see *one* apple?”) to predict the numerus of a noun, but that even children as old as five years do not show an adult-like ability to predict the phonology of words based on determining articles (“Can you see *a* ball?”, “Can you see *an* ice-cream?”).

In sum, a growing body of visual-world studies shows that even young children can leverage what they know to generate predictions during online sentence processing. This could explain their proficient handling of language, such that the speed and accuracy of language processing derive, at least in part, from prediction (e.g., Gambi et al., 2021; Mahr et al., 2015; Mani & Huettig, 2014). Thus, prediction boosts language comprehension because children can process (features of) predictable words even before they are presented, thereby speeding up recognition of such words. Evidence is robust that children rely on semantic cues to predict language from as early as two years of age. In contrast, confirmation for prediction based at higher level linguistic cues (e.g., number and gender marked determiners) is less consistent in children. In some studies, such higher-level cues were only used to predict by older but not by younger children or not at all by children relative to adults. Thus, sometimes, prediction seems to vary among populations of different age. Potential reasons for this are considered below.

1.5.1 Individual Differences in Language Prediction

Beyond the cognitive perspective, a variation in the usage of prediction between younger children, older children, and adults could stem from the fact that language experience (e.g., Borovsky & Creel, 2014; Huettig, 2015; Rabagliati et al., 2016) and working memory capacity (e.g., Cowan et al., 2010; Johnson et al., 2014; Kharitonova et al., 2015) — two factors that are still developing across childhood — may influence prediction.

Influence of Language Experience. Numerous studies highlight a positive relation between children's prediction skills and their verbal abilities (for a review, see Pickering & Gambi, 2018). That is, children with higher productive vocabulary (Brouwer et al., 2017a; Mani & Huettig, 2012; Mani et al., 2016), receptive vocabulary (Borovsky & Creel, 2014; Borovsky et al., 2012; Prescott et al., 2022), word reading skills (Mani & Huettig, 2014), grammar skills (Smolík & Bláhová, 2019), or such without (relative to such with) language impairment (Andreu et al., 2013) show improved prediction in form of faster and more anticipatory fixations of potential target objects in visual contexts than other children. There are several explanations for such findings, for instance, that children who form predictions could use the mismatch between their predictions and the input to update their vocabulary size (Chang et al., 2006) or that children who quickly generate predictions have a processing advantage, sparing up resources for vocabulary acquisition (Gambi et al., 2021). Both of this would be consistent with accounts that view prediction as one mechanism that can drive language learning (e.g., Chang et al., 2006; Dell & Chang, 2014; Elman, 1990; Rabagliati et al., 2016). Equally, in line with the cognitive view that prediction is based on predictive cues leading to the pre-activation of mental representations in long-term memory (Huettig, 2015; Mani et al., 2016; Pickering & Gambi, 2018; Pickering & Garrod, 2013), increased language experience could foster stronger prediction skills due to higher knowledge of the associated regularities between linguistic units (Bar, 2009). This, in turn, might shape the positive relation of children's verbal skills and their speed and extent of anticipatory fixations (Mani & Huettig, 2012). This view is supported by the fact that children with small (Theimann et al., 2021) or no (Brouwer et al., 2018) experience with semantic relations of words make fewer or no predictions based on semantic cues relative to other children, and by some studies that found an increase in prediction behavior in adults with higher versus smaller verbal skills (e.g., Favier et al., 2021; Hintz et al., 2017; Huettig & Brouwer, 2015; Mishra et al., 2012).

Influence of Working Memory (Capacity). There is initial evidence from visual-world eye-tracking studies that also working memory (capacity) plays a considerable role for predictive language processing. In a study of Ito et al. (2018a), adults who listened to predictive sentences and memorized a list of words at the same time showed a temporal delay in anticipatory target fixations relative to participants who did not perform an additional working memory task. This suggests a shared cognitive resource for prediction and working memory. Besides, working memory capacity has been shown to modulate prediction: Adults (Huettig & Janse, 2012, 2016) and children (Özkan et al., 2022; Zhang & Knoeferle, 2012) with smaller working memory capacity (assessed with span or word ordering tasks) reveal less anticipatory target fixations than individuals with higher working memory capacity (for similar results for adults, see Koch et al., 2021; but see Otten & van Berkum, 2009). Thus, despite much more evidence is needed, working memory seems to be involved in visually situated language prediction. Variations in prediction behavior across younger children, older children, and adults could therefore stem from the fact that working memory, which is required to pre-update predictable input, is still under development from early child- to adulthood (e.g., Cowan et al., 2010; Johnson et al., 2014; Kharitonova et al., 2015).

1.5.2 Summary and Implications

Like adults, children can rely on various linguistic cues to anticipatorily guide their eyes to prediction options in visual scenes. However, in some cases, prediction behavior has been shown to be weaker and/or slower in children versus adults. This could originate in children having less language experience and a smaller working memory capacity than adults since both of these factors have been shown to influence prediction. As more evidence is needed for this consideration, it would be interesting to compare prediction behavior of children and adults in situations where they are exposed to additional working memory demands, while also controlling for the influence of language experience. Additional demands on working memory can be implied by manipulating the visual context in which predictive language is encountered. How this can be operationalized is shown in Chapter 1.7.

In the next section, we first explain that the presentation of visual contexts next to predictable sentences is not only a method to investigate which linguistic cues comprehenders use to predict target words. Visual contexts as such can modulate the prediction process.

1.6 Language Prediction in the More Complex Visual World

The role of visual contexts in our understanding of prediction (in young children) is still unclear. Most studies examining the prediction boost in language processing with the Visual World Paradigm focus on the relation of the words in the input (i.e., on the question which linguistic cues comprehenders use to predict target words). It is less clear so far, how the characteristics of visual contexts as such can influence what a comprehender predicts. We therefore aim at contributing to the question of how visual contexts may modulate children's and adults' prediction behavior. One way of doing this is manipulating the number of visual objects in a scene that are consistent with the constraints of a predictive linguistic cue.

Studies with children and adults mostly presented them with only one visual referent of a prediction option next to a single distractor (see Figure 1, e.g., 2018; Mani & Huettig, 2012, 2014; Prescott et al., 2022). One may therefore assume that only in such simple visual scenarios it could be possible for visual information to be processed fast enough to keep pace with prediction (Reuter et al., 2020). However, this does not seem to be the case. Prediction can also be observed in more complex visual contexts for both children and adults. When listening to predictive sentences (e.g., "The man *milks* the cow"), adults as well as children three years and older have been shown to anticipatorily fixate the single visual referent (e.g., cow) that is consistent with the semantic verb constraints more than three distractor objects (Andreu et al., 2013; Borovsky et al., 2012; Lee et al., 2022; Nation et al., 2003). In a similar study with naturalistic photographs as stimuli, adults and preschoolers have been shown to anticipatorily fixate the single appropriate visual referent out of more than fifteen distractors (Reuter et al., 2020). These findings suggest that adults and even young children efficiently integrate complex visual scenarios into prediction.

1.6.1 Integration of Multiple Visual Prediction Options

However, in the real world, language processing usually takes place in even more complex visual environments (Huettig & Mani, 2016; Reuter et al., 2020), whereby we refer to complexity as the number of visual stimuli that are consistent with the constraints of the linguistic input (Ankener et al., 2018; Sikos et al., 2021). Here, adults and children could follow a *multiple predictions pattern*. They might show anticipatory fixations of multiple visual cues that are consistent with the input received so far. This would suggest that, when processing language, comprehenders can pre-process the mental representations of multiple prediction options in parallel. Given that correct predictions boost the speed of language processing, maintaining multiple prediction options may allow comprehenders to be maximally prepared for the upcoming input and thus to maximally benefit from prediction during language processing (Kamide, 2008; Kuperberg & Jaeger, 2016; Pickering & Garrod, 2007; Smith & Levy, 2013; van Petten & Luka, 2012). Otherwise, adults and children could follow a one-only prediction pattern and integrate only a single visual prediction option into language processing, implying that they avoid possible costs of unfulfilled predictions (Gambi et al., 2021; Kuperberg & Jaeger, 2016; van Petten & Luka, 2012).

From a theoretical point of view, it is conceivable that children and adults can maintain multiple visual prediction options. According to cognitive models on visually situated prediction (e.g., Huettig & Janse, 2016; Huettig et al., 2011a, 2011b; Magnuson, 2019; Özkan et al., 2022), the linguistic input and the visual context both lead to the pre-activation of related mental representations which are then integrated with each other in working memory. For all prediction options, for which there is an overlap among the mental representations activated by the input of the different modalities, this could cause an increase in the level of pre-processing, thus lead to anticipatory fixations of the related visual objects. This can be explained by the example of the sentence fragment "The girl *eats* …" presented in a visual scene of two edible objects (e.g., cake, apple). Here, the linguistic input could lead to the activation of the mental representations of the semantic category "edible". The visual scene could lead to the activation of the mental representations of the objects "cake" and "apple". The representations activated by the input of both modalities (linguistic, visual) could then be integrated with each other. In case of an overlap among their properties (e.g., edible — cake, edible — apple), this could lead to an increased level of activity for those prediction options (e.g., for the nouns "cake" and "apple"), resulting in the generation of an online model of the predicted sentence with both prediction options. As a result, children and adults could anticipatorily fixate both visual prediction options (e.g., cake and apple).

From an empirical perspective children and adults have already been shown to integrate multiple visual stimuli in parallel into *general* sentence processing. Two pioneer studies of online sentence processing presented children (5 years) and adults with unpredictive verbal instructions like “Put the frog […] in the box” while participants looked at two stuffed frogs and two other objects in front of them (Snedeker & Trueswell, 2004; Trueswell et al., 1999). When hearing the first part of the instruction (“Put the frog”), both age groups looked at the two visual referents of the linguistic input (two frogs) to the same extent (and more than at the other objects). This suggests that even children, like adults, identify multiple visual stimuli that represent the meaning of the linguistic input and integrate them in parallel into sentence processing.

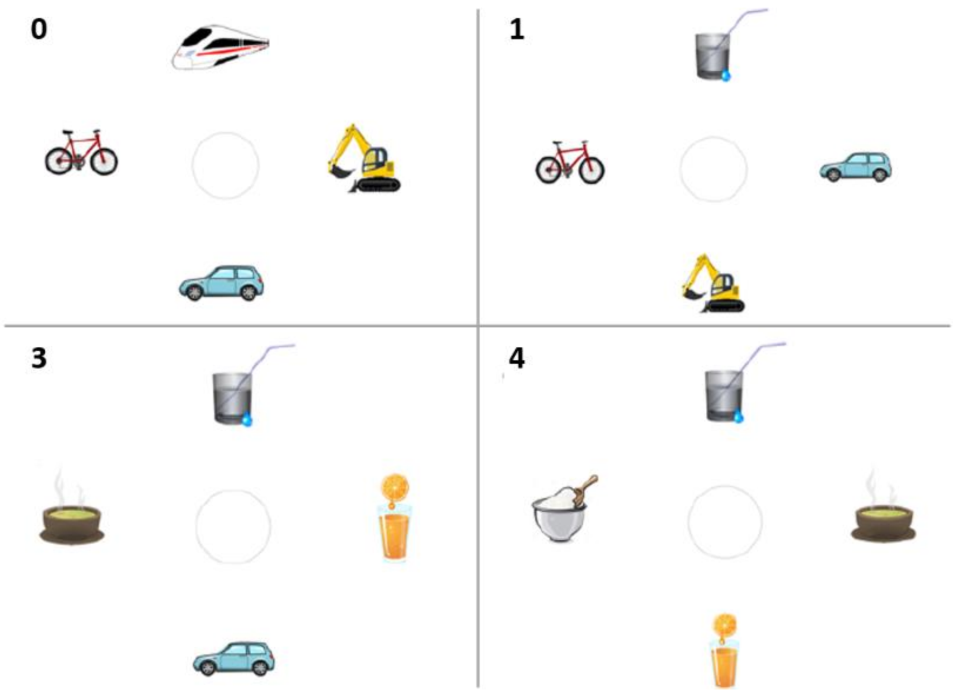
Whether this holds true in the context of prediction, i.e., whether children and adults anticipatorily fixate multiple visual prediction options when predicting input, was shown for adults by Ankener et al. (2018) with a complex Visual World Paradigm.[5] In this eye-tracking study, adults listened to sentences with semantically constraining verbs (e.g., “The man *spills* soon the water”) while they looked at complex visual scenes of four objects (see Figure 2). The novelty of this study was that the scenes varied in predictability: Across four conditions, either 0, 1, 3, or 4 objects were consistent with the verb’s semantic constraints, while the other objects onscreen were not, and therefore considered as visual distractors (4, 3, 1, or 0, respectively). Notably, adults’ anticipatory object fixations revealed initial evidence for a multiple predictions pattern: After hearing the constraining verb (“spill”) and before hearing the target word (“water”), adults anticipatorily fixated the target object (water) more often than all other objects in the highly predictive 1-consistent condition, where only one object

[5] Wherever we cite Ankener et al. (2018), we refer to the fourth Experiment of this publication.

onscreen was consistent with the verb constraints. This result refers to the classic prediction effect often reported for adults (e.g., Andreu et al., 2013; Borovsky et al., 2012, Nation et al., 2003). In the less predictive 3- and 4-consistent conditions, in contrast, adults fixated the target object less often than in the 1-consistent condition upon hearing the verb because they anticipatorily fixated also the two (lemonade, soup) or three (bowl, lemonade, soup) competitors that were also consistent with the verb constraints. In the 0-consistent control condition, adults fixated all four distractors to the same extent after verb presentation. These results suggest that adults can integrate more than one visual prediction option when they predict language, i.e., that they may follow a multiple predictions rather than a one-only approach to prediction in the complex visual world (for similar results, see Sikos et al., 2021).

Figure 2

Example Stimulus of the Study of Ankener et al. (2018)



Note. From left to right and top to bottom: There are 0, 1, 3, or 4 visual prediction options for the sentence “The man *spills* soon the water”. Adapted with permission from the authors.

Whether also children can adapt their prediction behavior to the predictability of visual contexts in such a way that they integrate either one or multiple visual prediction options is yet unclear. Evidence for this was provided by Mani et al. (2016). They presented 2-year-olds with semantically constraining sentences (e.g., “The boy *reads* something”) and visual scenes of two objects that both matched the verb constraints. Notably, the target object (e.g., book) was rated to be more strongly related to the verb than the competitor (e.g., letter). Although children generally fixated the target object more often than the competitor after hearing the verb, this preference decreased the lower the rated strength of the association between the target and the verb. Since the authors analyzed the proportion of fixations to the target object relative to the competitor, it can be concluded that, in these cases, children also fixated the competitor. This is a first sign that children can consider two visual prediction options in parallel when anticipating language. Similar results are reported by Gambi et al. (2021) who presented children (2–5 years) with predictable sentences (e.g., “The *dog chews* on the bone”) and visual scenes of three objects. After the constraining part of the sentences, children anticipatorily fixated a highly predictable target object (e.g., bone) and a mildly predictable competitor (e.g., slippers) more often than an unpredictable distractor (e.g., pajamas). Finally, Borovsky et al. (2012) presented 3- to 10-year-olds with visual scenes that displayed four objects of which only two objects (e.g., ship, treasure) were consistent with the semantically constraining agents (e.g., pirate) of auditorily played sentences (e.g., “The *pirate* chases the treasure”). Upon hearing the agents, children showed an increase in anticipatory fixations of both consistent objects, thus maintained two prediction options.

Although these studies indicate that children can maintain two visual prediction options, it remains unclear how they predict language in more complex visual contexts with a varying number of objects that are consistent with the linguistic constraints. In this regard, the current work examined whether children, comparably to adults, can identify and maintain even multiple suitable prediction options when predicting language in complex visual contexts where the number of visual prediction options varies.

1.6.2 Summary and Implications

When predicting language, young children and adults do not only rely on linguistic cues but also on visual contexts that allow for more specific predictions. While adults can integrate up to four, a few studies have shown that children can integrate two visual prediction options in parallel into predictive processing. From a cognitive perspective, this means that comprehenders can integrate multiple visual prediction options with the sentence input received so far and may maintain multiple online models of potential sentence continuations in working memory. An increasing number of visual prediction options can therefore be seen as manipulation of working memory load. Given this, it would be interesting whether also children, despite their limited working memory capacity relative to adults (e.g., Cowan et al., 2010; Johnson et al., 2014), can follow a multiple predictions pattern in complex visual contexts with an increasing number of visual prediction options.

Besides, when considering the presentation of multiple visual prediction options as working memory manipulation, and under the premise that children and adults can follow a multiple predictions pattern, it is interesting whether they engage more cognitive resources to pre-update multiple versus only a single prediction option in working memory (e.g., Kamide, 2008; Kutas et al., 2011; Pickering & Garrod, 2007). We turn to this point in the next chapter.

1.7 Cognitive Load and Predictive Language Processing

Although prediction allows comprehenders to extract the meaning of linguistic input more rapidly, and thus enables a more efficient language comprehension and a more fluent communication (e.g., Kamide, 2008; Reuter, 2020), predictive processing is far from being effortless (Liu et al., 2022). We illustrate this using the example of prediction in a visual-world scenario. Here, the predictive linguistic cue and the visual context both enter the cognitive system. This allows for the pre-activation of related long-term memory representations which are then transferred into working memory where they are integrated with the input received so far while an online model of the predicted input is formed, maintained, and compared with the continuously incoming signal until the prediction is either confirmed or falsified (Huettig & Janse, 2016; Huettig et al., 2011b; Özkan et al., 2022). Since working memory is a capacity limited system that involves more cognitive resources to process more information (e.g., Baddeley, 2000, 2003; Cowan, 2010; Green, 2017), it is reasonable that additional working memory resources are engaged when comprehenders do not only process the input received so far, but also do some pre-processing of the predicted input (Kuperberg et al., 2020; Ness & Meltzer-Asscher, 2018, 2021). Thus, working memory resources may be expended when predictions are formed and maintained (for empirical evidence, see above Chapter 1.5.1).

According to this, some psycholinguistic authors argue that the generation of linguistic predictions involves cognitive load (Ankener et al., 2018; Ness and Meltzer-Asscher, 2018, 2021; Sikos et al., 2021). Such works typically define cognitive load as the amount of mental resources a cognitive system exerts to perform a given task, including the task of language comprehension (e.g., Frank, 2013; Just et al., 2003; Vogels et al., 2018). While cognitive load is a multifaceted construct (Piolat et al., 2004) for which a unified definition or theory is still missing (for a review, see Westbrook & Braver, 2015), there is wide consensus that cognitive load strongly interacts with working memory load engaged in a task (e.g., Just & Carpenter, 1993; Kahneman, 1973; Vogels et al., 2018; Westbrook & Braver, 2015; Zou et al., 2022). In this sense, and in line with other works (e.g., Barrouillet et al., 2007; Sweller, 1988; Weber et al., 2021), we refer to *cognitive load* as the amount of cognitive resources expended by working memory in a given task. How cognitive load, which we also refer to as *processing load*, may be engaged in predictive language processing is outlined below.

It has already been evidenced in purely linguistic contexts that cognitive load is engaged when comprehenders generate predictions. Thus, reading times, a measure of cognitive load during language comprehension (e.g., Just & Carpenter, 1993; King & Just, 1991; Lewis et al., 2006), have been shown to increase for words that allow to predict upcoming input (e.g., Cutter et al., 2021; Frank, 2013; Lowder et al., 2018). This can be viewed as readers requiring additional processing time for predictive linguistic cues because they may not only process the cues themselves, but also do some pre-processing of the predictable input (Koornneef, 2021; Lau et al., 2013; Ness & Meltzer-Asscher, 2018, 2021).

Besides, two studies of Ness and Meltzer-Asscher (2018, 2021) provide neurophysiological evidence that processing predictive input can result in additional cognitive load. Here, adult participants read semantically constraining (e.g., "The librarian helped him find the book") and unconstraining (e.g., "He couldn't find the book") sentences. They showed larger P600 amplitudes on the verb (e.g., "find") for the constraining versus the unconstraining sentences. The P600 is a positive deflection in the ERP signal with centro-parietal and sometimes frontal distributions that reaches its maximum around 600 ms post stimulus (e.g., Brouwer et al., 2012; Delogu et al., 2019). The P600 is considered to reflect mechanisms of an integration of the mental representations of predicted word candidates into working memory (Brouwer et al., 2012; Delogu et al., 2019; Kaan et al., 2000; Ness & Meltzer-Asscher, 2018, 2021; Sikos et al., 2021). The authors therefore interpreted the above reported P600 effect as follows. In the constraining condition (i.e., when the sentence allowed to predict a specific noun), cognitive resources were not only deployed to process the input received so far, but also to integrate the predicted noun into working memory and to form an online model of the predicted sentence (i.e., for pre-updating). This could have resulted in an increase in processing load for the verb in the constraining condition. In turn, a smaller P600, indicating less cognitive load for the verb, was observed in the unconstraining condition where the context did not allow to pre-process a particular noun. Of note, that the increase in cognitive load for the verb in the constraining condition was more prominent in adults with higher (versus such with smaller) working memory capacity, is in line with the idea that working memory resources were involved in pre-updating processes (Ness & Meltzer-Asscher, 2018, 2021).

Notably, Ness and Meltzer-Asscher (2018) report another interesting finding. Adults with higher (versus such with smaller) working memory capacity showed smaller P600 amplitudes when the constraining sentences were finally completed with a predictable noun (e.g., “book”). Thus, there was a trade-off in cognitive load across the different parts of the sentences. Comprehenders with high working memory capacity engaged additional cognitive load when the input allowed for predictions. In turn, their cognitive load decreased for words that they could already process in advance. Similar findings are reported in Maess et al. (2016). Here, adults showed larger N400 amplitudes on verbs that were semantically constraining (e.g., “He *conducts* the orchestra”) versus unconstraining (e.g., “He *leads* the orchestra”). However, their N400 amplitudes were smaller when the noun (“orchestra”) finally revealed in the constraining versus the unconstraining condition (for similar results, see Freunberger & Roehm, 2017). Given that the N400 reflects the extent to which the cognitive system is engaged in retrieving information from semantic long-term memory (Kutas & Federmeier, 2000, 2011; van Berkum, 2009), this can be interpreted as comprehenders having pre-processed the mental representations of the predicted input, resulting in less semantic processing when the predicted words finally revealed.[6] In sum, cognitive load seems to increase when comprehenders process predictive linguistic cues that allow to pre-process predictable input. In turn, less cognitive load is required to process words that could already be pre-processed.

Notably, visual contexts can put additional constraints on predictive linguistic cues (Reuter et al., 2020; Venhuizen et al., 2019), thereby allowing for more specific predictions (Ankener et al., 2018). Thus, visual contexts that are in line with predictive linguistic cues are assumed to allow comprehenders to not only pre-activate but also to pre-update prediction options in working memory (Huettig, 2011a, 2011b; Özkan et al., 2022; see Chapter 1.3). Therefore, an interesting question is in how far visual contexts that allow for the highly specific prediction of one or the less specific prediction of multiple sentence continuations differ in the amount of cognitive load required to a) form those (multiple) predictions and b) to process a predictable target word when it finally cashes out. In the next chapter, we turn to the question of how visual contexts can affect the cognitive load engaged in maintaining (multiple) prediction options. Thereafter, we address the question of how visual contexts can influence the processing load for cashed out predictions.

1.7.1 Does Forming Predictions in the Visual World Induce Cognitive Load?

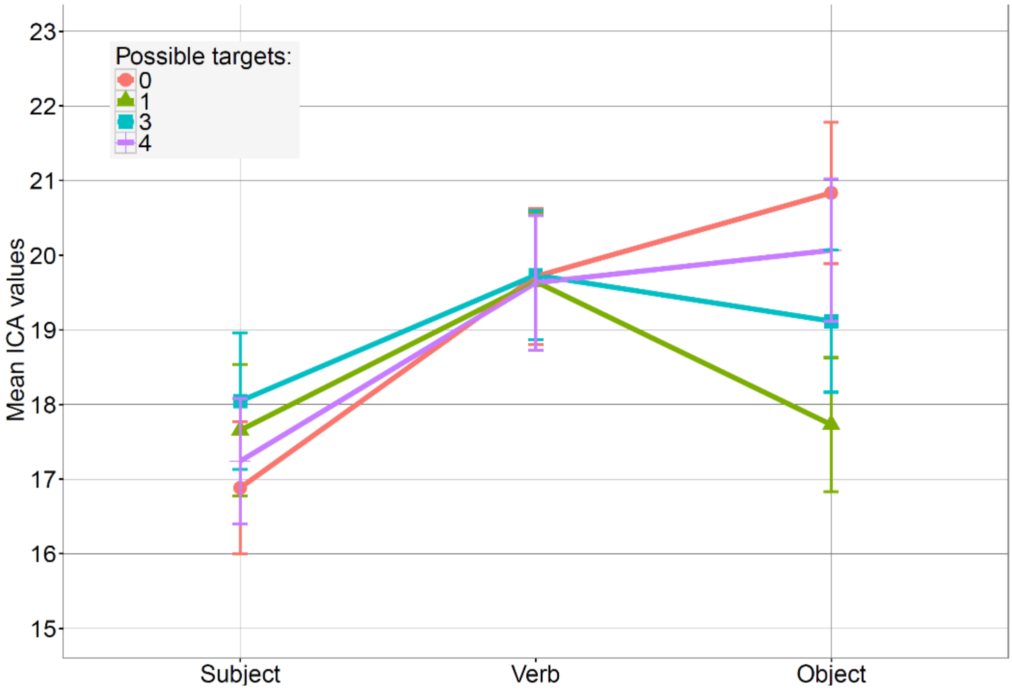
There is only little evidence of how visual contexts may influence the cognitive load engaged in the generation of predictions. Ankener et al. (2018) provide a first answer to this question. Here, adults listened to sentences with semantically constraining verbs (e.g., “The man *spills* soon the water”) while inspecting visual scenes of four objects of which either 0, 1, 3, or 4 were consistent with the verb constraints (e.g., spillable). The verb and the scenes in

[6] One could argue that attenuated P600 and N400 amplitudes for predictable versus unpredictable target nouns do not reflect effects of prediction but indicate that the predictable words were easier to integrate in the sentence context (see Chapter 1.2). However, the smaller P600 and N400 amplitudes for predictable target words were accompanied by larger P600 and N400 amplitudes for the preceding constraining cues. Thus, these results are more indicative of a processing advantage for the predictable nouns due to pre-processing.

the different visual conditions allowed to predict one (1-consistent), three (3-consistent) or four (4-consistent) particular noun candidates (see Figure 2, page 24). In the 0-consistent condition, only the semantic characteristics of the noun could be predicted (as the scenes displayed four verb-inconsistent visual distractors). The authors measured adults' Index of Cognitive Activity (ICA), a pupillometric measure of cognitive load engaged in cognitively demanding tasks such as language processing (for details, see Box 1, page 40). Higher ICA values indicate that more cognitive load is involved in a task (e.g., Demberg & Sayeed, 2016; Marshall, 2000). Adults' ICA values in the verb region of the sentences (i.e., in the region where predictions could be formed) did not differ among all visual conditions. Thus, adults engaged the same amount of cognitive load to predict either one (1-consistent) or multiple (3- and 4-consistent) noun candidates (see Figure 3). This is surprising as adults' anticipatory object fixations in the different visual conditions showed that they truly integrated either a single or multiple visual prediction options into prediction (see Chapter 1.6.1).

Figure 3

ICA Values of the Adult Participants in the Study of Ankener et al. (2018)



Note. Adult participants' ICA values in the study of Ankener et al. (2018) in three regions (unpredictive subject, constraining verb, predictable noun) of the sentences (e.g., "The man *spills* soon the water") across four visual conditions (0-, 1-, 3-, and 4-consistent). In the subject and verb region, the ICA values do not differ among the conditions. In the noun region, the ICA values are smaller in 1-consistent than in the 3- and 4-consistent conditions, but highest in the 0-consistent condition. Values do not differ among the 3- and 4-consistent conditions. Reprinted with permission from the authors.

A related study by Sikos et al. (2021) reports similar findings. Here, adults inspected visual scenes in four conditions that allowed, together with a noun, to predict either 0, 1, 3, or 4 potential sentence continuations. Adults' cognitive load (i.e., their ICA values) measured in the region of prediction (here: the noun region) did not vary as a function of the number of visual prediction options. Collectively, these findings suggest that adults' cognitive load does not increase when they maintain multiple versus a single visual prediction option.

These results are not consistent with the cognitive view that encountering a predictive linguistic cue together with a single or multiple visual prediction options should lead to the pre-updating of either a single or multiple prediction options in working memory which should, because working memory expends more cognitive resources to process more stimuli (e.g., Cowan, 2010; Just & Carpenter, 1993), lead to higher cognitive load for the pre-updating of multiple versus single prediction options (e.g., Huettig & Janse, 2016; Huettig et al., 2011a, 2011b; Lau et al., 2013; Ness & Meltzer-Asscher, 2018; Ito et al., 2018a; Slevc & Novick, 2013). Ankener et al. (2018) speculated that the ICA measure could just not be sensitive to the type of cognitive load engaged in visually situated prediction of single versus multiple word candidates. Alternatively, these results could mean that comprehenders do not incur any cognitive costs to form predictions in the visual world.

Of note, Ankener et al. (2018) and Sikos et al. (2021) measured comprehenders' ICA values in the region of the sentences that contained the predictive cue (e.g., the constraining verb). However, it may take some time until individuals have processed a predictive linguistic cue and a complex visual scene, stored and rehearsed the visual and linguistic input in working memory, pre-activated the related long-term memory representations, transferred them into working memory, and integrated them with the previous input in such a way that predictions are formed and maintained. As a result, possible effects in the ICA could rather reveal on a spill-over word after the predictive word than on the predictive word itself. This is likely as effects of prediction have often been shown to spill over from critical words to subsequent words in the input (e.g., Aurnhammer et al., 2021; Koornneef & van Berkum, 2006; Smith & Levy, 2013; Vela-Candelas et al., 2022). Based on this, this work aimed to examine whether effects of increased cognitive load to maintain multiple versus single prediction options may first become apparent in spill-over regions that succeed a predictive cue (e.g., "soon" given the sentence "The man *spills* soon the water").

1.7.2 Does Processing (Un-)Predicted Target Words in the Visual World Induce Cognitive Load?

Another question arising from the above findings is how visual contexts that contribute to the generation of predictions may influence the cognitive load required to process a prediction option when it is finally presented. In line with cognitive models of prediction, some electrophysiological studies have already shown that comprehenders experience little cognitive load (reflected in small P600 amplitudes) to process words that could be predicted highly specifically, and thus be pre-updated highly thoroughly in working memory by purely linguistic contexts (Ness and Meltzer-Asscher, 2018, 2021). However, also visual contexts can contribute to prediction in such a way that they induce the mechanism of pre-updating (Huettig & Janse, 2016; Huettig et al., 2011b; Özkan et al., 2022). It is therefore interesting how visual contexts that allow for more versus less specific predictions, and thus for a more or less thoroughly pre-processing of predictable target words may influence the cognitive load engaged to process these target words.

Ankener et al. (2018) provide an initial answer to this question as well. In the study outlined above, they also inspected adults' cognitive load in the noun region (e.g., "water") of the sentences (e.g., "The man *spills* soon the water"). That is, they examined whether adults' cognitive load differed across the visual conditions (0-, 1-, 3-, and 4-consistent) in the moment when the prediction was cashed out. Interestingly, adults showed less processing load (smaller ICA values) for the noun when the visual and the linguistic context jointly allowed for a more specific prediction of that noun (see Figure 3). Thus, when only one noun candidate (water) could be predicted (1-consistent condition), adults' cognitive load for the noun (water) was smaller than when multiple noun candidates could be predicted (3- and 4-consistent condition). Adults' cognitive load for the noun was highest when an unspecific prediction of only the semantic noun category could be formed (0-consistent condition).

These results can be interpreted as follows. When a visual scene and a linguistic cue jointly allow to predict a particular noun candidate, comprehenders require small cognitive load for that noun when it cashes out, because exactly that noun was been pre-processed. When multiple noun candidates can be predicted, more cognitive load is involved when the noun reveals, because the noun was pre-processed together with other noun candidates, and thus less thoroughly (Rommers & Federmeier, 2018). Finally, when only the semantic noun category can be predicted, comprehenders still engage a great amount of cognitive load for the noun since it was pre-processed at less specific levels of representations (at the level of semantics, not at word level). Similar results are reported in Sikos et al. (2021).[7]

In line with this, Tourtouri et al. (2015) found a processing advantage for words that could be predicted more specifically by the visuo-linguistic constraints. Here, adults showed more semantic processing (larger N400 amplitudes) for the noun ("bowl") of a sentence ("Find the *yellow* bowl") that was presented in a visual scene of two (yellow bowl, yellow can) versus only one (yellow bowl) color-consistent object. Since the N400 has been argued to reflect semantic retrieval of linguistic representations from long-term memory (Delogu et al., 2019; Lau et al., 2008; for reviews, see Kutas & Federmeier, 2011; Kutas et al., 2014), this suggests that the joint constraints of the adjective and the 1- versus 2-consistent scenes initiated the retrieval of long-term representation of either one or two noun candidates. Thus, the mental representations of either one or two nouns may have been pre-processed in working memory. In the 2-consistent condition, pre-processing resources may then have been distributed among two noun candidates which is why the noun ("bowl") was pre-processed less thoroughly and engaged more semantic processing when it finally revealed.

[7] Note that the results for the predictable target word are logical although both Ankener et al. (2018) and Sikos et al. (2021) found no effects on the preceding predictive cue. This could mean that processing the predictive cue and the target word in the different conditions (1-, 3-, 4-consistent) engaged different cognitive processes and that the ICA varies in its sensitivity to these processes. That is, the ICA may not reflect changes in cognitive load due to the prediction of single versus multiple words (cf. Ankener, 2018), but may be sensitive to the processing load of predictable target words that are shown among multiple distractor or competitor objects.

Together with the findings of Ankener et al. (2018) and Sikos et al. (2021) this result suggests that the more specifically a word can be predicted by the visuo-linguistic constraints, the more thoroughly it can be pre-processed by the cognitive system, and the less cognitive load is required when the word finally reveals. Beyond the cognitive view that pre-processing prediction options is strongly related to working memory capacity, the question arises of how a processing advantage for words that can be pre-processed more or less thoroughly by the visuo-linguistic constraints may emerge in comprehenders with limited working memory capacity such as (young) children.

Do Visual Contexts Influence Children's Processing Load for (Un-)Predictable Target Words? So far it has been evidenced that children's processing load for linguistic input can vary depending on the potential a visual context alone supplies to the anticipation of that input. This has been shown with the measure of pupil size, another pupillometric indicator of cognitive load engaged in language processing: Comprehenders' pupil size increases with increasing cognitive load (Schmidtke, 2017; Sirois & Brisson, 2014; for details, see Box 1, page 40). Besides, this has been shown with the ERP component N400 which has been validated also for children as measure for semantic incongruency processing (Benau et al., 2011; Friedrich & Friederici, 2004).

Krüger et al. (2020), for instance, have shown that children (1–6 years), like adults, involve less cognitive load (i.e., show smaller pupil sizes) when they hear animal sounds that could (e.g., mooing) versus could not (e.g., meowing) be anticipated by the picture of an animal (e.g., cow). Besides, as indicated by their pupil sizes and N400 amplitudes, children (14–30 months up to 9 years) who look at the picture of a familiar object (e.g., a fish) involve less processing load for the correct ("fish") versus incorrect ("car") label of the object (Bell et al., 2019; Csink et al., 2021; Friedrich & Friederici, 2005) as well as for the correct ("*fɪʃ*") versus incorrect ("*zɪʃ*") pronunciation of the object label (Fritsche & Höhle, 2015; Mani et al., 2012; Tamási et al., 2017, 2019). Thus, processing load is higher for an unpredicted label of an object. This could, according to cognitive models of prediction, be explained as follows (Huettig & Janse, 2016; Huettig et al., 2011b; Özkan et al., 2022). The visual object (e.g., picture of a fish) was stored and rehearsed in the visual component of working memory (Baddeley, 2000, 2003; Just & Carpenter, 1993). By this, related linguistic long-term memory representations were pre-activated and transferred into working memory, where they were integrated with the visual input in such a way that the word "fish" was pre-processed at several linguistic levels (e.g., semantical and phonological level). This resulted in a processing advantage for words that were consistent in semantic meaning or pronunciation with the representations pre-activated and pre-processed by the picture, but in higher cognitive load for picture-inconsistent words ("car") or morphemes ("*zɪʃ*").

Finally, there is first evidence that also the joint constraints of a visual and a linguistic context can modulate children's processing load for upcoming words. Children (2–3 years), like adults, who look at visual scenes (e.g., a yellow bus) and listen to German sentences with adjectives that either are (e.g., "A yellow$_{masc}$ bus$_{masc}$") or are not (e.g., "A yellow$_{fem}$ bus$_{masc}$") consistent with the target nouns (e.g., "bus") show less cognitive load (smaller pupil sizes) for the nouns in the consistent versus the inconsistent condition (Süss et al., 2018). Originally, this result was considered as comprehenders being sensitive to adjectival agreement, but it could also mean the following. In both conditions, the visual scene (yellow bus) activated the long-term representations of the target noun "bus" and its pre-processing. The masculine adjective may have elevated, and the female adjective may have reduced further pre-processing of the masculine noun "bus" causing additional cognitive load (larger pupil sizes) for the noun when it was finally presented in the inconsistent condition. Although the effect in pupil sizes was weaker in children versus adults, this result could mean that even young children show less cognitive load to process predictable words when the visuo-linguistic context allows for a more specific pre-processing of these words. This is also evidenced indirectly by a study of Gambi et al. (2021). Here, children (2–5 years) listened to constraining sentences (e.g., “The *dog chews* on the slippers”) while looking at highly (bone) versus mildly predictable objects (slippers). Moreover, they listened to unconstraining sentences with the same target nouns (e.g., “The girl looks for the slippers”) in the same visual contexts. Children took longer to guide their eyes to the target object (slippers) after the target noun (e.g., “slippers”) was named in the constraining versus the unconstraining condition. This could be explained as follows. In the constraining condition they may rather have pre-updated the highly predictable noun (“bone”) than the mildly predictable noun (“slippers”), causing additional recognition load for the actual target word (“slippers”). These results suggest that inaccurate predictions formed by visuo-linguistic constraints can hinder target word processing in children (which indirectly means that correct predictions should facilitate word processing).

1.7.3 Summary and Implications

Although prediction can make language comprehension fast and accurate, anticipating input requires cognitive resources. There is a trade-off in cognitive load engaged in predictive processing: Comprehenders involve higher cognitive load to process input that does (versus such that does not) allow to form predictions (since resources are not only engaged to process the actual but also the upcoming input). In turn, comprehenders show small cognitive load for words that could be predicted (since they have already been pre-processed). Because visual contexts can influence which prediction options are pre-updated in working memory, they may also affect the cognitive load engaged in predictive processing.

Surprisingly, adults *do not* experience higher cognitive load when a visual scene allows to pre-update multiple versus only single prediction options. This is not in line with the view that pre-updating an increasing number of sentence continuations should cause an increase in cognitive load because working memory typically engages more resources to manipulate more stimuli. In the current work, we tested two possible explanations for this result. First, we examined whether prior works did not find such results for adults as they did not account for spill-over effects. Second, children typically have smaller cognitive capacity than adults (e.g., Cowan et al., 2010; Johnson et al., 2014), which usually goes along with more processing load engaged in a task (e.g., Baddeley, 2000, 2003; Johnson et al., 2014). Based on this, we applied a developmental approach to test whether only children (with smaller cognitive capacity) but not adults would show effects of higher cognitive load to maintain multiple versus single prediction options. This would mean that variations in cognitive load among multiple versus single predictions are too small to reveal in adults given their higher cognitive capacity.

Besides, adults and children show small cognitive load for words that could (versus such that could not) be predicted by visual or visuo-linguistic contexts. While there is first indication that word processing load can decrease the more specifically a word could be predicted by the visuo-linguistic constraints, it remains interesting how sensitive this relation is to variations in working memory capacity. This is because individuals with small cognitive capacity may not be able to pre-update (one or multiple) prediction options as thoroughly as others, thus may not have/have smaller processing benefits for these prediction options when they finally reveal. Examining variations in children's and adults' cognitive load for target words that could be predicted highly specifically versus rather unspecifically by the visuo-linguistic context could reveal answers to this question due to children's limited cognitive capacity.

Box 1

Assessment of Cognitive Load With Pupillometry

Pupillometry. This is a physiological technique that captures changes in the size of the pupil, the circular opening of the iris in the center of the human eye (Sirois & Brisson, 2014). Smaller pupil sizes index that a small amount, larger pupil sizes index that a high amount of cognitive load is engaged in a task (e.g., Gavas et al., 2017; Just & Carpenter, 1993; Zekveld et al., 2018). Pupil size has been shown to increase (up till resource overload) as the number of stimuli such as digits (e.g., Granholm et al., 1996), words (e.g., Kahneman & Beatty, 1966), or visual items (e.g., Kursawe & Zimmer, 2015) maintained in working memory increases (for a review, see Zekveld et al., 2018). Besides, individuals show larger pupil sizes when they process more (object-relative sentences, difficult words) versus less (subject-relative sentences, easy words) complex linguistic input (Farmer et al., 2016; Just & Carpenter, 1993; Wendt et al., 2016). Thus, pupil size can measure cognitive load of language comprehension (e.g., Schmidtke, 2017; Sirois & Brisson, 2014).

This is because two antagonistic muscles control pupil size: The *sphincter* pupillae contracts, the *dilator* pupillae dilates the pupil. The sphincter is under control of the parasympathetic nervous system and regulates the pupillary light reflex (i.e., ensures that the pupil contracts with increasing light exposure to protect the retina from damage). Besides, the sphincter and dilator both obtain input from the sympathetic nervous system which regulates autonomic and cognitive functioning. The balance of the activation of both muscles dictates pupil size and is modulated by the *locus coeruleus* (LC), a small nucleus in the brainstem that releases norepinephrine (NE), a neurotransmitter in the central nervous system involved in autonomic and cognitive processes. The LC/NE system is active when individuals are exposed to arousing or cognitively demanding situations (Gilzenrat et al., 2010; Sara, 2009; Sara & Bouret, 2012; for reviews, see Eckstein et al., 2017; Samuels & Szabadi, 2008). Under such demands, activity in the LC leads to NE secretion which inhibits sleepiness and promotes wakefulness, thereby enables a shift of attention and cognitive functioning, thus a shift of mental resources to a situation (for reviews, see Eckstein et al., 2017; Samuels & Szabadi, 2008). Activity in the LC/NE system is associated in strength and time with pupil size (Eckstein et al., 2017). Thus, electrical micro stimulations of the LC in monkeys have been shown to result in pupil dilations (Joshi et al., 2016). Besides, single cell and MRI recordings have shown that LC activity in arousing or demanding tasks are tightly correlated with monkey's pupil dilations (Rajkowski et al., 2004; Varazzani et al., 2015).

Note. Box continued on the next page.

Box 1 (Continued)

Assessment of Cognitive Load With Pupillometry

That LC/NE activity and pupil dilations are related is based on two cascades. First, NE emission due to LC activity activates innervating sympathetic fibers connected with the dilator muscle, resulting in pupil dilation. Second, via noradrenergic fibers, LC activity initiates the inhibition of the sphincter muscle, causing the inhibition of pupil constriction which, in turn, causes pupil dilation (for reviews, see Eckstein et al., 2017; Zhang & Emberson, 2020). Such a relation of LC/NE activity and pupil size also exists in human adults. Using fMRI it has been shown that firing of LC neurons of adults who work on cognitively demanding tasks (e.g., oddball task) is related with pupil dilation (e.g., Alnæs et al., 2014; Murphy et al., 2014). In sum, increasing demands on the cognitive system cause activity in the LC/NE, leading to pupil dilation.

Pupil size can be recorded with eye-trackers over the time course of a task with high temporal resolution (Zhang & Emberson, 2020). There are different types of pupil size measures (for a review, see Mahanama et al., 2022). In this work we refer to two of them as measures of cognitive load.

Baseline Corrected Pupil Size. This measure compares the overall pupil size during a task with the overall pupil size measured immediately prior to the task. Since baseline corrections supply to the extraction of light-influences from pupil size (Weber et al., 2021) and since absolute pupil size varies across different populations (e.g., pupil size at rest decreases from child- to adulthood, Eckstein et al., 2017; Johnson et al., 2014), this allows for more reliable and comparable conclusions about task-evoked cognitive load (van Engen & McLaughlin, 2018). Baseline corrected pupil size is validated as measure of cognitive load in direct tasks of working memory: Adults and children (7–14 years) who maintain an increasing number of items for later recall show an increase in corrected pupil size (Johnson et al., 2014; Karatekin, 2004, Karatekin et al., 2004, 2007; Weber et al., 2021). Besides, baseline corrected pupil size reflects the amount of mental resources involved in indirect tasks of cognitive load such as task-switching (Katidioti, 2014; Rondeel et al., 2015; Wang et al., 2022) or arithmetical problem solving (e.g., Krejtz et al., 2018; Landgraf et al., 2010; Throndsen et al., 2022). Baseline corrected pupil size also reflects the cognitive load engaged in language comprehension. Children's (6–11 years, Lum et al., 2017; McGarrigle et al., 2017) and adults' (e.g., Chapman & Hallowell, 2021; Just & Carpenter 1993; Koelewijn et al., 2012; Piquado et al., 2010; Stanners et al., 1972; Zekveld et al., 2010) corrected pupil sizes are larger when they process sentences of high versus low complexity (induced by sentence structure, length, or intelligibility).

Note. Box continued on the next page.

Box 1 (Continued)

Assessment of Cognitive Load With Pupillometry

Baseline corrected pupil size also indexes the cognitive load of predictive processing: Children (6 month, Hochmann & Papeo, 2014) and adults (Häuser et al., 2018, 2019) show smaller baseline corrected pupil sizes to process predicted versus unpredicted linguistic input. In sum, baseline corrected pupil size is validated for adults and children to reflect the cognitive load engaged in a variety of tasks such as (predictive) language processing (for reviews, see Schmidtke, 2017; Sirois & Brisson, 2014; van der Wel & van Steenbergen, 2018; Zekveld et al., 2010).

Index of Cognitive Activity (ICA). The ICA is a more novel measure of cognitive load that is also based on the fact that the pupil dilates in response to the amount of mental resources involved in a task (Marshall, 2000). The ICA is a micro-level measure of pupil dilations that does not relate cognitive load to overall changes in pupil size but counts the frequency of rapid and small pupil dilations that are caused by a task and independent of light-influences. Thus, the ICA is a more dynamic and fine-grained measure of cognitive load than overall pupil size (Demberg, 2013; Demberg & Sayeed, 2016; Demberg et al., 2013; Marshall, 2000). To extract the ICA from the pupil data recorded with an eye-tracker, a wavelet analysis discards larger light-induced oscillations and extracts abrupt and small task-evoked oscillations, called ICA events. This is possible as the sphincter muscle of the pupil contracts when responding to light, while the dilator muscle contracts in a smaller and rapider way in response to cognitive activity and independent of changes in luminance (Beatty, 1982; Marshall, 2000). This procedure is patented by Marshall (U.S. Patent No. 6,090,051; Marshall, 2000) and can be conducted with the EyeWorks Cognitive Workload Module software (EyeTracking, 2016). A small number of ICA events indexes a small amount of cognitive load (Demberg, 2013; Demberg & Sayeed, 2016; Demberg et al., 2013; Marshall, 2000). To our knowledge, no study to date has used the ICA with children. For adults, the ICA is validated as measure of cognitive load in several cognitive tasks like arithmetical (Marshall, 2002) or spatial (Fehringer, 2021) problem solving and simulated driving (Dlugosch et al., 2013; Schwalm et al., 2008). Besides, the ICA can index cognitive load of language processing. Thus, adults show less ICA events when reading easy (subject-relative) versus complex (object-relative) sentences or when encountering predicted versus unpredicted input (Demberg & Sayeed, 2016; Demberg et al., 2013).

Note. This box provides background information for the assessment of cognitive load with pupillometric measures such as baseline corrected pupil size and the ICA.

1.8 Research Questions and Hypothesis of the Present Work

In line with the theoretical background outlined above, the purpose of this work was fivefold. Below, our research questions and hypothesis are delineated. We then provide an overview about the empirical part of this dissertation.

1.8.1 RQ 1: Do Children, Like Adults, Follow a Multiple Predictions Pattern?

We first aimed to examine whether children and adults who listen to constraining sentences and inspect complex visual scenes that allow to predict either one or multiple sentence continuations can follow a multiple predictions pattern, that is, whether they can maintain not only one but also multiple visual prediction options in parallel. In line with prior works (Ankener et al., 2018; Sikos et al., 2021) we assumed to find a multiple predictions pattern for adults. We expected them to predict a single sentence continuation when the visual context presents a single visual prediction option, but to predict multiple sentence continuations when a visual context shows multiple visual prediction options. Since children from an early age identify and anticipatorily rely on a single visual prediction option next to one or multiple distractors (e.g., Andreu et al., 2013; Mani & Huettig, 2012, 2014) or on two visual prediction options in parallel (Borovsky et al., 2012; Gambi et al., 2021; Mani et al., 2016), young children could be able to follow a multiple predictions pattern. They could maintain multiple visual prediction options in parallel, thereby rely on the visuo-linguistic constraints as efficiently as adults. However, children have limited working memory capacity relative to adults (e.g., Cowan et al., 2010; Johnson et al., 2014; Kharitonova et al., 2015), thus could be less capable than adults to maintain multiple prediction options in parallel. This question is addressed in Experiment 1.

1.8.2 RQ 2: Is Children's Prediction in the Complex Visual World Influenced by Language Experience?

Our next goal was to investigate whether children's ability to predict language in the complex visual world is associated with their language experience. Visually situated prediction involves the pre-activation of the mental representations of prediction options in long-term memory (e.g., Huettig et al., 2011b; Ness & Meltzer-Asscher, 2018, 2021), while the content of long-term memory is shaped by past (language) experience (Mani & Huettig, 2012; Zhang et al., 2020). Since language experience varies among children (e.g., Borovsky & Creel, 2014; Rabagliati et al., 2016), and in line with past research (e.g., Borovsky & Creel, 2014; Borovsky et al., 2012), we expected children's usage of prediction in the complex visual world to increase with increasing language experience. This question is considered in Experiment 1.

1.8.3 RQ 3: Does Cognitive Load Increase When Multiple Predictions are Maintained?

Under the assumption that children and adults can maintain multiple visual prediction options, we examined how visual contexts that allow to predict either one or multiple sentence continuations can affect comprehenders' cognitive load in the moment they pre-update these sentence continuation(s). Predicting input in the visual world involves the pre-updating of prediction options in working memory (e.g., Huettig et al., 2011b; Ness & Meltzer-Asscher, 2018, 2021). Since working memory typically expends more cognitive resources to process more stimuli (Johnson et al., 2014; Just & Carpenter, 1993), we expected children's and adults' cognitive load to increase when they pre-update multiple versus only a single prediction option. Past research with adults did not reveal such an effect in the constraining part of a sentence (Ankener et al., 2018; Sikos et al., 2021). However, it could take some time until changes in cognitive load reflect in measures of cognitive load. We extended prior comparable studies by also focusing on the cognitive load in the spill-over region following a constraining linguistic cue. We expected that effects of increases in cognitive load to pre-update multiple versus only single sentence continuations could reveal in such a spill-over region. Besides, we expected this effect to be more pronounced in children versus adults due to children's limited working memory capacity (e.g., Cowan et al., 2010; Johnson et al., 2014), and since individuals with smaller cognitive capacity engage more cognitive resources in a task than other individuals (e.g., Johnson et al., 2014; Just & Carpenter, 1993). This question is examined in Experiments 1 and 2.

1.8.4 RQ 4: Does Cognitive Load Increase for Specific Predictions

We fourth aimed to examine how children's and adults' cognitive load to form predictions varies among situations where a visual context does versus does not contribute to prediction. If a visual scene shows one or more visual prediction options that are consistent with the semantic constraints of a predictive linguistic cue, this enables the prediction of particular sentence continuation(s). In contrast, if only visual distractors that do not match the semantic constraints of the predictive linguistic cue are presented, this only allows to predict the semantic features of the upcoming input. We compared children's and adults cognitive load among the above conditions.

Here, two results were plausible: In line with cognitive models, input that does not allow to commit to specific prediction options (e.g., when only semantic properties can be predicted) could not induce that level of pre-activation needed to initiate the mechanism of pre-updating (e.g., Ness & Meltzer-Asscher, 2018, 2021). This could result in less cognitive load relative to situations in which a visual scene allows to pre-update particular words. Otherwise, when only semantic word properties can be predicted since the visual context is not consistent with the linguistic constraints (e.g., with the semantic constraints of a verb) comprehenders could engage additional cognitive load to resolve the ambiguity of the visual and linguistic input. This question is investigated in Experiments 1 and 2.

1.8.5 *RQ 5: Do More Specific Predictions Facilitate Target Word Processing?*

Finally, we investigated how a visual context that, together with a linguistic cue, allows to predict either one or multiple sentence continuations can contribute to the processing load for a sentence continuation when it finally cashes out. Adults engage small processing load for words that could be pre-processed by a purely linguistic context (e.g., Freunberger & Roehm, 2017; Maess et al., 2016; Ness & Meltzer-Asscher, 2018, 2021). Besides, adults' cognitive load decreases for words that could be predicted more specifically, thus be pre-processed more thoroughly by the joint visual and linguistic constraints (Ankener et al., 2018; Sikos et al., 2021; Tourtouri et al., 2015). Also children show small processing load for input that could be predicted by a visual context alone (e.g., Bell et al., 2019; Fritsche & Höhle, 2015) or by the joint visuo-linguistic constraints (Gambi et al., 2021; Süss et al., 2018). We therefore expected both children and adults to engage less processing load for words that could be predicted more (versus less) specifically and thus be pre-processed more (versus less) thoroughly by the visuo-linguistic constraints. This question is addressed in Experiments 1 and 2.

1.9 General Methods and Overview of the Present Work

To provide answers to the above questions we present data from two experiments in which we presented children and adults with predictable sentences containing semantically constraining verbs and predictable target nouns (e.g., "The girl *eats* soon the cake") together with visual scenes of four objects each. The scenes varied in predictability in such a way that they showed either 0, 1, 3, or 4 visual prediction options (e.g., edible objects). How the sentences and visual scenes were generated and pre-tested for our purposes with a group of young children (4–6 years) is outlined in Chapter 2.

We then present Experiment 1, which was an eye-tracking study in which young children (5–6 years) and adults listened to the sentences and inspected the visual scenes. The presentation of this experiment is divided into two separate chapters. To examine whether children and adults follow a multiple predictions pattern (RQ 1) and whether children's prediction behavior varies as a function of their language experience (RQ 2) we inspected participants' anticipatory fixations of the visual objects. This is presented in Chapter 3. To identify how visual scenes of varying predictability affect cognitive load during predictive processing (RQ 3, 4, and 5), we relied on participants' ICA and pupil size values across the time course of the sentences. This is presented in Chapter 4.

With Experiment 2 (see Chapter 5) we aimed to extend our findings of Experiment 1 for the questions of how visual contexts of varying predictability influence cognitive load during predictive processing (RQ 3, 4, and 5). This was done with a measure of cognitive load that is not based on pupillometry (i.e., with processing times) and with another input channel of the constraining linguistic sentences. In doing so, literate children (8–12 years) and adults read the predictable sentences in a word-by-word self-paced reading fashion while inspecting the visual scenes in the four conditions. Word processing times indicated processing load for the different parts of the sentences. This could provide first indication whether predictive processing load is also influenced by visual predictability when the linguistic signal is received in written form. While word reading puts demands on several components of working memory (e.g., phonological loop), it also puts demands on visual working memory (Pham & Hasson, 2014; Swanson, 2000, 2010; Swanson & Jerman, 2007). Thus, reading fluency and reading comprehension are positively associated with visual working memory capacity (Bayliss et al., 2003, 2005; Goff et al., 2005; Pham & Hasson, 2014; Swanson & Howell, 2001). We therefore examined the effect of visual predictability on predictive processing load in situations where the constraining linguistic signal is presented in written form, thus enters the cognitive system through a visual stream of perception. Here, not only the visual context, but also the words in the input may be temporarily stored and manipulated in visual working memory. Since visual working memory is capacity limited (Cowan et al., 2011; Luck & Vogel, 2013), this could cause a resource overload which, in turn, could mask variations in cognitive load among conditions of different visual predictability. Otherwise, in case Experiment 2 replicates (some) results of Experiment 1, this would mean that variations in cognitive load due to visual predictability are robust against changes in input-modality and strong enough to reveal despite additional load of visual working memory during reading. Besides, comparable findings among both studies would yield further evidence for research questions 3, 4, and 5 with another measure of cognitive load (i.e., with the behavioral measure of processing times), and thus make our results more generalizable. Finally, since Experiment 2 was conducted with older children the comparison of both studies can provide first indication how effects of visual predictability on predictive processing load develop across childhood.

As a closure of this dissertation, the results of Experiments 1 and 2 are integrated and discussed beyond the above theoretical background. This is presented in Chapter 6.

2 Stimulus Generation

In order to address the research questions formulated in Chapter 1.8, we created stimulus materials that are suitable to be used with children and adults. All stimuli of the present work consisted of a sentence that was presented together with a visual scene. We generated an initial set of 44 stimuli, each consisting of a semantically constraining sentence (e.g., “The father *eats* the waffle”) and a visual scene of four object pictures in four different visual conditions (0-, 1-, 3-, and 4-consistent). To ensure that children were familiar to the visual objects and would identify them as plausible arguments of the respective verbs, we conducted a pretest of the stimulus materials. This resulted in a final set of 32 items that were used in our Experiments. Below, we first explain the characteristics of the sentences and the visual scenes. Then, we present a pretest that was conducted to obtain the final set of stimuli.[8]

2.1 Linguistic Stimuli

Each stimulus consisted of a five-word German independent main clause (e.g., “Der Vater *verschlingt* die Waffel”; see Table 1). For ease of comprehension we refer to the approximate English translation of the example sentence from this point on (“The father *eats* the waffle”). All sentences followed the same syntactic structure (noun phrase — verb — noun phrase). Each sentence described the action (e.g., “eat”) of an agent (e.g., “father”) to or with an object (e.g., “waffle”). The agents of the sentences were easy to understand and uniformly distributed in terms of female and male characters (e.g., “father”, “mother”). Besides, it was ensured that the agents were plausible with respect to the upcoming sentence content. The verb “to iron”, for instance, is typically associated with adult agents which was considered in the respective stimulus. Nevertheless, the agents did not provide any clues for sentence continuation, i.e., they did not allow for the prediction of a particular verb or object

[8] Parts of this chapter were copied or adapted from a published manuscript (Sommerfeld et al., 2022).

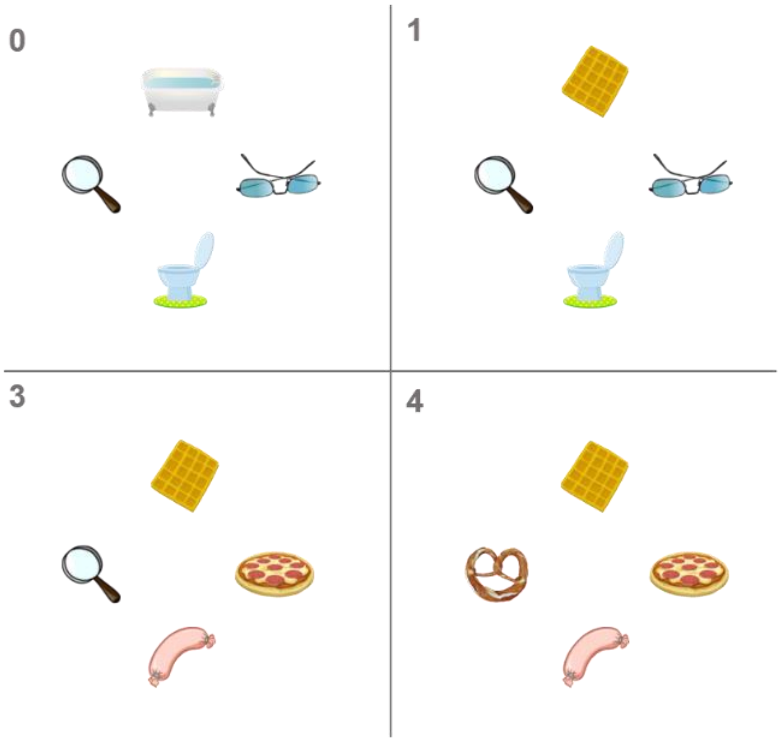
noun. Both of this was assured based on face-validity by three German native speaking experts of our psycholinguistic research group. The verbs (e.g., “to eat”) were semantically constraining, i.e., they allowed only for a limited number of plausible arguments (e.g., edible objects). Most of the verbs were taken from previous studies on predictive language processing in which they were associated with high cloze probabilities (cf. Altmann & Kamide, 1999; Andreu et al., 2013; Ankener, 2019; Brouwer et al., 2017b; Mani & Huettig, 2012). The remaining verbs were selected and rated as semantically constraining on the basis of face-validity by the experts of our research group. The target nouns (e.g., “waffle”) were plausible arguments of the constraining verbs (e.g., a waffle can be eaten). This was validated in a pretest (see Chapter 2.3).

2.2 Visual Stimuli

Each sentence was presented in combination with a visual scene that consisted of four object pictures arranged around the center of a white background screen (for an example, see Figure 4). All object pictures were colored cliparts because this ensured the ecological validity of the visual stimuli, i.e., the extent to which the pictures resemble the real-world objects they represent (Moreno-Martínez & Montoro, 2012; Reuter et al., 2020; Saryazdi et al., 2018). All cliparts were collected online from free clipart libraries, while some of them were created by combining and modifying existing cliparts. To avoid differences in processing across the pictures, they all depicted concrete and inanimate objects. For better recognizability, some pictures were re-inked and all pictures were cleared of shadows and lettering. Any editing was done with Microsoft Paint 3D (Microsoft Corporation, 2017). More details as well as the cliparts can be obtained from Sommerfeld et al. (2022).

Figure 4

Example of a Visual Scene in all Four Conditions



Note. From left to right and top to bottom: 0, 1, 3, or 4 objects are consistent with the semantic verb constraints of the sentence “The father *eats* the waffle” (or 4, 3, 1, or 0 objects match the verb constraints of the yoked sentence “The mother *cleans* the magnifier”).

For each sentence, four visual scenes in four different visual conditions were created (0-, 1-, 3-, or 4-consistent). Across the four conditions, the number of visual objects being consistent with the semantic verb constraints of the sentence (e.g., the number of edible objects) was manipulated: Either 0, 1, 3, or 4 of the objects were plausible arguments of the verb (see Figure 4). A scene in the 3-consistent condition, for instance, was made up of one target object (a picture of the target noun), two competitors (pictures of objects being also consistent with the verb constraints), and one distractor (a picture of an object being inconsistent with the verb constraints). We refer as “visual prediction options” to all objects that were plausible arguments of the constraining verbs (i.e., the target and the competitor objects). Objects that are not consistent with the verb constraints are considered as “visual distractors”. One of the visual prediction options shown in the 1-, 3-, and 4-consistent conditions each was the picture of the target noun of the corresponding sentence (e.g., a waffle picture). A pretest verified that young children truly identify the visual prediction options as plausible arguments of the constraining verbs.

The visual scenes were counterbalanced across the sentences in such a way that, for instance, a 0-consistent scene of one sentence served as a 4-consistent scene for a yoked sentence (see Figure 4). Given that grammatical gender allows for article based prediction in German (e.g., Bobb & Mani, 2013; Haeuser et al., 2020), we only used objects of the same grammatical gender within the visual scenes. The position of targets, competitors, and distractors was rotated across the stimuli between top, right, bottom, and left.

2.3 Pretest

According to the above requirements, we generated an initial set of 44 stimuli, each consisting of a semantically constraining sentence (e.g., “The father *eats* the waffle”) and a visual scene of four object pictures in four different visual conditions (0-, 1-, 3-, and 4-consistent). A total of 176 object pictures was used to generate the scenes. Given that the present work aimed at investigating visually situated predictive language processing in children and because children vary extensively in their language experience (e.g., Borovsky & Creel, 2014; Huettig, 2015; Rabagliati et al., 2016), a pretest of the stimulus materials was conducted to validate their usage with young children.

We presented children (4–6 years) with the semantically constraining verbs of the 44 sentences and with the corresponding visual scenes in the different conditions. They worked on two tasks. First, to control how familiar our visual stimuli were in general to young children, the participants were asked to name the object pictures of the visual scenes (naming task). Second, they were asked for each visual scene to classify those objects that are plausible arguments of the corresponding constraining verb (categorization task). We retained only those stimuli for our experiments, for which a considerable number of children correctly identified the visual prediction options.

2.3.1 Participants

The pretest was conducted with forty young children who did not participate in Experiments 1 and 2 (*M* = 5.50 years, *SD* = 0.82, *range* = 4.00–6.80 years, 16 boys and 24 girls). This age group was chosen because stimulus materials are typically normed for the youngest age group of interest of a given study (cf. Borovsky & Creel, 2014; Nation et al., 2003; Vergilova et al., 2020) which was 5 to 6 years for the present work (see Experiment 1). As confirmed by their parents, all children were German native speakers without any reading or writing experience. The pretest was conducted at Saarland University in 2019. Families were recruited via flyer and newspaper advertisement and received 10 Euro as compensation. Parents gave informed consent.

2.3.2 Materials

In the pretest, children were presented with each of the 44 constraining verbs in combination with the corresponding visual scenes in the 4- and 1-consistent conditions (see Figure 4). Showing the 4-consistent scenes allowed us to reveal for each stimulus whether young children truly identify the target object (e.g., waffle) and all three competitor objects (e.g., pizza, sausage, pretzel) as visual prediction options (e.g., as edible). Presenting the 1-consistent scenes enabled us to uncover whether children identify the target object as single plausible verb argument next to three distractor objects (e.g., magnifier, glasses, toilet). This also allowed inferences on whether the three competitors (shown in the 4-consistent scenes) and the single distractor (shown in the 1-consistent scenes) of the 3-consistent condition were correctly identified.

In sum, the pretest consisted of 44 semantically constraining verbs, each paired with one scene in the 4-consistent and one scene in the 1-consistent condition (88 visual scenes in total). To ensure that each child worked on each verb only in one of these two conditions, we divided the 88 visual scenes in two stimulus lists. Each list collected ratings for all 44 verbs and consisted of 22 scenes in the 4-consistent and 22 scenes in the 1-consistent condition. A verb that was tested in the 4-consistent condition in one list was tested in the 1-consistent condition in the other list.

To mask the design of the pretest, each list also consisted of 16 filler scenes to each of which we randomly assigned four additional object pictures that were not used in the item scenes. Filler scenes appeared after each second to fifth item scene. Each list started with seven additional practice pictures of objects shown individually on the screen to familiarize children with the study. For the filler scenes and the practice pictures, only the naming task was conducted. In sum, each list consisted of a total of 60 visual scenes of four objects each (44 item scenes, 16 filler scenes) and seven individual practice pictures. To each list, 20 children were assigned randomly.

2.3.3 Naming Task

For the naming task, children were instructed to name the objects of the scenes to the best of their knowledge while the experimenter pointed to them one after the other, starting with the top object, followed by the next three objects in a clockwise direction. If no name was given for an object, the experimenter asked again "What could that be?" until children produced a name or negated to know it. The naming task took as long as children needed to name all four objects (or to negate to know them) which they did rather quickly after the experimenter pointed at them.

As indicator of object recognition, we assessed children's name agreement for each object picture. Name agreement is defined as the extent to which participants agree on a particular name to refer to an object (Bonin et al., 2003). High name agreement ensures that an object is familiar to children and that its pictorial appearance elicits the intended object representation (Borovsky & Creel, 2014; Borovsky et al., 2012). For each object picture, name agreement was calculated as the percentage of children who named the object with the intended name, while higher values indicate greater object recognition (cf. Bonin et al., 2003; Pompéia et al., 2001; Snodgrass & Vanderwart, 1980). We accepted synonyms as correct names since general recognition of the objects and their properties was of greater interest for the present work than the production of the most common object name. Wrong names (being neither the correct name nor a synonym) and "do not know answers" were rated as "wrong answers". The intended names were determined a-priori by the three experts of our research group. For instance, regarding the "waffle" picture, all three experts distinguished a-priori that the intended name is "waffle". Post-hoc, the same experts determined all answers that were not identical with the intended names as either a synonym or a wrong name. As synonym we defined an official other correct name of the intended picture name (verified by https://www.duden.de/synonyme, last access: December 14, 2021). In addition, we rated diminutives (e.g., "Würstchen" instead of "Wurst"), accurate other names (e.g., "Longboard" instead of "Skateboard"), and over-informative names (e.g., "Holztür" instead of "Tür") as correct synonyms. All remaining answers were rated as wrong names. Plural names and mispronunciations were rated as correct names. For an overview on children's given answers, see Sommerfeld et al. (2022).

2.3.4 Semantic Categorization Task

Only for the item scenes, the naming task was followed by the categorization task. Here, children were asked if none, one, or more objects of a scene were suitable arguments of a particular verb. For example, regarding the scenes in the 1- and 4-consistent conditions of Figure 4, children were asked "Can one eat one or more of those objects and if so, which one(s)?”. The children could answer in the order of their choice by pointing at or naming those object(s) they rated as plausible verb arguments. If children hesitated, the experimenter engaged them to try an answer only once. After children confirmed to be ready, the next scene appeared on the screen.

We extracted a categorization score for each object of the visual scenes. The categorization score was defined as the percentage of children who rated an object as plausible argument of a verb. For instance, in case that all children classified the picture of the waffle as edible, this would indicate a categorization score of 100%. For each verb, we calculated one categorization score for the target object (e.g., for the waffle which was shown in the 1- and 4-consistent conditions), one score for each of the three competitor objects (e.g., for the pizza, sausage, and pretzel which were shown in the 4-consistent condition), and one score for each of the three distractor objects (e.g., for the magnifier, glasses, and toilet which were shown in the 1-consistent condition). This resulted in a total of seven categorization scores for each of the 44 verbs. Since the verbs were tested for their consistency with the target object in both stimulus lists (the waffle, for instance, was presented in the 1-consistent scene in the one list and in the 4-consistent scene in the other list), the categorization scores for the target objects based on the ratings of all 40 children. Since the verbs were tested for their fit with the competitor and distractor objects in only one of the stimulus lists, the scores for the competitors and distractors each based on the ratings of 20 children.

2.3.5 Procedure

Children were tested individually in one-hour sessions. Parents gave informed consent and filled in a form about their children’s age, gender, and mother tongue. Then, the child and the experimenter entered the laboratory. After a short adaptation phase, children were asked to take a seat in front of a computer and to look at the picture scenes presented onscreen one after the other. For each trial, children worked first on the naming task, for the item stimuli, they also worked on the categorization task. After each trial, children's answers were noted in a protocol and the experimenter started the next trial. Children were given a short break after each third of stimuli (i.e., after 20 scenes). In the end, families received the compensation.

2.3.6 Stimulus Selection Based on the Pretest Results

Based on children's pretest ratings in the categorization task, we selected 32 items for our experiments. The items consisted of a sentence paired with four different visual scenes (0-, 1-, 3-, and 4-consistent condition) and met the following requirements. For each item, at least 70% of the children correctly classified the target object and the three competitor objects as plausible arguments of the verb. For instance, regarding the item shown in Figure 4, at least 70% of the children classified the target object "waffle" and each of the three competitor objects "pizza", "sausage", and "pretzel" as "edible". This ensured that children truly considered the target object and the three competitors as visual prediction options of the given verb. For the final 32 items, the 32 target objects were classified as plausible verb arguments by 97% of the children on average (*SD* = 4.97, *range* = 78–100). The 96 competitor objects were rated as potential verb arguments by 95% of the children on average (*SD* = 6.18, *range* = 70–100). The semantic categorization scores for all object pictures of the final stimuli are presented in Appendix B (Table B1).[9]

To control for each item that the distractor objects of the scenes in the 1-consistent condition were truly rated as inconsistent with the given verb constraints (e.g., that children truly rated the "magnifier" as inconsistent argument of the verb "to eat", see Figure 4), we inspected the percentage of incorrect categorizations of each distractor object. We only

[9] At first, only 26 stimuli met our requirements. To retain a considerable number of stimuli, we conducted a second pretest with twelve stimuli that did not met our demands in the first pretest but only needed some small changes to improve their quality (*n* = 4 replacing the verb, *n* = 10 replacing the picture). The second pretest was run 2019 at Georg-August University of Göttingen with 15 illiterate German native speakers (*M* = 4.93 years, *SD* = 0.57, *range* = 4.06–5.91 years, 5 boys and 10 girls). Parents gave informed consent. The colored visual scenes were presented in paper form. Children worked on the naming and categorization task. As the test contained only one stimulus list, all rating scores based on the answers of all 15 children. For six stimuli of the second pretest the ratings met the requirements, resulting in a final set of 32 stimuli.

included items into our final set of stimuli for which at most 30% of the children incorrectly rated one or more of the three distractors as plausible argument of the corresponding verb. For instance, given the stimulus shown in Figure 4, we allowed at most 30% of the children to classify the “magnifier”, “lamp”, or “toilet” as “edible”. For the final 32 stimuli, only 14 of all 96 distractors were mismatched with a verb at all (*mean percentage of incorrect categorizations of the 96 distractors* = 1.06, *SD* = 3.31, *range* = 0–20).

To control how familiar our final set of stimuli was in general to young children we also inspected children’s name agreement scores. The scores for all object pictures of the final stimuli are presented in Appendix B (Table B1). The name agreement scores of the 32 target objects (e.g., “waffle” in Figure 4) were at 88.56% on average (*SD* = 15.04, *range* = 50–100). This does not only provide evidence that young children recognized the target objects in the visual scenes as intended but also that most young children should be familiar with the final words of the 32 item sentences (e.g., “waffle” in the sentence “The father eats the waffle”). Twenty-nine target objects were named as intended by at least 70% of the children. Three target objects were named as intended by only 50% of the children, but those were still correctly classified as plausible verb arguments by 100% of the children. For 28 of the 32 target words of the sentences we could verify child familiarity with age of acquisition norms provided by Birchenough et al. (2017) and Schröder et al. (2012). This was not possible for the four remaining target words as there are no corresponding age of acquisition norms.

The name agreement scores of the 96 competitor objects (e.g., pizza, sausage, and pretzel in Figure 4) were at 84.56% on average (*SD* = 20.64, *range* = 15–100). This provides evidence that, overall, also the competitor objects were of great familiarity to young children and that their pictorial appearance elicited the intended object representations. Fifteen competitors were named as intended by less than 70% of the children. However, those competitors were still categorized correctly as arguments of the corresponding verbs by at least 85% of the children.

2.4 Final Set of Stimuli

The final set of stimuli for Experiments 1 and 2 consisted of 32 items, each made up of a sentence with a semantically constraining verb that was presented in combination with a visual scene of four object pictures in four visual conditions (0-, 1-, 3-, and 4-consistent). This resulted in a total of 128 visual scenes. See Table 1 for a list of the final item sentences and Appendix B (Table B2) for all final visual scenes.

Table 1

Final Set of the Sentences

Nr.	German sentence	English translation
1	Der Vater **verschlingt** die **Waffel**.	The father **eats** the **waffle**.
2	Die Mutter **putzt** die **Lupe**.	The mother **cleans** the **magnifier**.
3	Der Onkel **schneidet** die **Banane**.	The uncle **cuts** the **banana**.
4	Die Tante **repariert** die **Lampe**.	The aunt **repairs** the **lamp**.
5	Der Enkel **probiert** die **Erdbeere**.	The grandson **tastes** the **strawberry**.
6	Die Enkelin **befüllt** die **Gießkanne**.	The granddaughter **fills** the **ewer**.
7	Der Vater **nascht** die **Himbeere**	The father **nibbles** the **raspberry**.
8	Die Mutter **entzündet** die **Rakete**.	The mother **ignites** the **rocket**.
9	Der Großvater **gießt** die **Tomate**.	The grandfather **waters** the **tomato**.
10	Die Großmutter **spielt** die **Trompete**.	The grandmother **plays** the **trumped**.
11	Der Bruder **hört** die **Gitarre**.	The brother **hears** the **guitar**.
12	Die Schwester **kaut** die **Pizza**.	The sister **chews** the **pizza**.
13	Der Bruder **schließt** die **Tasche**.	The brother **closes** the **case**.
14	Die Schwester **pflückt** die **Kirsche**.	The sister **grabs** the **cherry**.
15	Der Mann **verschüttet** den **Sprudel**.	The man **spills** the **soda**.
16	Die Frau **startet** den **Computer**.	The woman **starts** the **computer**.
17	Der Opa **backt** den **Kuchen**.	The grandpa **bakes** the **cake**.
18	Die Oma **bremst** den **Roller**.	The grandma **brakes** the **scooter**.
19	Der Mann **erntet** den **Salat**.	The man **harvests** the **salad**.
20	Die Frau **näht** den **Handschuh**.	The woman **sews** the **glove**.
21	Der Onkel **sammelt** den **Tannenzapfen**.	The uncle **picks** the **fir cone**.
22	Die Tante **trinkt** den **Kakao**.	The aunt **drinks** the **cocoa**.
23	Der Großvater **zerbricht** den **Pokal**.	The grandfather **breaks** the **trophy**.
24	Die Großmutter **strickt** den **Schal**.	The grandmother **knits** the **scarf**.
25	Der Vater **wirft** den **Ball**.	The father **throws** the **ball**.
26	Die Mutter **parkt** den **Bus**.	The mother **parks** the **bus**.
27	Der Opa **genießt** das **Ei**.	The grandpa **enjoys** the **egg**.
28	Die Oma **wäscht** das **Glas**.	The grandma **washes** the **glass**.
29	Der Mann **futtert** das **Bonbon**.	The man **guzzles** the **drop**.
30	Die Frau **fährt** das **Skateboard**.	The woman **drives** the **skateboard**.
31	Der Onkel **baut** das **Vogelhaus**.	The uncle **builds** the **bird house**.
32	Die Tante **bügelt** das **T-Shirt**.	The aunt **irons** the **T-shirt**.

Note. List of all item sentences used in Experiments 1 and 2 with the semantically constraining verbs and the target nouns in bold and the item number in the first column. For purposes of comprehension, the approximate English translations are presented.

3 Experiment 1: Prediction of Auditory Sentence Input in the Complex Visual World

3.1 Rationale & Design

One main goal of Experiment 1 was to examine children's and adults' prediction behavior in the more complex visual world. Experiment 1 involved an eye-tracking task in combination with a Visual World Paradigm. We presented young children (5–6 years) and adults with visual scenes of four object pictures each while they listened to German sentences with semantically constraining verbs and predictable target nouns (e.g., "Der Vater *verschlingt* gleich die Waffel"). Across four conditions (0-, 1-, 3-, and 4-consistent), the scenes varied in predictability in such a way that either 0, 1, 3, or 4 of the visual objects onscreen were consistent with the semantic verb constraints of the sentences and thus considered as visual prediction options. We assessed participants' fixations of the visual objects across the time course of the sentences because they can provide answers to the question of whether children and adults follow a multiple predictions pattern. To test whether children's prediction behavior in the complex visual world is influenced by their language experience, children's anticipatory object fixations during verb presentation were related with their performance in a test of receptive vocabulary size. To control for the cognitive abilities of our sample participants worked on a series of psychometric tests.[10]

[10] Parts of this chapter were copied or adapted from a published manuscript (Sommerfeld et al., 2023).

3.2 Research Questions and Hypotheses

3.2.1 Hypothesis 1: Children and Adults Follow a Multiple Predictions Pattern

First, we examined whether not only adults, but also young children can follow a multiple predictions pattern in the complex visual world. For adults, we expected that they can maintain multiple prediction options during visually situated prediction as this has already been shown by prior research (Ankener et al., 2018; Sikos et al., 2021). Since young children anticipatorily fixate visual referents consistent with the linguistic constraints, given both a single consistent visual referent (e.g., Mani & Huettig, 2012, 2014; Nation et al., 2003) or even two consistent visual referents (Borovsky et al., 2012; Gambi et al., 2021; Mani et al., 2016), we expected that young children would similar to adults anticipatorily fixate multiple visual referents that are semantically consistent with the linguistic context. In other words, we hypothesized that even young children can maintain multiple prediction options in parallel when processing language. Alternatively, multiple suitable referents may challenge children's capabilities to come up with prediction and they might not be able to maintain multiple prediction options. Since children have limited working memory capacity relative to adults (e.g., Cowan et al., 2010; Johnson et al., 2014), they could be less capable to pre-update multiple prediction options in working memory. As a result, they could pick out only one visual referent as potential prediction option, or none at all. We expected the following results pattern.

(a) In each of the four visual conditions, both age groups should fixate all four objects to the same extent upon hearing the agent of the sentences (e.g., "The father"). This is because the unconstraining agent did not allow for a discrimination of the visual scene.

(b) Upon hearing the constraining verb (e.g., "eat"), both age groups were expected to fixate the target object (e.g., waffle) more often than all other objects in the highly predictive 1-consistent condition where only the target object was consistent with the verb constraints.

This would reflect the classic prediction effect often reported for children (e.g., Mani & Huettig, 2014) and adults (e.g., Altmann & Kamide, 1999).

(c) In the less predictive 3- and 4-consistent conditions, where either three (pizza, sausage, waffle) or four (pretzel, pizza, sausage, waffle) visual objects were consistent with the verb constraints, we expected both age groups to show fewer anticipatory fixations of the target object (e.g., waffle) than in the 1-consistent condition. This is because here, listeners should also fixate some of the competitors, suggesting that they can integrate multiple visual prediction options into anticipatory sentence processing.

(d) To further examine whether children's and adults' prediction behavior is related to the exact number of visual prediction options, we compared the 3- and 4-consistent conditions. Upon hearing the verb, both age groups may show more target fixations in the 3-consistent versus the 4-consistent condition. This is because anticipatory fixations in the 3-consistent condition may fall on all three prediction options (and rarely on the single distractor), while those in the 4-consistent condition may be distributed among all four prediction options. A results pattern like this would suggest that children and adults even adapt their prediction behavior to the exact number of visual prediction options.

(e) Finally, when hearing the target word (e.g., "waffle"), we expected children and adults to fixate the target object (e.g., waffle) as the single visual referent of the target word more often than all other objects in the 1-, 3-, and 4-consistent conditions. This is because listeners typically guide their eyes to that object in a visual scene that refers to a word in the input (Allopenna et al., 1998; Cooper, 1974).

(f) In the 0-consistent control condition, where none of the visual objects was consistent with the verb constraints or the target word, both age groups were expected to similarly fixate all four distractors across the whole sentence.

3.2.2 Hypothesis 2: Prediction in Children Depends on Language Experience

We next examined whether children's language experience, indexed by their receptive vocabulary size, is associated with their prediction behavior in complex visual contexts. In line with previous studies reporting a positive relation of language skills and the use of prediction (e.g., Borovsky & Creel, 2014; Borovsky et al., 2012), we expected children's receptive vocabulary size to be positively associated with their prediction behavior also in our complex visual scenes. Hence, larger receptive vocabulary size in children should go along with more anticipatory fixations to single or multiple visual prediction options.

3.3 Methods

3.3.1 Participants

The sample consisted of *n* = 26 children (*M* = 5.74 years, *SD* = 0.51, *range* = 4.80–6.70 years, 13 boys and 13 girls) and *n* = 37 adults (*M* = 24.19 years, *SD* = 4.53, *range* = 19–40 years, 15 men and 22 women). Data of four additional children were excluded due to problems with the eye-tracker calibration. Parents and adult participants filled in a form about their (children's) age, gender, mother tongue, and vision or hearing impairment. All participants were German native speakers with reported (corrected-to) normal vision and hearing. For children, we used the form to ensure that they were not literate and asked for problems with language comprehension (*n* = 0) or general development (*n* = 5 sigmatism). All children attended kindergarten (and not yet school).[11] Adults were asked about their highest academic degree (*n* = 32 Abitur, *n* = 3 Bachelor's degree, *n* = 2 Master's degree), current employment (*n* = 33 student, *n* = 2 scientist, *n* = 1 caregiver, *n* = 1 no value), and whether they had some challenges with language comprehension that could affect their

[11] As in similar studies on prediction in the visual world (e.g., Andreu et al., 2013; Borovsky & Creel, 2014; Borovsky et al., 2012; Reuter et al., 2020) children were aged 5 to 6 years to ensure that they know the meaning of all verbs and recognize all visual objects of the eye-tracking task while being illiterate, thus differing in language experience from the adult sample.

outcome in the study (*n* = 0). Participants were recruited via newspapers and flyers. All parents and adult participants gave informed consent and received 10 Euro as compensation. The study was conducted at Saarland University in 2019 and 2020.

3.3.2 Cognitive Tests

We applied two standardized tests to control for the cognitive abilities of our sample (for a description of the tests, see Appendix C). In the Semantic Verbal Fluency Task, a test of cognitive functioning (e.g., Tröger et al., 2019), children's performance (*M* = 10.24, *SD* = 4.72, *range* = 1–22) was comparable to other typically developing children of the same age (Prigatano et al., 2008; Tallberg et al., 2011; van der Elst et al., 2011). Adults showed slightly higher verbal fluency (*M* = 26.12, *SD* = 4.92, *range* = 11–38) than in other studies (Martins et al., 2007; Rosselli et al., 2002; Troyer et al., 1997; Zimmermann et al., 2014) and performed significantly better than the child sample as revealed by a Welch *t*-test for independent samples with unequal variances and sample sizes ($t(46.69) = 12.17$, $p < .001$, *CI* [13.59, 18.98]). In the Color Naming Task, a child-friendly test of processing speed (Karbach et al., 2011; Kray et al., 2006; Vergilova et al., 2021), children (*M* = 28.83, *SD* = 7.02, *range* = 12–40) performed in line with other samples of young children (Karbach et al., 2011; Kray et al., 2006) but significantly worse than adults (*M* = 70.63, *SD* = 9.16, *range* = 46–84) as verified by another Welch *t*-test ($t(59.65) = 19.99$, $p < .001$, *CI* [38.09, 46.57]). Both of this suggests that the cognitive capacity of our child sample was comparable to other typically developing young children participating in empirical studies but smaller than those of the adults.

3.3.3 Experimental Task

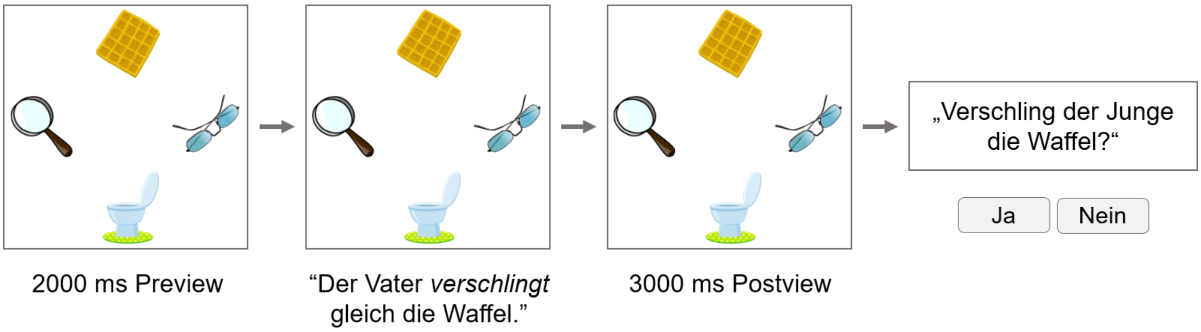
The eye-tracking task was programmed with Experiment Builder (SR Research, 2019a). Binocular data were recorded at 500 Hz via EyeLink 1000+ in remote mode. As recommended for remote mode, the eye-tracker was positioned below the experimental computer (26.77 inch) at a distance of 50 cm from the participants (SR Research, 2017). This distance slightly varied due to participants' body movements as usual for remote eye-tracking mode. Before testing, participants' gaze was calibrated in a nine-point calibration procedure in which an attention-getter appeared in every position of a 3 x 3 grid of calibration points.

The task took about 20 minutes. The trial procedure is shown in Figure 5. Participants sat in front of the computer and were instructed to look at all four objects of a visual scene for 2000 ms.[12] Then, a pre-recorded sentence (e.g., "Der Vater *verschlingt* gleich die Waffel") was played auditorily (M_{length} = 5401 ms). The visual scene remained on the screen during the sentence and for a postview of 3000 ms. Each trial was followed by a comprehension question referring to the sentence's subject, verb, or object (e.g., "Verschlingt der *Junge* die Waffel?") or to the visual scene (e.g., "War die Waffel *oben* zu sehen?"). Questions were correctly answered with either a "yes" or "no" response in 50% of cases each. Adults read the questions on the screen and answered via button press on a button box. Children were presented auditorily with the pre-recorded questions and responded verbally by saying "yes" or "no" to avoid potential issues related to the button press response at this age. The experimenter gave the button press response for them using the button box. Participants could take a short break after each trial. The next trial started after a one-point calibration.

[12] The preview time was about twice as long as in prediction studies with adults (e.g., Ankener et al., 2018; Sikos et al., 2018), but comparable to preview times in studies with children (e.g., Brouwer et al., 2018; Prescott et al., 2022). This was because we planned the preview time based on the time young children might need to encode the objects and their locations in the complex visual scenes (cf. Reuter et al., 2020).

Figure 5

Trial Procedure of the Eye-Tracking Task



Note. Example of a trial: Preview of the scene, followed by the auditorily presented sentence, the postview, and the question.

The task consisted of the 32 items validated with the pretest (see Chapter 2). Each item was made up of a semantically constraining sentence presented together with a visual scene in one out of four different visual conditions (0-, 1-, 3-, and 4-consistent). The conditions varied in such a way that either 0, 1, 3, or 4 objects of the scene were suitable arguments of the semantically constraining verb of the sentence (see Figure 4, page 51).

The visual scenes were identical to those selected as final visual stimuli in Chapter 2. They are shown in Appendix B (Table B2). The 32 sentences were slightly adapted from those selected as final linguistic stimuli in Chapter 2 (see Table 1). While no changes were made to the agents, verbs, articles, and target nouns, we inserted the spill-over word "gleich" after the verb of each sentence (e.g., "Der Vater *verschlingt* gleich die Waffel"). This did not add substantial information to the sentences or mismatch their content but allowed us to account for spill-over effects after the verb. Like in other eye-tracking studies on prediction, the spill-over region consisted of a single word (e.g., Andreu et al., 2013; Ankener et al., 2018; Mani & Huettig, 2012). No spill-over word was inserted after the target noun because the postview of the visual scene accounted for spill-over effects of the noun. The sentences were recorded by a female German native speaker slowly and in a voice appropriate for children with Audacity (Audacity Team, 2019). See Appendix A (Table A1) for a list of all item sentences.

We used a latin square design to ensure that each of the 32 sentences was presented in each of the four visual conditions (0-, 1-, 3-, and 4-consistent) while no participant should experience a sentence in more than one condition. This resulted in four different lists of 32 items, each with eight items per condition. Participants were randomly assigned to one list. To mask the study design, we also presented eight filler trials that introduced variation to the visual stimuli as two objects were consistent with the verb constraints for each filler (for the filler sentences, see Appendix A, Table A1; for the filler scenes, see Appendix B, Table B3). The order of presentation of trials was randomized. Finally, we included three practice trials at the beginning of the task. The filler and practice trials were the same for all participants.

3.3.4 Vocabulary Assessment

We applied the Peabody Picture Vocabulary Test (Lenhard et al., 2015) that measures receptive vocabulary size (e.g., Borovsky & Creel, 2014; Borovsky et al., 2012; Lenhard et al., 2015; Vergilova et al., 2022). Participants were presented with a sheet of four colored pictures of nouns, verbs, or adjectives. The experimenter vocalized the label of one of these nouns, verbs, or adjectives and asked participants to point at the picture representing the named item. There was no time limit and participants could ask for the label to be repeated. They received no feedback (aside from four practice trials) and were asked to guess the answer if necessary. The experimenter noted the answer and participants could take a short break before the next trial began. The test consisted of 19 trial sets, each of which consisted of 12 trials. The trial sets were arranged in order of ascending difficulty and the participant’s age determined which set the test began with. If participants finished the last set or made eight mistakes or more within a set, the test was completed. The test lasted about 15 minutes.

As in comparable studies (e.g., Borovsky & Creel, 2014; Vergilova et al., 2022) we extracted the raw test score that was calculated as suggested in the manual: The number of the last trial minus the total number of mistakes. Higher scores indicate higher receptive vocabulary size.

3.3.5 Procedure

Each participant was tested alone in a one-hour session. After the consent form and questionnaire were filled in, participants completed the Semantic Verbal Fluency Task and the Color Naming Task. To work on the experimental task, they moved to another desk with the eye-tracking setup. Here, they sat on a height-adjustable chair in front of the computer. The experimenter stuck a reference sticker on their forehead as required for remote eye-tracking (SR Research, 2017). Successful camera adjustment and calibration were followed by written instructions for adults and verbal instructions for children. After participants completed the practice trials and confirmed to understand the task, a divider was placed between participant and experimenter to avoid distraction. Then the experimental task began. Finally, participants worked on the Peabody Picture Vocabulary Task and were compensated.

3.3.6 Data Analysis

Data of all 26 children and 37 adults entered the analysis. Fixations were extracted with EyeLink Data Viewer (SR Research, 2019b). Since all objects were the same size (650 mm x 650 mm) and appeared in the same four positions onscreen (see Figure 4), we set up the same areas of interest for all visual scenes. We used only fixations that fell within these areas and excluded fixations shorter than 60 ms (cf. Mani et al., 2016). Since the eye-tracker provided an estimate of where participants looked at each time stamp during a trial, with one data point every 2 ms, we aggregated the data into 20-ms bins so that each 20-ms bin coded where participants fixated. As the onset of the different words varied across the items we aggregated the 20-ms bins in three time windows separately for each item. The *baseline* *window* included all fixations within 2000 ms before the onset of the verb. The *verb window* contained all fixations from 200 ms after verb onset until noun onset. The *noun window* included all fixations from 200 ms after noun onset until 2000 ms after noun onset. The delayed onset of the verb and noun window ensured that we only considered fixations that can reliably be attributed to the auditory input (cf. Mani & Huettig, 2014; Mani et al., 2016).

The dependent variable was the proportion of fixations to the target object relative to all other objects onscreen. This allowed us to show how fixations to the target and, due to the proportional character of this measure, to the other objects evolved across the trials. The target object was defined as the pictorial representation of the target noun of each sentence. As there was no target object shown onscreen in the 0-consistent condition, we specified one distractor of this condition each as a pseudo-target in a counterbalanced way (i.e., the position of the pseudo-target was rotated across participants and items). Thus, we could use the proportion of target fixations as dependent variable in the 0-consistent condition as well.

To reveal whether children and adults can adapt their fixation behavior to the exact number of visual prediction options we ran a control analysis in addition. We created a target advantage score (cf. Borovsky & Creel, 2014) which is defined as the difference between the proportion of fixations to the target minus each of the other objects in a scene. A score of 0 reflects an equal proportion of fixations to the target and the respective object. A positive score reflects a higher proportion of fixations to the target than the respective object. Hence, by comparing the target advantage scores of two objects we can determine whether they were fixated to a similar or different extent. We computed three target advantage scores for each condition. In the 0- and 1-consistent conditions, the scores were defined as difference of the fixations to the (pseudo-)target and the first, second, or third distractor. In the 3-consistent condition, the scores were computed as difference between fixations to the target and the first or second competitor, or the single distractor. In the 4-consistent condition, the scores were defined as difference between fixations to the target and the first, second, or third competitor.

3.4 Results

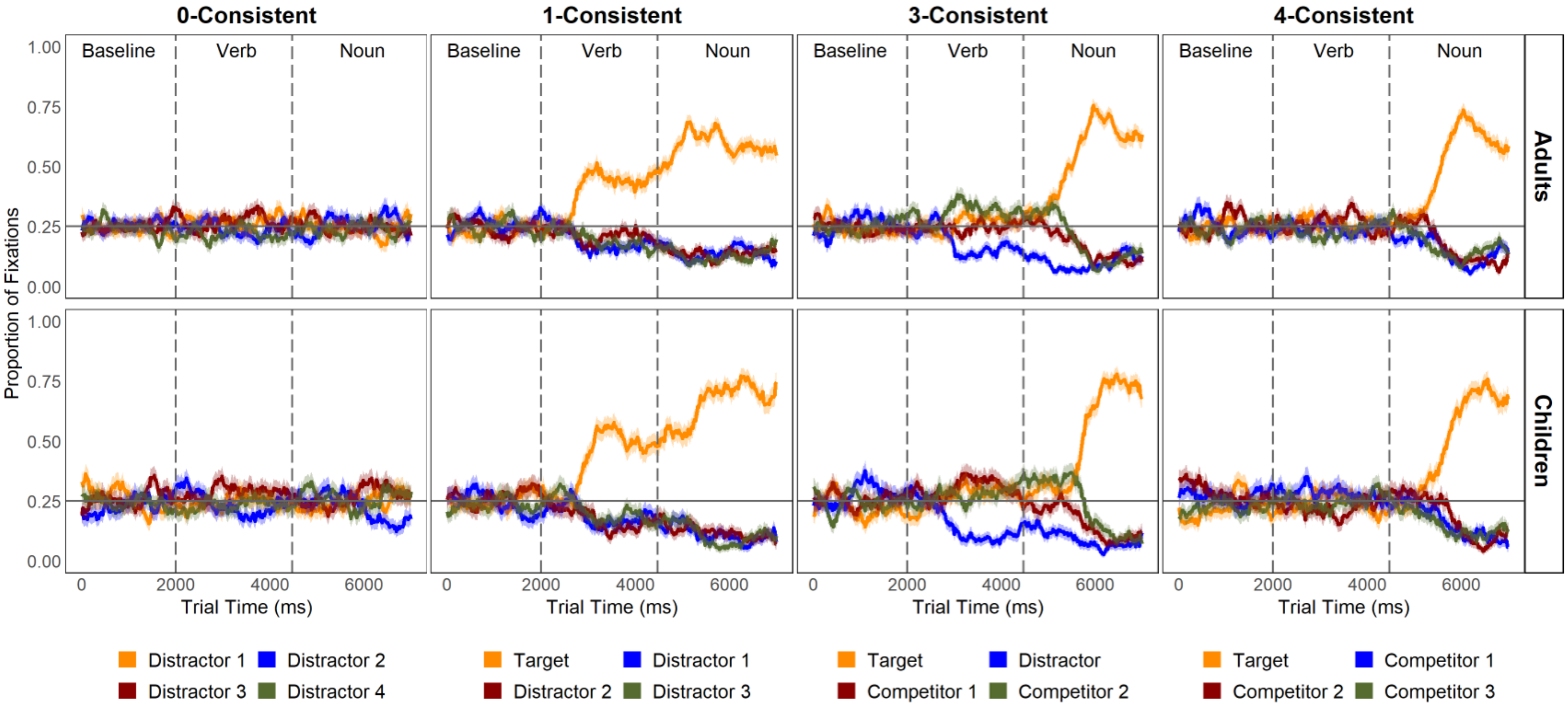
This section first explains why we used linear mixed effects models to analyze the data and briefly explains this statistical method. To provide an overview for the fixation data, we then describe the fixation pattern of both age groups across the different visual conditions and time windows (see Figure 6). Afterwards, we report the results of the analyses for the proportion of target fixations and the target advantage score. We also present two post-hoc analyses which ensure that a) the results do not derive from the fixation behavior of particular participants or items and b) that listeners truly fixated multiple prediction options prior to target noun presentation. Finally, we present the analysis of the relation between prediction and receptive vocabulary.

3.4.1 Use of Linear Mixed Effects Models

All data were analyzed using RStudio (RStudio Team, 2018). A significance criterion of $p < .05$ was applied and we report 95% confidence intervals of the mean. For all statistical analyses we applied linear mixed effects models, due to their advantage to consider fixed and random effects in the data at the same time (e.g., Baayen et al., 2008; Cunnings et al., 2012; Winter, 2013, 2020). These models do not only average data across participants or items when estimating the effects of the visual condition (0-, 1-, 3-, 4-consistent) on the fixation behavior (as it would be the case with more traditional analyses like ANOVA designs). Instead, linear mixed effects models allow for the analysis of random effects, thus they consider variability in the data deriving from different participants and items. The random intercepts in the models take into account that multiple observations of a single participant across different items are not independent from each other (e.g., the mean proportion of target fixations could be 70% for one participant and 50% for another participant). The similar logic can be applied to different items. Besides, the random slopes in the models account for the fact that different participants and items typically vary in their sensitivity to the manipulation (e.g., proportional target fixations could largely vary due to the visual condition for one participant but only slightly for another participant). Hence, by including random intercepts and random slopes for participants and items in our statistical models we could estimate any effect of the visual condition on the fixation behavior over and above the variability of the data across participants and items.

Figure 6

Proportion of Object Fixations Across the Averaged Trials



Note. Adults' and children's proportional fixations to the targets, distractors, and competitors in all conditions. The baseline is shown in the first 2000 ms, the verb window between the two dashed lines, and the noun window after the second dashed line. Error bars indicate the standard error of the mean.

3.4.2 Description of the Fixation Pattern

As visualized in Figure 6, there was few variation in the fixation pattern of children and adults across the visual conditions and time windows (for the numerical values, see Table 2). In the baseline of all visual conditions, both age groups fixated all four objects. In the verb window of the 1-consistent condition, fixations to the single target object increased while those to the three distractors decreased. In the verb window of the 3-consistent condition, fixations to the target object and the two competitors slightly increased while those to the single distractor decreased. In the verb window of the 4-consistent condition, fixations fell on the target object and on the three competitors. In the noun window of all conditions with a target object (1-, 3-, and 4-consistent), fixations to the target object increased while those to all other objects decreased. In the 0-consistent condition, both age groups‘ fixations fell at all four distractors in all time windows. Our statistical analyses confirm this descriptive pattern.

3.4.3 Analysis of the Proportion of Target Fixations

We conducted a linear mixed effects model (lme4 library, Bates et al., 2015) on the proportion of target fixations with the factors condition (0-, 1-, 3-, 4-consistent), time window (baseline, verb window, noun window), and age group (adults, children). The factors were effect coded in a planned structure. For the factor condition, we defined three contrasts of most theoretical interest. First, we compared the 0-consistent versus all other conditions. Second, we compared the 1-consistent versus the 3- and 4-consistent conditions. Finally, we compared the 3-consistent versus the 4-consistent condition. Time window contrasts were coded for two comparisons. First, we contrasted the baseline versus the verb window. Then, we contrasted the verb versus the noun window. The age group contrast compared children and adults. The contrasts were added to the model including their interaction terms.

To consider variability across participants (also named “subjects” in the statistical sections below) and items, the model structure included random intercepts for subjects and items. Moreover, we included by-subject random slopes for the factors window and condition as well as by-item random slopes for the factors window, condition, and age group. The interactions of the random slopes were also added. Non-converging models were simplified with the least-variance approach (Barr et al., 2013). We estimated *p*-values with the Satterthwaite degrees of freedom method (lmerTest library, Kuznetsova et al., 2017) and confidence intervals with the stats library (RStudio Team, 2018). For improved readability, we report only effects that are relevant for our research questions. Table 3 shows the model with all results.[13]

Baseline Versus Verb Window. There was an overall increase in proportional target fixations from the baseline to the verb window ($\beta = -.22$, $SE = .01$, $t(5525.88) = -23.21$, $p < .001$, *CI* [-.24, -.20]) which varied across conditions. The increase of proportional target fixations from the baseline to the verb window varied between the 0-consistent and all other conditions ($\beta = .30$, $SE = .02$, $t(5525.38) = 13.59$, $p < .001$, *CI* [.26, .34]), irrespective of the age group ($p = .185$). In the 0-consistent condition, there were no changes in both age groups' fixations to the pseudo-target from the baseline (25%) to the verb window (26%, see Table 2). There also was a difference in the increase of proportional target fixations from the baseline to the verb window among the 1-consistent versus the 3- and 4-consistent conditions ($\beta = -.12$, $SE = .02$, $t(5527.76) = -5.35$, $p < .001$, *CI* [-.17, -.08]), independent of the age group ($p = .271$). In the 1-consistent condition, both age groups' proportional target fixations increased from 25% in the baseline to 43% in the verb window (see Table 2). In the 3- and 4-consistent conditions, proportional target fixations changed from 24% and 23% in the baseline to 27% and 25% in the verb window. Independent of the age group ($p = .731$), this small change did not vary significantly among the 3- and 4-consistent conditions ($p = .442$).

[13] We also computed the same model using only those items whose comprehension questions were answered correctly (i.e., we excluded 2 items for adults and 72 items for children). This did not change the pattern of results. We therefore report the model including all cases.

Verb Versus Noun Window. There was an overall increase in proportional target fixations from the verb to the noun window (β = -.32, *SE* = .02, *t*(75.09) = -21.03, *p* < .001, *CI* [-.35, -.29]) that varied across conditions. Irrespective of the age group (*p* = .863), the proportional target fixations evolved significantly different from the verb to the noun window in the 0-consistent versus all other conditions (β = .44, *SE* = .02, *t*(5530.62) = 20.02, *p* < .001, *CI* [.40, .49]). In the 0-consistent condition, both age groups' proportional pseudo-target fixations did not change from the verb (26%) to the noun (24%) window (see Table 2). In all other conditions, proportional target fixations increased from the verb to the noun window (see Figure 6, orange line). Here, the 3- and 4-consistent conditions did not differ from each other (*p* = .667), irrespective of the age group (*p* = .331). Both age groups' proportional target fixations increased about 34% from the verb to the noun window in the 3-consistent (27% to 61%) and 4-consistent (25% to 58%) conditions (see Table 2). In the 1-consistent condition, contrasted to the 3- and 4-consistent conditions, the proportion of target fixations increased significantly less (19%) from the verb (43%) to the noun (62%) window (β = .07, *SE* = .02, *t*(5530.10) = 3.05, *p* < .001, *CI* [.03, .12]). This difference interacted with the factor age group (β = .11, *SE* = .05, *t*(5529.99) = 2.40, *p* = .017, *CI* [.02, .20]). Post-hoc linear mixed effects models with the same maximal random slope structure as in the last model separately for each age group (for the models with their results, see Table 4) showed that adults' proportional target fixations increased significantly more in the 3-consistent (34%) and 4-consistent (32%) conditions compared to the 1-consistent condition (15%) from the verb to the noun window (β = .13, *SE* = .03, *t*(3124.93) = 4.46, *p* < .001, *CI* [.07, .18]). Children's increase in target fixations from the verb to the noun window did not differ between the 3-consistent (32%) and 4-consistent (34%) conditions compared to the 1-consistent (24%) condition (*p* = .680).

Table 2

Averaged Proportion of Object Fixations

Condition	Group	Window	Object			
			Pseudo-target	Distractor 1	Distractor 2	Distractor 3
0-consistent	Adults	Baseline	.25 (.26)	.26 (.24)	.25 (.24)	.24 (.24)
		Verb	.27 (.22)	.24 (.22)	.26 (.22)	.23 (.21)
		Noun	.24 (.25)	.28 (.27)	.23 (.26)	.26 (.28)
	Children	Baseline	.25 (.26)	.23 (.24)	.27 (.26)	.25 (.26)
		Verb	.24 (.21)	.23 (.20)	.28 (.23)	.25 (.23)
		Noun	.25 (.27)	.21 (.23)	.29 (.27)	.25 (.25)
			Target	Distractor 1	Distractor 2	Distractor 3
1-consistent	Adults	Baseline	.25 (.26)	.26 (.25)	.25 (.25)	.25 (.26)
		Verb	.43 (.30)	.18 (.19)	.20 (.21)	.19 (.19)
		Noun	.58 (.36)	.15 (.23)	.15 (.24)	.12 (.20)
	Children	Baseline	.25 (.27)	.25 (.27)	.27 (.27)	.24 (.26)
		Verb	.44 (.29)	.19 (.19)	.17 (.19)	.20 (.20)
		Noun	.68 (.34)	.10 (.21)	.12 (.24)	.09 (.20)
			Target	Competitor 1	Competitor 2	Distractor
3-consistent	Adults	Baseline	.25 (.25)	.24 (.25)	.25 (.25)	.26 (.25)
		Verb	.27 (.23)	.25 (.21)	.31 (.25)	.17 (.20)
		Noun	.61 (.31)	.15 (.20)	.14 (.20)	.10 (.17)
	Children	Baseline	.23 (.25)	.25 (.27)	.24 (.24)	.28 (.28)
		Verb	.28 (.24)	.28 (.26)	.29 (.24)	.15 (.19)
		Noun	.60 (.30)	.15 (.21)	.18 (.23)	.08 (.13)
			Target	Competitor 1	Competitor 2	Competitor 3
4-consistent	Adults	Baseline	.24 (.23)	.24 (.24)	.27 (.26)	.25 (.25)
		Verb	.25 (.24)	.23 (.23)	.27 (.24)	.24 (.23)
		Noun	.57 (.31)	.14 (.18)	.14 (.19)	.15 (.21)
	Children	Baseline	.21 (.25)	.27 (.28)	.28 (.27)	.23 (.24)
		Verb	.24 (.24)	.27 (.25)	.24 (.24)	.25 (.24)
		Noun	.58 (.29)	.13 (.19)	.15 (.19)	.14 (.19)

Note. The averaged proportion of fixations to the (pseudo-)targets, distractors, and competitors are shown for each condition, time window, and age group. Standard deviations are presented in parentheses.

Table 3

Results of the Model on Both Age Groups' Proportional Target Fixations

Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Intercept	.35	.01	61.62	30.75	.000	.33, .37
v1	-.14	.02	28.37	-8.33	.000	-.17, -.11
v2	.07	.02	45.84	4.87	.000	.04, .10
v3	.02	.01	5557.74	2.47	.014	.00, .04
t1	-.22	.01	5525.88	-23.21	.000	-.24, -.20
t2	-.32	.02	75.09	-21.03	.000	-.35, -.29
a	.00	.02	59.30	0.15	.882	-.03, .03
v1:t1	.30	.02	5525.38	13.59	.000	.26, .34
v1:t2	.44	.02	5530.62	20.02	.000	.40, .49
v2:t1	-.12	.02	5527.76	-5.35	.000	-.17, -.08
v2:t2	.07	.02	5530.10	3.05	.002	.03, .12
v3:t1	-.02	.03	5524.40	-0.77	.442	-.07, .03
v3:t2	-.01	.03	5526.68	-0.43	.667	-.06, .04
v1:a	.02	.02	5549.42	1.34	.181	-.01, .05
v2:a	-.05	.02	56.46	-1.97	.053	-.09, .00
v3:a	.00	.02	5551.95	-0.10	.922	-.04, .04
t1:a	.03	.02	5526.22	1.56	.119	-.01, .07
t2:a	.05	.03	75.08	1.55	.125	-.01, .11
v1:t1:a	-.06	.04	5525.57	-1.33	.185	-.15, .03
v1:t2:a	-.01	.04	5529.84	-0.17	.863	-.09, .08
v2:t1:a	.05	.05	5528.07	1.10	.271	-.04, .14
v2:t2:a	.11	.05	5529.99	2.40	.017	.02, .20
v3:t1:a	-.02	.05	5524.41	-0.34	.731	-.12, .09
v3:t2:a	-.05	.05	5526.61	-0.97	.331	-.16, .05

Note. The model on the proportional target fixations (tf) covered the factors condition, time window, and age group. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second window contrasts. a is the age group contrast. The converged model:

```
lmer(tf(v1+v2+v3)*(t1+t2)*a+(1+(v2)*(t2)||participant)+(1+(v1+v2)||item))
```

Table 4

Results of the Models on the Proportional Target Fixations per Age Group

Age group	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Adults	Intercept	.35	.01	51.67	27.04	.000	.32, .38
	v1	-.13	.02	45.88	-5.80	.000	-.17, -.08
	v2	.05	.01	27.08	3.61	.001	.02, .08
	v3	.02	.01	3124.31	1.97	.049	.00, .05
	t1	-.21	.02	35.43	-10.50	.000	-.25, -.17
	t2	-.30	.03	35.23	-10.95	.000	-.35, -.24
	v1:t1	.27	.04	34.23	6.65	.000	.19, .35
	v1:t2	.44	.05	33.42	8.83	.000	.34, .53
	v2:t1	-.10	.03	3122.58	-3.54	.000	-.16, -.04
	v2:t2	.13	.03	3124.93	4.46	.000	.07, .18
	v3:t1	-.03	.03	3121.06	-0.97	.332	-.10, .03
	v3:t2	-.04	.03	3121.75	-1.25	.213	-.11, .02
Children	Intercept	.35	.01	38.23	25.47	.000	.32, .38
	v1	-.15	.02	30.94	-6.88	.000	-.19, -.11
	v2	.10	.02	29.82	4.37	.000	.05, .14
	v3	.03	.01	2208.74	1.76	.078	.00, .06
	t1	-.24	.01	2159.79	-16.05	.000	-.27, -.21
	t2	-.35	.01	2160.06	-23.34	.000	-.37, -.32
	v1:t1	.33	.03	2161.27	9.65	.000	.26, .40
	v1:t2	.45	.03	2159.99	13.15	.000	.38, .52
	v2:t1	-.15	.04	2159.61	-4.19	.000	.32, .38
	v2:t2	.01	.04	2161.54	0.41	.680	-.17, -.08
	v3:t1	-.01	.04	2158.11	-0.26	.797	.02, .08
	v3:t2	.02	.04	2157.97	0.39	.694	.00, .05

Note. The model on the proportional target fixations (tf) per age group covered the factors condition and time window. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second window contrasts. The converged models:

Adults: lmer(tf~(v1+v2+v3)*(t1+t2)+(1+(v1)*(t1+t2)||subject)+(1+(v1+v2)||item))
Children: lmer(tf~(v1+v2+v3)*(t1+t2)+(1+(v1+v2)||subject)+(1+(v1+v2)||item))

In sum, there were no differences in children’s and adults’ anticipatory fixations of the target object. In the 1-, 3-, and 4-consistent conditions, both age groups fixated the target object as soon as the verb was played. However, they revealed more anticipatory target fixations in the 1-consistent than in the 3- and 4-consistent conditions. Given the proportional character of the dependent variable, this suggests that children and adults must have fixated also some other objects upon hearing the verb in the 3- and 4-consistent conditions (for an analysis of which objects other than the target they fixated, see the next section).

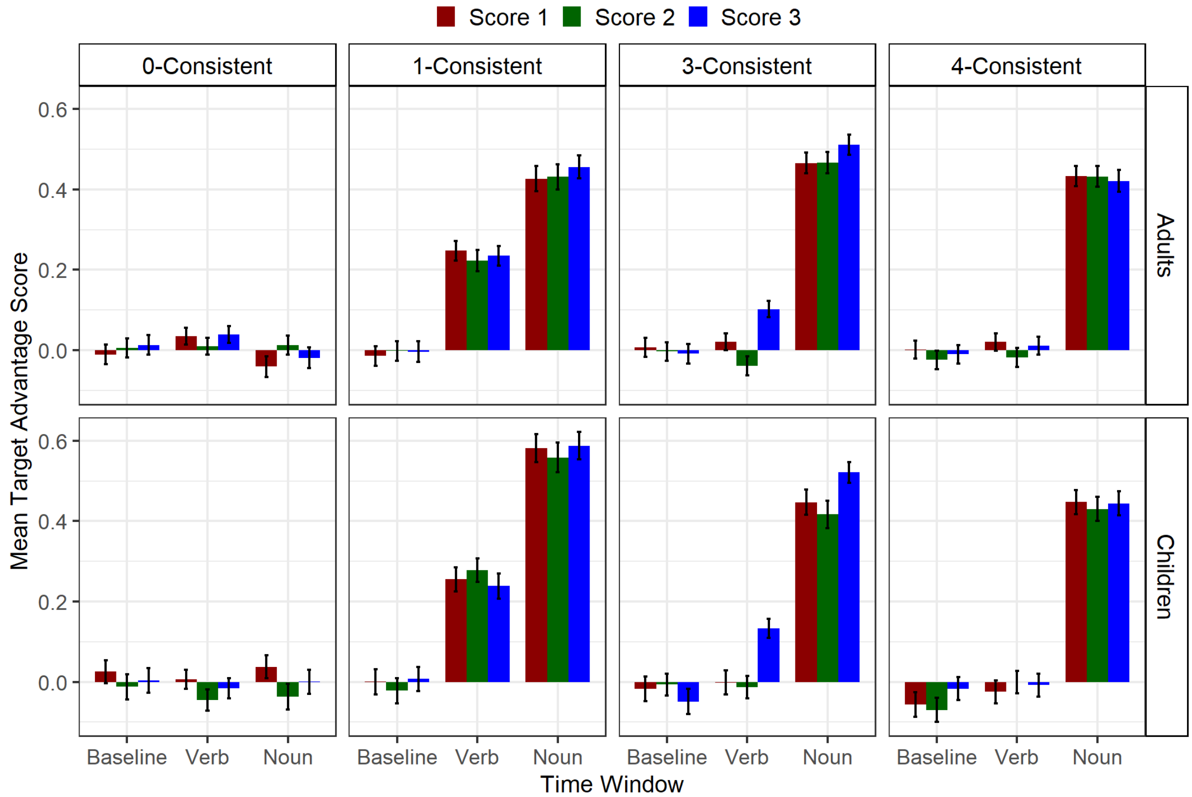
3.4.4 Analysis of the Target Advantage Score

To reveal whether distractors and/or competitors were fixated to a similar or different extent across the trials, i.e., whether participants adapted their prediction behavior to the exact number of visual prediction options, we analyzed the target advantage score. Figure 7 shows the averaged target advantage scores of both age groups across all conditions and time windows (for the numerical data, see Appendix D). A small target advantage score means that the target and the respective object were fixated to a similar extent. A positive score means that the target was fixated more often than the respective object. A different score between the objects means that they were fixated to a different extent.

We ran one linear mixed effects model on the target advantage score for each of the four conditions (0-, 1-, 3-, 4-consistent). The effect coded factors object, time window, and age group were included. Object contrasts were coded for two comparisons: We contrasted the first and the second object of a visual scene versus the third one. Then, we contrasted the first versus the second object. This was because in the 3-consistent condition the first and second object both were competitors while the third one was a distractor. Time window and age group contrasts were the same as in the analysis reported above. Contrasts were added to the models with their interaction terms. We included a) random intercepts for subjects and items and b) by-subject random slopes for the factors object and window as well as by-item random slopes for the factors object, window, and age group. The interactions of the random slopes were also added. Non-converging models were simplified with the least-variance approach (Barr et al., 2013).

Figure 7

Averaged Target Advantage Scores per Condition, Time Window, and Age Group



Note. In the 0- and 1-consistent conditions, scores 1 to 3 reflect the difference of fixations on the (pseudo-)target minus each distractor. In the 3-consistent condition, scores 1 and 2 reflect the difference of fixations on the target minus each competitor and score 3 (blue bar) reflects the difference of fixations on the target minus the distractor. In the 4-consistent condition, scores 1 to 3 reflect the difference of fixations on the target minus each competitor. Error bars indicate the standard error of the mean.

In the 0-, 1-, and 4-consistent conditions both age groups' target advantage scores did not differ between the three objects in all time windows ($p > .05$, for the models with their results, see Table 5). This means that all three distractors in the 0- and 1-consistent conditions and that all three competitors in the 4-consistent condition were fixated to a similar extent across the trials (see Figure 7).

In the 3-consistent condition, both age groups' target advantage scores did not vary among the two competitors in all time windows (*p* = .240). Thus, participants fixated the two competitors to a similar extent across the trials in the 3-consistent condition (for the model with its results, see Table 6). However, both age groups' score was significantly higher for the competitors than for the distractor (β = .06, *SE* = .01, *t*(4048.99) = 4.61, *p* < .001, *CI* [.03, .08]). This interacted with the time window contrast comparing the baseline and the verb window (β = -.16, *SE* = .03, *t*(4048.99) = -4.63, *p* < .001, *CI* [-.23, -.09]). Thus, we ran post-hoc linear mixed effects models with the same maximal random slope structure as in the last model separately for the baseline and the verb window (for the models with their results, see Table 6). In the baseline, the target advantage score of both age groups did not differ among the competitors and the distractor (*p* = .284). In the verb window, both age groups' score was significantly smaller for the competitors than for the distractor (β = .13, *SE* = .02, *t*(1336.70) = 6.28, *p* < .001, *CI* [.09, .16]). This means that children and adults fixated the competitors and the distractor to a similar extent in the baseline. In the verb window, in contrast, the two competitors were fixated more often than the distractor and as often as the target. This was because the target advantage scores of the competitors were close to zero while the score of the distractor was positive (see Figure 7).

In sum, the analysis of the target advantage score showed no age differences in how children and adults fixated potential visual referents of the sentence context. Although we found no difference in the proportion of target fixations between the 3- and 4-consistent conditions, the analysis of the target advantage score shows that both age groups adapted their prediction behavior to the number of visual prediction options: Upon hearing the verb they anticipatorily fixated the two competitors and the target object in the 3-consistent condition or all three competitors and the target object in the 4-consistent condition.

Table 5

Results of the Models on the Target Advantage Scores (0-, 1-, and 4-Consistent Conditions)

Condition	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
0-consistent	Intercept	.00	.02	55.88	-0.24	.811	-.04, .03
	o1	.00	.01	4159.47	0.38	.703	-.02, .03
	o2	-.02	.02	60.22	-1.12	.267	-.06, .02
	t1	.01	.02	4163.16	0.25	.800	-.03, .04
	t2	.02	.02	4163.02	0.94	.347	-.02, .06
	a	.03	.03	65.48	0.87	.386	-.03, .08
	o1:t1	.00	.03	4159.47	0.12	.906	-.06, .07
	o1:t2	.01	.03	4159.47	0.41	.680	-.05, .08
	o2:t1	.02	.04	4164.25	0.44	.660	-.06, .10
	o2:t2	-.02	.04	4164.73	-0.49	.627	-.10, .06
	o1:a	.01	.02	4159.47	0.36	.719	-.04, .06
	o2:a	.07	.04	60.22	1.93	.059	.00, .15
	t1:a	-.01	.04	4163.01	-0.24	.807	-.09, .07
	t2:a	.06	.04	4163.74	1.60	.110	-.01, .14
	o1:t1:a	.02	.07	4159.47	0.29	.771	-.12, .16
	o1:t2:a	.03	.07	4159.47	0.40	.688	-.11, .16
	o2:t1:a	-.03	.08	4164.25	-0.37	.711	-.19, .13
	o2:t2:a	-.12	.08	4164.73	-1.47	.142	-.27, .04
1-consistent	Intercept	.24	.03	67.07	8.14	.000	.18, .30
	o1	.01	.01	3977.66	0.52	.603	-.02, .03
	o2	-.01	.01	3977.66	-0.35	.725	-.03, .02
	t1	-.50	.04	79.22	-11.20	.000	-.58, -.41
	t2	-.52	.05	81.84	-9.96	.000	-.62, -.42
	a	-.04	.05	81.21	-0.83	.408	-.13, .05
	o1:t1	.01	.03	3977.66	0.27	.785	-.06, .08
	o1:t2	-.03	.04	3977.66	-0.89	.372	-.10, .04
	o2:t1	.00	.04	3977.66	0.02	.986	-.08, .08
	o2:t2	.01	.04	3977.66	0.22	.830	-.07, .09
	o1:a	.01	.02	3977.66	0.32	.751	-.04, .06
	o2:a	.00	.03	3977.66	0.17	.863	-.05, .06
	t1:a	.08	.09	74.09	0.93	.357	-.09, .25
	t2:a	.16	.10	79.16	1.64	.106	-.03, .35
	o1:t1:a	-.04	.07	3977.66	-0.61	.544	-.18, .09
	o1:t2:a	.00	.07	3977.66	-0.04	.971	-.14, .13
	o2:t1:a	.06	.08	3977.66	0.73	.464	-.10, .22
	o2:t2:a	-.04	.08	3977.66	-0.55	.579	-.20, .11

Note. Table continued on the next page.

Table 5 (Continued)

Results of the Models on the Target Advantage Scores (0-, 1-, and 4-Consistent Conditions)

Condition	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
4-consistent	Intercept	.13	.02	62.53	6.66	.000	.09, .17
	o1	.01	.01	3990.17	0.77	.442	-.01, .03
	o2	-.01	.01	3990.17	-0.84	.399	-.04, .02
	t1	-.32	.03	74.48	-1.46	.000	-.37, -.26
	t2	-.59	.04	76.56	-16.13	.000	-.66, -.52
	a	.01	.03	74.33	0.43	.670	-.05, .08
	o1:t1	.03	.03	3990.17	0.83	.404	-.04, .10
	o1:t2	.03	.03	3990.17	0.72	.470	-.04, .09
	o2:t1	-.02	.04	3990.17	-0.38	.703	-.09, .06
	o2:t2	-.01	.04	3990.17	-0.15	.878	-.08, .07
	o1:a	-.02	.02	3990.17	-0.79	.428	-.07, .03
	o2:a	-.02	.03	3990.17	-0.69	.494	-.07, .04
	t1:a	.03	.06	80.29	0.50	.621	-.08, .14
	t2:a	.05	.07	76.56	0.72	.475	-.09, .20
	o1:t1:a	-.05	.07	3990.17	-0.76	.446	-.19, .08
	o1:t2:a	.00	.07	3990.17	-0.07	.944	-.14, .13
	o2:t1:a	.02	.08	3990.17	0.19	.847	-.14, .17
	o2:t2:a	-.07	.08	3990.17	-0.90	.371	-.23, .09

Note. The models on the target advantage scores (tas) in the 0-, 1-, and 4-consistent conditions covered the factors object, time window, and age group. o1 and o2 are the first and second object contrasts. t1 and t2 are the first and second window contrasts. a is the age group contrast. The converged models:

```
0-consistent:   lmer(tas~(o1+o2)*(t1+t2)*a+(1+(o2)||subject)+(1+a||item))
1-consistent:   lmer(tas~(o1+o2)*(t1+t2)*a+(1+(t1+t2)||subject)+(1+(t1+t2)*a||item))
4-consistent:   lmer(tas~(o1+o2)*(t1+t2)*a+(1+(t1+t2)||subject)+(1+(t1)*a||item))
```

Table 6

Results of the Models on the Target Advantage Scores (3-Consistent Condition)

Region	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Overall	Intercept	.18	.02	66.49	10.84	.000	.15, .22
	o1	.06	.01	4048.99	4.61	.000	.03, .08
	o2	-.02	.01	4048.99	-1.17	.240	-.04, .01
	t1	-.40	.03	76.95	-12.36	.000	-.47, -.34
	t2	-.62	.04	83.08	-13.79	.000	-.71, -.53
	a	.01	.03	74.06	0.19	.850	-.05, .06
	o1:t1	-.16	.03	4048.99	-4.63	.000	-.23, -.09
	o1:t2	-.02	.03	4048.99	-0.64	.526	-.09, .05
	o2:t1	.03	.04	4048.99	0.86	.392	-.04, .11
	o2:t2	.00	.04	4048.99	-0.09	.927	-.08, .07
	o1:a	-.02	.02	4048.99	-0.64	.523	-.06, .03
	o2:a	-.01	.03	4048.99	-0.46	.647	-.07, .04
	t1:a	.04	.07	76.95	0.63	.530	-.09, .17
	t2:a	-.01	.08	73.42	-0.09	.927	-.17, .15
	o1:t1:a	.08	.07	4048.99	1.20	.231	-.05, .22
	o1:t2:a	.06	.07	4048.99	0.81	.419	-.08, .19
	o2:t1:a	-.02	.08	4048.99	-0.21	.835	-.17, .14
	o2:t2:a	-.09	.08	4048.99	-1.09	.277	-.24, .07
Baseline	Intercept	-.02	.02	52.44	-0.99	.325	-.06, .02
	o1	-.02	.02	1345.09	-1.07	.284	-.07, .02
	o2	.00	.03	1345.09	0.02	.985	-.05, .05
	a	.03	.04	52.81	0.66	.509	-.06, .11
	o1:a	.03	.04	1345.09	0.59	.558	-.06, .11
	o2:a	-.02	.05	1345.09	-0.42	.676	-.12, .08
Verb	Intercept	.08	.02	48.57	3.27	.002	.03, .12
	o1	.13	.02	1336.70	6.28	.000	.09, .16
	o2	-.04	.02	1336.70	-1.54	.124	-.08, .01
	a	-.02	.04	56.33	-0.57	.569	-.09, .05
	o1:a	-.03	.04	1336.70	-0.73	.467	-.11, .05
	o2:a	-.05	.05	1336.70	-1.05	.296	-.14, .04

Note. The models on the target advantage scores (tas) in the baseline and verb window of the 3-consistent condition covered the factors object and age group; the overall model covered the factor time window as well. o1 and o2 are the first and second object contrasts. t1 and t2 are the first and second window contrasts. a is the age group contrast. The converged models:

Overall: lmer(tas~(o1+o2)*(t1+t2)*a+(1+(t1+t2)||subject)+(1+(t2)*a||item))
Baseline: lmer(tas~(o1+o2)*a+(1|subject)+(1+a||item))
Verb: lmer(tas~(o1+o2)*a+(1|subject)+(1+a||item))

3.4.5 Post-Hoc Tests of Deviance

We conducted some tests of deviance for our statistical models (e.g., Winter, 2020). This was to ensure that the above results were robust against variability in the data stemming from different participants and items, and thus did not derive from averaging data across participants or items. Deviance tests can reveal whether there is specific variance among participants or items that could affect the influence of a fixed effect (e.g., visual condition) on a dependent variable (e.g., proportion of target fixations).

For the deviance tests regarding the proportion of target fixations we considered three linear mixed effects models. We first considered the “maximal” model with random effects for subjects and items (this is the model reported in Chapter 3.4.3). Second, the same model was run without random effects for subjects. Third, we ran the same model without random effects for items. We then compared the models with a likelihood ratio test (e.g., Winter, 2020). The “maximal” model differed significantly from the two other models in its fit with the underlying data ($\chi^2(6) = 279.01$, $p < .001$). The Akaike Information Criterion (AIC) was smaller for the “maximal” model than for the two other models (see Table 7). This means that the random effects “subject” and “item” contributed significantly to the model. Thus, there was particular variability in the influence of the visual condition of the proportion of target fixations that went back on subjects and items. Notably, we considered this variability as we included random effects for subjects and items into our reported model (e.g., Winter, 2020).

The same procedure was performed for the four models on the target advantage score (one per condition). For each condition we first considered the “maximal” model with random effects for subjects and items (these are the models reported in Chapter 3.4.4). Second, we ran the same model without random effects for subjects. Third, the same model was run without random effects for items. For each condition, the deviance tests showed the same: The “maximal” model each had a significant better fit with the data structure than the two other models (for the results of the deviance tests, see Table 7). Thus, there was particular variance in the influence of the visual condition on the target advantage score deriving from subjects and items. Notably, we considered this variance because we included random effects for subjects and items into our reported models (e.g., Winter, 2020).

Table 7

Results of the Deviance Tests

Measure	Condition	Model	AIC	Deviance test
Proportion of target fixations	All conditions	"Maximal"	891	$\chi^2(6) = 279.01$, $p < .001$
		No subject random effects	1158	
		No item random effects	1135	
Target advantage score	0-consistent	"Maximal"	4033	$\chi^2(2) = 93.82$, $p < .001$
		No subject random effects	4137	
		No item random effects	4123	
	1-consistent	"Maximal"	4576	$\chi^2(3) = 604.07$, $p < .001$
		No subject random effects	5174	
		No item random effects	4987	
	3-consistent	"Maximal"	4297	$\chi^2(3) = 219.09$, $p < .001$
		No subject random effects	4510	
		No item random effects	4403	
	4-consistent	"Maximal"	4165	$\chi^2(3) = 244.75$, $p < .001$
		No subject random effects	4404	
		No item random effects	4307	

Note. The column "Model" specifies whether the AIC values refer to the "maximal model" (with random effects for subjects and items) or to the model with only item or only subject random effects. The degrees of freedom of the chi-square tests vary depending on how the model structure was simplified with the least variance approach (Barr et al., 2013).

3.4.6 Post-Hoc Descriptive Data Analysis

To further underline our results, we determined manually the proportion of cases in which listeners fixated more than one/all visual prediction options in the verb window of the 3- and 4-consistent conditions. To do so, we first defined objects as "fixated" when at least 5% of all fixations of a given participant and item fell on them. Then, we extracted the number of cases for which more than one/all visual prediction options were fixated and related this to the total number cases. In the 3-consistent condition, adults anticipatorily fixated multiple (i.e., more than one) verb-consistent objects in 89% and children in 84% of all cases. In the 4-consistent condition, adults anticipatorily fixated multiple verb-consistent objects in 91% and children in 88% of all cases (see Table 8). Two chi-square tests showed that these percentages each did not differ between the age groups (*p*-values > .05).

Table 8

Exploratory Data Inspection

			Multiple objects		All consistent objects	
Group	Condition	Observations	Number	Percentage	Number	Percentage
	1-consistent	296	-	-	254	86
Adults	3-consistent	296	264	89	148	50
	4-consistent	296	269	91	107	36
	1-consistent	199	-	-	171	86
Children	3-consistent	199	168	84	93	47
	4-consistent	198	175	88	62	31

Note. Data address the verb window. “Observations” is the number of cases per condition across participants and items. The value 296 stems from 37 adults encountering 8 items per condition (37 x 8). For the 26 children, there should be 208 cases per condition (26 x 8), but for some cases children’s eye data could not be extracted. “Multiple objects” first shows the number of cases where more than one verb-consistent object was fixated. Next, this value is set in relation to the respective number of cases to get a percentage (e.g., 264 divided by 296 is 89). The cells of the 1-consistent condition are empty as here only the target object was consistent with the verb. “All consistent objects” shows the number and percentage of cases where all verb-consistent objects were fixated.

Second, we determined in how many cases listeners anticipatorily fixated all visual prediction options. In the verb window of the 1-consistent condition, the target object was fixated by 86% of the children and adults each. In the 3-consistent condition adults fixated all three visual prediction options (one target, two competitors) upon hearing the verb in 50% and children in 47% of all cases. In the 4-consistent condition adults anticipatorily fixated all four visual prediction options (one target, three competitors) in 36% and children in 31% of all cases (see Table 8). Three chi-square tests showed that these percentages each did not vary among the age groups (*p*-values > .05).

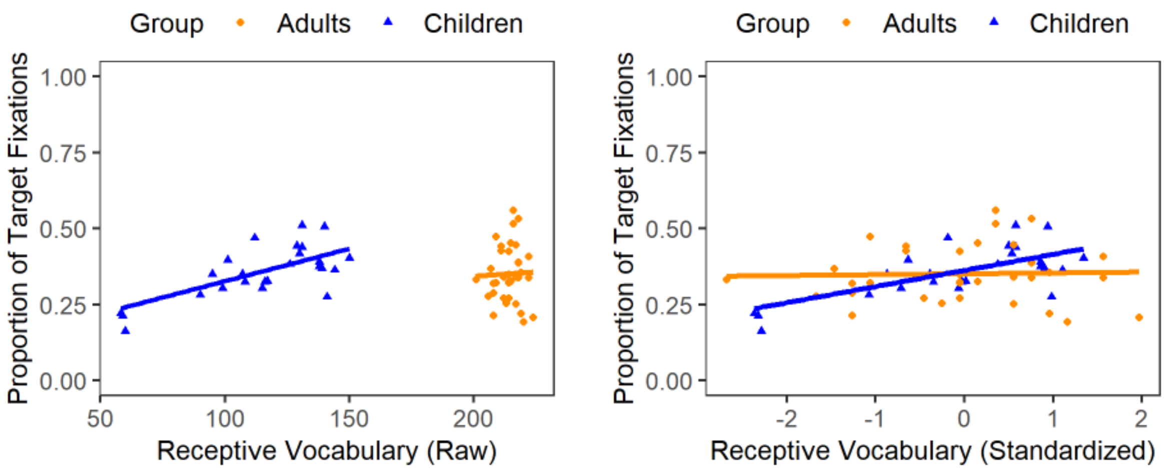
3.4.7 Influence of Vocabulary Size

Children's averaged raw score in the Peabody Picture Vocabulary Test (*M* = 116.62, *SD* = 24.86, *range* = 58–150) corresponded to a *T*-value of 52 (*CI* [47, 57]) as reported by the test manual for children aged 5.70 years (which was the mean age of our child sample). Thus, the receptive vocabulary size of our child sample was common for their age. Adults' raw scores in the Peabody Picture Vocabulary Test (*M* = 214.32, *SD* = 4.91, *range* = 201–224) corresponded to a *T*-value of 63 (*CI* [61, 65]) as reported by the manual for 17-year-olds (the German test version is only normed upon age 17). Thus, adults' receptive vocabulary size was above the norms of 17-year-olds. A Welch *t*-test showed that adults' raw scores in the test were significantly higher than those of the children ($t(25.19) = 18.39$, $p < .001$, *CI* [88.17, 110.41]. Figure 8 visualizes this variation. One child and one adult did not perform the test.

To analyze the interplay of prediction and receptive vocabulary size, the raw scores in the Peabody Picture Vocabulary Test were *z*-standardized separately for each age group. They were included in a linear mixed effects model on the proportion of target fixations in the verb window of the 1- and 3-consistent conditions. This is because more fixations to the target than to the other objects in the verb window directly reflect prediction and because only the 1- and 3-consistent scenes allowed for an anticipatory discrimination of the visual objects (in the 0- and 4-consistent scenes either all or none of the objects were in line with prediction). We added the effect coded factors condition (1-consistent versus 3-consistent), age group (adults versus children), and the continuous variable receptive vocabulary size with their interaction terms to the model. We also added random intercepts for subjects and items as well as by-subject random slopes for the factor condition and by-item random slopes for the factors condition, age group, and receptive vocabulary size (including their interactions). In case of a non-converging model, it was simplified with the least variance approach (Barr et al., 2013). The model and its results are shown in Table 9.

Figure 8

Relation of Anticipatory Target Fixations and Receptive Vocabulary Size



Note. Both age groups' proportional target fixations in the verb window of the 1- and 3-consistent conditions related to their raw (left side) and *z*-standardized (right side) Peabody Picture Vocabulary Test scores. Children's anticipatory fixations increased with increasing receptive vocabulary.

The model revealed a significant main effect of receptive vocabulary size (β = .03, *SE* = .01, *t*(61.11) = 2.76, *p* = .008, *CI* [.01, .05]) which was independent of the condition (*p* = .688) but interacted with the age group (β = -.05, *SE* = .02, *t*(63.06) = -2.42, *p* = .018, *CI* [-.09, -.01]). Model splits by age group (see Table 9) showed that children's anticipatory target fixations increased with increasing receptive vocabulary size (β = .05, *SE* = .01, *t*(26.72) = 4.14, *p* < .001, *CI* [.03, .08]). This was independent of the condition (*p* = .611) and not the case for adults (*p* > .05). Figure 8 visualizes this pattern of results.

Table 9

Results of the Models on the Relation of Prediction and Vocabulary Size

Age Group	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Overall	Intercept	.36	.02	40.08	21.16	.000	.32, .39
	v	.16	.02	37.46	6.87	.000	.11, .20
	a	-.01	.02	56.80	-0.42	.674	-.05, .03
	vocZ	.03	.01	61.11	2.76	.008	.01, .05
	v:a	-.01	.04	54.21	-0.18	.860	-.09, .07
	v:vocZ	.01	.02	58.74	0.40	.688	-.03, .05
	a:vocZ	-.05	.02	63.06	-2.42	.018	-.09, -.01
	v:a:vocZ	-.01	.04	61.04	-0.31	.757	-.09, .07
Adults	Intercept	.35	.02	39.24	17.62	.000	.31, .39
	v	.16	.03	29.69	5.09	.000	.10, .22
	vocZ	.00	.02	34.73	0.23	.823	-.03, .03
	v:vocZ	.00	.03	33.80	0.05	.960	-.05, .06
Children	Intercept	.36	.02	22.65	19.81	.000	.33, .40
	v	.16	.03	15.57	5.54	.000	.11, .22
	vocZ	.05	.01	26.72	4.14	.000	.03, .08
	v:vocZ	.01	.03	25.64	0.51	.611	-.04, .07

Note. The models on the proportional target fixations (tf) in the verb window per age group covered the factors condition and vocabulary size; the overall model covered the factor age group as well. c is the condition contrast. a is the age group contrast. vocZ are the z-standardized scores in the test of vocabulary size. The converged models:

Overall: lmer(tf~c*a*vocZ+(1+c||subject)+(1+c||item))
Adults: lmer(tf~c*vocZ +(1+c||subject)+(1+c||item))
Children: lmer(tf~c*vocZ +(1+c||subject)+(1+c||item))

3.5 Discussion

The main goals of this part of Experiment 1 were to examine age differences in predicting sentence continuations based on semantically constraining verbs in complex visual environments and whether the use of prediction is associated with individual differences in vocabulary size. We used eye-tracking in combination with a Visual World Paradigm and presented 5- to 6-year-olds and adults with sentences consisting of semantically constraining verbs, while they were looking at complex visual scenes showing four objects of which either 0, 1, 3, or 4 were consistent with the sentence context. Moreover, participants completed the Peabody Picture Vocabulary Test as a measure of receptive vocabulary size. Comparing children and adults in their fixation pattern across the experimental conditions, we found overall similarity but also some differences in the increase and decrease of fixations to the different objects across the sentences. Our results revealed four noteworthy findings that we discuss in more detail.

3.5.1 No Age Differences in Predictive Processing in Complex Visual Scenes

The first finding is that children and adults predict sentence input even in complex visual contexts. This claim is supported by several results of our analysis. Considering first the predictive 1-, 3-, and 4-consistent conditions in which either one, three, or four visual objects were consistent with the sentence context. Here, children and adults first fixated all four objects during the unpredictive beginning of the sentences (baseline). Upon hearing the constraining verb, they fixated the target object most frequently in the 1-consistent condition (while fixations to all three distractors decreased). In the 3- and 4-consistent conditions, both age groups fixated the target object less often than in the 1-consistent condition. This was because here they also fixated some of the other objects being consistent with the verb constraints (fixations to the single distractor decreased in the 3-consistent condition).

For the 1-consistent condition presenting one target next to multiple distractors our finding is in line with the results of numerous eye-tracking studies in the visual world and replicates the classic prediction effect often reported for children and adults (cf. Altmann & Kamide, 1999; Andreu et al., 2013; Ankener et al., 2018; Borovsky & Creel, 2014; Borovsky et al., 2012; Mani & Huettig, 2014; Nation et al., 2003). Regarding the more novel 3- and 4-

consistent conditions with multiple visual prediction options, our results are consistent with the fixation pattern of adults reported by Ankener et al. (2018) but show that even young children predict sentence input also in complex visual environments.

Considering next the non-predictive control condition in which none of the four visual objects was consistent with the sentence context. In the 0-consistent condition, children and adults fixated all four distractors across all time windows. This is also in line with the fixation based findings of Ankener et al. (2018) and shows that visual objects not consistent with linguistic constraints do not elicit children's and adults' anticipatory attention during sentence processing. In turn, this verifies that the fixation patterns in the predictive conditions (1-, 3-, and 4-consistent) were not due to chance but indicate that participants predicted the sentence input based on the constraining verbs in those other conditions.

3.5.2 Children and Adults Consider Multiple Prediction Options

A second important finding is that even young children integrate multiple visual prediction options into anticipatory sentence processing. Upon hearing the verb, both age groups fixated one object when only one object was consistent with the verb constraints (1-consistent condition). They anticipatorily fixated multiple objects (i.e., more than one) when multiple objects were consistent with the verb constraints (3- and 4-consistent conditions). This can be concluded since children and adults showed less anticipatory target fixations in the 3- and 4-consistent conditions versus the 1-consistent condition. Given that we analyzed the proportion of fixations to the target object relative to all other objects, this shows that both groups also fixated some of the other visual prediction options in the 3- and 4-consistent conditions. This suggests that children, like adults, integrate multiple visual prediction options into anticipatory sentence processing and argues against the possibility that they follow a one-only prediction fashion.

This finding presents considerable advance on previous research which so far only provides few indication that children can integrate two visual cues into online sentence processing (Snedeker & Trueswell, 2004; Trueswell et al., 1999) as well as two visual prediction options into anticipatory sentence processing (Borovsky et al., 2012; Gambi et al., 2021; Mani et al., 2016). The ability to track multiple prediction options may result from children's real-world experiences. Here, they experience language together with complex visual information that can even facilitate their linguistic processing (Knoeferle et al., 2005; van Rij, 2012; Weighall & Altmann, 2011; Zhang & Knoeferle, 2012) but also varies in its consistency with the linguistic input (Reuter et al., 2020). Such experiences may help children to learn to integrate multiple visual cues into language processing. Given that predicting upcoming input has positive associations with the speed of language processing, one option may be that children's maintaining of multiple options allows them to benefit from their predictions to a greater extent (Kuperberg & Jaeger, 2016; Mani & Huettig, 2014; van Petten & Luka, 2012). Since the benefits of prediction are, moreover, thought to exceed its costs (Kuperberg & Jaeger, 2016), we conclude that already young children may be able to integrate much of the predictive information of complex visual contexts into language processing, thereby processing language efficiently in the complex visual world.

Regarding next the question whether children and adults can adapt their prediction behavior to the exact number of probable target objects in the visual scenes. Admittedly, we found no statistical difference in the proportion of target fixations between the 3- and 4-consistent conditions. However, our analysis on the target advantage scores revealed that upon hearing the verb in the 3-consistent condition both age groups fixated the two competitors to a similar extent but more often than the single distractor. Here, the target advantage scores of the two competitors were close to zero indicating that they were fixated to a similar extent as the target. In the verb window of the 4-consistent condition, the target advantage scores of all three competitors did not differ from each other and were close to zero. Hence, adults and children fixated all competitors and the target to a similar extent upon hearing the verb in this condition. These results indicate that both age groups can adapt their prediction behavior to the exact number of probable target objects in the visual scenes.

One may argue that the results in the 3- and 4-consistent conditions emerged from averaging data across participants and items. In principle, it could be the case that different participants only fixated one competitor but each of them a different one on a particular item. Averaging across those participants would result in the same fixation pattern. The same argument can be made at item level. However, we included random effects for participants and items into our models which improved their fit with the underlying data structure as shown by the post-hoc deviance tests. Thus, participant and item specific variance had an impact on both types of data (proportional target fixations and target advantage scores). However, all statistical results revealed under the control of such participant and item specific variability since we added random effects for participants and items to our models (e.g., Baayen et al., 2008; Cunnings et al., 2012; Winter, 2013, 2020). Thus, we can rule out such an explanation for our data.

The post-hoc descriptive data analysis further underlines our statistical results. Here, we manually determined the amount of cases in which participants anticipatorily fixated multiple or all visual prediction options. This inspection showed for the 3- and 4-consistent conditions that both age groups a) anticipatorily fixated more than one visual prediction option in most cases and b) adapted their prediction behavior to the exact number of visual prediction options for many cases (as also indicated by the results of our control analysis of the target advantage scores). This shows that children and adults both can follow a multiple predictions pattern in the complex visual world. Notably, the post-hoc chi-square tests showed that the proportion of cases in which multiple or all visual prediction options were fixated anticipatorily did not differ across the age groups. This emphasizes again that young children's prediction behavior in the complex visual world is very similar to that of adults.

At first glance, it is remarkable that children did not differ from adults in their ability to maintain multiple predictions. Given their limited language experience (e.g., Huettig, 2015; Rabagliati et al., 2016), it would have been logical that children do not recognize the semantic properties of all visual prediction options (Bar, 2009; Mani & Huettig, 2012) and thus do not fixate all visual prediction options anticipatorily. However, the pretest of our stimuli verified that all visual objects are typically recognized as potential arguments of the verbs by 5- to 6-year-olds. Besides given their limited cognitive capacity (e.g., Cowan et al., 2010; Johnson et al., 2014) it is also remarkable that children were in charge of a sufficient extent of cognitive resources to pre-update even multiple noun candidates. This could have derived from the fact that the visual scenes were displayed for the whole course of the sentence which may have facilitated the continuous updating of the visual prediction options (Sikos et al., 2021).

3.5.3 Children Process Target Words Differently Than Adults

Next, our study provides evidence how children process dissolving sentence endings, i.e., the target word at the end of a constraining sentence, when combined with complex visual contexts. Indeed, it has often been shown that children use the target word at the end of a sentence to guide their eyes to the referent of this word in a visual scene (Andreu et al., 2013; Borovsky et al., 2012; Nation et al., 2003). This study revealed that they even do so when the other given objects were potentially relevant for language processing at an earlier stage of the sentence. After the target noun was named, children's and adults' fixations of the target object increased in the 3- and 4-consistent conditions, indicating that their competitor fixations now decreased as we analyzed the proportion of target fixations. This shows that even young children can inhibit previously integrated visual prediction options once a sentence is resolved and reveals how fast they can update their visual attention focus when processing language in the visual world.

However, the fixation patterns of children and adults differed slightly in the noun window: Adults' target fixations increased significantly less in the 1-consistent (15%) compared to the 3- and 4-consistent conditions (about 33%). Adults still paid more attention to the three distractors (about 14% each, see Table 2) than children (about 10% each) after the target was named in the 1-consistent condition. Children, in turn, fixated the target object almost exclusively in the noun window of the 1-consistent condition (68%). However, we note that children's increase in target fixations varied *descriptively* between the 1-consistent (24%) versus the 3- and 4-consistent (about 33%) conditions, although this difference was not statistically significant.

We explain this finding with recourse to the demands of language learning or inhibition of attention. Some accounts suggest that children use prediction to improve their language skills by comparing what they predicted with the actual input they received (e.g., Chang et al., 2006; Fazekas et al., 2020; Ramscar et al., 2013). Thus, they may have paid most attention to the fulfilled visual prediction option in the 1-consistent condition to memorize the connection between the semantically constraining verb and the semantical and perceptual properties of the target object. Adults, in contrast, with their more advanced prediction and language comprehension skills may not need to memorize the resolved combination of the linguistic and visual information. This would explain their reduced attention to the resolved target object in the 1-consistent condition.

3.5.4 Relations Between Prediction Skills and Vocabulary Size

Finally, we also found evidence for age differences in the association between participants' prediction skills and vocabulary size. The Peabody Picture Vocabulary Test score was positively associated with children's — but not adults' — anticipatory target fixations. Children with higher receptive vocabularies showed increased anticipatory fixations to the target object in the 1- and 3-consistent conditions. For the 1-consistent condition, this is in line with results from a few other studies showing that the size of children's receptive vocabularies is associated with their anticipatory fixations towards single visual prediction options (Borovsky & Creel, 2014; Borovsky et al., 2012; Prescott et al., 2022). For the 3-consistent condition this suggests, for the first time, that children with higher receptive vocabularies also integrate complex visual contexts with more than one visual prediction option more strongly into predictive sentence processing than those with smaller vocabularies. This finding might be explained in terms of increased vocabulary knowledge itself being associated with increased knowledge of the associations between verbs and their arguments (Bar, 2009), resulting in an increased pre-activation of prediction options.

It should, however, be noted that the size of our child sample (*n* = 26) was relatively small to examine the influence of individual differences in vocabulary size on prediction. Our finding that children with higher vocabularies were more efficient at predicting how constraining sentences proceed than children with smaller vocabularies should therefore be treated with caution. However, we do note that, together with comparable studies (Borovsky & Creel, 2014; Borovsky et al., 2012; Prescott et al., 2022), we provide further indication that children's language prediction may be positively related to their receptive vocabulary size.

Surprisingly, we did not find a modulation of prediction by receptive vocabulary size in our adult sample, although the studies by Borovsky and colleagues showed that adults' prediction in the visual world is also associated with their performance in the English version of the Peabody Picture Vocabulary Test (Borovsky & Creel, 2014; Borovsky et al., 2012). Possibly, this divergence is based on the restricted variance in the performance of our adult sample in this test, as the adult scores in the Peabody Picture Vocabulary Test were very homogeneous and in the upper range, hinting at a ceiling effect.

3.5.5 Limitations

We do, however, note another limitation of the present study. We found that young children anticipatorily integrate more complex visual environments into sentence processing. As the visual contexts varied in predictability, they were ecologically more valid than in other prediction research with children (e.g., Andreu et al., 2013; Mani & Huettig, 2012, 2014; Nation et al., 2003). At the same time, it must be noted that four colored pictorial objects that are arranged in static visual scenes with systematically varying predictability are indeed closer to but still far away from the real world (cf. Reuter et al. 2020). As typical for research on prediction (e.g., Mani & Huettig, 2012, 2014; Nation et al., 2003), our linguistic stimuli were artificial as well (32 sentences of the same structure). Thus, while we provide a first clue that children are able to integrate complex visual contexts into predictive processing, this remains to be investigated in more realistic situations such as daily conservations in daily visual environments (Huettig & Mani 2016). This could be possible, for instance, with head-mounted eye trackers (Reuter, 2020; Tanenhaus & Brown-Schmidt, 2008).

Finally, it must be noted that our focus was on semantical prediction. Most recent work on prediction in children also relied on semantic (e.g., Mani & Huettig, 2012, 2014; Tribushinina & Mak, 2016) or morphosyntactic (e.g., Bosch et al., 2022; Deevy et al., 2017; Smolík & Bláhová, 2019) cues. However, individuals predict input also by non-linguistic information such as communicative cues (disfluencies: Bosker et al., 2014; Kidd et al., 2011, prosody and communicative behavior: Casillas & Frank, 2012, 2013) and social cues (agent identity: Borovsky et al., 2012; speaker identity: Borovsky & Creel, 2014). Further developmental work is needed to reveal how individuals integrate linguistic, visual, and social cues to form predictions during online language comprehension (Kidd et al., 2011; Reuter, 2020).

3.5.6 Conclusion

Our study revealed that young children (5–6 years), like adults, predict sentence input based on semantically constraining verbs even in complex visual contexts with more than one potential target object. In particular, we showed that children and adults can integrate multiple visual prediction options into anticipatory sentence processing, which in children was associated with receptive vocabulary size.

4 Experiment 1: Cognitive Demands of Predictive Processing in the Complex Visual World

4.1 Rationale and Design

The aim of this experiment was to examine whether the predictability of visual context information would influence children's and adults' cognitive load engaged in predictive language processing. The data presented in this chapter were collected in the experiment that was reported previously in Chapter 3. This is because the visual-world eye-tracking study that presented young children (5–6 years) and adults with semantically constraining sentences (e.g., "Der Vater *verschlingt* gleich die Waffel") in four visual conditions (0-, 1-, 3-, and 4-consistent) also allowed to draw conclusions about the cognitive load individuals engage during visually situated predictive processing. We measured participants' Index of Cognitive Activity (ICA) and pupil sizes across the time course of the sentences as an indicator of cognitive load. Assessing these pupillometric measures could reveal whether children and adults engage higher cognitive load to predict multiple (3- and 4-consistent) versus only single sentence continuations (1-consistent). In addition, this could show whether they engage smaller cognitive load to predict input in visual scenes that do (1-consistent) versus do not contribute to prediction (0-consistent). Finally, these measures could uncover whether target word processing is facilitated for target words that could be predicted more (1-consistent) versus less specifically (3- and 4-consistent) by the visuo-linguistic constraints.

4.2 Research Questions and Hypothesis

4.2.1 Hypothesis 1: Additional Cognitive Load to Maintain Multiple Predictions

We investigated whether children and adults engage higher cognitive load to maintain multiple versus only single visual prediction options. The joint constraints of a visual and a linguistic context allow to pre-update prediction options in working memory (e.g., Huettig & Janse, 2016; Huettig et al., 2011b). Besides, the amount of cognitive load usually increases with the number of stimuli updated in working memory (e.g., Karatekin, 2004; Karatekin et al., 2004). Since this might also apply for the mechanism of *pre*-updating, we expected children and adults to show higher cognitive load indicated by higher ICA and pupil size values in the 3- and 4-consistent conditions (where multiple nouns could be pre-updated) versus the 1-consistent condition (where one noun could be pre-updated). Since effects of predictive processing often spill over from critical words to subsequent words in the input (e.g., Koornneef & van Berkum, 2006; Smith & Levy, 2013) and since pupil size has a slower latency than the ICA measure (e.g., Vogels et al., 2018), the above effect could, in particular for the pupil size measure, first reveal in the spill-over region “gleich” that succeeded the constraining verb (and not yet on the verb itself). Since pre-updating involves working memory resources (e.g., Ness & Meltzer-Asscher, 2018, 2021), while individuals with smaller working memory capacity typically engage more cognitive load in a task than other individuals (e.g., Johnson et al., 2014; Just & Carpenter, 1993), the above effect could be more pronounced in children versus adults due to their smaller working memory capacity (e.g., Johnson et al., 2014; Kharitonova et al., 2015).

4.2.2 Hypothesis 2: Predictive Visual Contexts Facilitate the Forming of Predictions

We next examined whether children’s and adults’ cognitive load to form predictions differs among situations where a visual context does versus does not contribute to prediction.

The 1-consistent scenes allow to predict one specific noun candidate because they display one visual prediction option. The 0-consistent scenes show four verb-inconsistent visual distractors, therefore do not contribute to prediction. Here, individuals can only rely on the verb (e.g., “eat”) to predict the semantic category of a noun (e.g., edible) while the prediction of a particular noun candidate is not possible. Here, two alternative results are plausible: In line with cognitive models, input that does not allow a commitment to a specific prediction option (e.g., when only a semantic category can be predicted) may not induce that level of pre-activation of the predicted input required to initiate the resource demanding mechanism of pre-updating (e.g., Ness & Meltzer-Asscher, 2018, 2021). Thus, the 0-consistent scenes may not induce pre-updating and participants could show smaller ICA and pupil size values in the constraining part of the sentences (verb, spill-over word) in the 0-consistent versus the 1-consistent condition (where pre-updating is possible). Otherwise, the 0-consistent scenes are ambiguous with the linguistic input as they show four visual distractors that are inconsistent with the verb constraints. Here, listeners could try to resolve that ambiguity and/or inhibit the visual distractors which could result in additional cognitive load (i.e., in high ICA and pupil size values) in the constraining part of the sentences (verb, spill-over word).

4.2.3 Hypothesis 3: More Specific Predictions Facilitate Target Word Processing

Finally, we investigated whether children and adults show less cognitive load to process target words that could be predicted more (versus less) specifically and thus be pre-updated more (versus less) thoroughly by the visuo-linguistic constraints. As the constraints of visual and linguistic contexts are integrated in working memory in such a way that an online model of the predicted sentence is pre-updated (e.g., Huettig & Janse, 2016; Huettig et al., 2011b) and since working memory is capacity limited (Cowan, 2010; Johnson et al., 2014), resources for the mechanism of pre-updating should be limited. Thus, the constraining verbs (e.g., “eat”) together with the visual scenes in the different conditions could affect pre-updating of the target noun (e.g., “waffle”) as follows. In the 1-consistent scenes listeners could predict only the target noun, which could then be pre-updated highly thoroughly. In the 3- and 4-consistent conditions multiple noun candidates could be predicted, thus each noun could be pre-updated less thoroughly. In the 0-consistent condition, listeners could only predict the semantic noun category, thus the target word could not be pre-updated. In line with prior findings for adults (Ankener et al., 2018; Sikos et al., 2021) and since also children show less cognitive load for linguistic input that could be pre-updated by a visual context (e.g., Csink et al., 2021; Fritsche & Höhle, 2015) or by the joint constraints of a visual and linguistic context (Gambi et al., 2021; Süss et al., 2018) we expected children and adults to engage less cognitive load (i.e., to show smaller ICA values and pupil sizes) for the target noun in the 1-consistent versus the 3- and 4-consistent conditions. The highest load was expected in the 0-consistent condition. This results pattern could, due to its slower latency (e.g., Vogels et al., 2018), reveal later in the pupil size versus the ICA measure.

4.3 Methods

These research questions were addressed in an eye-tracking study in the visual world with young children and adults. We extracted participants’ ICA values and pupil sizes during visually situated predictive processing as measures of cognitive load. Note that these data were collected within the same study that is reported in Chapter 3. The methods that we applied to collect the ICA and pupil size data are therefore identical to and can be obtained from the methods presented in Chapter 3.3.

4.4 Analysis and Results of the ICA Data

We now present the analysis and results of the ICA data. Then, we show the analysis and results of the pupil size data (see Chapter 4.5). Finally, the results of both measures are discussed (see Chapter 4.6).

4.4.1 Preparation of the ICA Data

For technical problems, ICA values could not be extracted for three children and one adult. Thus, data of n = 23 children and n = 36 adults entered the analysis reported below. The ICA values were extracted from the data recorded with the eye-tracker using the EyeWorks Cognitive Workload Module (EyeTracking, 2016). This software uses a wavelet analysis to discard larger light-induced pupil oscillations, while extracting abrupt and small task-evoked pupil oscillations, called ICA events (for details, see Box 1, page 40). ICA events were extracted separately for each participant, eye, and trial. Data of all practice and filler trials were omitted. Each ICA event was obtained with its time of occurrence during an item in seconds with two decimals which we converted into milliseconds to allow for a more accurate annotation of the ICA events. That is, we annotated for each ICA event in which part of an item it occurred (e.g., during verb presentation). For each part of an item we summarized the ICA values in bins of 100 ms. The 100 ms bins of the left and the right eye were summarized (cf. Ankener et al., 2018).

We then generated four regions of interest. They all were non-overlapping and 600 ms in length (cf. Ankener et al., 2018). First, the *baseline region* included the ICA values obtained during the presentation of the unpredictive agent of a sentence. It served as a control for the visual manipulation as there was no constraining linguistic input here. The baseline region started 500 ms after the onset of the subject to remove that part of the baseline region in which participants got used to the change from purely visual input to visual combined with auditory input. All other regions started after the first 100 ms of the respective word as it takes about 100 ms to recognize a word (e.g., Grosjean, 1980; Gwilliams et al., 2018; Lash et al., 2013). Second, the *verb region* included the ICA values recorded during verb presentation. In contrast to other studies that examined visually situated language processing load with the ICA (Ankener et al., 2018; Sikos et al., 2021), we also considered the *spill-over region*. This region contained the ICA values obtained during the presentation of the spill-over word “gleich” and was included since effects of constraining input often spill over to subsequent words (e.g., Koornneef & van Berkum, 2006; Smith & Levy, 2013). Fourth, the *noun region* included the ICA values recorded during target noun presentation but did not include the preceding article since we controlled that only objects of the same grammatical gender were presented in each visual scene.

We excluded all observations for which children (17%) and adults (< 1%) responded incorrectly to the comprehension questions.[14] With two chi-square tests, we tested whether the number of incorrect answers differed among the four visual conditions. This was neither the case for children ($\chi^2(3) = 0.77$, $p = .857$) nor for adults ($\chi^2(3) = 1.57$, $p = .666$). Thus, correct answers to the questions did not depend on the visual condition.

Outliers were detected with the interquartile range method. Here, any data point with a value 1.5 times greater or less than the interquartile range is declared as outlier, while the interquartile range is the difference between the 25th and 75th percentile of a data set (Walfish, 2006). We used this method as it extracts outliers based on values coming from the middle half of the distribution, thus is unlikely to be influenced by outliers themselves (Ramsey & Ramsey, 2007). Because the ICA values averaged across all regions and conditions varied significantly among children ($M = 34.80$, $SD = 12.30$) and adults ($M = 33.10$, $SD = 12.20$) as verified by a Welch *t*-test ($t(4861.60) = 5.66$, $p < .001$, *CI* [1.15, 2.36]), we detected outliers

[14] Since we used the ICA as a measure of cognitive load during visually situated prediction and since incorrect answers to the questions could be associated with general challenges in storing and processing the visuo-linguistic input in working memory, we excluded items that were answered incorrectly. The same was done for the pupil size measure which was also used as indicator of cognitive load (see Chapter 4.5).

separately for each age group and condition. For each region outliers were replaced with the median of the respective age group and condition (< 1% of all observations for adults and children). Finally, the ICA values were rounded as required for generalized linear mixed effects models. The ICA values in the four regions (baseline, verb, spill-over, noun) entered the analyses.

4.4.2 Analysis of the ICA Data

Data were analyzed using RStudio (RStudio Team, 2018). A significance criterion of $p < .05$ was applied and we report 95% confidence intervals of the mean. Since the ICA is a count variable, the lme4 library (Bates et al., 2015) was used to conduct a generalized linear mixed effects model with Poisson distribution on participants' ICA values (cf. Ankener et al., 2018; Vogels et al., 2018). The factors condition (0-, 1-, 3-, 4-consistent), region (verb, spill-over, noun), and age group (adults, children) were added to the model. They were effect coded in a planned structure of most theoretical interest. The factor condition consisted of three comparisons. First, we contrasted the 0-consistent versus the 1-consistent condition (i.e., conditions that allow for an unspecific versus a highly specific prediction). Second, we compared the 1-consistent versus the 3- and 4-consistent conditions (i.e., conditions that allow to predict a single versus multiple noun candidates). Finally, we contrasted the 3-consistent versus the 4-consistent condition (i.e., conditions that allow to predict three versus four noun candidates). The region factor was coded for two comparisons. First, we contrasted the verb and the spill-over region. Second, we compared the spill-over and the noun region. The age group factor compared children versus adults.[15]

[15] We ran a control model on the ICA values in the baseline region with the factors condition and age group which found no evidence for baseline differences in the ICA across the conditions prior to the constraining linguistic input. To control for spill-over effects of the target noun, we ran the same model on the ICA values in the first 600 ms after noun offset which did not reveal any spill-over effects (see Appendix E, Table E1).

All factors were added to the model with their interaction terms. To consider specific variance of subjects and items (Bayeen et al., 2008; Cunnings et al., 2012; Winter, 2020), the model contained random intercepts for subjects and items. Besides, we added by-subject random slopes for the factors condition and region as well as by-item random slopes for the factors condition, region, and age group. The interactions of the random slopes were also added (for details on the usage of random effects, see Chapter 3.4.1). Non-converging models were simplified with the least-variance approach (Barr et al., 2013). We estimated *p*-values with the Satterthwaite degrees of freedom method (lmerTest library, Kuznetsova et al., 2017) and confidence intervals with the stats library (RStudio Team, 2018). We only report results that are relevant for our research questions.

4.4.3 Results for the ICA Data

The model with its results is shown in Table 10. Both age groups' ICA values did not differ between the 1- and 0-consistent conditions (p = .501). However, independent of the region or age group (p-values > .05), the ICA values were significantly higher in the 3-consistent versus the 4-consistent condition (β = 0.05, SE = 0.02, z = 3.02, p = .003, *CI* [0.02, 0.08]) which is visualized in Figure 9 (for the numerical ICA values, see Appendix E, Table E2). Although the ICA values did not vary among the 1-consistent versus the 3- and 4-consistent conditions (p = .106), this interacted with the second region contrast (spill-over versus noun region) and the age group (β = 0.13, SE = 0.06, z = 2.21, p = .027, *CI* [0.02, 0.25]). Given this three-way-interaction, we re-ran the above model separately for each age group. The results of these models are reported below and shown in Table 11.

Adults' ICA values did not differ across the visual conditions (p-values > .05). For the comparison of the 1-consistent versus the 3- and 4-consistent conditions, this interacted with the contrast comparing the spill-over and the noun region (β = 0.11, SE = 0.04, z = 2.89, p = .004, *CI* [0.03, 0.18]). Model splits by region (see Table 11) showed that adults' ICA values in the spill-over region were significantly smaller in the 1-consistent versus the 0-consistent condition (β = -0.08, *SE* = 0.04, *z* = -2.03, *p* = .042, *CI* [-0.16, 0.00]). In the noun region, their ICA values were significantly smaller in the 1-consisent versus the 3- and 4-consistent conditions (β = -0.06, *SE* = 0.02, *z* = -2.67, *p* = .008, *CI* [-0.11, -0.02]). For children, the model only showed that their ICA values in all regions were significantly higher in the 3-consistent versus the 4-consistent condition (β = 0.08, *SE* = 0.03, *z* = 2.88, *p* = .004, *CI* [0.03, 0.14]). This pattern of results is reflected visually in Figure 9.

In sum, we found age differences in children's and adults' ICA values across the different conditions and regions. Children's ICA values did not differ between the visual conditions in the separate regions but were generally higher in the 3- versus the 4-consistent condition. Surprisingly, adults' ICA values did not vary among conditions that allowed to predict multiple (3- and 4-consistent) versus single (1-consistent) noun candidates in either the verb or the spill-over region. However, adults' ICA values were higher in the spill-over region when the scene did (1-consistent) versus did not (0-consistent) show a visual prediction option. Besides, in line with our expectation, adults' ICA values in the noun region were higher when the scene allowed for multiple (3- and 4-consistent) versus only single (1-consistent) predictions.

Table 10

Results of the Model on the ICA Values With All Factors

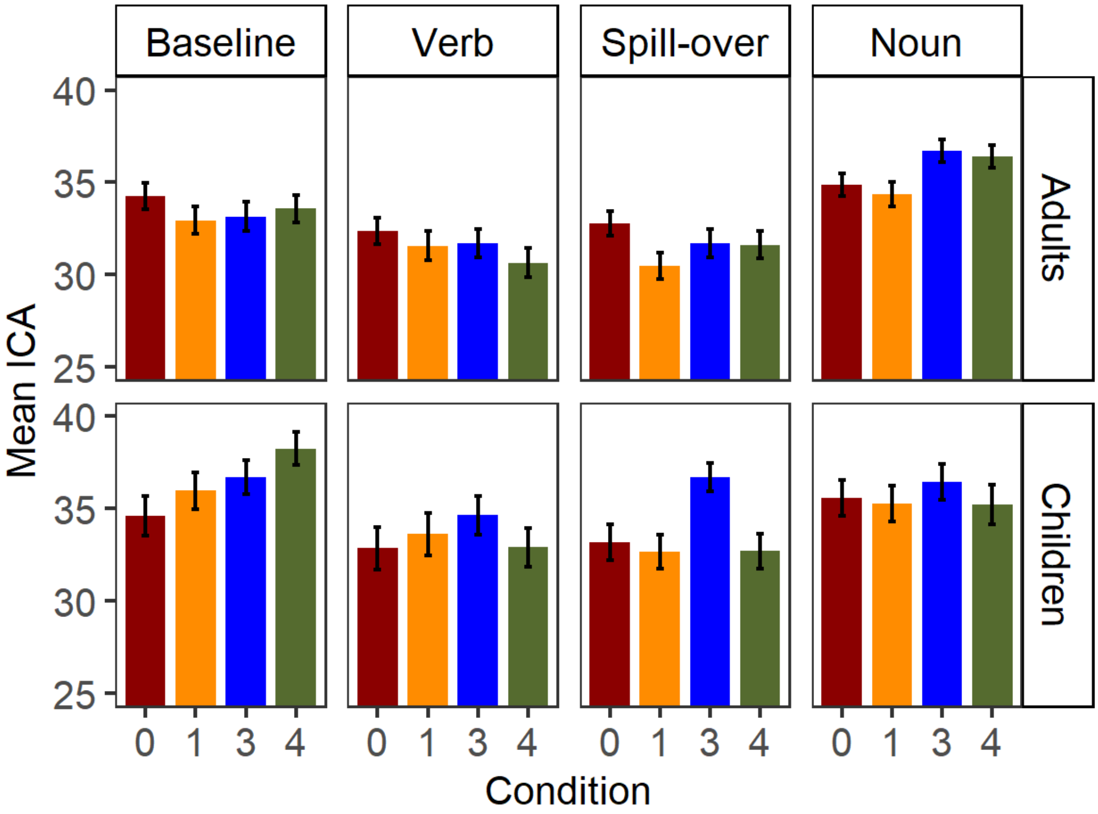
Comparison	β	*SE*	*z*	*p*	95% *CI*
Intercept	3.48	0.02	161.49	.000	3.44, 3.52
v1	-0.01	0.02	-0.67	.501	-0.05, 0.02
v2	-0.02	0.02	-1.62	.106	-0.06, 0.01
v3	0.05	0.02	3.02	.003	0.02, 0.08
t1	-0.06	0.02	-2.45	.014	-0.10, -0.01
t2	-0.12	0.02	-6.10	.000	-0.16, -0.08
a	-0.03	0.04	-0.76	.446	-0.12, 0.05
v1:t1	0.00	0.04	0.00	.999	-0.09, 0.09
v1:t2	-0.03	0.05	-0.57	.566	-0.12, 0.07
v2:t1	0.05	0.05	1.04	.299	-0.04, 0.14
v2:t2	0.04	0.03	1.19	.233	-0.03, 0.11
v3:t1	0.00	0.06	0.02	.986	-0.11, 0.11
v3:t2	0.04	0.04	0.98	.327	-0.04, 0.12
v1:a	-0.06	0.04	-1.45	.147	-0.13, 0.02
v2:a	0.03	0.03	1.06	.290	-0.03, 0.09
v3:a	-0.07	0.04	-1.85	.064	-0.14, 0.00
t1:a	-0.02	0.05	-0.38	.706	-0.11, 0.07
t2:a	-0.05	0.04	-1.28	.199	-0.12, 0.03
v1:t1:a	-0.03	0.10	-0.31	.756	-0.23, 0.17
v1:t2:a	-0.13	0.10	-1.36	.174	-0.32, 0.06
v2:t1:a	0.07	0.10	0.68	.494	-0.13, 0.27
v2:t2:a	0.13	0.06	2.21	.027	0.02, 0.25
v3:t1:a	0.07	0.11	0.63	.527	-0.14, 0.28
v3:t2:a	-0.06	0.08	-0.75	.451	-0.23, 0.10

Note. The model on the ICA values (ica) covered the factors condition, region, and age group. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second region contrasts. a is the age group contrast. The converged model:

ica~(v1+v2+v3)*(t1+t2)*a+(1+(v1+v2+v3)*(t1+t2)||subject)+(1+(v1+v2+v3)*(t1+t2)*a||item))

Figure 9

Averaged ICA Values per Condition, Region, and Age Group



Note. Each panel presents the averaged ICA values for the respective non-overlapping 600 ms regions of the sentences. Error bars indicate the standard error of the mean.

Table 11

Results of the Models on the ICA Values per Age Group

Group	Model	Comparison	β	*SE*	*z*	*p*	95% *CI*
Adults	Overall	Intercept	3.47	0.03	123.87	.000	3.41, 3.52
		v1	-0.04	0.02	-1.82	.069	-0.09, 0.00
		v2	-0.01	0.02	-0.55	.580	-0.04, 0.02
		v3	0.01	0.02	0.71	.477	-0.02, 0.05
		t1	-0.07	0.03	-2.13	.033	-0.13, -0.01
		t2	-0.15	0.02	-6.24	.000	-0.19, -0.10
		v1:t1	-0.01	0.06	-0.27	.790	-0.12, 0.09
		v1:t2	-0.09	0.05	-1.68	.093	-0.20, 0.02
		v2:t1	0.08	0.06	1.31	.191	-0.04, 0.21
		v2:t2	0.11	0.04	2.89	.004	0.03, 0.18
		v3:t1	0.03	0.06	0.60	.552	-0.08, 0.15
		v3:t2	0.01	0.05	0.19	.849	-0.09, 0.11
	Spill-over	Intercept	3.43	0.03	105.70	.000	3.36, 3.49
		v1	-0.08	0.04	-2.03	.042	-0.16, 0.00
		v2	0.00	0.03	0.07	.942	-0.06, 0.07
		v3	0.00	0.04	0.05	.964	-0.07, 0.07
	Noun	Intercept	3.54	0.03	136.56	.000	3.49, 3.59
		v1	0.01	0.03	0.19	.849	-0.05, 0.06
		v2	-0.06	0.02	-2.67	.008	-0.11, -0.02
		v3	0.01	0.03	0.35	.729	-0.05, 0.07
Children	Overall	Intercept	3.50	0.03	109.46	.000	3.45, 3.58
		v1	0.02	0.03	0.58	.565	-0.07, 0.06
		v2	-0.04	0.03	-1.53	.126	-0.09, 0.03
		v3	0.08	0.03	2.88	.004	0.03, 0.14
		t1	-0.05	0.03	-1.41	.160	-0.13, 0.01
		t2	-0.10	0.03	-3.15	.002	-0.16, -0.02
		v1:t1	0.02	0.08	0.21	.835	-0.13, -0.01
		v1:t2	0.04	0.09	0.51	.608	-0.17, 0.13
		v2:t1	0.02	0.07	0.30	.765	-0.17, 0.18
		v2:t2	-0.03	0.06	-0.45	.651	-0.14, 0.16
		v3:t1	-0.03	0.10	-0.31	.758	-0.14, 0.08
		v3:t2	0.08	0.06	1.25	.213	-0.15, 0.21

Note. The overall models on the ICA values (ica) per age group covered the factors condition and region. For adults, the spill-over and the noun model only covered the factor condition. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second region contrasts. The converged models:

Adults, overall: ica~(v1+v2+v3)*(t1+t2)+(1+(v1+v2+v3)*(t1+t2)||subject)+(1+(v1+v2+v3)*(t1+t2)||item))
Adults, spill-over: ica~(v1+v2+v3)+(1+(v1+v2+v3)||subject)+(1+(v1+v2+v3)||item))
Adults, noun: ica~(v1+v2+v3)+(1+(v1+v2+v3)||subject)+(1+(v1+v2+v3)||item))
Children: ica~(v1+v2+v3)*(t1+t2)+(1+(v1+v2+v3)*(t1+t2)||subject)+(1+(v1+v2+v3)*(t1+t2)||item))

4.5 Analysis and Results of the Pupil Size Data

4.5.1 Preparation of the Pupil Size Data

Data of two children for whom we could not extract a reliable number of observations were excluded. The analysis below therefore included data of 24 children and 37 adults. Participants' pupil sizes among the time course of the trials were extracted from the eye-tracking data using EyeLink Data Viewer (SR Research, 2019b). The software extracted the mean pupil size of the left and the right eye in diameter separately for each participant and trial, while there was one estimate of mean pupil size every 2 ms of a trial. We excluded observations for which only the pupil size of one eye was obtained.[16] After discarding all practice and filler trials, we annotated for each such mean pupil size value in which region of an item it was obtained (e.g., during verb presentation). We then generated five non-overlapping regions of interest. The *baseline region* included all pupil size values obtained during the last 1000 ms prior to verb presentation. This region served as baseline because here participants already got used to the change from purely visual to visual combined with linguistic input while no constraining language (only the unpredictive agent) was presented. The *verb region* contained all pupil size values recorded during verb presentation. The *spill-over region* included all values obtained during the presentation of the spill-over word "gleich" and the article. Values recorded during article presentation were included because pupil size has a slow latency (e.g., Olivia, 2019) and effects could spill over not only to the spill-over word but also to the article. The *noun region* contained pupil size values obtained during noun presentation. The *postview region* included values acquired during the first 1000 ms after noun offset and should account for spill-over effects from the noun to the postview.

[16] As a control, all models reported below were also conducted with a dataset for which pupil size was extracted with the cyclopean mode. Here, the pupil size of one eye was used as mean pupil size when only data from one eye were obtained (for details, see Data Viewer User's Manual, SR Research, 2019b). As this did not change the results pattern, we report the more conservative analysis based on the mean pupil size of both eyes.

All 2-ms time stamps for which a blink event was registered were identified. We then excluded all items for which 70% or more of the time stamps in the baseline occurred during a blink (data loss = 6.17% for children and 5.32% for adults) and all items for which 50% or more of the time stamps in one of the other regions (verb, spill-over, noun, postview) occurred during a blink (additional data loss = 7.18% for children and 1.62% for adults). For the remaining items, we excluded all blink events (additional data loss = 2.96% for children and 4.59% for adults).

Next, we excluded extreme values. This was done separately for each participant since pupil size can strongly vary among individuals (e.g., Eckstein et al., 2017; Johnson et al., 2014). Extrema were detected with the interquartile range method (Ramsey & Ramsey, 2007; Walfish, 2006), where any data point with a value 1.5 times greater or smaller than the interquartile range was declared as extreme value (for more details, see Chapter 4.4.1). This caused an additional data loss of 2.19% for children and 2.29% for adults. Since extrema exclusion caused a lack of baseline values for one item of one child and one adult each, these two items were omitted. We also excluded all items for which the comprehension questions were answered incorrectly (17% of the items for children, < 1% of the items for adults).

Separately for each participant and item, we calculated the proportional change in pupil size from the baseline to the verb, spill-over, noun, and postview region. To do so, the mean pupil size in the baseline and in all other regions was calculated for each participant and item. Then, separately for each participant and item, we subtracted the mean pupil size in the baseline from the mean pupil size in each of the other regions, divided this difference through the mean pupil size in the baseline and multiplicated the ratio with 100 (cf. Piu et al., 2019). For each participant and item this resulted in one estimate of baseline pupil size and one estimate of proportional pupil size in all other regions (verb, spill-over, noun, postview). These values entered our analysis.

4.5.2 Analysis of the Pupil Size Data

Data were analyzed using RStudio (RStudio Team, 2018). A significance criterion of $p < .05$ was applied and we report 95% confidence intervals of the mean. We conducted a linear mixed effects model (lme4 library, Bates et al., 2015) on the proportional pupil sizes with the factors condition (0-, 1-, 3-, 4-consistent), region (verb, spill-over, noun, postview), and age group (adults, children). The factors were effect coded in a planned structure. For the factor condition, we defined three contrasts of most theoretical interest. First, we contrasted the 0-consistent versus the 1-consistent condition (i.e., conditions that allow for an unspecific versus a highly specific prediction). Next, we compared the 1-consistent versus the 3- and 4-consistent conditions (i.e., conditions that allow to predict a single versus multiple noun candidates). Finally, we contrasted the 3-consistent versus the 4-consistent condition (i.e., conditions that allow to predict three versus four noun candidates). The factor region consisted of three contrasts. First, the verb and the spill-over region were compared. Next, we contrasted the spill-over versus the noun region. Third, we compared the noun and the postview region. The age group factor compared adults versus children.[17]

All factors were added to the model with their interaction terms. To consider specific variance of the subjects and items, the model contained random intercepts for subjects and items (Bayeen et al., 2008; Cunnings et al., 2012; Winter, 2020). We also included by-subject random slopes for the factors condition and region as well as by-item random slopes for the factors condition, region, and age group. The random slopes were added with their interaction terms (for details on the usage of random effects, see Chapter 3.4.1). Non-converging models were simplified with the least-variance approach (Barr et al., 2013). We estimated *p*-values

[17] We ran a control model on the pupil sizes in the baseline region with the factors condition and age group which provided no evidence for baseline differences in the pupil sizes across the conditions prior to the constraining linguistic input (see Appendix F, Table F1). However, this model showed that baseline pupil sizes were larger in children (M = 3284, SD = 587) versus adults (M = 2747, SD = 864) across all conditions (β = -518.60, SE = 189.63, $t(59.75) = -2.74$, $p = .008$, CI [-889.91, -147.15]).

with the Satterthwaite degrees of freedom method (lmerTest library, Kuznetsova et al., 2017) and confidence intervals with the stats library (RStudio Team, 2018). We only report results that are relevant for our research questions.

4.5.3 Results for the Pupil Size Data

The model with its results is shown in Table 12. There was no difference in both age groups' proportional pupil sizes among the visual conditions (*p*-values > .05). However, for the comparison of the 1- and 0-consistent conditions this interacted with the contrast of the noun and the postview region (β = 3.88, *SE* = 1.32, *t*(6143.48) = 2.94, *p* = .003, *CI* [1.29, 6.47]). Model splits by region (see Table 13) showed that both age groups' proportional pupil sizes did not vary among the conditions in the noun region (*p*-values > .05). In the postview region, however, both age groups' proportional pupil sizes were significantly smaller in the 1-consistent versus the 0-consistent condition (β = -3.20, *SE* = 1.04, *t*(73.68) = -3.08, *p* = .003, CI [-5.23,-1.17]). This results pattern is reflected visually in Figure 10. As we expected that differences in pupil sizes among the visual conditions could vary between the age groups, we conducted the above model again but separately for adults (see Table 14) and children (see Table 15). The results of these models are reported below.

Adults' proportional pupil sizes did not vary among the 1- and 0-consistent conditions (*p* = .118). However, this interacted with the contrast of the noun and the postview region (β = 3.37, *SE* = 1.60, *t*(4096.88) = 2.10, *p* = .036, *CI* [0.23, 6.51]). Model splits by region (see Table 14) showed that adults' proportional pupil sizes in the noun region were smaller in 1-consistent versus the 3- and 4-consistent conditions (β = -1.94, *SE* = 0.88, *t*(987.34) = -2.21, *p* = .027, *CI* [-3.65, -0.22]). In the postview region, adults' proportional pupil sizes were smaller in the 1-consistent versus the 0-consistent condition (β = -3.24, *SE* = 1.29, *t*(40.02) = -2.51, *p* = .016, *CI* [-5.77, -0.75]).

Also children's proportional pupil sizes did not vary among the 1- and 0-consistent conditions (*p* = .454). However, this interacted with the contrast of the noun and the postview region (β = 4.39, *SE* = 1.88, *t*(1972.78) = 2.33, *p* = .020, *CI* [0.71, 8.07]). Model splits by region (see Table 15) showed that children's pupil sizes did not vary among the visual conditions in the noun region (*p*-values > .05). However, their proportional pupil sizes in the postview region were significantly smaller in the 1-consistent versus the 0-consistent condition (β = -3.17, *SE* = 1.46 *t*(25.81) = -2.18, *p* = .039, *CI* [-6.08, -0.28]). This pattern of results is reflected visually in Figure 10 (for the numerical pupil size values, see Appendix F, Table F2).

In sum, neither children's nor adults' proportional pupil sizes varied among the visual conditions in the predictive part of the sentences, which was expected for the verb, but not for the spill-over region. As expected, both age groups' showed smaller pupil sizes after the noun was played in the 1-consistent versus the 0-consistent condition. However, only the adults showed the expected result of smaller pupil sizes after noun presentation in the 1-consistent versus the 3- and 4-consistent conditions.

Table 12

Results of the Model on the Proportional Pupil Sizes With All Factors

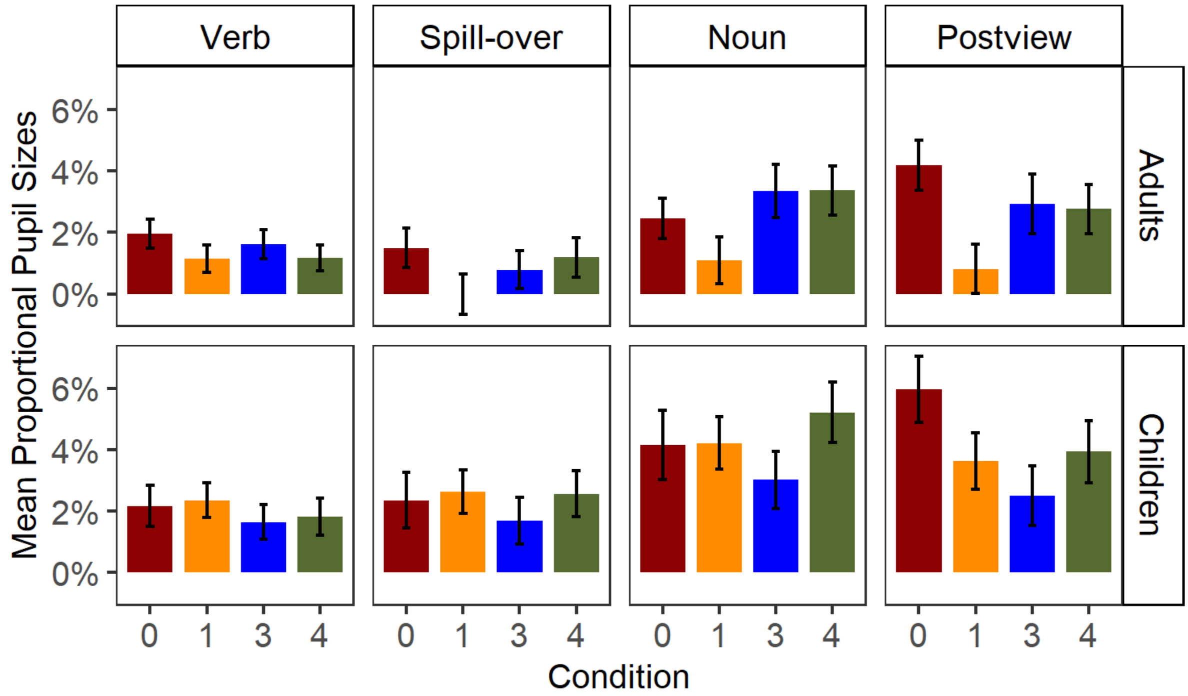
Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Intercept	2.39	0.48	73.17	4.96	.000	1.44, 3.33
v1	-1.42	0.94	62.57	-1.51	.137	-3.26, 0.43
v2	0.16	0.55	63.00	0.29	.775	-0.91, 1.23
v3	-0.40	0.96	64.45	-0.42	.678	-2.28, 1.48
t1	-1.37	0.52	6142.81	-2.62	.009	-2.39, -0.35
t2	-3.21	0.60	6143.04	-5.32	.000	-4.39, -2.03
t3	-1.98	0.52	6143.35	-3.79	.000	-3.01, -0.96
a	-1.20	0.91	65.03	-1.32	.193	-2.99, 0.59
v1:t1	1.33	1.32	6142.83	1.01	.312	-1.25, 3.92
v1:t2	2.31	1.53	6143.03	1.52	.129	-0.67, 5.30
v1:t3	3.88	1.32	6143.48	2.94	.003	1.29, 6.47
v2:t1	0.69	1.08	6142.90	0.64	.523	-1.42, 2.80
v2:t2	0.68	1.25	6143.34	0.55	.584	-1.76, 3.12
v2:t3	-1.30	1.08	6143.17	-1.21	.228	-3.42, 0.81
v3:t1	1.38	1.28	6143.05	1.08	.279	-1.12, 3.88
v3:t2	1.21	1.48	6143.76	0.82	.411	-1.68, 4.11
v3:t3	0.16	1.28	6143.06	0.12	.901	-2.35, 2.66
v1:a	-0.64	1.67	62.09	-0.39	.701	-3.93, 2.63
v2:a	-1.43	1.10	63.00	-1.30	.198	-3.57, 0.72
v3:a	1.08	2.06	70.89	0.52	.603	-2.96, 5.13
t1:a	1.84	1.04	6142.81	1.77	.078	-0.20, 3.89
t2:a	1.69	1.21	6143.04	1.40	.162	-0.67, 4.05
t3:a	0.65	1.05	6143.35	0.63	.532	-1.39, 2.7
v1:t1:a	0.03	2.64	6142.83	0.01	.989	-5.13, 5.20
v1:t2:a	-1.31	3.05	6143.03	-0.43	.667	-7.28, 4.66
v1:t3:a	-1.04	2.64	6143.48	-0.39	.695	-6.21, 4.14
v2:t1:a	1.86	2.16	6142.91	0.86	.389	-2.37, 6.08
v2:t2:a	3.13	2.49	6143.38	1.25	.210	-1.75, 8.01
v2:t3:a	1.80	2.16	6143.21	0.83	.404	-2.43, 6.03
v3:t1:a	-1.20	2.56	6143.04	-0.47	.639	-6.20, 3.80
v3:t2:a	-2.74	2.95	6143.69	-0.93	.355	-8.51, 3.05
v3:t3:a	-0.74	2.56	6142.99	-0.29	.772	-5.75, 4.27

Note. The model on the proportional pupil sizes (p) covered the factors condition, region, and age group. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second region contrasts. a is the age group contrast. The converged model:

lmer(p~v1+v2+v3)*(t1+t2+t3)*a+(1+(v1+v2+v3)||subject)+(1+(v1+v3)*a||item))

Figure 10

Averaged Proportional Pupil Sizes per Condition, Region, and Age Group



Note. Error bars indicate the standard error of the mean.

Table 13

Results of the Models on the Proportional Pupil Sizes in the Noun and Postview Region

Model	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Noun	Intercept	3.02	0.57	72.27	5.30	.000	1.90, 4.14
	v1	-0.42	0.98	68.47	-0.43	.667	-2.34, 1.48
	v2	-0.83	0.84	83.81	-0.99	.326	-2.47, 0.81
	v3	-0.99	1.25	64.89	-0.79	.430	-3.44, 1.45
	a	-1.86	1.14	72.27	-1.63	.108	-4.09, 0.38
	v1:a	-0.57	1.95	68.47	-0.29	.771	-4.40, 3.24
	v2:a	-2.18	1.68	83.81	-1.30	.199	-5.47, 1.11
	v3:a	2.04	2.50	64.89	0.82	.418	-2.85, 6.93
Postview	Intercept	3.39	0.69	71.09	4.91	.000	2.03, 4.74
	v1	-3.20	1.04	73.68	-3.08	.003	-5.23, -1.17
	v2	0.83	0.82	82.82	1.01	.315	-0.77, 2.42
	v3	-0.48	1.37	65.08	-0.35	.729	-3.16, 2.20
	a	-1.56	1.27	72.66	-1.23	.221	-4.04, 0.93
	v1:a	-0.10	2.08	73.93	-0.05	.964	-4.17, 3.96
	v2:a	-2.40	1.63	82.66	-1.47	.145	-5.59, 0.79
	v3:a	1.18	2.74	65.08	0.43	.668	-4.17, 6.54

Note. The models on the proportional pupil sizes (p) for the noun and the postview region covered the factors condition and age group. v1, v2, and v3 are the first, second, and third condition contrasts. a is the age group contrast. The converged models:

```
Noun:       lmer(p~(v1+v2+v3)*a+(1+(v1+v2+v3)||subject))
Postview:   lmer(p~(v1+v2+v3)*a+(1+(v1+v2+v3)||subject)+(1|item))
```

Table 14

Results of the Models on Adults' Proportional Pupil Sizes

Model	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Overall	Intercept	1.80	0.56	41.84	3.22	.003	0.69, 2.90
	v1	-1.67	1.05	45.56	-1.59	.118	-3.75, 0.40
	v2	-0.58	0.83	45.72	-0.70	.488	-2.23, 1.07
	v3	0.14	1.29	43.75	0.11	.916	-2.42, 2.70
	t1	-0.45	0.64	4096.91	-0.71	.481	-1.69, 0.80
	t2	-2.37	0.73	4097.10	-3.22	.001	-3.81, -0.93
	t3	-1.65	0.64	4097.08	-2.60	.009	-2.90, -0.41
	v1:t1	1.35	1.60	4096.90	0.84	.401	-1.79, 4.49
	v1:t2	1.65	1.85	4096.92	0.89	.372	-1.97, 5.28
	v1:t3	3.37	1.60	4096.88	2.10	.036	0.23, 6.51
	v2:t1	1.62	1.31	4096.81	1.24	.216	-0.94, 4.18
	v2:t2	2.25	1.51	4096.90	1.49	.137	-0.71, 5.21
	v2:t3	-0.39	1.31	4096.90	-0.30	.766	-2.96, 2.17
	v3:t1	0.79	1.54	4096.76	0.51	.611	-2.23, 3.81
	v3:t2	-0.15	1.78	4096.68	-0.08	.934	-3.64, 3.34
	v3:t3	-0.21	1.54	4096.72	-0.13	.894	-3.23, 2.82
Noun	Intercept	2.09	0.72	43.79	2.92	.006	0.67, 3.51
	v1	-0.66	1.09	39.37	-0.61	.545	-2.84, 1.48
	v2	-1.94	0.88	987.34	-2.21	.027	-3.65, -0.22
	v3	0.00	1.65	36.21	0.00	.998	-3.27, 3.28
Postview	Intercept	2.61	0.83	43.44	3.16	.003	1.01, 4.21
	v1	-3.24	1.29	40.02	-2.51	.016	-5.77, -0.75
	v2	-0.38	0.94	944.94	-0.40	.690	-2.22, 1.47
	v3	0.14	1.87	31.56	0.08	.941	-3.57, 3.84

Note. The overall model on adults' proportional pupil sizes (p) covered the factors condition and region. The noun model was run on the values in the noun region, the postview model on the values in the postview region. The two latter models covered the factor condition. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second region contrasts. The converged models:

Overall: lmer(p~(v1+v2+v3)*(t1+t2+t3)+(1+(v1+v2+v3)||subject)+(1+(v1+v2+v3)||item))
Noun: lmer(p~(v1+v2+v3)+(1+(v1+v3)||subject))
Postview: lmer(p~(v1+v2+v3)+(1+(v1+v3)||subject)+(1+(v3)||item))

Table 15

Results of the Models on Children's Proportional Pupil Sizes

Model	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Overall	Intercept	2.98	0.79	30.86	3.76	.001	1.40, 4.54
	v1	-1.21	1.60	33.14	-0.76	.454	-4.38, 1.96
	v2	0.94	0.92	29.04	1.02	.318	-0.89, 2.78
	v3	-1.00	1.40	34.35	-0.72	.477	-3.78, 1.76
	t1	-2.29	0.74	1972.50	-3.09	.002	-3.74, -0.84
	t2	-4.05	0.86	1972.58	-4.72	.000	-5.72, -2.37
	t3	-2.30	0.74	1972.70	-3.10	.002	-3.76, -0.85
	v1:t1	1.31	1.88	1972.51	0.70	.484	-2.36, 4.99
	v1:t2	2.96	2.17	1972.59	1.37	.172	-1.28, 7.21
	v1:t3	4.39	1.88	1972.78	2.33	.020	0.71, 8.07
	v2:t1	-0.24	1.54	1972.58	-0.16	.874	-3.25, 2.76
	v2:t2	-0.89	1.78	1972.81	-0.50	.618	-4.36, 2.59
	v2:t3	-2.20	1.54	1972.68	-1.43	.152	-5.21, 0.81
	v3:t1	1.99	1.82	1972.67	1.09	.275	-1.58, 5.56
	v3:t2	2.60	2.11	1973.06	1.23	.218	-1.52, 6.72
	v3:t3	0.54	1.83	1972.67	0.30	.768	-3.03, 4.11
Noun	Intercept	3.92	0.88	28.87	4.44	.000	2.16, 5.67
	v1	-0.19	1.70	25.73	-0.11	.913	-3.58, 3.19
	v2	0.24	1.11	507.09	0.22	.830	-1.94, 2.41
	v3	-2.08	1.49	21.37	-1.39	.178	-5.04, 0.92
Postview	Intercept	4.14	1.00	30.00	4.13	.000	2.14, 6.12
	v1	-3.17	1.46	25.81	-2.18	.039	-6.08, -0.28
	v2	2.05	1.12	490.82	1.84	.067	-0.14, 4.25
	v3	-1.15	1.45	20.82	-0.80	.435	-4.01, 1.77

Note. The overall model on children's proportional pupil sizes (p) covered the factors condition and region. The noun model was run on the values in the noun region, the postview model on the values in the postview region. The two latter models covered the factor condition. v1, v2, and v3 are the first, second, and third condition contrasts. t1 and t2 are the first and second region contrasts. The converged models:

```
Overall:    lmer(p~(v1+v2+v3)*(t1+t2+t3)+(1+(v1+v2+v3)||subject)+(1+(v1+v2+v3)||item))
Noun:       lmer(p~(v1+v2+v3)+(1+(v1+v3)||subject))
Postview:   lmer(p~(v1+v2+v3)+(1+(v1+v3)||subject)+(1|item))
```

4.5.4 Relation of Proportional Pupil Sizes and ICA Values

Finally, we inspected the relation of participants' ICA values and proportional pupil sizes (cf. Demberg, 2013). Since the ICA measure was, to our knowledge, not applied with children prior to this study, we did so separately for each age group. For children, the correlations of the two pupillometric measures in the verb region (r = .40, t(525) = 10.02, p <.001, *CI* [.33, .47]) and spill-over region (r = .10, t(525) = 2.36, p = .018, *CI* [.02, .19]) were small to moderate and significant. The same was true for adults regarding the verb region (r = .28, t(1064) = 9.67, p < .001, *CI* [.23, .34]) and spill-over region (r = .07, t(1064) = 2.31, p = .021, *CI* [.01, .13]). However, for both children (p = .185) and adults (p = .265) the ICA and pupil size values did not correlate in the noun region. Note that the postview region was not included in this correlation analysis because ICA values obtained in the postview region were not included in the ICA analysis due to the short latency of that measure (see Chapter 4.4.1).

4.6 Discussion of the ICA and Pupil Size Based Findings

The aim of this part of Experiment 1 was to uncover age differences in the way visual contexts can affect processing load engaged in predictive processing. When listeners receive a predictive linguistic cue, their cognitive load might increase since they do not only process that cue but also the predictable input. This, in turn, can cause a processing benefit for input that could be pre-processed (e.g., Maess et al., 2016; Ness & Meltzer-Asscher, 2018, 2021). Since language is typically accompanied by visual contexts (Reuter et al., 2020) we were interested in how visual contexts can affect children's and adults' processing load for predictive cues and predictable target words. We conducted an eye-tracking study with children (5–6 years) and adults who listened to predictable sentences (e.g., "The father *eats* soon the waffle") and inspected visual scenes of four objects with varying predictability. Either 0, 1, 3, or 4 of the objects were consistent with the verb constraints (e.g., edible) and thus considered as visual prediction options. Comparing children's and adults' ICA values and proportional pupil sizes as measures of cognitive load across the visual conditions revealed some commonalities and differences among the age groups and measures. We discuss four notable findings below.

Notably, we found no evidence for differences in both age groups' ICA values and pupil sizes across the visual conditions in the baseline region where the unpredictive agent, but no constraining linguistic input was yet presented. Thus, there is no reason to assume that the findings below may not derive from the experimental manipulation.

4.6.1 No Evidence for Additional Cognitive Load to Maintain Multiple Predictions

We first found that both age groups' ICA values and proportional pupil sizes in the verb and spill-over region of the sentences did not differ between the 1-consistent versus the 3- and 4-consistent conditions. If this null-effect is statistically valid, this would mean that neither children nor adults engaged higher cognitive load when they could pre-update multiple versus only single noun candidates by the visuo-linguistic constraints.

For the verb region this result is consistent with our hypothesis and with comparable findings for adults with the ICA measure (Ankener et al., 2018; Sikos et al., 2021). For the spill-over region this result disconfirms our hypothesis. We expected listeners to engage higher cognitive load when the visuo-linguistic constraints allow to pre-update multiple versus only single noun candidates. This is because constraining linguistic and visual input is integrated in working memory which results in the pre-updating of prediction options by the use of working memory resources (e.g., Huettig et al., 2011a, 2011b; Özkan et al., 2022). Since working memory engages more resources to maintain more stimuli (e.g., Johnson et al., 2014; Just & Carpenter, 1993) and since effects of predictive processing often first reveal in spill-over regions (e.g., Koornneef & van Berkum, 2006; Smith & Levy, 2013), we predicted additional cognitive load in the spill-over region (but not yet in the verb region) when multiple nouns could be pre-updated. Thus, we expected higher ICA values and pupil sizes in the spill-over region of the 3- and 4-consistent versus the 1-consistent condition. Due to their smaller cognitive capacity (e.g., Johnson et al., 2014; Kharitonova et al., 2015) we expected this effect to be more pronounced in children. However, neither children's nor adults' ICA values or pupil sizes were larger in the spill-over region of the 3- and 4-consistent versus the 1-consistent condition.

We reject the conclusion that predictive processing was just not affected by the visuo-linguistic constraints. This is because both age groups' anticipatory object fixations show that they relied on the visuo-linguistic constraints to predict one or multiple noun candidates (see Chapter 3.4). Moreover, that listeners' ICA and pupil size values *did* vary among the 1-, 3-, and 4-consistent conditions in the noun region (see below, Chapter 4.6.3) shows that the joint constraints of the verb and the scenes must have affected predictive processing.

If the null-result in the spill-over region is statistically valid, it could be explained by effects of cognitive capacity. Most of the adults were students (n = 33) and thus possibly had high cognitive capacity. This is also indicated because their scores in the Semantic Verbal Fluency Task (a test of cognitive functioning) were higher than in other adult samples. That adults' cognitive load in the spill-over region did not vary among the 1-, 3-, and 4-consistent conditions could mean that, given their high cognitive capacity, they did not invest notably more resources to maintain multiple candidates. For children, in contrast, their small cognitive capacity (e.g., Cowan et al., 2010; Johnson et al., 2014) could be the reason why their cognitive load in the spill-over region did not vary among the conditions. Children answered 17% of the comprehension questions incorrectly (this was <1% for adults). Thus, the experimental task may generally have been more demanding for them, causing a similar extent of cognitive load among the conditions. This view is supported by the fact that pupillometric measures of cognitive load do not increase further when individuals reach their limit of cognitive capacity (e.g., Karatekin, 2004; Karatekin et al., 2004).

Otherwise, that listeners' cognitive load did not vary among the 1-, 3- , and 4-consistent conditions, could mean that they did not only engage cognitive resources to pre-update multiple nouns in the 3- and 4-consistent scenes but also in the 1-consistent scenes to inhibit the three visual distractors and their mental representations that were pre-activated by the scenes earlier. Thus, cognitive load may not have varied among the conditions because they each involved different, resource demanding processes (pre-updating, inhibition).

In sum, we can only speculate whether our finding was due to methodological limitations or whether pre-updating multiple candidates truly does not require additional cognitive resources. Since it was once speculated that pupillometric measures like the ICA are insensitive to that type of load engaged in the pre-updating of information (Ankener et al., 2018; Sikos et al., 2021), ERP studies might provide answers here. The ERP component P600, for instance, is sensitive to the integration of predicted words into working memory (e.g., Brouwer et al., 2012; Delogu et al., 2019), thus can reflect pre-updating costs (Ness & Meltzer-Asscher, 2018; 2021). In case the P600 increases when individuals can pre-update multiple ("One member of The Rolling Stones is …") versus only single ("The first man on the moon was …") sentence continuations this would show that pre-updating is more demanding for multiple candidates.

4.6.2 Predictive Visual Contexts Can Facilitate the Generation of Predictions — But Only in Measures of Short Latency

We next examined whether comprehenders' cognitive load to form predictions differs among situations where a visual context does versus does not contribute to prediction. Here, our findings are not consistent among both age groups and measures. With the ICA measure, we found that adults engaged higher cognitive load in the spill-over region of the 0-consistent versus the 1-consistent condition. Thus, prediction was facilitated in case a specific noun versus only a semantic noun category could be predicted. This is not in line with cognitive models of prediction where input that does not allow a commitment to a specific prediction option is argued to not induce the resource demanding process of pre-updating (e.g., Ness & Meltzer-Asscher, 2018, 2021).

Nevertheless, adults' additional cognitive load in the spill-over region of the 0-consistent condition can be explained: Here, the constraining verb allowed to predict a semantic noun category while none of the visual objects was consistent with that category. Thus, the visual scene was ambiguous with the linguistic input. After the verb was played, adults may have engaged additional cognitive load in the 0-consistent scenes to resolve that ambiguity and/or to inhibit the visual distractors. This is possible since both the resolving of ambiguity (Beck et al., 2008; Gillis et al., 2014; Kadem et al., 2020) and the inhibition of information irrelevant to linguistic processing (e.g., Gambi et al., 2021; Kuperberg & Jaeger, 2016; van Petten & Luka, 2012) are resource demanding.

This reasoning is supported by prior research. Thus, ICA values have been shown to increase for ambiguous linguistic input, but to decrease when that input is disambiguated (Demberg, 2013; Demberg et al., 2013). Moreover, Ankener et al. (2018, reported in Staudte et al., 2021) found with their comparable eye-tracking task that adults' amplitude of the N400 was larger in the verb region of the 0-consistent versus the 1-consistent condition. Since the N400 can index the semantic fit of a stimulus with a context (e.g., Huettig, 2015; Kutas et al., 2011) or even ambiguity detection (Kutas & Federmeier, 2011), this contributes to the view that adults may have engaged additional cognitive load to resolve the visuo-linguistic ambiguity in the 0-consistent scenes (for similar results, see Sikos et al., 2021, Experiment 2).

Children's ICA values, in contrast, did not differ in the spill-over region of the 0- and 1-consistent conditions. If this null-effect is statistically valid, this could mean that children did not engage higher cognitive load in the spill-over region of the 0- versus the 1-consistent condition, possibly since they engaged cognitive load for ambiguity resolving (0-consistent scenes) but also to predict one noun while inhibiting three distractors (1-consistent scenes). Otherwise, the absence of the effect in children could mean that the ICA is not suitable to assess cognitive load in children. However, we can only speculate about this because little is known about the usage of the ICA with children.

Notably, our results varied among both measures. Neither children's nor adults' pupil sizes in the verb or spill-over region varied among the 1- and 0-consistent conditions. While adults' pupil sizes in the spill-over region at least trended to be larger in the 0-consistent condition, this was not the case for children. If this null-results is statistically valid, this could mean that pupil sizes were less sensitive than the ICA for variations in that type of cognitive load engaged in ambiguity resolving. However, this is inconsistent with prior studies which found that children's and adults' pupil size can increase during ambiguity resolving (e.g., Demberg et al., 2013; Kadem et al., 2020; Krüger et al., 2020; Süss et al., 2018). We speculate that the slower latency of the pupil size measure (e.g., Vogels et al., 2018) meant that effects of ambiguity detection in the 0-consistent scenes did not translate in pupil sizes before the noun was played. That latency in pupil size changes is slower in children versus adults (e.g., Karatekin et al., 2007; Zhang & Emberson, 2020; Zhang et al., 2018), could explain why only adults showed a trend for larger pupil sizes in the 0-consistent condition.

4.6.3 More Specific Predictions Facilitate Target Word Processing

Adults and children have a processing benefit for words that can (versus such that cannot) be predicted by purely linguistic (e.g., Cutter et al., 2023; Ness & Meltzer-Asscher, 2018 Vergilova et al., 2022; Wassenburg et al., 2015) or purely visual contexts (Friedrich & Friederici, 2005; Fritsche & Höhle, 2015). Besides, adults show less cognitive load for words that can be predicted more versus less specifically by the visuo-linguistic constraints (e.g., Ankener et al., 2018, Sikos et al., 2021, Tourtouri et al., 2015). In line with this, we found with both pupillometric measures that target noun processing was facilitated in adults when the visual scenes allowed for a more (versus less) specific prediction of the target noun. That is, adults showed smaller ICA values and pupil sizes in the 1-consistent versus the 3- and 4-consistent conditions after the target noun (e.g., “waffle”) was played. This result is novel for the pupil size measure, but consistent with prior findings for the ICA (Ankener et al., 2018; Sikos et al., 2021) and could be explained as follows. The 1-consistent scenes allowed to predict the target noun as single sentence continuation. Here, listeners could commit to and highly thoroughly pre-update the target noun shortly after the verb was played (Ness & Meltzer-Asscher, 2018, 2021). Thus, they engaged small cognitive load for the noun. The 3- and 4-consistent scenes, in contrast, allowed to predict multiple nouns. Here, the limited cognitive resources available for pre-updating (Ness & Meltzer-Asscher, 2018, 2021) were also engaged to pre-update the competitors. As a result, the target noun was pre-updated less thoroughly, causing additional load when it revealed (Freunberger & Roehm, 2017).

Otherwise, also the demands of inhibition could explain additional costs to process the target noun in the 3- and 4-consistent conditions. When hearing the noun, listeners needed to inhibit further processing of the two (3-consistent) or three (4-consistent) competitors that were relevant earlier during sentences processing. As inhibiting unfulfilled prediction options is cognitively demanding (e.g., Kuperberg & Jaeger, 2016; van Petten & Luka, 2012) adults may have shown higher cognitive load in the 3- and 4-consistent versus the 1-consistent condition (where no competitors were inhibited).

For children, in contrast, the above finding was not obtained. Neither children’s pupil sizes nor their ICA values varied among the 1-, 3-, and 4-consistent conditions after the noun was played. If these null-results were statistically valid, this would mean that children do not have a processing benefit for target words that could be predicted more relative to less specifically by a visuo-linguistic context. For the pupil size measure, this would be inconsistent with studies that found a processing benefit in form of smaller pupil sizes in children for words that are more versus less predictable by the visuo-linguistic constraints (Süss et al., 2018). However, our experimental task of visually situated language processing may generally have been highly demanding for the sample of young children, masking potential variations of their pupil sizes across the visual conditions. This claim is supported by the observation that children showed higher cognitive load (in form of larger pupil sizes) than adults in the baseline region of the sentences and since they answered a large number of comprehension questions incorrectly (17%).

While this explanation might be true for the ICA measure as well, the null-result in children's ICA values in the noun region could also result from the following. The ICA values of adults who perform a task have sometimes been shown to decrease in case they work on a secondary task (e.g., Demberg et al., 2013b; Vogels et al., 2018). Besides, it is assumed that children rely on resolved predictions to improve their language skills by comparing what they predicted with the input they received (e.g., Chang et al., 2006; Fazekas et al., 2020). Thus, children's ICA values could not have varied among the 1-, 3-, and 4-consistent conditions because they compared the target noun with the one, three, or four nouns they predicted next to the task of visually situated sentence processing. However, it could also be that the effect did not reveal due to methodological reasons, for instance, because the ICA may not be appropriate for the usage with children.

4.6.4 Ambiguous Visual Contexts Turn Target Word Processing More Demanding

We also found that children and adults engaged higher cognitive load, reflected in larger pupil sizes, to process the target noun (e.g., "waffle") when it was played in the 0-consistent versus the 1-consistent scenes. This could be explained as follows. While the 1-consistent condition allowed to pre-update one particular noun (Ness & Meltzer-Asscher, 2018, 2021), the 0-consistent condition only allowed to predict a semantic noun category. Both age groups may have engaged higher processing load for the noun when it revealed in the 0-consistent scenes because here the noun was not pre-updated. This is in line with some studies that found a processing benefit for words that could be pre-updated by a purely linguistic context relative to such that could not (e.g., Ness & Meltzer-Asscher, 2018, 2021).

Otherwise, this result could be explained with respect to the demands of ambiguity resolving. In the 0-consistent condition, the target noun was presented together with four visual distractors. Here, listeners may have engaged additional load to process the ambiguity of the noun and the scene. This is in line with findings of larger pupil sizes in children and adults when a visual context is (versus is not) ambiguous with the linguistic signal (Csink et al., 2021; Fritsche & Höhle, 2015; Krüger et al., 2020; Tamási et al., 2017, 2019).

In sum, by measuring pupil sizes, we found that target word processing is facilitated in children and adults when a visual context displays an appropriate target object, rather than being ambiguous with the linguistic input. Surprisingly, we did not find such an effect with the ICA measure. Both age groups' ICA values in the noun region did not vary among the 1- and 0-consistent conditions. If this result holds statistically valid, it would not be in line with Ankener et al. (2018) who found larger ICA values on the noun for adults in the equivalent 0-consistent versus 1-consistent scenes. We note that we recorded the sentences more slowly than Ankener et al. (2018) to make them comprehensible to children. This may be the reason why we found adults' ICA values to be larger in the 0-consistent versus the 1-consistent condition already in the spill-over region (see Chapter 4.6.2). Thus, ICA-based effects of ambiguity resolving and/or inhibition in the 0-consistent condition may have revealed prior the noun in our study. That this was not the case for the pupil size measure is conceivable due to the slower latency of that measure (e.g., Vogels et al., 2018). However, further validation of the ICA as measure of cognitive load of ambiguity resolving and/or inhibition is needed.

4.6.5 Inhibiting Unfulfilled Predictions is Cognitively Demanding

Our last finding was not part of our hypothesis but strengthens the view that not only pre-updating but also inhibition processes induce cognitive load during visually situated language processing, and that the ICA may be sensitive to that type of load. In all regions of the sentences, children's and adults' ICA values (but not their pupil sizes) were higher in the 3-consistent versus the 4-consistent condition.

For the verb and spill-over regions this could mean the following. The 4-consistent scenes showed four visual prediction options and thus allowed to pre-update four noun candidates in working memory (Huettig et al., 2011a, 2011b). In the 3-consistent condition, listeners may have faced additional load as they discriminated among the objects (one distractor, three prediction options) and inhibited further processing of the distractor while pre-updating the prediction options. That we found higher ICA values for the 3- versus the 4-consistent condition also in the noun region could mean the following. After the noun was played, listeners may have inhibited the unfulfilled predictions in the 3- and 4-consistent conditions. This may have been more demanding in the 3-consistent condition where listeners had pre-updated only three noun candidates (instead of four) and thus made a stronger commitment to and more thoroughly pre-updated each candidate.

This finding suggest that cognitive demands of inhibition can reflect in the ICA measure not only for adults but also for children. Besides, we show that cognitive load of visually situated predictive processing is not only engaged in the pre-updating of prediction options but also in the inhibition of visually distracting information. This indicates that visual contexts can have a major impact on the ease of language processing which is relevant given current interests in factors that can facilitate or impede language comprehension for children in particular (e.g., Gambi et al., 2021; Mani & Huettig, 2016).

4.6.6 Limitations

It is crucial consider that a considerable number of pupil size data of both age groups could not be recorded or were omitted prior to analysis due to poor data quality. This could be because pupil size was recorded with the eye-tracker in remote mode. That is, no chin-rest was used, but a reference sticker was put on participants' foreheads so that they could still move their heads relatively freely. Remote mode was applied in order to make it more comfortable for the young children to sit still during the whole task (about 20 minutes). However, head movements can affect the quality of pupil size data (e.g., Zhang & Emberson, 2020). This way of data collection may have affected the quality of the ICA data as well, since the ICA data were extracted from the pupil size data.

Moreover, we excluded 17% of the items for children due to incorrect answers to the comprehension questions (for adults this was < 1%). This suggests that the experimental task was more difficult for children than for adults which can be explained with children's smaller cognitive capacity (e.g., Cowan et al., 2010; Johnson et al., 2014). This view is supported by the fact that children showed higher cognitive load (larger pupil sizes) than adults in the baseline region (although, this could also be because pupil size is generally larger in children versus adults, Eckstein et al., 2017; Johnson et al., 2014). However, children also showed higher ICA values than adults among all conditions and regions. Finally, children performed significantly worse than adults in the Semantic Verbal Fluency Task, a test of cognitive functioning, which also suggests smaller cognitive abilities for the children. Both of this (data omission, high cognitive load) could have caused small variability in children's pupillometric data among the visual conditions. Future works on (children's) cognitive load, should minimize the demands deriving from factors other than the experimental manipulation.

There are two more general aspects that need to be considered when interpreting pupillometric data. Pupil size does not only change due to the cognitive demands of a task but also due to changes in luminance (e.g., Eckstein et al., 2017; Sirois & Brisson, 2014). Within the scope of our possibilities, we controlled for this: Light conditions were identical among participants (the laboratory was darkened with curtains) and pupil size values in the regions of interest were relativized at baseline pupil sizes which supplies to the extraction of light-influences (Weber et al., 2021). Besides, our results for both pupillometric measures varied only slightly and the measures correlated significantly in most regions of interest. As the ICA is argued to be independent from light influences (e.g., Ankener et al., 2018; Vogels et al., 2018), this suggests that the results of our pupil size analysis were not strongly affected by luminance (although we cannot fully rule out that possibility).

However, there is a limitation of the ICA as well. The ICA is a patented measure (Marshall, 2000) of low transparency for which an independent verification is not possible (Mahanama et al., 2022; Weber et al., 2021). Studies using the ICA cannot fully disclose the internals of that measure due to intellectual property reasons. That is, we cannot reconstruct how exactly the ICA values were extracted from the raw data collected with the eye-tracker (Mahanama et al., 2022). In addition, and possibly for this reason, only few studies have used the ICA with adults and we are (to our knowledge) the first applying the ICA with children. We therefore cannot draw strong conclusions about the validity of the ICA based results of the child sample. It is crucial to validate the ICA for the usage in future studies (and with children in particular). Otherwise, to improve control of data, researchers could apply the Index of Pupillary Activity (IPA), a newer pupillometric measure inspired by the ICA that has a similar underlying concept but discloses the internals of its process (for details, see Duchowski et al., 2018; Mahanama et al., 2022; Weber et al., 2021).

4.6.7 Conclusion

With two pupillometric measures (ICA, pupil sizes) we could not show that children and adults engage additional cognitive load to pre-update multiple versus only single word candidates by the visuo-linguistic constraints. As predicted, we found a processing advantage in adults for predictable target words in case a visual scene allowed for a highly specific prediction of that word (and not also for the prediction of other word candidates). Unexpectedly, this result did not reveal for children. However, for both age groups, we found that visual contexts that are ambiguous with the linguistic input can put additional cognitive demands on both — the prediction of words and the processing of predictable words. We finally provide indication that the ICA may be suitable to assess cognitive load engaged in inhibition rather than (pre-)updating processes.

Since Experiment 1 focused on pupillometric measures of cognitive load that have some disadvantages, we extended on the above research with a more transparent measure of cognitive load that is validated for children and unaffected by lighting conditions. The next chapter presents Experiment 2 where children's and adults' word processing times were used as measure of cognitive load while they read constraining sentences in complex visual scenes.

5 Experiment 2: Prediction of Written Sentence Input in the Complex Visual World

5.1 Rationale and Design

Like Experiment 1, this study aimed to examine how children's and adults' cognitive load to process predictive linguistic cues and predictable target words is affected by complex visual contexts of varying predictability. Contrary to our first study (auditory sentence input), the sentences were shown onscreen in written form to a group of older children (8–12 years) and adults. To our knowledge, this is the first study that examines predictive processing in combination with visual contexts in a reading paradigm. However, this is critical because the input modality of the linguistic input could influence the cognitive load engaged in predictive processing. This is because written words enter the cognitive system via the visual input channel and are then processed in visual components of working memory (e.g., Goff et al., 2005; Pham & Hasson, 2014; Swanson & Jerman, 2007). Thus, when presenting the visual scenes *and* the constraining sentences visually, all prediction-relevant information may be temporarily stored and manipulated in visual working memory. We were interested in how these specific demands on visual working memory could affect the cognitive load that children and adults engage during visually-situated predictive processing. In case Experiment 2 reveals similar variations in processing load among the visual conditions as Experiment 1, this would mean that such effects are robust against changes in input-modality and strong enough to reveal regardless of reading-induced demands on visual working memory.

In addition, Experiment 2 may provide answers to our research questions that have so far only been examined with pupillometry (e.g., Ankener et al., 2018; Sikos et al., 2021) using a different measure of cognitive load (i.e., processing times). This could be useful, since we found partly inconsistent evidence in Experiment 1 with the ICA and the pupil size measure. Investigating the influence of visual predictability on predictive processing load with an additional measure of cognitive load could yield more generalizable findings. Finally, comparing the results of Experiment 1 (5- to 6-year-olds) and Experiment 2 (8- to 12-year-olds) could indicate how effects of visual predictability on cognitive load during predictive processing may develop across childhood.

Experiment 2 involved an online self-paced reading task with additional visual contexts. Literate children (8–12 years) and adults inspected visual scenes while reading German sentences. The stimuli were nearly the same as in Experiment 1: The sentences were semantically constraining (e.g., “Der Vater *verschlingt* am frühen Sonntagmorgen die Waffel auf dem Balkon”) and paired with visual contexts in four different conditions (0-, 1-, 3-, and 4-consistent), showing either 0, 1, 3, or 4 visual prediction options. The sentences were read in a word-by-word self-paced reading procedure. Each word appeared individually onscreen and a button press revealed the next word. We assessed the time it took participants from the start of the presentation of a word until they pressed the button to reveal the next word. This processing time was considered as an index of processing load resulting from the words shown in the visual scenes in the different conditions. To control for the cognitive and verbal abilities of the sample, we additionally applied a series of psychometric tests.

5.2 Research Questions and Hypotheses

5.2.1 Hypothesis 1: Additional Cognitive Load to Maintain Multiple Predictions

With Experiment 2, we aimed to examine whether children and adults experience higher cognitive load when a semantically constraining verb and a visual scene allow to predict multiple versus only a single sentence continuation (for a detailed explanation, see Chapter 4.2.1). Since working memory expends more resources when more stimuli are updated (e.g., Cowan, 2010; Johnson et al., 2014), and because individuals could pre-update multiple nouns in the 3- and 4-consistent but only one noun in the 1-consistent condition, we expected both age groups to show higher cognitive load (i.e., longer processing times) in the constraining part of the sentences (i.e., after verb presentation) for the 3- and 4-consistent versus the 1-consistent condition. Since effects of prediction in reading typically reveal in spill-over regions (e.g., Smith & Levy, 2013; Vela-Candelas et al., 2022), the above effect should first become apparent for the spill-over words after the constraining verb. Given their small working memory capacity (e.g., Cowan et al., 2010; Johnson et al., 2014) which goes along with the use of more mental resources in a task (Johnson et al., 2014; Just & Carpenter, 1993) this effect could be more pronounced in children than adults.

5.2.2 Hypothesis 2: Predictive Visual Contexts Facilitate the Forming of Predictions

This study next aimed to reveal whether children's and adults' cognitive load to form predictions differs among situations where a visual scene does versus does not contribute to prediction (for a detailed explanation, see Chapter 4.2.2). The visuo-linguistic constraints in the 1-, 3-, and 4-consistent conditions allowed to predict one, three, or four particular noun candidates. The 0-consistent scenes only allowed to predict a semantic noun category as they displayed four verb-inconsistent distractors. In Experiment 1, adults showed less cognitive load when a visual scene allowed to predict one specific noun candidate (1-consistent condition) relative to situations where the visual scene was ambiguous with the linguistic input (0-consistent condition). With Experiment 2, we aimed to extend on this finding. We expected both age groups to show a processing benefit for prediction-consistent over prediction-inconsistent visual scenes also if the prediction-consistent scenes allow for *multiple* specific predictions. This is because the 0-consistent scenes which are not consistent with the prediction may cause readers to resolve the visuo-linguistic ambiguity and/or to inhibit the visual distractors, causing additional cognitive load in the constraining part of the sentences. We therefore expected children and adults to reveal shorter processing times in the spill-over region after the verb in the 1-, 3-, and 4-consistent versus the 0-consistent condition.

5.2.3 Hypothesis 3: More Specific Predictions Facilitate Target Word Processing

Finally, Experiment 2 aimed to uncover whether children's and adults' processing of a predictable target noun is facilitated when the visuo-linguistic constraints allow for a highly versus less specific prediction of that noun (for a detailed explanation, see Chapter 4.2.3). We expected less cognitive load for target nouns in the 1-consistent versus the 3- and 4-consistent conditions. This is because the 1-consistent scenes allowed to pre-update one particular noun candidate, probably highly thoroughly. This should result in less processing load for that noun compared to the 3- and 4-consistent conditions where the target noun was pre-updated together with multiple competitors and thus less thoroughly. As typical for reading studies (e.g., Smith & Levy, 2013; Vela-Candelas et al., 2022), we expected this effect to be present in the spill-over region after the noun. Since adults and children both engage less cognitive load for input that could be pre-updated by the constraints of a purely visual (e.g., Csink et al., 2021; Friedrich & Friederici, 2005; Fritsche & Höhle, 2015; Mani et al., 2012; Tamási et al., 2017, 2019) or visuo-linguistic (Ankener et al., 2018; Sikos et al., 2021; Süss et al., 2018; Tourtouri et al., 2015) context, we expected similar results among both age groups.

5.3 Methods

5.3.1 Participants

The final sample of this study consisted of *n* = 77 children (*M* = 10.20 years, *SD* = 1.38, *range* = 8–12 years, 41 girls and 36 boys) and *n* = 63 adults (*M* = 26.30 years, *SD* = 6.64, *range* = 19–58 years, 48 women and 15 men). All participants were German native speakers without any reported reading or writing disorders. This was verified with a questionnaire filled in by the children's parents or the adult participants. Data of four additional children were excluded due to a diagnosed reading and writing disorder (*n* = 1) or because they did not produce enough reasonable observations in the self-paced reading task (*n* = 3; for details, see Chapter 5.3.6). All parents and adult participants gave informed consent and received 10 Euro as compensation. Participants were recruited with flyers and online recruiting platforms (for children: https://kinderschaffenwissen.de, last access: June 28, 2023; for adults: https://studien-saarland.com, last access: June 28, 2023).

5.3.2 Overview of the Experiment

The experiment was programmed and executed online in LabVanced, a platform for online studies (Finger et al., 2017). Data were collected in 2020 for children and in 2021 for adults. For participation, individuals needed access to a computer with internet connection, a loudspeaker, and a microphone. The study started with a questionnaire about participants' age, gender, reading abilities, and writing abilities. Then, participants worked on five tasks: Self-paced reading task, Semantic Verbal Fluency Task, Phonemic Verbal Fluency Task, Digit Symbol Substitution Task, and Peabody Picture Vocabulary Test. While the tasks were identical for both age groups, the instructions were slightly adjusted for each age group.

5.3.3 Self-Paced Reading Task

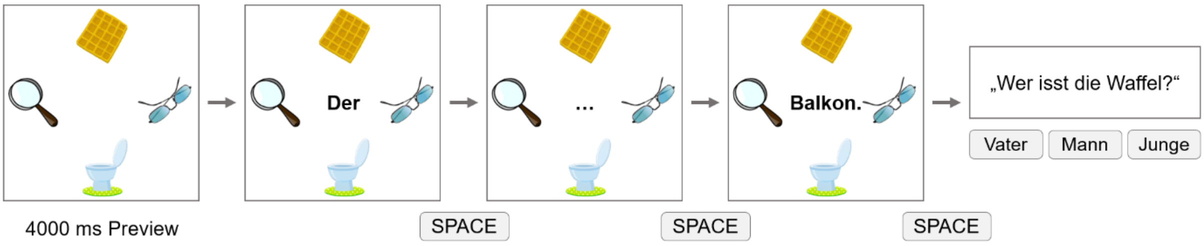
The first task was a self-paced reading paradigm with additional visual context that took about 20 minutes. Figure 11 visualizes the trial procedure. Participants were instructed to sit in front of the computer and to look carefully at all four objects of a visual scene that appeared on the screen. This preview of the scene lasted for 4000 ms.[18] Then, the scene remained on the screen while a sentence was displayed in the middle of the scene in a word-by-word fashion (e.g., "Der Vater *verschlingt* am frühen Sonntagmorgen die Waffel auf dem Balkon"). That is, the first word of the sentence (e.g., "Der") appeared in the middle of the screen. Participants were asked to read that word quickly but carefully and to press the space bar to reveal the next word. After the button press, the current word disappeared. All subsequent words of the sentence then appeared one after the other after button press each. The visual scene remained on the screen until the last word of the sentence (e.g., "Balkon") was presented and the space bar was pressed.

After each trial, a comprehension question appeared on the screen and should be answered by clicking at one out of three response options with the computer mouse. The questions were either related to the sentence (e.g., *Wer* isst gleich die Waffel? — Vater, Mann, Junge) or to the visual scene (e.g., *Wo* wurde die Waffel gezeigt? — oben, unten, links). They were answered correctly with the left, middle, or right response button in about 33% of cases each. The next trial began after participants confirmed to be ready via button click. Participants could familiarize with the trial procedure in three practice trials at the beginning of the task.

[18] The preview was twice as long as in Experiment 1. This was to ensure that readers could inspected all visual objects prior to sentence presentation (as we were the first to apply a self-paced reading task in a Visual World Paradigm, we could not guaranty that readers would inspect the scenes after the sentence started).

Figure 11

Trial Procedure of the Self-Paced Reading Task



Note. Example of a trial: The preview of the visual scene is followed by the sentence (“Der Vater *verschlingt* am frühen Sonntagmorgen die Waffel auf dem Balkon”) presented in a word-by-word reading fashion. Pressing the space bar reveals the next word. The trial ends with a comprehension question with three response options.

The task consisted of the 32 items validated with the pretest (see Chapter 2), each consisting of a semantically constraining sentence together with a visual scene in four visual conditions (0-, 1-, 3-, and 4-consistent). The conditions varied in such a way that either 0, 1, 3, or 4 objects of the scene were suitable arguments of the constraining verb of the sentence (see Figure 4, page 51). The visual scenes were identical to those selected after the pretest.

The sentences validated with the pretest were slightly adapted for the self-paced reading task. The agents, verbs, and target nouns were not modified. Since effects of predictive processing in reading typically spill over from critical words to subsequent words (e.g., Koornneef & van Berkum, 2006; Smith & Levy, 2013; Vela-Candelas et al., 2022) we enriched each sentence with two spill-over regions. We inserted one spill-over region after the constraining verb and one after the target noun. Both spill-over regions consisted of three words and did neither add substantial semantical information to the sentence nor mismatch its content. The first spill-over region was placed after the verb and constituted an adverbial phrase of time (e.g., “am frühen Montagmorgen”). Across all 32 sentences, the content of this phrase was rotated across the adjectives “early” and “late”, the seven days of a week, and the time of day (morning, noon, evening). The second spill-over region followed after the target noun and was an adverbial phrase of place (e.g., “auf dem Balkon”). Across all 32 sentences, the content of this phrase rotated among 16 different places. For example, the sentence “Der Vater verschlingt die Waffel” was modified to the sentence “Der Vater verschlingt *am frühen Sonntagmorgen* die Waffel *auf dem Balkon*”. Three German native speaking experts of our research group checked a-priori and on the basis of face-validity that the places were plausible continuations of the sentences (e.g., a waffle can be eaten on a balcony) and that the phrase of time matched the sentence context (e.g., a waffle can be eaten early Sunday morning). This resulted in 32 sentences of eleven words in the same syntactic structure (noun phrase — verb phrase — adverbial phrase — noun phrase — adverbial phrase). Appendix A (Table A2) shows all sentences of the self-paced reading task.

Since testing time is limited with children (Brewer, 2013), each participant only read 16 sentences, four per condition (0-, 1-, 3-, 4-consistent). We divided the 32 sentences into two sets of 16 sentences each. To ensure that each sentence was presented in all four visual conditions in each set while none of the participants should read one sentence in multiple conditions, we used a latin square design. In doing so, the 16 sentences of each set were divided to four item lists, each consisting of 16 sentences (four per condition). Since this procedure was conducted for both item sets, this resulted in a total of eight item lists of 16 sentences each. Each list consisted of four sentences of each condition. Participants were randomly assigned to the lists.

To mask the study design and to avoid monotony in sentence reading, each list also included 16 filler trials. Fillers introduced variation to the sentences since most of them followed another syntactic structure and because only half of them included semantically constraining verbs (for the filler sentences, see Appendix A, Table A3). Fillers also contributed variation to the visual scenes since there was no 0-consistent but a 2-consistent condition for the fillers (for the filler scenes, see Appendix B, Table B4). The fillers and three additional practice trials were the same for all participants. Each list started with the practice trials, followed by the filler and item trials in randomized order.

5.3.4 Psychometric Measures

To control for the cognitive and verbal abilities of our sample, we applied three psychometric tests (for a detailed description of each test, see Appendix C). In the Semantic Verbal Fluency Task, a test of cognitive functioning (e.g., Tröger et al., 2019), children's performance (M = 18.00, SD = 3.96, *range* = 8–28) was comparable to other typically developing children of the same age with German (Vergilova et al., 2022) or other native languages (Kavé, 2006; Moura et al., 2014). Adults outperformed children ($t(82.18) = 9.65$, $p < .001$, CI [7.56, 11.48]) and showed higher semantic verbal fluency (M = 27.80, SD = 6.14, *range* = 15–41) than in other studies with healthy adult participants (Martins et al., 2007; Rosselli et al., 2002; Troyer et al., 1997; Zimmermann et al., 2014). However, their performance was comparable to the adults in Experiment 1 (see Chapter 3.3.2).[19]

The Phonemic Verbal Fluency Task measures specific aspects of executive functioning (Martins et al., 2007; Moura et al., 2014) and is related to literacy (Kavé, 2006; Kavé & Sapir-Yogev, 2023). Children (M = 9.59, SD = 4.59, *range* = 2–22) were outperformed by adults (M = 19.40, SD = 4.62, *range* = 10–35) in this task ($t(106.40)$ = 11.28, $p < .001$, CI [8.06, 11.50]), but their phonemic fluency was comparable to other studies with typically developing children of the same age with German (Vergilova et al., 2022) or other native languages (Brandeker & Thordardottir, 2023; Kavé, 2006; Moura et al., 2014; Oliveira et al., 2016). Adults' phonemic fluency was slightly higher than in other

[19] Children of Experiment 2 outperformed the younger children of Experiment 1 in the Semantic Verbal Fluency Task which was verified with a Welch t-test ($t(31.89) = -7.07$, $p < .001$, CI [-10.53, -5.82]). The adult samples of both studies did not vary in their verbal fluency ($p = .242$).

studies with healthy adult participants (Martins et al., 2007; Rosselli et al., 2002; Troyer et al., 1997).

Participants also worked on an online version of the Digit Symbol Substitution Task, a test of perceptual processing speed (e.g., Karbach & Kray, 2007). Adults (*M* = 1280, *SD* = 213, *range* = 945–2118) performed comparable to other samples of adults with a similar mean age (Häuser et al., 2018, 2019) but significantly better than children ($t(103.98) = -8.10$, $p < .001$, *CI* [-664.28, -403.07]). Children's performance (*M* = 1816, *SD* = 506, *range* = 62–2711) could not be compared with other samples since there is, to our knowledge, no study that reports the performance of a child sample in the online version of this test.

To measure their receptive vocabulary size (Lenhard et al., 2015), both age groups worked on an online version of the Peabody Picture Vocabulary Test. This test was identical to the original version (see Chapter 3.3.4) except for the fact that the picture labels were played by the computer microphone. Labels were recorded a-priori by a female German native speaker with accurate pronunciation. In addition, participants answered via mouse click instead of pointing at the pictures. Children's averaged raw score in the Peabody Picture Vocabulary Test (*M* = 178.00, *SD* = 25.40, *range* = 85–209) corresponded to a *T*-value of 52 (*CI* [49, 55]) as reported by the test manual for children aged 10.20 years (which was the mean age of our child sample). Thus, the receptive vocabulary size of our child sample was common for their age. Adults outperformed children ($t(83.59) = 11.76$, $p < .001$, *CI* [29.27, 41.18]). Their raw scores (*M* = 213.00, *SD* = 5.62, *range* = 195–224) corresponded to a *T*-value of 61 (*CI* [59, 63]) as reported by the manual for 17-year-olds (the German test version is only normed upon age 17). Thus, adults' receptive vocabulary size was above the norms of 17-year-olds but comparable to that of the adults in Experiment 1 (see Chapter 3.4.7).[20]

[20] Children of Experiment 2 outperformed the younger children of Experiment 1 in the Peabody Picture Vocabulary Test which was verified with a Welch *t*-test ($t(39.19) = -10.35$, $p < .001$, *CI* [-75.11, 50.55]). The adult samples of both studies did not vary in their receptive vocabulary size ($p = .275$).

In sum, children's scores in both Verbal Fluency Tasks and the Peabody Picture Vocabulary Test suggest that their cognitive and verbal abilities were comparable to other typically developing children of the same age. Adults scores in both Verbal Fluency Tasks but not in the Digit Symbol Substitution Task suggest that their cognitive abilities were higher than in other samples of adults.

5.3.5 Procedure

Children and adults participated in the study online via a link from LabVanced. The study lasted about one hour. The study first asked for a permission allowing LabVanced to access the microphone of the computer (audio recording was needed for the Verbal Fluency Tasks). Parents or adult participants gave informed consent and filled in the questionnaire. Then, they were asked to set the loudspeaker of the computer at an adequate volume (audio files were played in the Peabody Picture Vocabulary Test). Parents were asked not to help their children with the tasks, but only with technical issues or comprehension questions. Then, the self-paced reading task began with three practice trials (children were allowed to work on these trials together with their parents). After participants confirmed to understand the task, the experimental trials were shown. Afterwards, participants worked on the Semantic Verbal Fluency Task, the Phonemic Verbal Fluency Task, the Digit Symbol Substitution Task, and the Peabody Picture Vocabulary Test. They could take a break after each task. Finally, the compensation was paid.

5.3.6 Data Analysis

Data were extracted from the .csv file recorded with LabVanced. For each participant and trial, the data file provided the processing times for each of the eleven words of a sentence in milliseconds. Processing times were defined as the time from the beginning of the presentation of a word until participants pressed the space bar to reveal the next word. Data of all practice and filler trials were excluded. Data of one additional child that responded correctly to less than 66% of the comprehension questions of the items were omitted. All items in which the comprehension questions were answered incorrectly were omitted (cf. Gibson & Levy, 2016), causing a data loss of 10.05% for children and 8.37% for adults.

For each sentence (e.g., “Der Vater *verschlingt* am frühen Sonntagmorgen die Waffel auf dem Balkon”), processing times were extracted for five regions of interest (see Figure 12). The *baseline region* covered the processing times of the agent (e.g., “Vater”) and was considered as a control of the manipulation since here no constraining input was presented. The article preceding the subject (e.g., “Der”) was not included in the baseline region since here participants got used to the change from purely visual to visual combined with linguistic input. The *verb region* included the processing times of the verb (e.g., “verschlingt”). The *post-verb region* summarized the processing times of the three words after the verb (e.g., “am frühen Sonntagmorgen”) and should account for spill-over effects of the verb. The *noun region* contained the processing times of the noun (e.g., “Waffel”). The article preceding the noun (e.g., “die”) was not included because the grammatical gender of all visual objects in a scene was identical. The *post-noun region* summarized the processing times of the three words after the noun (e.g., “auf dem Balkon”) and should account for spill-over effects of the target noun.

Figure 12

Regions of Interest of the Self-Paced Reading Task

x	Baseline	Verb	Post-Verb	x	Noun	Post-Noun
Der *The*	Vater *father*	*verschlingt* *eats*	am frühen Sonntagmorgen *early Sunday morning*	die *the*	Waffel *waffle*	auf dem Balkon. *on the balcony.*

Note. Illustration of the five regions of interest for an example sentence. The name of the region is shown at the top. The "X" indicates that no region was assigned to a word.

Separately for each region extremely low and extremely high processing times were omitted. For regions that covered processing times of only one word (baseline, verb, noun) all values below 100 and above 3000 ms were omitted (cf. Gibson & Levy, 2016; Jaffe et al., 2018; Linzen & Jaeger, 2016). Since the post-verb and the post-noun region summarized processing times of three words each, threshold values were tripled here: Processing times below 300 and above 9000 ms were omitted. To rely on a reasonable number of observations for each participant, data from two children with less than 66% remaining data points were omitted (resulting in the final sample size of 77 children). Regarding all other participants the extrema exclusion caused a data loss of 2.68% for children and 0.59% for adults.

To reduce right skewness of the data, processing times were log-transformed separately for each region and age group (e.g., Baayen & Milin, 2010; Blumenthal-Dramé, 2021; Cutter et al., 2021; Gardini et al., 2021; Linzen & Jaeger, 2016; Loeys et al., 2011; Rouder et al., 2015; van Breukelen, 2005; van der Linden, 2006). Outliers in the log-transformed processing times were detected with the interquartile range method (for details, see Chapter 4.4.1). Any data point with a value 1.5 times greater or less than the interquartile range was declared as outlier (Ramsey & Ramsey, 2007; Walfish, 2006). Outliers were omitted separately for each region, age group, and condition resulting in a data loss of 0.24% for children and 1.16% for adults. The log-transformed processing times in the five regions of interest (baseline, verb, post-verb, noun, post-noun) were included in the analyses.

5.4 Results

All data were analyzed using RStudio (RStudio Team, 2018). A significance criterion of $p < .05$ was applied and we report 95% confidence intervals of the mean. We first specify the statistical models and then report their results.

5.4.1 Statistical Models

As typical for self-paced reading data, we conducted one linear mixed effects model (lme4 library, Bates et al., 2015) on the log-transformed processing times separately for each of the five regions of interest (e.g., Gibson & Levy, 2016; Haeuser et al., 2020; Tucker et al., 2015). The factors condition (0-, 1-, 3-, 4-consistent) and age group (adults, children) were included as fixed effects. They were effect coded in a planned structure. For the factor condition, three contrasts of most theoretical interest were defined: First, we compared the 0-consistent condition to all other conditions. Second, we contrasted the 1-consistent versus the 3- and 4-consistent conditions. Finally, we compared the 3-consistent versus the 4-consistent condition. The age group contrast compared children and adults.

The condition and the age group contrasts were added to the models together with their interaction terms. Since we aimed to consider variability across participants and items (for details, see Chapter 3.4.1), all models contained random intercepts for subjects and items as well as by-subject random slopes for the factor condition and by-item random slopes for the factors condition and age group (together with their interaction).

As it is necessary to control for known effects when isolating new effects (Jaffe et al., 2018), we added three control variables to all models. We first controlled for the length of a region (cf. Demberg, 2013; Haeuser et al., 2022; Jaffe et al., 2018; Monsalve et al., 2012) which was defined as the number of characters of the word(s) contained in this region. For instance, given the verb region "verschlingen" the length was twelve, while it was twenty for the post-verb region "am frühen Montagmorgen". We also added trial number (Blumenthal-Dramé, 2021; Haeuser et al., 2020, 2022; Scheffler et al., 2022) and processing time of the previous region (Tucker et al., 2015) as a control.[21] All control variables were scaled and included as additive variables (i.e., without their product terms) as no interactions between control variables and fixed effects (condition, age group) were expected. Non-converging models were simplified with the least-variance approach (Barr et al., 2013). We estimated *p*-values with the Satterthwaite degrees of freedom method (lmerTest library, Kuznetsova et al., 2017) and confidence intervals with the stats library (RStudio Team, 2018).

5.4.2 Statistical Results

We only report results that are relevant for our research questions. The models with their results are shown in Table 16. Independent of the visual condition, children processed each region of interest slower than adults (p-values < .001). While processing times of both age groups did not vary among the visual conditions in the baseline, the verb, and the noun region (p-values > .05), they did so in the post-verb region. Here, processing times of both age groups were significantly slower in the 0-consistent than in all other conditions ($\beta = 0.04$, $SE = 0.01$, $t(1547.21) = 2.97$, $p = .003$, CI [0.01, 0.07]). Besides, both age groups processed the post-verb region significantly faster in the 1-consistent versus the 3- and 4-consistent

[21] The region preceding the baseline was the first article of the sentence. The region preceding the verb was the baseline. The region preceding the post-verb was the verb. The region preceding the noun was the second article of the sentence. The region preceding the post-noun was the noun. Before entering the processing times of both articles, their values were trimmed for extrema (i.e., values below 100 and above 3000 ms were omitted), log-transformed, and scaled separately for each age group.

conditions (β = -0.03, *SE* = 0.01, *t*(163.17) = -2.29, *p* = .023, *CI* [-0.06, 0.00]). In the post-noun-region, both age groups' processing times (*p* = .670) were significantly shorter in the 1-consistent versus the 3- and 4-consistent conditions (β = -0.03, *SE* = 0.02, *t*(160.17) = -2.05, *p* = .042, *CI* [-0.06, 0.00]). This pattern of results is visualized in Figure 13 (for the numerical raw and log-transformed values, see Appendix G).

In sum, and in line with our expectations, processing times did not vary among the visual conditions in the baseline, verb, and noun region. As expected, both age groups read the post-verb region slower in the 0-consistent versus all other conditions. In line with our expectations, the post-verb and the post-noun region were read faster in the 1-consistent than in the 3- and 4-consistent conditions by both age groups.

Table 16

Results of the Models on the Processing Times per Region

Region	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Baseline	Intercept	6.41	0.04	157.03	173.25	.000	6.34, 6.48
	v1	-0.01	0.02	1603.84	-0.58	.564	-0.05, 0.02
	v2	0.00	0.02	44.32	0.20	.842	-0.03, 0.04
	v3	0.03	0.02	1603.28	1.20	.230	-0.02, 0.07
	a	-0.53	0.07	152.73	-7.23	.000	-0.67, -0.38
	length	0.03	0.02	27.62	1.93	.063	0.00, 0.06
	trial	0.00	0.01	28.40	0.05	.963	-0.02, 0.03
	previous	0.03	0.01	1692.11	3.20	.001	0.01, 0.05
	v1:a	0.05	0.04	1601.65	1.26	.209	-0.03, 0.12
	v2:a	0.00	0.04	1600.33	-0.08	.938	-0.07, 0.07
	v3:a	0.02	0.04	1606.54	0.48	.635	-0.06, 0.10
Verb	Intercept	6.61	0.03	154.89	259.01	.000	6.56, 6.66
	v1	0.01	0.02	68.00	0.78	.438	-0.02, 0.05
	v2	0.02	0.02	171.06	1.40	.164	-0.01, 0.05
	v3	0.02	0.02	1626.71	1.30	.193	-0.01, 0.06
	a	-0.63	0.05	138.54	-13.48	.000	-0.72, -0.54
	length	0.05	0.01	27.51	4.23	.000	0.03, 0.07
	trial	0.02	0.01	25.45	1.93	.065	0.00, 0.04
	previous	0.33	0.01	1758.90	25.43	.000	0.30, 0.36
	v1:a	0.00	0.03	1608.95	-0.02	.985	-0.06, 0.06
	v2:a	-0.01	0.03	171.14	-0.44	.659	-0.08, 0.05
	v3:a	0.02	0.04	1634.08	0.61	.543	-0.05, 0.10
Post-Verb	Intercept	7.75	0.03	147.22	272.03	.000	7.69, 7.80
	v1	0.04	0.01	1547.21	2.97	.003	0.01, 0.07
	v2	-0.03	0.01	163.17	-2.29	.023	-0.06, 0.00
	v3	0.01	0.02	133.92	0.54	.589	-0.02, 0.04
	a	-0.53	0.06	134.40	-9.51	.000	-0.64, -0.42
	length	0.01	0.01	26.41	1.55	.132	0.00, 0.03
	trial	0.01	0.01	24.27	0.71	.485	-0.01, 0.02
	previous	0.22	0.01	1903.00	18.69	.000	0.20, 0.25
	v1:a	0.00	0.03	1547.52	0.00	.998	-0.06, 0.06
	v2:a	0.00	0.03	163.14	0.00	.997	-0.06, 0.06
	v3:a	0.01	0.03	133.85	0.19	.847	-0.06, 0.08

Note. Table continued on the next page.

Table 16 (Continued)

Results of the Models on the Processing Times per Region

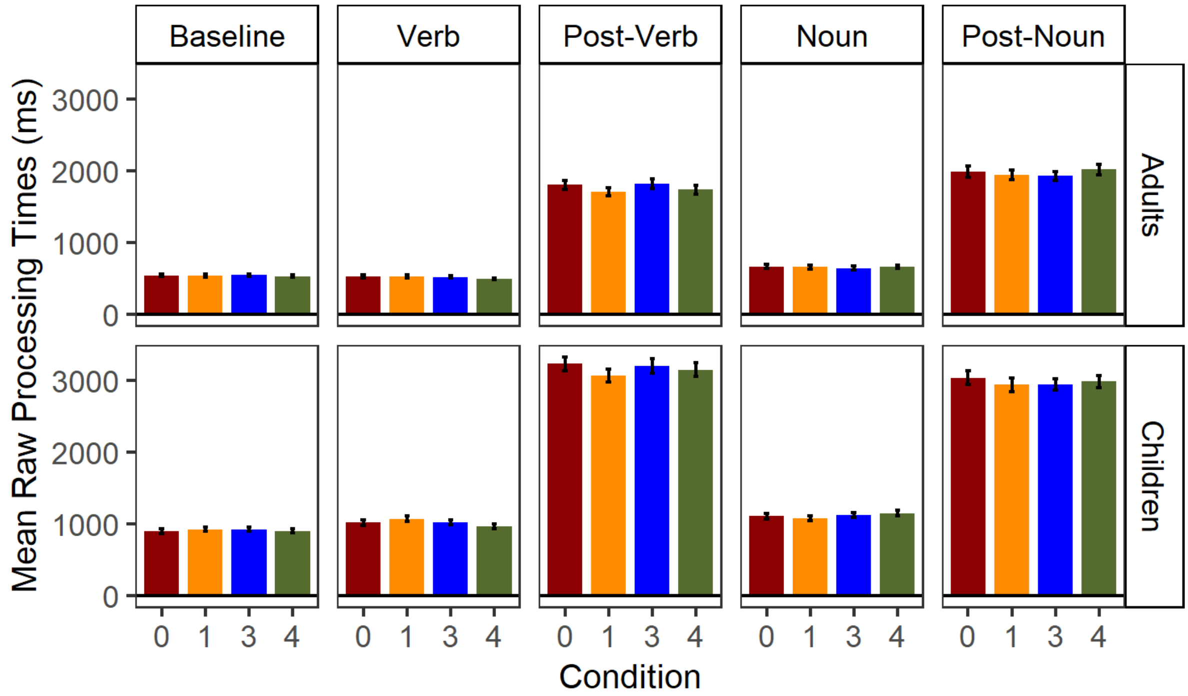
Region	Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Noun	Intercept	6.70	0.03	162.26	227.69	.000	6.64, 6.76
	v1	0.02	0.02	1504.28	1.00	.317	-0.02, 0.06
	v2	-0.03	0.02	167.87	-1.57	.118	-0.07, 0.01
	v3	-0.01	0.02	138.71	-0.65	.515	-0.06, 0.03
	a	-0.55	0.06	149.36	-9.57	.000	-0.66, -0.44
	length	0.04	0.01	24.70	4.02	.000	0.02, 0.06
	trial	0.00	0.01	26.75	-0.36	.723	-0.02, 0.02
	previous	0.21	0.02	1893.97	12.49	.000	0.17, 0.24
	v1:a	-0.02	0.04	1503.57	-0.61	.543	-0.10, 0.05
	v2:a	0.07	0.04	167.79	1.64	.103	-0.01, 0.15
	v3:a	0.01	0.05	138.72	0.26	.797	-0.08, 0.10
Post-Noun	Intercept	7.71	0.03	143.14	271.21	.000	7.65, 7.76
	v1	0.02	0.02	1631.49	1.40	.161	-0.01, 0.05
	v2	-0.03	0.02	160.17	-2.05	.042	-0.06, 0.00
	v3	-0.01	0.02	1638.02	-0.52	.604	-0.04, 0.03
	a	-0.39	0.06	141.89	-6.95	.000	-0.50, -0.28
	length	0.03	0.01	26.69	3.78	.001	0.01, 0.04
	trial	0.01	0.01	28.44	0.91	.372	-0.01, 0.02
	previous	0.15	0.01	1895.47	13.12	.000	0.13, 0.17
	v1:a	0.01	0.03	1630.42	0.19	.848	-0.05, 0.07
	v2:a	-0.01	0.03	160.28	-0.43	.670	-0.07, 0.05
	v3:a	-0.04	0.04	1637.54	-1.19	.233	-0.11, 0.03

Note. The models on the log-transformed processing times (logrt) per region covered the factors age group and condition as well as the control variables length of a region, trial number, and processing time of the previous region. v1, v2, and v3 are the first, second, and third condition contrasts. a is the age group contrast. The converged models:

Baseline: lmer(logrt~(v1+v2+v3)*a+length+trial+previous+(1|subject)+(1+(v2)||item))
Verb: lmer(logrt~(v1+v2+v3)*a+length+trial+previous+(1+(v2)||subject)+(1+(v1)||item))
Post-verb: lmer(logrt~(v1+v2+v3)*a+length+trial+previous+(1+(v2+v3)||subject)+(1|item))
Noun: lmer(logrt~(v1+v2+v3)*a+length+trial+previous+(1+(v2+v3)||subject)+(1+a||item))
Post-noun: lmer(logrt~(v1+v2+v3)*a+length+trial+previous+(1+(v2)||subject)+(1+(v3)*a||item))

Figure 13

Averaged Raw Processing Times per Condition, Region, and Age Group



Note. For ease of comprehension the averaged raw (not log-transformed) processing times are shown. Values in the post-verb and post-noun region are higher than in the other regions as they summarize the processing times of three words each. Error bars indicate the standard error of the mean.

5.5 Discussion

The aim of Experiment 2 was to examine whether the potential a visual contexts contributes to prediction can affect children's and adults' cognitive load to process predictive linguistic cues and predictable target nouns. In this experiment, the linguistic input was presented in written form and thus could enter the comprehenders' cognitive systems via the visual input channel. We conducted a self-paced reading study with literate children (8–12 years) and adults. Participants read sentences with semantically constraining verbs and predictable target nouns in a word-by-word self-paced reading fashion. At the same time, they were presented with visual scenes of four objects each. Across four conditions the scenes varied in predictability: They consisted of 0, 1, 3, or 4 visual prediction options that were consistent with the verb constraints. Word processing times were used as indicator of cognitive load. Variations in processing times across the course of the sentences and the conditions were similar across the age groups. We discuss three notable findings below.

First, it must be noted that children showed longer processing times than adults in all regions of interest. Thus, reading the sentences in the visual scenes may generally have been more demanding for children than for adults. This is not surprising as children usually have smaller reading skills than adults and because children had less language experience than adults (as indicated by their scores in the test of receptive vocabulary). However, this difference in processing times between the age groups was not modulated by the visual conditions and thus does not conflict with our findings. Notably, we found no evidence for variations in both age groups' processing times among the visual conditions in the baseline region where the unpredictive agent, but no constraining linguistic input was yet shown. Thus, there is no reason to assume that the findings below could not derive from the experimental manipulation.

5.5.1 Increase in Cognitive Load to Maintain Multiple Predictions

We found that both age groups engaged higher cognitive load when they could pre-update multiple versus only single noun candidates by the visuo-linguistic constraints. Children and adults showed longer processing times in the spill-over region after the verb (where they still could form and maintain predictions) in the 3- and 4-consistent versus the 1-consistent condition. This is in line with the cognitive view that visual and linguistic contexts are integrated in working memory while an overlap among that information can result in the commitment to and the resource demanding pre-updating of prediction options (e.g., Huettig & Janse, 2016; Özkan et al., 2022). Since working memory involves more resources to update more stimuli (e.g., Johnson et al., 2014; Karatekin, 2004), it is conceivable that we found additional cognitive load in form of longer processing times when readers could pre-update multiple noun candidates. That this effect revealed in the post-verb region (and not yet on the verb) is plausible since effects of predictive processing in reading typically first reveal in spill-over regions (e.g., Smith & Levy, 2013; Vela-Candelas et al., 2022).

Unexpectedly, the above result was not affected by age group. Given their smaller cognitive capacity (e.g., Cowan et al., 2010; Johnson et al., 2014), we expected children to show a higher increase in cognitive load to pre-update multiple nouns than adults. This is because individuals with small cognitive capacity typically engage more cognitive load in a task than others (e.g., Johnson et al., 2014; Just & Carpenter, 1993). We suspect that potential variations among the age groups were not measured sensitively enough to reveal statistical significance since we collected data in a self-monitored online study and could not fully control how carefully participants worked on the task. This may have been particularly crucial for children who may not have been familiar with online studies.

Interestingly, the above result is not consistent with our findings of Experiment 1 or with other studies that used a comparable design (Ankener et al., 2018; Sikos et al., 2021). Here, adults and children did *not* engage higher cognitive load (measured with the ICA and overall pupil sizes) when they could pre-update multiple versus single sentence continuations. This could mean that processing times are more sensitive to that type of cognitive load engaged in pre-updating than pupillometric measures (cf. Ankener et al., 2018). Processing times have often been shown to increase for words that allow to predict input (e.g., Cutter et al., 2021; Frank, 2013; Lowder et al., 2018), and the present study indicates that they also do so when an increasing amount of information can be pre-updated. For pupillometric measures this is rather unclear. So far, they were most often used to show processing benefits for target words that are more or less predictable (e.g., Demberg & Sayeed, 2016; Hochmann & Papeo, 2014), while their sensitivity to the processing costs of constraining input itself was only focused in a few visual world studies (Ankener et al., 2018; Sikos et al., 2021). To reveal whether increases in cognitive load for the pre-updating of multiple versus single prediction options are unique to the measure of processing times or whether this effect holds true also for other measures of cognitive load (except pupillometry), future research could rely on ERP techniques. Since the ERP component P600 can reflect mechanisms of an integration of the mental representations of predicted words into working memory (e.g., Delogu et al., 2019; Kaan et al., 2000), variations in the P600 among situations where one versus multiple words can be pre-updated by the visuo-linguistic constraints could yield answers here.

The inconsistent results among Experiments 1 and 2 could also derive from the different study designs. As it may take some time for a cognitive system to integrate a predictive verb and a complex visual scene in working memory, to pre-activate the related mental representations, and to pre-update one or multiple prediction options, we speculate the following. In Experiment 2 the spill-over region contained three words (e.g., “am frühen Sonntagmorgen”), thus effects of visual condition had sufficient time to translate into processing times prior to target word presentation. In Experiment 1, the sentences included only one spill-over word (“gleich”). Together with the fact that changes in pupil size have a slow latency (Olivia, 2019), this may have prevented effects of pre-updating load to reveal in the pupillometric measures prior to the target word.

Besides, our studies varied in the preview time of the scenes prior to the sentence (2000 ms in Experiment 1, 4000 ms in Experiment 2). Thus, in Experiment 2, the visual objects could be encoded longer in working memory and their mental representations could be manipulated more thoroughly prior to the sentence input. Given this, and in line with the view that only a sufficiently long preview allows for higher level predictive processing (Hintz et al., 2017; Huettig & Guerra, 2019), pre-updating could be carried out at a deeper level of processing in Experiment 2, causing variability in cognitive load among the conditions.

Finally, the different study results could derive from the different modalities of the linguistic input. In Experiment 1 the sentences were played auditorily, while they were shown in written form in Experiment 2. Originally, we expected that effects of higher load to pre-update multiple prediction options would rather reveal for the auditory input. This is because word reading puts demands on visual components of working memory (Pham & Hasson, 2014; Swanson, 2000, 2010; Swanson & Jerman, 2007). Thus, readers in Experiment 2 could temporarily process not only the visual scenes but also the sentences in the visual working memory. Since visual working memory is capacity limited (Cowan et al., 2011; Luck & Vogel, 2013), we expected that those high demands on visual working memory could mask potential condition effects in Experiment 2.

However, the opposite results pattern was found: Effects of increased cognitive load to make multiple predictions did only reveal for the written input modality (Experiment 2). We speculate the following. The auditory sentence input in Experiment 1 may have caused generally higher cognitive load because the linguistic input and the visual contexts entered the cognitive system via different input channels (auditory, visual). Thus, prediction-relevant pieces of information were temporarily maintained in different working memory components before they were transferred to integrative components to form predictions (Özkan et al., 2022). This may have caused a general high amount of cognitive load, thereby masking possible condition effects. In Experiment 2, in contrast, the sentences and the scenes were shown onscreen, thus both entered the cognitive system via the visual input channel. Here, any prediction-relevant information could be temporarily stored and integrated in the same visual components of working memory (e.g., Huettig et al., 2011a, 2011b; Pham & Hasson, 2014; Swanson, 2000, 2010). As a result, effects of increased load to pre-update multiple prediction options could become visible. This reasoning is supported by Sikos et al. (2021) who also found higher cognitive load in adults who pre-updated multiple versus single target words in the equivalent visual conditions when the sentences were shown onscreen, but not when they were played auditorily.

To conclude, at least the results of our self-paced reading study provide indication that cognitive load can increase in children and adults when they pre-update multiple versus only single sentence continuations by the joint visual and linguistic constraints. This strengthens the view that predicting input in the visual world engages working memory resources (e.g., Huettig & Janse, 2016; Özkan et al., 2022) and shows that prediction, which enables a fast and accurate language comprehension (e.g., Huettig, 2015; Mani & Huettig, 2014), is far from being effortless (Liu et al., 2022).

5.5.2 Predictive Visual Contexts Facilitate the Generation of (Multiple) Predictions

We next found that forming predictions was facilitated in children and adults when the visual scenes did versus did not contribute to prediction. Both age groups showed shorter processing times in the post-verb region of the 1-, 3-, and 4-consistent conditions (where specific noun candidates could be predicted) versus the 0-consistent condition (where only the semantic noun category could be predicted). This result extends our findings of Experiment 1 and those of Sikos et al. (2021). Here, adults engaged higher cognitive load to predict input in the 0-consistent scenes (that were ambiguous with the linguistic input) versus the 1-consistent scenes (that allowed to predict one specific sentence continuation). With Experiment 2 we showed for the first time that this processing benefit also holds true in visual scenes that allow to predict multiple sentence continuations which could be explained as follows. The 0-consistent scenes displayed four distractors that were inconsistent with the predicted semantic category. Readers may have tried to resolve that ambiguity and thus engaged higher cognitive load in the 0-consistent versus the 1-, 3-, and 4-consistent scenes (that were not ambiguous with the linguistic input but contributed to prediction). That this result truly reflects the demands of ambiguity resolving is supported by a finding of Ankener et al. (2018, reported in Staudte et al., 2021) who found larger N400 amplitudes for adults on the constraining verbs in the equivalent 0-consistent versus 1-consistent scenes, while the N400 is considered to reflect semantic ambiguity detection (Kutas & Federmeier, 2011).

Notably, with Experiment 2 we show for the first time that also children expended more cognitive load to predict input in ambiguous versus predictive visual contexts. This is notable as children could, due to their limited cognitive capacity (e.g., Cowan et al., 2010; Johnson et al., 2014), engage high cognitive load in both situations (i.e., when resolving ambiguity and when maintaining one or multiple prediction options). However, we found that ambiguity resolving is more resource demanding than pre-updating in children as well. That we did not find such a result in Experiment 1 could derive from the fact that, independent of the visual condition, processing the visuo-linguistic input could have been more difficult for the 5- to 6-year-olds in Experiment 1 than for the 8- to 12-year-olds in Experiment 2.

5.5.3 More Specific Predictions Facilitate Target Word Processing

Third, this study replicated some results of Experiment 1 and of two related studies (Ankener et al., 2018; Sikos et al., 2021) for the predictable target noun. We found that children and adults engaged less cognitive load for the target noun when the visual and linguistic context jointly allowed for a more versus less specific prediction of that noun. That is, both age groups showed shorter processing times in the post-noun region of the 1-consistent versus the 3- and 4-consistent conditions. That this result first revealed in the post-noun region is in line with the fact that effects in reading typically spill over from critical to subsequent words (e.g., Smith & Levy, 2013; Vela-Candelas et al., 2022).

This result can be considered as follows. The 1-consistent scenes allowed for a highly specific prediction of the target noun. Here, readers could make a commitment to the target noun and pre-update it highly thoroughly, resulting in less cognitive load when it is finally revealed (e.g., Freunberger & Roehm, 2017; Maess et al., 2016; Ness & Meltzer-Asscher, 2018, 2021). In the 3- and 4-consistent conditions, in contrast, the visual scenes allowed for a less specific prediction of multiple noun candidates. Here, readers may have engaged higher cognitive load to process the target noun because it was pre-updated less thoroughly together with multiple competitors. This reasoning is in line with prior research that found increases in processing load in children and adults for target words when a visual context did versus did not allow to pre-update that target word (Csink et al., 2021; Fritsche & Höhle, 2015; Krüger et al., 2020; Tamási et al., 2017, 2019). Besides, also Ness and Meltzer-Asscher (2018) showed that adults who pre-update a target word by a predictive linguistic cue engaged less cognitive load (smaller P600 amplitudes) when the anticipated word finally cashed out. In contrast, processing load was higher for words that could not be pre-updated and thus still needed to be processed more thoroughly when they revealed (see also Maess et al., 2016; Freunberger & Roehm, 2017).

However, the above presented result could also be explained with recourse to the demands of inhibition (Gambi et al., 2021; Kuperberg & Jaeger, 2016; van Petten & Luka, 2012). That is, both age groups could have engaged additional processing load for the target noun in the 3- and 4-consistent conditions because they inhibited the competitors that were pre-updated earlier during sentences processing. We cannot rule out whether our results derived from the benefits of pre-updating or from the costs of inhibition, but this could be addressed in ERP studies. While the ERP component P600 is considered to reflect updating processes (e.g., Ness & Meltzer-Asscher, 2018, 2021; Sikos et al., 2021), the late frontal positivity is argued to reflect the inhibition of anticipated input (e.g., DeLong et al., 2011; Höltje et al., 2019; van Petten & Luka, 2012). Considering these ERP components could reveal whether processing costs for target words that could be predicted highly specifically or along with other word candidates may derive from the benefits of pre-updating or from the costs of inhibition.

5.5.4 Limitations

When interpreting our findings, it is important to consider that the present study was conducted online. Therefore, we could not fully control how carefully participants worked on the self-paced reading task. Within the scope of our possibilities, we monitored this: Each sentence was followed by a simple comprehension question and accuracy scores were used to exclude individuals who may not have worked carefully on the task. Besides, we omitted unreasonably high or low processing times. Our results should therefore be as reliable as in other online self-paced reading studies that recently became relevant in the prediction literature (e.g., Haeuser & Kray, 2022; Scheffler et al., 2022). However, it would be useful to conduct future studies, in particular those with children who quickly lose interest and motivation in a study (Brewer, 2013), in laboratory settings.

Notably, Experiment 2 is one of the first studies that applied a self-paced reading task in a Visual World Paradigm. It is therefore not surprising that comprehenders' processing times (independent of the visual condition) were longer (nearly twice as long) than in other self-paced reading studies without visual contexts (e.g., Gibson & Levy, 2016; Haeuser & Kray, 2022; Haeuser et al., 2020). However, this is in line with a finding of Scheffler et al. (2022). Here, adults either were or were not presented with visual input (colored emojis) while they read sentences in a self-paced reading procedure. Their processing times in the visual condition were about twice as long as in the non-visual condition. This was reasoned on the basis that word processing times in a visual world scenario may reflect not only the processing time of that word, but also the time spent on processing the visual input (Scheffler et al., 2022). Thus, for the present study, the dependent variable processing times may index the processing difficulty of the words in the visual scenes rather than only the processing difficulty of the words. However, since we were interested in cognitive load of visually situated predictive processing and since word reading (e.g., Just & Carpenter, 1993; King & Just, 1991; Lewis et al., 2006) and visual information processing (e.g., Huettig & Janse, 2016; Magnuson, 2019; Özkan et al., 2022) should both involve working memory resources, this does not discount our findings.

Finally, and most crucially, we cannot guarantee that children and adults truly relied on the visual and linguistic contexts to form predictions. We did not record eye-movements, therefore cannot refer to anticipatory object fixations as an indicator of prediction. However, participants in Experiment 1 were presented with the same sentences (with shorter spill-over regions) and the identical visual scenes. Here, children's and adults' anticipatory object fixations showed that both age groups predicted one or multiple noun candidates by the joint visual and linguistic constraints (see Chapter 3.4). Since the children in Experiment 1 (5–6 years) were even younger than in the current study (8–12 years) and because it is a robust phenomenon that children from an early age rely on semantical constraints to predict language in the visual world (e.g., Mani & Huettig, 2012, 2014), we assume that the results of our reading study, which are also in line with cognitive models of prediction, are likely to derive from prediction.

5.5.5 Conclusion

This study applied a self-paced reading study in combination with the Visual World Paradigm and used processing times as indicator of cognitive load. We found that both predicting input and processing predictable target words is facilitated in children and adults when a visual context is unambiguous with the linguistic constraints and contributes to prediction. This is even more remarkable because the sentences in Experiment 2 were presented in written form and thus may have been maintained along with the visual scenes in visual components of working memory. That we still found a processing benefit for the constraining verbs and predictable target nouns in the predictive visual conditions suggests that such effects are robust to changes in linguistic input modality and strong enough to show up despite the reading-induced additional demands on visual working memory. However, it remains unclear which cognitive mechanisms these processing benefits are based on (pre-updating, ambiguity resolving, inhibition). This needs to be examined in further studies that should preferably be conducted in laboratory settings and with other/additional measures of cognitive load. However, we can conclude that visual environments have a major impact on the ease of predictive language processing in children and adults. This is relevant given current interests in factors that can facilitate or impede language comprehension for children in particular (e.g., Gambi et al., 2021; Mani & Huettig, 2016).

6 General Discussion

In every-day life, language is usually processed in visual environments. In this way, visual contexts can affect a comprehenders' predictive processing. For instance, listening to the sentence fragment "The girl will *eat* the …" while inspecting a picture of an apple can lead comprehenders to predict "apple" as a possible sentence continuation. This thesis aimed to examine how the prediction of language is influenced by visual contexts that vary in predictability, i.e., in how strongly they contribute to prediction. Specifically, we focused on two main goals. First, we aimed to examine how visual contexts with varying predictability can influence the prediction of sentence continuations. Second, we aimed to reveal how such visual contexts can affect the cognitive demands involved in predictive language processing.

These questions where examined in a developmental testing bed. That is, we compared children and adults in their prediction behavior and cognitive load during visually situated predictive processing. This is because children have less language experience and cognitive capacity than adults, two factors that are — according to cognitive models of prediction — of particular relevance for prediction as they influence the pre-activation of prediction options in long-term memory and the pre-updating of prediction options in working memory.

6.1 Summary of the Findings of Experiments 1 and 2

We conducted two experiments that presented children and adults with predictable sentences (e.g., "The father *eats* the waffle") in visual scenes of four objects that varied in predictability: Either 0, 1, 3, or 4 visual objects were consistent with the semantic verb constraints (e.g., edible). Depending on the visual condition either one (1-consistent), three (3-consistent), or four (4-consistent) noun candidates could be predicted. The 0-consistent condition, only allowed to predict the semantic noun category (e.g., edible). In Experiment 1, the sentences were played auditorily while young children (5–6 years) and adults inspected the scenes. We recorded their eye fixations of the visual objects (as index of prediction behavior) as well as their ICA values and pupil sizes (as index of cognitive load). Experiment 2 used another approach to measure cognitive load. Here, the visual scenes were presented onscreen while literate children (8–12 years) and adults read the sentences in a word-by-word self-paced reading fashion. Word processing times indexed cognitive load.

6.1.1 Even Young Children Make Multiple Predictions in the Complex Visual World

Prior research has consistently shown that children, like adults, rely on semantic cues to predict language. However, this has usually been shown in simple visual scenarios that only allow to predict a single sentence continuation (e.g. Altmann & Kamide, 1999; Mani & Huettig, 2012, 2014). Experiment 1 showed that young children (5–6 years), comparably to adults, follow a *multiple predictions pattern* when predicting language in the complex visual world. As indexed by their anticipatory object fixations, children and adults relied on the joint constraints of the predictive verbs and the visual scenes to predict not only one but even multiple noun candidates. Both age groups even adapted their prediction behavior to the exact number of visual prediction options: They anticipatorily fixated (i.e., predicted) one noun candidate, when the visual scene showed one prediction option, but three or four noun candidates, when the visual scene showed three or four prediction options. This result is in line with prior research for adults (e.g., Ankener et al., 2018), but shows for the first time that young children rely on the visuo-linguistic constraints to predict not only one or two (e.g., Borovsky et al., 2012; Gambi et al., 2021; Mani & Huettig, 2012, 2014; Mani et al., 2016) but even multiple sentence continuations in parallel.

Notably, children did not differ from adults in their ability to make multiple predictions. It could be interesting whether this changes in case they are presented with more than four visual prediction options. Possibly, such even more complex visual scenes can reveal differences in the multiple predictions pattern of children and adults. This could be because children might not, but adults might, due to their higher cognitive abilities, be able to maintain even more prediction options.

Importantly, variations in comprehenders' multiple predictions patterns are not only relevant regarding the comparison of children and adults. From a developmental perspective, it remains unclear how the ability to make multiple predictions develops across childhood with increases in language experience and cognitive capacity. An intriguing question is, for instance, how much prior language experience is needed to be able to predict multiple prediction options and not only the most likely one. In addition, it is yet unclear whether the ability to make multiple predictions increases linearly with increases in cognitive abilities.

While these are some developmental questions that remain open in the field of prediction in the more complex visual world, this study, however, showed that young children aged 5 to 6 years are able to integrate complex visual scenes with a varying number of prediction options into predictive processing as efficiently as adults. This contributes to the view that children do not only passively receive language, but actively rely on input of different modalities to be as prepared as possible for upcoming input. This could be one factor promoting the accurate and rapid language comprehension upon early childhood (e.g., Huettig & Mani, 2016; Mani et al., 2016; van Alphen et al., 2021).

6.1.2 Language Experience Influences Children's Prediction Behavior

In line with a growing interest in factors that modulate prediction (e.g., Huettig, 2015; Pickering & Gambi, 2018), Experiment 1 also examined how children's usage of prediction in the complex visual world is modulated by their language experience. We found that children's receptive vocabulary size was positively associated with their prediction of single and multiple sentence continuations. This is in line with prior studies that found children's prediction ability for single sentence continuations to be positively related with vocabulary size (e.g., Borovsky & Creel, 2014; Borovsky et al., 2012) and other measures of language experience (Mani & Huettig, 2014; Smolík & Bláhová, 2019). However, we show for the first time that children's ability to make *multiple* predictions is also positively associated with their language experience.

6.1.3 Making Multiple Predictions Can Increase Cognitive Load

Another purpose of this thesis was to examine how visual contexts affect children's and adults' cognitive load engaged in predictive processing. Given that visual contexts enable comprehenders to pre-update prediction options in working memory (e.g., Huettig et al., 2011a, 2011b) and since working memory is a capacity limited system that engages more resources when more stimuli are updated (e.g., Johnson et al., 2014; Just & Carpenter, 1993), we examined in two studies whether cognitive load increases when the visuo-linguistic constraints allow to pre-update not only one but multiple sentence continuations.

Contrary to our expectation, but in line with prior research for adults (Ankener et al., 2018; Sikos et al., 2021), Experiment 1 found no indication with two pupillometric measures that children's or adults' cognitive load could vary among situations where a constraining verb and a scene allowed to pre-update one versus multiple target nouns. In Experiment 2, however, children's and adults' cognitive load in form of their processing times increased when they could pre-update multiple noun candidates. Given these inconsistent results, we speculated that pupillometric measures could be less sensitive to the type of cognitive load engaged in the pre-updating of prediction options than processing times (cf. Ankener et al., 2018).

Besides, we suspected that the different modalities of the linguistic input (auditory, visual) in our studies caused the different results. Possibly, listeners in Experiment 1 engaged a general high amount of cognitive load since the visual scenes and the sentences were stored in different working memory components before being transferred to integrative components to form predictions (Özkan et al., 2022). In Experiment 2, in contrast, any prediction-relevant input may have been stored in the same working memory components, causing generally small cognitive load, allowing effects of increases in cognitive load to predict multiple nouns to reveal in the measure of cognitive load. This reasoning hints at the need to examine the influence of input modality on prediction in general and on the cognitive load of prediction. This has, to our knowledge, rarely been focused in the prediction literature so far. Notably, children have small experience with written language, which increases continuously from school age onwards. Developmental studies with literate children of increasing age would therefore be an optimal testing bed to examine the influence of input modality on prediction.

To conclude, while it remains open whether this applies only to the measure of processing times or to situations in which the input channel is the same for all prediction-relevant information, we provide a first indication that children and adults engage higher cognitive load when the visuo-linguistic constraints allow for the pre-updating of multiple versus only single sentence continuations. This shows that visual contexts do not only facilitate prediction but, under certain conditions, also increase the cognitive demands of prediction. This is also indicated by our next finding.

6.1.4 Ambiguous Visual Contexts Make Prediction More Resource Demanding

We found an increase in children's and adults' cognitive load to predict input when the visual scenes were ambiguous with the sentence context (i.e., showed only visual distractors that were inconsistent with the semantic verb constraints) compared to visual scenes that contributed to prediction (i.e., showed visual prediction options). Thus, forming predictions was more demanding when the visual scenes were ambiguous with the received and the predictable input. This additional cognitive load could have resulted from comprehenders trying to resolve the visuo-linguistic ambiguity (Gillis et al., 2014; Kadem et al., 2020) and/or to inhibit the processing of the visual distractors (e.g., Kuperberg & Jaeger, 2016; van Petten & Luka, 2012).

Interestingly, this finding was obtained for adults in both experiments, but only in Experiment 2 for children as well. This could mean that the task of visually situated language processing was generally more demanding for the children in Experiment 1, leading to a cognitive overload which then masked the potential effects of visual condition. While we cannot directly compare children's cognitive load among both studies, we found at least indication that the task of visually situated sentence processing was generally more demanding for children than for adults in Experiment 1 (indicated because children showed larger ICA values and pupil sizes across all visual conditions and regions of the sentences). That children may have experienced a cognitive overload in Experiment 1 could derive from the fact the here, the children were younger (5–6 years) than in Experiment 2 (8–12 years), and thus possibly had less experience with (visually situated) language processing. This view is supported by the fact that children in Experiment 1 had significantly smaller receptive vocabulary size (one measure of language experience) than children in Experiment 2 (see Footnote 20). Besides, it could be that the children in Experiment 1 had smaller cognitive capacity as typical for younger compared to older children. Our empirical data support this reasoning (see Footnote 19). That is, the performance in the Semantic Verbal Fluency Task, a test of cognitive functioning, was significantly worse for the younger (Experiment 1) compared to the older children (Experiment 2). Finally, the task of visually situated language processing could have been more demanding for the children in Experiment 1, since here the sentences and the scenes entered the cognitive system via different input channels. A discussion of this point is provided in Chapters 5.5.1 and 6.1.3.

In sum, we provide an indication that not only the predictability of a linguistic cue but also visual contexts that contribute to prediction can affect the ease of prediction in children and adults. While this possibly depends on cognitive abilities and/or input modalities regarding children, prediction seems to be facilitated by visual signals that are consistent but not ambiguous with the linguistic signal.

6.1.5 More Specific Predictions Facilitate Target Word Processing

Finally, we aimed to uncover how visual contexts affect the processing of predictable target words. To date, we know that children and adults have a processing benefit for words that are predictable over such that are not predictable by purely linguistic (e.g., Cutter et al., 2023; Vergilova et al., 2022; Wassenburg et al., 2015) or purely visual contexts (e.g., Friedrich & Friederici, 2005; Fritsche & Höhle, 2015). We extend these findings. In both of our studies, we found a processing benefit for target words that could be predicted highly specifically versus only less specifically among multiple competitors by the visuo-linguistic constraints. That is, children and adults engaged less cognitive load for a predictable noun when that particular noun versus multiple noun candidates could be predicted. This could mean that the target noun was pre-processed more thoroughly when it was pre-updated alone instead of with multiple competitors. Otherwise, it could be easier to process the noun when there was no need to inhibit multiple competitors that were relevant earlier during sentences processing. However, independent of whether it derived from the benefits of pre-updating or the costs of inhibition, we show with different measures of cognitive load and with different modalities of the linguistic input (auditory, written) that processing is facilitated for target words that can be predicted highly specifically versus those that can be predicted less specifically among multiple competitors by the visuo-linguistic constraints.

While this finding ties in with prior research for adults (e.g. Ankener et al., 2018, Sikos et al., 2021, Tourtouri et al., 2015), it is novel for children. This is notable, because children have limited cognitive capacity and thus could generally experience high cognitive load to process target nouns (independent of visual contexts). In line with this reasoning, we found the above processing benefit only for the older children in Experiment 2 (8–12 years) but not for the younger children in Experiment 1 (5–6 years) who may have been in charge of even smaller cognitive abilities. So, children's processing load for upcoming nouns may be sensitive to visuo-linguistic constraints in case they are already in charge of a certain amount of cognitive abilities. However, for generalizable conclusions, this needs to be validated directly in studies with different age groups and under control of cognitive abilities.

While it remains to be examined which cognitive processes exactly benefit from visual contexts (e.g., pre-updating, inhibition), we found that visual contexts have a great influence on the ease to process predictable target words, even for children. This is of particular relevance for developmental research, because interest in factors that can facilitate language processing in children is growing (e.g. Gambi et al., 2021; Huettig, 2015; Huettig & Mani, 2016).

6.2 Contributions and Future Directions

The findings summarized above provide new insights in the field of predictive language processing in children and adults. In addition, they draw attention to some aspects that still need to be examined regarding the phenomenon of language prediction.

6.2.1 Validity of Cognitive Perspectives of Language Prediction

Recently, cognitive models of prediction became more relevant. According to this view, prediction consists of two mechanisms. When comprehenders receive a predictive linguistic cue, they pre-activate the mental representations of prediction options in long-term memory. When the level of pre-activation reaches a certain threshold (e.g., because commitment to a prediction option is possible), this prediction option is pre-updated in working memory. This means that an online model of the predicted input is maintained and integrated with the continuously incoming speech signal until the utterance is completed (e.g., Ness and Meltzer-Asscher, 2018, 2021; Özkan et al, 2022). This perspective of prediction is relatively novel, and it remains to be verified whether this view can explain the many facets of prediction during online language comprehension. As a first step, it could be investigated whether and under which conditions prediction involves the mechanisms of pre-activation and pre-updating, something that has received little attention to date.

Experiment 1 of this dissertation provides indication that prediction may involve *pre-activation*. Like others (e.g., Borovsky et al., 2012; Brouwer et al., 2017c; Mani & Huettig, 2012, 2014), we found children's prediction behavior to be positively associated with their language experience. Since language experience is defined as the number of linguistic representations stored and linked in long-term memory (e.g., Mani & Huettig, 2012; Zhang et al., 2020), this shows that long-term memory representations, and probably their activation, play a role for prediction. Thus, the children in our study seem to have pre-activated the prediction options onscreen soon after the predictive verb was played.

So far, there is no consensus on which facets of language experience are relevant for prediction. Some studies (like the current work) found effects of receptive (e.g., Borovsky & Creel, 2014; Borovsky et al., 2012), others of productive (Mani & Huettig, 2012) language skills. Future studies should include large test batteries of different facets of language experience to clarify which contents of long-term memory play a role in prediction. Besides, as language experience increases across childhood, while age as such is not an explanatory variable in developmental research (Kray et al., 2023), it is important to examine prediction ability in longitudinal studies. This could allow for more reliable conclusions on how prediction is affected by growth of long-term memory storage.

Finally, ERP studies could uncover whether pre-activation plays a role in prediction. The ERP component N400, for instance, is considered to reflect the semantic retrieval of linguistic representations from long-term memory (e.g., Delogu et al., 2019; Lau et al., 2008). However, most studies to date focused on N400 effects for more versus less predictable words (e.g., for reviews, see Kochari & Flecken, 2019; Nicenboim et al., 2020). Only few studies already reported higher N400 amplitudes for adults on semantically predictive versus unpredictive cues, thereby showing that the pre-activation of prediction options in long-term memory may be involved in making predictions (Freunberger & Roehm, 2017; Maess et al., 2016). To further emphasize the relevance of pre-activation for prediction, comparable studies are needed, preferably also with children or in longitudinal designs and with tests on different facets of language experience.

Notably, in the current dissertation, we also provide first indication that *pre-updating* is relevant for prediction, at least for prediction in the visual world. We found children and adults to engage higher cognitive load when a predictive verb and a visual scene allowed to predict not only one but multiple target nouns. Since working memory typically needs more mental resources when more stimuli are updated (e g., Johnson et al., 2014; Just & Carpenter, 1993), this indicates that the visual scenes caused comprehenders to pre-process (i.e., to pre-update) the prediction options in working memory. However, this finding did only reveal in Experiment 2, but not in Experiment 1 or in related studies (Ankener et al., 2018; Sikos et al., 2021). More research on the role of pre-updating for prediction is needed.

Since pre-updating is argued to be a working memory process (e.g., Ness & Meltzer-Asscher, 2018, 2021; Özkan et al., 2022), it could be helpful to focus on the relation of prediction and working memory (capacity). While we manipulated working memory load by varying the number of visual prediction options, future research could rely on more common working memory manipulations. One could measure prediction in comprehenders that either do or do not work on a secondary working memory task. In case prediction suffers from the secondary task, as has once been shown for adults in a visual world study (Ito et al., 2018a), this would suggest a shared cognitive resource for prediction and working memory. Besides, while we only relied on the assumption that working memory capacity is smaller in children versus adults, it would be more appropriate to control for working memory capacity with suitable measures (e.g., working memory span tasks). In case prediction behavior increases with increasing working memory capacity, as it has been shown for adults (Huettig & Janse, 2012, 2016) and children (Özkan et al., 2022; Zhang & Knoeferle, 2012) in few visual world studies, this would show that working memory capacity can modulate prediction. Both of the above would contribute to the view that pre-updating of prediction options in working memory can play a role for prediction.

Notably, also this line of research should be considered from a developmental perspective. This is because the ability to pre-update future input is argued to depend on working memory capacity, which is known to continuously increases across childhood (e.g., Alloway & Alloway, 2013; Swanson, 1999). It would therefore be interesting to examine whether increases in children's working memory capacity go along with increases in the ability to pre-update input. Such findings would provide further indication for the important role of cognitive resources for prediction.

Notably, the role of working memory for prediction was so far only examined in few visual world studies. It could, however, also be validated whether pre-updating is relevant in purely linguistic contexts, where visual contexts do not support the pre-updating process (Huettig et al., 2011b). Here, ERP studies could be helpful. The ERP component P600, for instance, can reflect the integration of mental representations of prediction options into working memory (Brouwer et al., 2012, 2017c; Delogu et al., 2019; Kaan et al., 2000). So far, one research group has shown that adults show an increase in the P600 when linguistic cues allow for specific predictions, i.e., to pre-update upcoming input (Ness and Meltzer-Asscher, 2018, 2021). This suggests that also purely linguistic contexts can induce pre-updating. But it remains open whether also children, despite their smaller working memory capacity, can pre-update input without supportive visual signals. Comparing children's P600 among situations where a purely linguistic context allows to pre-update specific input (e.g., In the *circus*, children *laugh* the most during the performance of the clown") or to only pre-activate unspecific input in long-term memory (e.g., "In the *circus*, children are most *amazed* during the performance of the clown") could provide insights here.

In sum, our findings are in line with and provide a slight indication for the validity of cognitive models of prediction. However, this still needs to be examined with a broader spectrum of research methods and paradigms. To draw even more reliable conclusions on the cognitive processes that adults, and children in particular, go through during prediction, it is necessary to assess the influence of individual differences in verbal and cognitive abilities on prediction.

6.2.2 Many Open Questions Regarding Language Prediction in Children

Although much research has been conducted on prediction in children some more general questions than those mentioned above are still open. Children predict language from an early age. This was mostly shown by studies using the Visual World Paradigm (e.g., Mani & Huettig, 2012, 2014), which is well suited to study prediction in children as it can be applied with individuals who are not yet literate but still able to predict. Besides, it is argued that it is more ecological valid to examine prediction in visual contexts (compared to purely linguistic contexts), because also real-world prediction is most often visually situated (e.g., Reuter et al., 2020; Venhuizen et al., 2019).

With the Visual World Paradigm most developmental research focused on the questions of *which* linguistic cues (e.g., semantical or morphosyntactical) children at *which* particular age use to predict input of *which* level of representations (e.g., semantical or morphosyntactical). Answers to these questions are reported in numerous studies (for reviews, see Huettig & Mani, 2016; Kray et al., 2023; Pickering & Gambi, 2018).

However, one question only recently received the focus of attention, namely *why* children make use of prediction. At a general level, predictions can facilitate language comprehension as they prepare individuals for and free up resources to process upcoming input, thereby making communication fast and accurate (e.g., Huettig, 2015; Kamide, 2008; Kutas et al., 2010). A deeper understanding of why children predict is, however, still needed. One line of reasoning views prediction as a mechanism that contributes to language learning (e.g., Dell & Chang, 2014; Rabagliati et al., 2016). Here, children are argued to compare their predictions with the actual input they receive. Unfulfilled predictions (i.e., prediction errors) are then used to update linguistic knowledge base and to tune the ability to make successful predictions in future (e.g., Rabagliati et al., 2016). This is how prediction-error based learning is theorized to work, but numerous questions still are open.

It is, for instance, largely unknown how much prior language learning is required before prediction can take place and support further language learning. This could be tested in studies that present children with predictable sentences with unexpected novel target words. Later recognition tests of these words could show whether prediction error contributed to target word recognition (cf. Borovsky et al., 2010, 2012; Vergilova et al., 2022). Here, it is also interesting to reveal how many exposures of a prediction error children need in order to recognize the novel input in later recognition tests. Do children, as it has been shown for adults (Borovsky et al., 2010, 2012), reveal effects of one-shot learning? It also remains open how much exposures of a prediction error children need so that mental representations of novel words are even stored in long-term memory. Finally, the above questions need to be examined under control of linguistic and/or general knowledge base as well as cognitive abilities. This is because children may only experience prediction error-based learning in case they have a certain level of knowledge that enables them to predict input or in case their cognitive skills allow them to maintain predictions until an utterance is completed and prediction error occurs (Kray et al., 2023). The above questions could be examined in longitudinal studies that intentionally induce prediction error at different levels of representations, while controlling for linguistic and/or general knowledge base as well as cognitive abilities in order to reveal at what age and under which conditions children can use prediction error to recognize novel words or to even store them in long-term memory.

Overall, new paradigms are needed to shed light on when children begin to predict language and under which conditions they improve knowledge base by prediction error based learning. This could add to our knowledge of why prediction during online language comprehension plays a role in children’s efforts to make sense of the diverse and complex linguistic input they receive in every-day life.

7 Conclusion

This dissertation showed that even young children, despite their limited language experience and cognitive capacity, integrate constraining linguistic cues with complex visual contexts as efficiently as adults to predict not only one but even multiple sentence continuations in parallel. Thus, children do not only passively receive language but rely on any information that can help to prepare for upcoming input. Since children's prediction behavior increased with increasing language experience, individual differences in verbal skills should be considered when studying prediction. While it depended on the modality of the linguistic input whether comprehenders engaged more mental resources to predict more sentence continuations, we found cross-modal evidence for a processing benefit in adults for words that can be predicted more versus less specifically by the visuo-linguistic constraints. Regarding children, this benefit was only obtained for an older sample, and thus may depend on individuals' general cognitive capabilities. In sum, we showed that prediction is possible early in life but seems to be affected by situational factors (e.g., properties of visual context, modality of linguistic input) and individual factors such as verbal abilities. More research is needed to bring light into these questions, in particular for children, because it is still not fully uncovered how children from early age comprehend language while continuously improving their linguistic abilities until they are in charge of a fully developed, adult-like language comprehension system. Examining the many facets of prediction could help to uncover how children process language and what factors they consider to facilitate language comprehension. This could contribute to our understanding of how children learn the complexity of language.

8 Annotations

8.1 Ethics and Funding

All studies were carried out in accordance with the recommendations of the American Psychological Association. All adult participants and all parents of the child participants gave written informed consent in accordance with the Declaration of Helsinki. The studies were approved by the ethics committee of Saarland University (protocol ID: 18-09) and of Deutsche Gesellschaft für Sprachwissenschaft (DGfS, protocol ID: 2017-07-180423).

8.2 Funding

All reported studies were funded by the Deutsche Forschungsgemeinschaft (DFG) — Sonderforschungsbereich 1102 (project number: 232722074).

8.3 Use of the Pronoun "We"

Since the reported studies were designed and conducted by a research group, the pronouns "we" and "our" were used in this work (as usual for English research manuscripts). However, this thesis was written, and data were analyzed individually by Linda Sommerfeld.

8.4 Connection With Other Publications

Parts of Chapters 1 and 3 are copied or adapted from Sommerfeld et al. (2023). Parts of Chapter 2 are copied or adapted from Sommerfeld et al. (2022).

Sommerfeld, L., Staudte, M., & Kray, J. (2022). Ratings of name agreement and semantic categorization of 247 colored clipart pictures by young German children. *Acta Psychologica*, *226*, 103558. https://doi.org/10.1016/j.actpsy.2022.103558

Sommerfeld, L., Staudte, M., Mani, N., & Kray, J. (2023). Even young children make multiple predictions in the complex visual world. *Journal of Experimental Child Psychology*, *235*, 105690. https://doi.org/10.1016/j.jecp.2023.105690

9 References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Alloway, T. P., & Alloway, R. G. (2013). Working memory across the lifespan: A cross-sectional approach. *Journal of Cognitive Psychology*, *25*(1), 84–93. https://doi.org/10.1080/20445911.2012.748027

Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., Van De Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*(4), 1–20. https://doi.org/10.1167/14.4.1

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Andreu, L., Sanz-Torrent, M., & Trueswell, J. C. (2013). Anticipatory sentence processing in children with specific language impairment: Evidence from eye movements during listening. *Applied Psycholinguistics*, *34*(1), 5–44. https://doi.org/10.1017/S0142716411000592

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

Arias-Trejo, N., Falcón, A., & Alva-Canto, E. A. (2013). The gender puzzle: Toddlers' use of articles to access noun information. *Psicológica*, *34*(1), 1–23.

Audacity Team. (2019). *Audacity, die kostenlose, quelloffene, plattformübergreifende Software zur Aufnahme und Bearbeitung von Klängen* (Version 2.3.2) [Computer software]. Audacity. https://www.audacityteam.org

Aumeistere, A., Bultena, S., & Brouwer, S. (2022). Wisdom comes with age? The role of grammatical gender in predictive processing in Russian children and adults. *Applied Psycholinguistics*, *43*(4), 867–887. https://doi.org/10.1017/S0142716422000170

Aurnhammer, C., Delogu, F., Schulz, M., Brouwer, H., & Crocker, M. W. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PLOS ONE, 16*(9), e0257430. https://doi.org/10.1371/journal.pone.0257430

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364–6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

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

Bayeen, H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Bayliss, D. M., Jarrold, C., Baddeley, A. D., & Gunn, D. M. (2005). The relationship between short-term memory and working memory: Complex span made simple? *Memory*, *13*(3/4), 414–421. https://doi.org/10.1080/09658210344000332

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92. https://doi.org/10.1037/0096-3445.132.1.71

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Beck, S. R., Robinson, E. J., & Freeth, M. M. (2008). Can children resist making interpretations when uncertain? *Journal of Experimental Child Psychology*, *99*(4), 252–270. https://doi.org/10.1016/j.jecp.2007.06.002

Becker, J. T., & Morris, R. G. (1999). Working Memory(s). *Brain and Cognition*, *41*(1), 1–8. https://doi.org/10.1006/brcg.1998.1092

Bell, N., Angwin, A. J., Arnott, W. L., & Wilson, W. J. (2019). Semantic processing in children with cochlear implants: Evidence from event-related potentials. *Journal of Clinical and Experimental Neuropsychology*, *41*(6), 576–590. https://doi.org/10.1080/13803395.2019.1592119

Benau, E. M., Morris, J. & Couperus, J. W. (2011). Semantic processing in children and adults: Incongruity and the N400. *Journal of Psycholinguistic Research*. *40*, 225–239. https://doi.org/10.1007/s10936-011-9167-1

Bialystok, E., & Poarch, G. J. (2014). Language experience changes language and cognitive ability. *Zeitschrift Für Erziehungswissenschaft*, *17*(3), 433–446. https://doi.org/10.1007/s11618-014-0491-8

Birchenough, J. M. H., Davies, R., & Connelly, V. (2017). Rated age-of-acquisition norms for over 3,200 German words. *Behavior Research Methods*, *49*(2), 484–501. https://doi.org/10.3758/s13428-016-0718-0

Blumenthal-Dramé, A. (2021). The online processing of causal and concessive relations: Comparing native speakers of English and German. *Discourse Processes*, *58*(7), 642–661. https://doi.org/10.1080/0163853X.2020.1855693

Bobb, S. C., Huettig, F., & Mani, N. (2016). Predicting visual information during sentence processing: Toddlers activate an object's shape before it is mentioned. *Journal of Experimental Child Psychology*, *151*, 51–64. https://doi.org/10.1016/j.jecp.2015.11.002

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Version 6.0.37) [Computer software]. https://www.praat.org

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 158–167. https://doi.org/10.3758/BF03195507

Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600–1613. https://doi.org/10.1037/a0035591

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289–296. https://doi.org/10.1016/j.cognition.2010.05.004

Borys, M., & Plechawska-Wójcik, M. (2017). Eye-tracking metrics in perception and visual attention research. *European Journal of Medical Technologies, 3*(16), 11–23.

Bosch, J. E., Chailleux, boshM., Yee, J., Guasti, M. T., & Arosio, F. (2022). Prediction on the basis of gender and number in Mandarin-Italian bilingual children. In D. Ayoun (Ed.), *The acquisition of gender: Crosslinguistic perspectives* (Vol. 63, pp. 243–272). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.63.10bos

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116. https://doi.org/10.1016/j.jml.2014.05.004

Brandeker, M. (2017). *Beyond vocabulary size: Depth of vocabulary in bilingual children and the role of language experience* [Doctoral dissertation, McGill University]. eScholarship@McGill. https://escholarship.mcgill.ca/concern/theses/8049g755c

Brandeker, M., & Thordardottir, E. (2023). Verbal fluency in bilingual school-age children: Looking at switching, clustering, and the effect of language experience. *International Journal of Bilingualism*, *27*(3), 374–393. https://doi.org/10.1177/13670069211055263

Brewer, R., Anthony, L., Brown, Q., Irwin, G., Nias, J., & Tate, B. (2013). Using gamification to motivate children to complete empirical studies in lab environments. In J. P. Hourcade, N. Sawhney, & E. Reardon (Eds.), *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 388–391). Association for Computing Machinery. https://doi.org/10.1145/2485760.2485816

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149. https://doi.org/10.1016/j.cognition.2014.10.017

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017c). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*, 1318-1352. https://doi.org/10.1111/cogs.12461

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Brouwer, S., Özkan, D., & Küntay, A. C. (2017b). Semantic prediction in monolingual and bilingual children. In E. Blom, L. Cornips, & J. Schaeffer (Eds.), *Cross-linguistic influence in bilingualism: In honor of Aafke Hulk* (Vol. 52, pp. 49–74). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.52.04bro

Brouwer, S., Özkan, D., & Küntay, A. C. (2018). Verb-based prediction during language processing: The case of Dutch and Turkish. *Journal of Child Language*, *46*(1), 80–97. https://doi.org/10.1017/S0305000918000375

Brouwer, S., Sprenger, S., & Unsworth, S. (2017a). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, *159*, 50–65. https://doi.org/10.1016/j.jecp.2017.01.007

Carnerero, J. J., & Pérez-González, L. A. (2014). Induction of naming after observing visual stimuli and their names in children with autism. *Research in Developmental Disabilities*, *35*(10), 2514–2526. https://doi.org/10.1016/j.ridd.2014.06.004

Casillas, M., & Frank, M. C. (2012). Cues to turn boundary prediction in adults and preschoolers. In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 61–69). SemDial. http://semdial.org/anthology/Z12-Casillas_semdial_0010.pdf

Casillas, M., & Frank, M. C. (2013). The development of predictive processes in children's discourse understanding. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 299–304). Cognitive Science Society. https://dblp.org/rec/conf/cogsci/CasillasF13

Chapman, L. R., & Hallowell, B. (2021). The unfolding of cognitive effort during sentence processing: Pupillometric evidence from people with and without aphasia. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4900–4917. https://doi.org/10.1044/2021_JSLHR-21-00129

Cholewa, J., Neitzel, I., Bürsgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: An eye-tracking study with monolingual German 3rd and 4th graders. *Frontiers in Psychology*, *10*, 2586. https://doi.org/10.3389/fpsyg.2019.02586

Clahsen, H., Lück, M., & Hahne, A. (2007). How children process over-regularizations: Evidence from event-related brain potentials. *Journal of Child Language*, *34*(3), 601–622. https://doi.org/10.1017/S0305000907008082

Coco, M. I., Keller, F., & Malcolm, G. L. (2016). Anticipation in real-world scenes: The role of visual context and visual memory. *Cognitive Science*, *40*(8), 1995–2024. https://doi.org/10.1111/cogs.12313

Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading*, *19*(2), 114–134. https://doi.org/10.1080/10888438.2014.943905

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science*, *14*(5), 1066–1074. https://doi.org/10.1111/j.1467-7687.2011.01060.x

Csink, V., Mareschal, D., & Gliga, T. (2021). Does surprise enhance infant memory? Assessing the impact of the encoding context on subsequent object recognition. *Infancy*, *26*(2), 303–318. https://doi.org/10.1111/infa.12383

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Cutter, M. G., Paterson, K. B., & Filik, R. (2023). Syntactic prediction during self-paced reading is age invariant. *British Journal of Psychology*, *114*(1), 39–53. https://doi.org/10.1111/bjop.12594

Debruille, J. B. (2007). The N400 potential could index a semantic inhibition. *Brain Research Reviews*, *56*(2), 472–477. https://doi.org/10.1016/j.brainresrev.2007.10.001

Deevy, P., & Leonard, L. B. (2018). Sensitivity to morphosyntactic information in preschool children with and without developmental language disorder: A follow-up study. *Journal of Speech, Language, and Hearing Research*, *61*(12), 3064–3074. https://doi.org/10.1044/2018_JSLHR-L-18-0038

Deevy, P., Leonard, L. B., & Marchman, V. A. (2017). Sensitivity to morphosyntactic information in 3-year-old children with typical language development: A feasibility study. *Journal of Speech, Language, and Hearing Research*, *60*(3), 668–674. https://doi.org/10.1044/2016_JSLHR-L-15-0153

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, *135*, 103569. https://doi.org/10.1016/j.bandc.2019.05.007

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207. https://doi.org/10.1111/j.1469-8986.2011.01199.x

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Demberg, V. (2013). Pupillometry: The Index of Cognitive Activity in a dual-task study. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society. https://dblp.org/rec/conf/cogsci/Demberg13

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Demberg, V., Kiagia, E., & Sayeed, A. (2013). The Index of Cognitive Activity as a measure of linguistic processing. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Cognitive Science Society. https://dblp.org/rec/conf/cogsci/DembergKS13

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLOS ONE, 11*(1), e0146194. https://doi.org/10.1371/journal.pone.0146194

Dlugosch, C., Conti, A. S., & Bengler, K. (2013, June 17–20). *Driver distraction through conversation measured with pupillometry* [Paper presentation]. International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, New York, United States of America. https://doi.org/10.17077/drivingassessment.1488

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice* (3rd ed.). Springer International Publishing. https://doi.org/10.1007/978-3-319-57883-5

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., & Giannopoulos, I. (2018). The Index of Pupillary Activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In R. Mandryk, M. Hancock, M. Perry, & A. L. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3173574.3173856

Dudenredaktion. (n.d.). Literaturverzeichnis. In *Duden online Synonyme*. Last access December 14, 2021, https://www.duden.de/synonyme

Dudenredaktion. (n.d.). Literaturverzeichnis. In *Duden online*. Last access December 9, 2021, https://www.duden.de

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, *74*, 1–15. https://doi.org/10.1016/j.jml.2014.03.006

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69–91. https://doi.org/10.1016/j.dcn.2016.11.001

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

EyeTracking. (2016). *EyeWorks Cognitive Workload Module* (Version 3.20.6108.25856) [Computer software]. EyeTracking Incorporation. https://www.eyetracking.com

Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2016). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *Quarterly Journal of Experimental Psychology*, *70*(3), 413–433. https://doi.org/10.1080/17470218.2015.1131310

Favier, S., Meyer, A. S., & Huettig, F. (2021). Literacy can enhance syntactic prediction in spoken language processing. *Journal of Experimental Psychology: General*, *150*(10), 2167–2174. https://doi.org/10.1037/xge0001042

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*, *7*(11), 180877. https://doi.org/10.1098/rsos.180877

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & 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. https://doi.org/10.1006/jmla.1999.2660

Fehringer, B. (2021). Optimizing the usage of pupillary based indicators for cognitive workload. *Journal of Eye Movement Research*, *14*(2). https://doi.org/10.16910/jemr.14.2.4

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology*, *60*(3), 190–217. https://doi.org/10.1016/j.cogpsych.2009.12.002

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, *27*(6), 443–448. https://doi.org/10.1177/0963721418794491

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469. https://doi.org/10.1037/a0036756

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494. https://doi.org/10.1111/tops.12025

Frank, S. L., 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. https://doi.org/10.1016/j.bandl.2014.10.006

Frank, S., & Thompson, R. (2012). Early effects of word surprisal on pupil size during reading. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th*

Annual Meeting of the Cognitive Science Society (pp. 1554–1559). Cognitive Science Society.

Freunberger, D., & Roehm, D. (2017). The costs of being certain: Brain potential evidence for linguistic preactivation in sentence processing. *Psychophysiology, 54*(6), 824–832. https://doi.org/10.1111/psyp.12848

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*, 1465–1477. https://doi.org/10.1162/0898929042304705

Friedrich, M., & Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, *17*(11), 1785–1802. https://doi.org/10.1162/089892905774589172

Friesen, D. C., Luo, L., Luk, G., & Bialystok, E. (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language, Cognition and Neuroscience*, *30*(3), 238–250. https://doi.org/10.1080/23273798.2014.918630

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 862–877. https://doi.org/10.1037/0278-7393.31.5.862

Fritzsche, T., & Höhle, B. (2015). Phonological and lexical mismatch detection in 30-month-olds and adults measured by pupillometry. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. The University of Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0339.pdf

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, *173*, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012

Gambi, C., Jindal, P., Sharpe, S., Pickering, M. J., & Rabagliati, H. (2021). The relation between preschoolers' vocabulary development and their ability to predict and recognize words. *Child Development*, *92*(3), 1048–1066. https://doi.org/10.1111/cdev.13465

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*, 340–351. https://doi.org/10.1016/j.cognition.2016.10.003

Gardini, A., Trivisano, C., & Fabrizi, E. (2021). Bayesian analysis of ANOVA and mixed models on the log-transformed response variable. *Psychometrika*, *86*(2), 619–641. https://doi.org/10.1007/s11336-021-09769-y

Garrido-Pozú, J. J. (2022). Predictive processing of grammatical gender: Using gender cues to facilitate processing in Spanish. *Lingua*, *278*, 103416. https://doi.org/10.1016/j.lingua.2022.103416

Gavas, R., Chatterjee, D., & Sinha, A. (2017, October 5–8). *Estimation of cognitive load based on the pupil size dilation* [Paper presentation]. International Conference on Systems, Man, and Cybernetics, Banff Center, Banff, Canada. https://doi.org/10.1109/SMC.2017.8122826

Gibson, E., & Levy, R. (2016). *An attempted replication of Hackl, Koster-Hale, Varvoutis (2012)*. arXiv: https://doi.org/10.48550/ARXIV.1605.00178

Gillis, R., & Nilsen, E. S. (2014). Cognitive flexibility supports preschoolers' detection of communicative ambiguity. *First Language*, *34*(1), 58–71. https://doi.org/10.1177/0142723714521839

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. https://doi.org/10.3758/CABN.10.2.252

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing*, *18*, 583–616. https://doi.org/10.1007/s11145-004-7109-0

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Green, C. (2017). Usage-based linguistics and the magic number four. *Cognitive Linguistics*, *28*(2), 209–237. https://doi.org/10.1515/cog-2015-0112

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283. https://doi.org/10.3758/BF03204386

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *The Journal of Neuroscience*, *38*(35), 7585–7599. https://doi.org/10.1523/JNEUROSCI.0065-18.2018

Haeuser, K. I., & Kray, J. (2022). How odd: Diverging effects of predictability and plausibility violations on sentence reading and word memory. *Applied Psycholinguistics*, *43*(5), 1193–1220. https://doi.org/10.1017/S0142716422000364

Haeuser, K. I., Kray, J., & Borovsky, A. (2020). Great expectations: Evidence for graded prediction of grammatical gender. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1157–1163). Cognitive Science Society. https://cognitivesciencesociety.org/cogsci-2020/

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899. https://doi.org/10.1162/089892903322370807

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children’s language development. *Journal of Cognitive Neuroscience*, *16*(7), 1302–1318. https://doi.org/10.1162/0898929041920504

Häuser, K. I., Demberg, V., & Kray, J. (2018). Surprisal modulates dual-task performance in older adults: Pupillometry shows age-related trade-offs in task performance and time-course of language processing. *Psychology and Aging*, *33*(8), 1168–1180. https://doi.org/10.1037/pag0000316

Häuser, K. I., Demberg, V., & Kray, J. (2019). Effects of aging and dual-task demands on the comprehension of less expected sentence continuations: Evidence from pupillometry. *Frontiers in Psychology*, *10*, 709. https://doi.org/10.3389/fpsyg.2019.00709

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1352–1374. https://doi.org/10.1037/xlm0000388

Hintz, F., Meyer, A. S., & Huettig, F. (2020). Visual context constrains language-mediated anticipatory eye movements. *Quarterly Journal of Experimental Psychology, 73*(3), 458–467. https://doi.org/10.1177/1747021819881615

Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological Science*, *25*(11), 2038–2046. https://doi.org/10.1177/0956797614547918

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*(2–3), 203–241. https://doi.org/10.1080/87565649209540525

Höltje, G., Lubahn, B., & Mecklinger, A. (2019). The congruent, the incongruent, and the unexpected: Event-related potentials unveil the processes involved in schematic encoding. *Neuropsychologia*, *131*, 285–293. https://doi.org/10.1016/j.neuropsychologia.2019.05.013

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/0267658315624960

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., & Verhaeghen, P. (2004). Adult age and Digit Symbol Substitution performance: A meta-analysis. *Psychology and Aging*, *19*(1), 211–214. https://doi.org/10.1037/0882-7974.19.1.211

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia-evidence from eye-tracking: Word reading and predictive language processing. *Dyslexia*, *21*(2), 97–122. https://doi.org/10.1002/dys.1497

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Huettig, F., & Janse, E. (2012, September 6–8). *Anticipatory eye movements are modulated by working memory capacity: Evidence from older adults* [Poster presentation]. 18th Conference on Architectures and Mechanisms for Language Processing, Riva del Garda, Italy. https://www.mpi.nl/publications/item_1563495

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011a). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Huettig, F., Rommers, J., & Meyer, A. S. (2011b). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, A., Corley, M., & Pickering, M. J. (2018a). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. https://doi.org/10.1017/S1366728917000050

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language,* *Cognition and Neuroscience*, *32*(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761

Ito, A., Pickering, M., & Corley, M. (2018b). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Jaffe, E., Shain, C., & Schuler, W. (2018). Coreference and focus in reading times. In A. Sayeed, C. Jacobs, T. Linzen, & M. Van Schijndel (Eds.), *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics* (pp. 1–9). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-0101

Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L., & Bunge, S. A. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in Psychology*, *5*, 218. https://doi.org/10.3389/fpsyg.2014.00218

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. https://doi.org/10.1016/j.neuron.2015.11.028

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310–339. https://doi.org/10.1037/h0078820

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kadem, M., Herrmann, B., Rodd, J. M., & Johnsrude, I. S. (2020). Pupil dilation is sensitive to semantic ambiguity and acoustic degradation. *Trends in Hearing*, *24*, 1–16. https://doi.org/10.1177/2331216520964068

Kamide, Y. (2008). Anticipatory processes in sentence processing: Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *International Journal of Psychophysiology*, *52*(1), 7–21. https://doi.org/10.1016/j.ijpsycho.2003.12.002

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*(2), 175–185. https://doi.org/10.1111/j.1469-8986.2004.00147.x

Karatekin, C., Marcus, D. J., & Couperus, J. W. (2007). Regulation of cognitive resources during sustained attention and working memory in 10-year-olds and adults. *Psychophysiology*, *44*(1), 128–144. https://doi.org/10.1111/j.1469-8986.2006.00477.x

Karbach, J., & Kray, J. (2007). Developmental changes in switching between mental task sets: The influence of verbal labeling in childhood. *Journal of Cognition and Development*, *8*(2), 205–236. https://doi.org/10.1080/15248370701202430

Karbach, J., Kray, J., & Hommel, B. (2011). Action-effect learning in early childhood: Does language matter? *Psychological Research*, *75*(4), 334–340. https://doi.org/10.1007/s00426-010-0308-1

Katidioti, I., Borst, J. P., & Taatgen, N. A. (2014). What happens when we switch tasks: Pupil dilation in multitasking. *Journal of Experimental Psychology: Applied*, *20*(4), 380–396. https://doi.org/10.1037/xap0000031

Kavé, G. (2006). The development of naming and word fluency: Evidence from Hebrew-speaking children between ages 8 and 17. *Developmental Neuropsychology*, *29*(3), 493–508. https://doi.org/10.1207/s15326942dn2903_7

Kavé, G., & Sapir-Yogev, S. (2023). Differences between semantic and phonemic verbal fluency in adolescents with reading disorders. *Archives of Clinical Neuropsychology*, *38*(1), 126–130. https://doi.org/10.1093/arclin/acac062

Kharitonova, M., Winter, W., & Sheridan, M. A. (2015). As working memory grows: A developmental account of neural bases of working memory capacity in 5- to 8-year old children and adults. *Journal of Cognitive Neuroscience*, *27*(9), 1775–1788. https://doi.org/10.1162/jocn_a_00824

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions: Toddlers use disfluencies to predict referential intentions. *Developmental Science*, *14*(4), 925–934. https://doi.org/10.1111/j.1467-7687.2011.01049.x

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of*

Experimental Psychology: General, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, *5*(1), 115–132. https://doi.org/10.22599/jesla.79

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. https://doi.org/10.1080/23273798.2018.1524500

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear & Hearing*, *33*(2), 291–300. https://doi.org/10.1097/AUD.0b013e3182310019

Koornneef, A. (2021). The processing signature of anticipatory reading: An eye-tracking study on lexical predictions. *Linguistics*, *59*(2), 449–479. https://doi.org/10.1515/ling-2021-0014

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*(4), 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development, 2*(1), 1–25. https://doi.org/10.1207/s15473341lld0201_1

Kray, J., Eber, J., & Karbach, J. (2008). *Verbal self-instructions in task switching: A compensatory tool for action-control deficits in childhood and old age? 11*(2), 223–236. https://doi.org/10.1111/j.1467-7687.2008.00673.x

Kray, J., Eenshuistra, R., Kerstner, H., Weidema, M., & Hommel, B. (2006). Language and action control: The acquisition of action goals in early childhood. *Psychological Science*, *17*(9), 737–741. https://doi.org/10.1111/j.1467-9280.2006.01774.x

Kray, J., Sommerfeld, L., Borovsky, A., & Haeuser, K. (2023). *Prediction error in language processing: A driving force for the development of language and memory?* [Manuscript submitted for publication]. Department of Psychology, Saarland University.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, *13*(9), e0203629. https://doi.org/10.1371/journal.pone.0203629

Krüger, M., Bartels, W., & Krist, H. (2020). Illuminating the dark ages: Pupil dilation as a measure of expectancy violation across the life span. *Child Development*, *91*(6), 2221–2236. https://doi.org/10.1111/cdev.13354

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kursawe, M. A., & Zimmer, H. D. (2015). Costs of storing colour and complex shape in visual working memory: Insights from pupil size and slow waves. *Acta Psychologica*, *158*, 67–77. https://doi.org/10.1016/j.actpsy.2015.04.004

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Landgraf, S., Meer, E., & Krueger, F. (2010). Cognitive resource allocation for neural activity underlying mathematical cognition: A multi-method study. *ZDM — Mathematics Education*, *42*(6), 579–590. https://doi.org/10.1007/s11858-010-0264-7

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and entropy in spoken word recognition: Effects of age and hearing acuity. *Experimental Aging Research*, *39*(3), 235–253. https://doi.org/10.1080/0361073X.2013.779175

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lee, R., Chambers, C. G., Huettig, F., & Ganea, P. A. (2022). Children's and adults' use of fictional discourse and semantic knowledge for prediction in language processing. *PLOS ONE*, *17*(4), e0267297. https://doi.org/10.1371/journal.pone.0267297

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition*, *26*(1), 21–45. https://doi.org/10.1080/10489223.2017.1391815

Lenhard, A., Lenhard, W., Suggate, S., Segerer, R. (2015). *Peabody Picture Vocabulary Test — 4. Ausgabe*. Pearson.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, *40*(6), 1382–1411. https://doi.org/10.1111/cogs.12274

Liu, Y., Hintz, F., Liang, J., & Huettig, F. (2022). Prediction in challenging situations: Most bilinguals can predict upcoming semantically-related words in their L1 source language when interpreting. *Bilingualism: Language and Cognition*, *25*(5), 801–815. https://doi.org/10.1017/S1366728922000232

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika*, *76*(3), 487–503. https://doi.org/10.1007/s11336-011-9211-y

Lowder, M. W., Choi, W., Ferreira, F., & Henderson, J. M. (2018). Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science*, *42*, 1166–1183. https://doi.org/10.1111/cogs.12597

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, *146*, 349–370. https://doi.org/10.1016/j.cognition.2015.10.012

Lum, J. A. G., Youssef, G. J., & Clark, G. M. (2017). Using pupillometry to investigate sentence comprehension in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1648–1660. https://doi.org/10.1044/2017_JSLHR-L-16-0158

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*, 591. https://doi.org/10.3389/fnhum.2016.00591

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(2), 113–139. https://doi.org/10.1007/s41809-019-00035-3

Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardena, G., Chukoskie, L., Snider, J., & Jayarathna, S. (2022). Eye movement and pupil measures: A review. *Frontiers in Computer Science*, *3*, 733531. https://doi.org/10.3389/fcomp.2021.733531

Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, *142*, 345–350. https://doi.org/10.1016/j.cognition.2015.05.009

Mani, N., Daum, M. M., & Huettig, F. (2016). “Proactive” in many ways: Developmental evidence for a dynamic pluralistic approach to prediction. *Quarterly Journal of Experimental Psychology*, *69*(11), 2189–2201. https://doi.org/10.1080/17470218.2015.1111395

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake — But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

Mani, N., Mills, D. L., & Plunkett, K. (2012). Vowels in early words: An event-related potential study. *Developmental Science*, *15*(1), 2–11. https://doi.org/10.1111/j.1467-7687.2011.01092.x

Marshall, S. P. (2000). *Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity* (U.S. Patent No. 6,090,051). U.S. Patent and Trademark Office.

Marshall, S. P. (2002). The Index of Cognitive Activity: Measuring cognitive workload. In J. J. Persensky, B. Hallbert, & H. Blackman (Eds.), *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants* (pp. 7-5–7-9). The Institute of Electrical and Electronic Engineers Incorporation. https://doi.org/10.1109/HFPP.2002.1042860

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*, 1079. https://doi.org/10.1038/s41598-018-19499-4

Martin, C. D., Thierry, G., Kuipers, J.-R., 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. https://doi.org/10.1016/j.jml.2013.08.001

Martins, I. P., Vieira, R., Loureiro, C., & Santos, M. E. (2007). Speech rate and fluency in children and adolescents. *Child Neuropsychology*, *13*(4), 319–332. https://doi.org/10.1080/09297040600837370

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6), 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95–112. https://doi.org/10.1016/j.jecp.2017.04.006

Microsoft Corporation. (2017). *Microsoft Paint 3D* (Version 6.2203.1037.0) [Computer software]. Microsoft Corporation. https://apps.microsoft.com/detail/9NBLGGH5FV99?hl=de-de&gl=DE

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability — Evidence from

Indian low and high literates. *Journal of Eye Movement Research*, *5*(1), 1–10. https://doi.org/10.16910/jemr.5.1.3

Monsalve, I. F., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. In W. Daelemans (Ed.), *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 398–408). Association for Computational Linguistics. https://aclanthology.org/E12-1000.pdf

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLOS ONE*, *7*(5), e37527. https://doi.org/10.1371/journal.pone.0037527

Moura, O., Simões, M. R., & Pereira, M. (2014). Executive functioning in children with developmental dyslexia. *The Clinical Neuropsychologist*, *28*, 20–41. https://doi.org/10.1080/13854046.2014.964326

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. https://doi.org/10.1002/hbm.22466

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329. https://doi.org/10.1016/j.jecp.2003.09.001

Ness, T., & Meltzer-Asscher, A. (2018). Predictive pre-updating and working memory capacity: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *30*(12), 1916–1938. https://doi.org/10.1162/jocn_a_01322

Ness, T., & Meltzer-Asscher, A. (2021). From pre-activation to pre-updating: A threshold mechanism for commitment to strong predictions. *Psychophysiology*, *58*(5), e13797. https://doi.org/10.1111/psyp.13797

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*, 107427. https://doi.org/10.1016/j.neuropsychologia.2020.107427

Nielsen, T. R., & Waldemar, G. (2016). Effects of literacy on semantic verbal fluency in an immigrant population. *Aging, Neuropsychology, and Cognition*, *23*(5), 578–590. https://doi.org/10.1080/13825585.2015.1132668

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, e33468. https://doi.org/10.7554/eLife.33468

Oliva, M. (2019). Pupil size and search performance in low and high perceptual load. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(2), 366–376. https://doi.org/10.3758/s13415-018-00677-w

Oliveira, R. M., Mograbi, D. C., Gabrig, I. A., & Charchat-Fichman, H. (2016). Normative data and evidence of validity for the Rey Auditory Verbal Learning Test, Verbal Fluency Test, and Stroop Test with Brazilian children. *Psychology & Neuroscience*, *9*(1), 54–67. https://doi.org/10.1037/pne0000041

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101. https://doi.org/10.1016/j.brainres.2009.07.042

Özkan, D., Küntay, A. C., & Brouwer, S. (2022). The role of verbal and working memory skills in Turkish-speaking children's morphosyntactic prediction. *Applied Psycholinguistics*, *43*(6), 1305–1328. https://doi.org/10.1017/S0142716422000388

Pham, A. V., & Hasson, R. M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, *29*(5), 467–477. https://doi.org/10.1093/arclin/acu024

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Piolat, A., Roussey, J.-Y., Olive, T., & Amada, M. (2004). Processing time and cognitive effort in revision: Effects of error type and of working memory capacity. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision Cognitive and Instructional Processes* (Vol. 13, pp. 21–38). Springer Science + Business Media New York 2004. https://doi.org/10.1007/978-94-007-1048-1_3

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569. https://doi.org/10.1111/j.1469-8986.2009.00947.x

Piu, P., Serchi, V., Rosini, F., & Rufa, A. (2019). A cross-recurrence analysis of the pupil size fluctuations in steady scotopic conditions. *Frontiers in Neuroscience*, *13*, 407. https://doi.org/10.3389/fnins.2019.00407

Pompéia, S., Miranda, M. C., & Bueno, O. F. A. (2001). A set of 400 pictures standardised for Portuguese: Norms for name agreement, familiarity and visual complexity for children and adults. *Arquivos de Neuro-Psiquiatria*, *59*(2B), 330–337. https://doi.org/10.1590/S0004-282X2001000300004

Prescott, K. E., Mathée-Scott, J., Reuter, T., Edwards, J., Saffran, J., & Ellis Weismer, S. (2022). Predictive language processing in young autistic children. *Autism Research*, *15*(5), 892–903. https://doi.org/10.1002/aur.2684

Python Software Foundation. (2019). *Python Programming Language* (Version 3.7.3) [Computer software]. https://www.python.org

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, *31*(1), 94–105. https://doi.org/10.1080/23273798.2015.1077979

Rajkowski, J., Majczynski, H., Clayton, E., & Aston-Jones, G. (2004). Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task. *Journal of Neurophysiology*, *92*(1), 361–371. https://doi.org/10.1152/jn.00673.2003

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, *89*(4), 760–793. https://doi.org/10.1353/lan.2013.0068

Ramsey, P. H., & Ramsey, P. P. (2007). Optimal trimming and outlier elimination. *Journal of Modern Applied Statistical Methods*, *6*(2), 355–360. https://doi.org/10.22237/jmasm/1193889660

Rayner, K., & Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, *80*(1), 4–9. https://doi.org/10.1016/j.biopsycho.2008.05.002

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. https://doi.org/10.3758/BF03214555

Reuter, T., Dalawella, K., & Lew-Williams, C. (2020). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience*, *36*(4), 474–490. https://doi.org/10.1080/23273798.2020.1839665

Rommers, J., & Federmeier, K. D. (2018). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, *101*, 16–30. https://doi.org/10.1016/j.cortex.2017.12.018

Rommers, J., Meyer, A. S., 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*(3), 437–447. https://doi.org/10.1016/j.neuropsychologia.2012.12.002

Rondeel, E. W. M., Van Steenbergen, H., Holland, R. W., & Van Knippenberg, A. (2015). A closer look at cognitive control: Differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience*, *9*, 494. https://doi.org/10.3389/fnhum.2015.00494

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*(2), 491–513. https://doi.org/10.1007/s11336-013-9396-3

RStudio Team. (2018). *RStudio: Integrated development for R* (Version 1.2.1335) [Computer software]. RStudio. http://www.rstudio.com.

Salthouse, T. A. (1992). What do adult age differences in the Digit Symbol Substitution Test reflect? *Journal of Gerontology*, *47*(3), 121–128. https://doi.org/10.1093/geronj/47.3.P121

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1), 35–54. https://doi.org/10.1016/S0301-0511(00)00052-1

Samuels, E., & 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. https://doi.org/10.2174/157015908785777229

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211–223. https://doi.org/10.1038/nrn2573

Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron*, *76*(1), 130–141. https://doi.org/10.1016/j.neuron.2012.09.011

Saryazdi, R., Bannon, J., Rodrigues, A., Klammer, C., & Chambers, C. G. (2018). Picture perfect: A stimulus set of 225 pairs of matched clipart and photographic images normed by Mechanical Turk and laboratory participants. *Behavior Research Methods*, *50*(6), 2498–2510. https://doi.org/10.3758/s13428-018-1028-5

Scheffler, T., Brandt, L., Fuente, M. D. L., & Nenchev, I. (2022). The processing of emoji-word substitutions: A self-paced-reading study. *Computers in Human Behavior*, *127*, 107076. https://doi.org/10.1016/j.chb.2021.107076

Schmidtke, J. (2017). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*(3), 529–549. https://doi.org/10.1017/S0272263117000195

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118. https://doi.org/10.1073/pnas.2105646118

Schröder, A., Gemballa, T., Ruppin, S., & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods*, *44*(2), 380–394. https://doi.org/10.3758/s13428-011-0164-y

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories — new models. *Interpreting*, *13*(2), 176–204. https://doi.org/10.1075/intp.13.2.02see

Sekerina, I. A. (2015). Predictions, fast and slow. *Linguistic Approaches to Bilingualism*, 5(4), 532–536. https://doi.org/10.1075/lab.5.4.16sek

Sekerina, I. A., Dragoy, O. V., & Ivanova, M. V. (2015). Time course processing of grammatical agreement information in Russian agrammatism. *Procedia Social and Behavioral Sciences*, 23, 57–58. https://doi.org/10.1016/j.sbspro.2011.09.168

Sikos, L., Stein, K., & Staudte, M. (2021). A rose by any other verb: The effect of expectations and word category on processing effort in situated sentence comprehension. *Frontiers in Psychology*, *12*, 661898. https://doi.org/10.3389/fpsyg.2021.661898

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Slevc, L. R., & Novick, J. M. (2013). Memory and cognitive control in an integrated theory of language processing. *Behavioral and Brain Sciences*, *36*(4), 373–374. https://doi.org/10.1017/S0140525X12002683

Smolík, F., & Bláhová, V. (2019). Czech 23-month-olds use gender agreement to anticipate upcoming nouns. *Journal of Experimental Child Psychology*, *178*, 251–265. https://doi.org/10.1016/j.jecp.2018.10.004

Smolík, F., & Bláhová, V. (2022). Here come the nouns: Czech two-year-olds use verb number endings to predict sentence subjects. *Cognition*, *219*, 104964. https://doi.org/10.1016/j.cognition.2021.104964

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sommerfeld, L., Staudte, M., Mani, N., & Kray, J. (2023). Even young children make multiple predictions in the complex visual world. *Journal of Experimental Child Psychology*, *235*, 105690. https://doi.org/10.1016/j.jecp.2023.105690

Sommerfeld, L., Staudte, M., & Kray, J. (2022). Ratings of name agreement and semantic categorization of 247 colored clipart pictures by young German children. *Acta Psychologica*, *226*, 103558. https://doi.org/10.1016/j.actpsy.2022.103558

SR Research (2017). *EyeLink 1000 Plus user manual* (Version 1.0.12). SR Research. https://www.sr-research.com

SR Research. (2019a). *EyeLink Experiment Builder* (Version 2.2.61) [Computer software]. SR Research. https://www.sr-research.com

SR Research. (2019b). *EyeLink Data Viewer* (Version 4.1.63) [Computer software]. SR Research. https://www.sr-research.com

Stanners, R. F., Headley, D. B., & Clark, W. R. (1972). The pupillary response to sentences: Influences of listening set and deep structure. *Journal of Verbal Learning and Verbal Behavior, 11*(2), 257–263. https://doi.org/10.1016/S0022-5371(72)80086-0

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425–436. https://doi.org/10.1037/0278-7393.32.2.425

Stone, K., Veríssimo, J., Schad, D. J., Oltrogge, E., Vasishth, S., & Lago, S. (2021). The interaction of grammatically distinct agreement dependencies in predictive processing. *Language, Cognition and Neuroscience*, *36*(9), 1159–1179. https://doi.org/10.1080/23273798.2021.1921816

Süss, A. M. A., Hendriks, P., Fritzsche, T., & Höhle, B. (2018). Acquisition of adjectival agreement in German: Sensitivity to grammar is reflected in 3-year-olds' pupil dilation. In A. B. Bertolini, & M. J. Kaplan (Eds.), *Proceedings of the 42nd annual Boston University Conference on Language Development* (pp. 722–735). Cascadilla Press. https://hdl.handle.net/11370/39a61dd4-98f8-4d46-8acd-7551b0e0e649

Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental Psychology*, *35*(4), 986–1000. https://doi.org/10.1037/0012-1649.35.4.986

Swanson, H. L. (2000). Are working memory deficits in readers with learning disabilities hard to change? *Journal of Learning Disabilities, 33*(6), 551–566. https://doi.org/10.1177/002221940003300604

Swanson, H. L. (2010). Does the dynamic testing of working memory predict growth in nonword fluency and vocabulary in children with reading disabilities? *Journal of Cognitive Education and Psychology*, *9*(2), 139–165. https://doi.org/10.1891/1945-8959.9.2.139

Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology*, 93(4), 720-734. https://doi.org/10.1037/0022-0663.93.4.720

Swanson, H. L., & Jerman, O. (2007). The influence of working memory on reading growth in subgroups of children with reading disabilities. *Journal of Experimental Child Psychology*, *96*(4), 249–283. https://doi.org/10.1016/j.jecp.2006.12.004

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. https://doi.org/10.1016/S0010-0277(99)00021-9

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297–314. https://doi.org/10.1016/j.jml.2012.12.002

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology*, *153*, 140–148. https://doi.org/10.1016/j.jecp.2016.07.014

Tamási, K., Mckean, C., Gafos, A., & Höhle, B. (2019). Children's gradient sensitivity to phonological mismatch: Considering the dynamics of looking behavior and pupil dilation. *Journal of Child Language*, *46*, 1–23. https://doi.org/10.1017/S0305000918000259

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1105–1122. https://doi.org/10.1098/rstb.2007.2162

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. https://doi.org/10.1126/science.7777863

Theimann, A., Kuzmina, E., & Hansen, P. (2021). Verb-mediated prediction in bilingual toddlers. *Frontiers in Psychology*, *12*, 719447. https://doi.org/10.3389/fpsyg.2021.719447

Throndsen, T. U., Lindskog, M., Niemivirta, M., & Mononen, R. (2022). Does mathematics anxiety moderate the effect of problem difficulty on cognitive effort? *Scandinavian Journal of Psychology*, *63*(6), 601–608. https://doi.org/10.1111/sjop.12852

Tourtouri, E. N., Delogu, F., & Crocker, M. W. (2015). ERP indices of situated reference in visual contexts. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2422–2427). Cognitive Science Society. https://dblp.org/rec/conf/cogsci/TourtouriDC15

Tribushinina, E., & Mak, W. M. (2016). Three-year-olds can predict a noun based on an attributive adjective: Evidence from eye-tracking. *Journal of Child Language*, *43*(2), 425–441. https://doi.org/10.1017/S0305000915000173

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation vs. Exploration — Computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, *131*, 53–61. https://doi.org/10.1016/j.neuropsychologia.2019.05.007

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology*, *6*, 347. https://doi.org/10.3389/fpsyg.2015.00347

Urbach, T. P., DeLong, K. A., Chan, W.-H., & Kutas, M. (2020). An exploratory data analysis of word form prediction during word-by-word reading. *Proceedings of the National Academy of Sciences*, *117*(34), 20483–20494. https://doi.org/10.1073/pnas.1922028117

Van Alphen, P., Brouwer, S., Davids, N., Dijkstra, E., & Fikkert, P. (2021). Word recognition and word prediction in preschoolers with (a suspicion of) a developmental language disorder: Evidence from eye tracking. *Journal of Speech, Language, and Hearing Research*, *64*(6), 2005–2021. https://doi.org/10.1044/2021_JSLHR-20-00227

Van Berkum, J. J. A., Brown, C. M., 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. https://doi.org/10.1037/0278-7393.31.3.443

Van Breukelen, G. J. P. (2005). Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika*, *70*(2), 359–376. https://doi.org/10.1007/s11336-003-1078-0

Van Der Elst, W., Hurks, P., Wassenberg, R., Meijs, C., & Jolles, J. (2011). Animal Verbal Fluency and Design Fluency in school-aged children: Effects of age, sex, and mean level of parental education, and regression-based normative data. *Journal of Clinical and Experimental Neuropsychology*, *33*(9), 1005–1015. https://doi.org/10.1080/13803395.2011.589509

Van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181–204. https://doi.org/10.3102/10769986031002181

Van Der Schoot, M., Reijntjes, A., & Van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665–1690. https://doi.org/10.1007/s11145-011-9337-4

Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research*, *369*, 56–66. https://doi.org/10.1016/j.heares.2018.04.013

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Van Rij, J. (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults* [Doctoral dissertation, University of Groningen]. s.n. https://hdl.handle.net/11370/55ce05ac-6c9c-4886-b8eb-bbbc8ce35dad

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *The Journal of Neuroscience*, *35*(20), 7866–7877. https://doi.org/10.1523/JNEUROSCI.0454-15.2015

Vela-Candelas, J., Català, N., & Demestre, J. (2022). Effects of world knowledge on the prediction of upcoming verbs: An eye-tracking study. *Journal of Psycholinguistic Research*, *51*(6), 1335–1345. https://doi.org/10.1007/s10936-022-09900-9

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Semantic entropy in language comprehension. *Entropy*, *21*(12), 1159. https://doi.org/10.3390/e21121159

Vergilova, Y., Jachmann, T. K., Mani, N., & Kray, J. (2022). Age-related differences in expectation-based novel word learning. *Psychophysiology*, *59*(8), e14030. https://doi.org/10.1111/psyp.14030

Vogels, J., Demberg, V., & Kray, J. (2018). The Index of Cognitive Activity as a measure of cognitive processing load in dual task settings. *Frontiers in Psychology*, *9*, 2276. https://doi.org/10.3389/fpsyg.2018.02276

Wang, Y., Zhou, X., Peng, X., & Hu, X. (2022). Task switching involves working memory: Evidence from neural representation. *Frontiers in Psychology*, *13*, 1003298. https://doi.org/10.3389/fpsyg.2022.1003298

Wassenburg, S. I., Beker, K., Van Den Broek, P., & Van Der Schoot, M. (2015). Children's comprehension monitoring of multiple situational dimensions of a narrative. *Reading and Writing*, *28*(8), 1203–1232. https://doi.org/10.1007/s11145-015-9568-x

Weber, P., Rupprecht, F., Wiesen, S., Hamann, B., & Ebert, A. (2021). Assessing cognitive load via pupillometry. In H. R. Arabnia, K. Ferens, D. De la Fuente, E. B. Kozerenko, J. A. O. Varela, & F. G. Tinetti (Eds.), *Advances in Artificial Intelligence and Applied Cognitive Computing: Proceedings from ICAI'20 and ACC'20* (pp. 1087–1096). Springer. https://doi.org/10.1007/978-3-030-70296-0_86

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation.

Weighall, A. R., & Altmann, G. T. M. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of Child Language*, *38*(3), 579–605. https://doi.org/10.1017/S0305000910000267

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology*, *7*, 345. https://doi.org/10.3389/fpsyg.2016.00345

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Winter, B. (2013). *A very basic tutorial for performing linear mixed effects analyses (Tutorial 2)* [Unpublished tutorial]. University of California, Merced, Cognitive and Information Sciences. https://bodo-winter.net/tutorial/bw_LME_tutorial2.pdf

Winter, B. (2013). *Linear models and linear mixed effects models in R: Tutorial 1*. arXiv: http://arxiv.org/pdf/1308.5499

Winter, B. (2020). *Statistics for linguists: An introduction using R* (1st ed.). Routledge. https://doi.org/10.4324/9781315165547

Wlotko, E. W., & Federmeier, K. D. (2012a). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension: Context use in young and older adults. *Psychophysiology*, *49*(6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x

Wlotko, E. W., & Federmeier, K. D. (2012b). So that’s what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356–366. https://doi.org/10.1016/j.neuroimage.2012.04.054

Zabrucky, K., & Ratner, H. H. (1986). Children’s comprehension monitoring and recall of inconsistent stories. *Child Development*, *57*(6), 1401–1418. https://doi.org/10.2307/1130419

Zargar, E., Adams, A. M., & Connor, C. M. (2020). The relations between children’s comprehension monitoring and their reading comprehension and vocabulary knowledge: An eye-movement study. *Reading and Writing*, *33*(3), 511–545. https://doi.org/10.1007/s11145-019-09966-3

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 1-25. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear & Hearing*, *31*(4), 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zhang, F., & Emberson, L. L. (2020). Using pupillometry to investigate predictive processes in infancy. *Infancy*, *25*(6), 758–780. https://doi.org/10.1111/infa.12358

Zhang, F., Jaffe-Dax, S., Wilson, R. C., & Emberson, L. L. (2018). Prediction in infants and adults: A pupillometry study. *Developmental Science*, *22*(4), e12780. https://doi.org/10.1111/desc.12780

Zhang, L., & Knoeferle, P. (2012). Visual context effects on thematic role assignment in children versus adults: Evidence from eye tracking in German. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 2593–2598). Cognitive Science Society. https://escholarship.org/uc/item/8hp1f9f7

Zhang, L.-L., Zhong, Y.-Q., Sun, J.-W., Chen, L., Sun, J.-Q., Hou, X.-Y., Chen, J.-W., & Guo, X.-T. (2020). Deficit of long-term memory traces for words in children with cochlear implants. *Clinical Neurophysiology*, *131*(6), 1323–1331. https://doi.org/10.1016/j.clinph.2020.03.012

Zhou, P., Zhan, L., & Ma, H. (2019). Predictive language processing in preschool children with autism spectrum disorder: An eye-tracking study. *Journal of Psycholinguistic Research*, *48*(2), 431–452. https://doi.org/10.1007/s10936-018-9612-5

Zimmermann, N., Parente, M. A. D. M. P., Joanette, Y., & Fonseca, R. P. (2014). Unconstrained, phonemic and semantic verbal fluency: Age and education effects, norms and discrepancies. *Psicologia: Reflexão e Crítica*, *27*(1), 55–63. https://doi.org/10.1590/S0102-79722014000100007

Zou, D. (2022). Multi-dimensional consideration of cognitive effort in translation and interpreting process studies. In S. Larocca, J. Campbell, J. Marciano, K. Savenkov, A. Yanishevsky, & S. Richardson (Eds.), *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas* (Volume 2, pp. 416–426). Association for Machine Translation in the Americas. https://aclanthology.org/2022.amta-upg.29.pdf

Appendix A — Linguistic Stimuli

Table A1

Item and Filler Sentences of Experiment 1

Nr.	Type	Subject	Verb	Spill-over	Target noun
1	Item	Der Vater *The father*	**verschlingt** *eats*	gleich die *now the*	**Waffel**. *waffle.*
2	Item	Die Mutter *The mother*	**putzt** *cleans*	gleich die *soon the*	**Lupe**. *magnifier.*
3	Item	Der Onkel *The uncle*	**schneidet** *cuts*	gleich die *soon the*	**Banane**. *banana.*
4	Item	Die Tante *The aunt*	**repariert** *repairs*	gleich die *soon the*	**Lampe**. *lamp.*
5	Item	Der Enkel *The grandson*	**probiert** *tastes*	gleich die *soon the*	**Erdbeere**. *strawberry.*
6	Item	Die Enkelin *The granddaughter*	**befüllt** *fills*	gleich die *soon the*	**Gießkanne**. *ewer.*
7	Item	Der Vater *The father*	**nascht** *nibbles*	gleich die *soon the*	**Himbeere**. *raspberry.*
8	Item	Die Mutter *The mother*	**entzündet** *ignites*	gleich die *soon the*	**Rakete**. *rocket.*
9	Item	Der Großvater *The grandfather*	**gießt** *waters*	gleich die *soon the*	**Tomate**. *tomato.*
10	Item	Die Großmutter *The grandmother*	**spielt** *plays*	gleich die *soon the*	**Trompete**. *trumped.*
11	Item	Der Bruder *The brother*	**hört** *hears*	gleich die *soon the*	**Gitarre**. *guitar.*
12	Item	Die Schwester *The sister*	**kaut** *chews*	gleich die *soon the*	**Pizza**. *pizza.*
13	Item	Der Bruder *The brother*	**schließt** *closes*	gleich die *soon the*	**Tasche**. *case.*
14	Item	Die Schwester *The sister*	**pflückt** *grabs*	gleich die *soon the*	**Kirsche**. *cherry.*
15	Item	Der Mann *The man*	**verschüttet** *spills*	gleich den *soon the*	**Sprudel**. *soda.*
16	Item	Die Frau *The woman*	**startet** *starts*	gleich den *soon the*	**Computer**. *computer.*
17	Item	Der Opa *The grandpa*	**backt** *bakes*	gleich den *soon the*	**Kuchen**. *cake.*
18	Item	Die Oma *The grandma*	**bremst** *brakes*	gleich den *soon the*	**Roller**. *scooter.*
19	Item	Der Mann *The man*	**erntet** *harvests*	gleich den *soon the*	**Salat**. *salad.*
20	Item	Die Frau *The woman*	**näht** *sews*	gleich den *soon the*	**Handschuh**. *glove.*

Note. Table continued on the next page.

Table A1 (Continued)

Item and Filler Sentences of Experiment 1

Nr.	Type	Subject	Verb	Spill-over	Target noun
21	Item	Der Onkel *The uncle*	**sammelt** *picks*	gleich den *soon the*	**Tannenzapfen**. *fir cone.*
22	Item	Die Tante *The aunt*	**trinkt** *drinks*	gleich den *soon the*	**Kakao**. *cocoa.*
23	Item	Der Großvater *The grandfather*	**zerbricht** *breaks*	gleich den *soon the*	**Pokal**. *trophy.*
24	Item	Die Großmutter *The grandmother*	**strickt** *knits*	gleich den *soon the*	**Schal**. *scarf.*
25	Item	Der Vater *The father*	**wirft** *throws*	gleich den *soon the*	**Ball**. *ball.*
26	Item	Die Mutter *The mother*	**parkt** *parks*	gleich den *soon the*	**Bus**. *bus.*
27	Item	Der Opa *The grandpa*	**genießt** *enjoys*	gleich das *soon the*	**Ei**. *egg.*
28	Item	Die Oma *The grandma*	**wäscht** *washes*	gleich das *soon the*	**Glas**. *glass.*
29	Item	Der Mann *The man*	**futtert** *guzzles*	gleich das *soon the*	**Bonbon**. *drop.*
30	Item	Die Frau *The woman*	**fährt** *drives*	gleich das *soon the*	**Skateboard**. *skateboard.*
31	Item	Der Onkel *The uncle*	**baut** *builds*	gleich das *soon the*	**Vogelhaus**. *bird house.*
32	Item	Die Tante *The aunt*	**bügelt** *irons*	gleich das *soon the*	**T-Shirt**. *T-shirt.*
33	Filler	Der Mann *The man*	**klebt** *sticks*	gleich das *soon the*	**Kanne**. *can.*
34	Filler	Die Frau *The woman*	**schluckt** *swallows*	gleich die *soon the*	**Milch**. *milk.*
35	Filler	Der Onkel *The uncle*	**reinigt** *tidies*	gleich die *soon the*	**Socke**. *sock.*
36	Filler	Der Opa *The grandpa*	**bemalt** *paints*	gleich den *soon the*	**Stuhl**. *chair.*
37	Filler	Die Oma *The grandma*	**löffelt** *spoons*	gleich den *soon the*	**Pudding**. *pudding.*
38	Filler	Der Vater *The father*	**lackiert** *varnishes*	gleich das *soon the*	**Boot**. *boat.*
39	Filler	Der Bruder *The brother*	**faltet** *folds*	gleich das *soon the*	**Papierflugzeug**. *paper plane.*
40	Filler	Die Schwester *The sister*	**kostet** *tastes*	gleich das *soon the*	**Sandwich**. *sandwich.*

Note. All item and filler sentences of Experiment 1 with their trial number. The constraining verbs and target nouns are presented in bold, approximate English translations in italics.

Table A2

Item Sentences of Experiment 2

Nr.	Subject	Verb	Post-Verb	Target noun	Post-Noun
1	Der Vater *The father*	**verschlingt** *eats*	am frühen Montagmorgen die *early Monday morning the*	**Waffel** *waffle*	auf dem Balkon. *on the balcony.*
2	Die Mutter *The mother*	**putzt** *cleans*	am späten Dienstagmorgen die *late Tuesday morning the*	**Lupe** *magnifier*	in der Garage. *in the garage.*
3	Der Onkel *The uncle*	**schneidet** *cuts*	am frühen Mittwochmorgen die *early Wednesday morning the*	**Banane** *banana*	in dem Wohnzimmer. *in the living room*
4	Die Tante *The aunt*	**repariert** *repairs*	am späten Donnerstagmorgen die *late Thursday morning the*	**Lampe** *lamp*	in der Waschküche. *in the laundry.*
5	Der Enkel *The grandson*	**probiert** *tastes*	am frühen Freitagmorgen die *early Friday morning the*	**Erdbeere** *strawberry*	auf dem Balkon. *on the balcony.*
6	Die Enkelin *The granddaughter*	**befüllt** *fills*	am späten Samstagmorgen die *late Saturday morning the*	**Gießkanne** *ewer*	in der Garage. *in the garage.*
7	Der Vater *The father*	**nascht** *nibbles*	am frühen Sonntagmorgen die *early Sunday morning the*	**Himbeere** *raspberry*	in der Küche. *in the kitchen.*
8	Die Mutter *The mother*	**entzündet** *ignites*	am späten Montagmittag die *late Monday noon the*	**Rakete** *rocket*	hinter dem Haus. *behind the house.*
9	Der Großvater *The grandfather*	**gießt** *waters*	am frühen Dienstagmittag die *early Tuesday noon the*	**Tomate** *tomato*	vor dem Haus. *in front of the house.*
10	Die Großmutter *The grandmother*	**spielt** *plays*	am späten Mittwochmittag die *late Wednesday noon the*	**Trompete** *trumped*	in dem Keller. *in the basement.*
11	Der Bruder *The brother*	**hört** *hears*	am frühen Donnerstagmittag die *early Thursday macnoon the*	**Gitarre** *guitar*	in dem Arbeitszimmer. *in the workroom.*

Note. Table continued on the next page.

Table A2 (Continued)

Item Sentences of Experiment 2

Nr.	Subject	Verb	Post-Verb	Target noun	Post-Noun
12	Die Schwester *The sister*	**kaut** *chews*	am späten Freitagmittag die *late Friday noon the*	**Pizza** *pizza*	in dem Wohnzimmer. *in the living room.*
13	Der Bruder *The brother*	**schließt** *closes*	am frühen Samstagmittag die *early Saturday noon the*	**Tasche** *case*	in dem Schlafzimmer. *in the bedroom.*
14	Die Schwester *The sister*	**pflückt** *grabs*	am späten Sonntagmittag die *late Sunday noon the*	**Kirsche** *cherry*	vor dem Haus. *in front of the house.*
15	Der Mann *The man*	**verschüttet** *spills*	am frühen Montagabend den *early Monday evening the*	**Sprudel** *soda*	in dem Kinderzimmer. *in the nursery.*
16	Die Frau *The woman*	**startet** *starts*	am späten Dienstagabend den *late Tuesday evening the*	**Computer** *computer*	in der Stadt. *in the city.*
17	Der Opa *The grandpa*	**backt** *bakes*	am frühen Mittwochabend den *early Wednesday evening the*	**Kuchen** *cake*	in der Küche. *in the kitchen.*
18	Die Oma *The grandma*	**bremst** *brakes*	am späten Donnerstagabend den *late Thursday evening the*	**Roller** *scooter*	auf der Straße. *in the street.*
19	Der Mann *The man*	**erntet** *harvests*	am frühen Freitagabend den *early Friday evening the*	**Salat** *salad*	in dem Garten. *in the garden.*
20	Die Frau *The woman*	**näht** *sews*	am späten Samstagabend den *late Saturday evening the*	**Handschuh** *glove*	in dem Kinderzimmer. *in the nursery.*
21	Der Onkel *The uncle*	**sammelt** *picks*	am frühen Sonntagabend den *early Sunday evening the*	**Tannenzapfen** *fir cone*	in dem Wald. *in the forest.*
22	Die Tante *The aunt*	**trinkt** *drinks*	am späten Montagmorgen den *late Monday morning the*	**Kakao** *cocoa*	auf der Terrasse. *on the terrace.*

Note. Table continued on the next page.

Table A2 (Continued)

Item Sentences of Experiment 2

Nr.	Subject	Verb	Post-Verb	Target noun	Post-Noun
23	Der Großvater *The grandfather*	**zerbricht** *breaks*	am frühen Dienstagmorgen den *early Tuesday morning the*	**Pokal** *trophy*	in dem Arbeitszimmer. *in the workroom.*
24	Die Großmutter *The grandmother*	**strickt** *knits*	am späten Mittwochmorgen den *late Wednesday morning the*	**Schal** *scarf*	auf der Terrasse. *on the terrace.*
25	Der Vater *The father*	**wirft** *throws*	am frühen Donnerstagmorgen den *early Thursday morning the*	**Ball** *ball*	in der Waschküche. *in the laundry.*
26	Die Mutter *The mother*	**parkt** *parks*	am späten Freitagmorgen den *late Friday morning the*	**Bus** *bus*	in der Stadt. *in the city.*
27	Der Opa *The grandpa*	**genießt** *enjoys*	am frühen Samstagmorgen das *early Saturday morning the*	**Ei** *egg*	hinter dem Haus. *behind the house.*
28	Die Oma *The grandma*	**wäscht** *washes*	am späten Sonntagmorgen das *late Sunday morning the*	**Glas** *glass*	in dem Garten. *in the garden.*
29	Der Mann *The man*	**futtert** *guzzles*	am frühen Montagmittag das *early Monday noon the*	**Bonbon** *drop*	in dem Wald. *in the forest.*
30	Die Frau *The woman*	**fährt** *drives*	am späten Dienstagmittag das *late Tuesday noon the*	**Skateboard** *skateboard*	auf der Straße. *in the street.*
31	Der Onkel *The uncle*	**baut** *builds*	am frühen Mittwochmorgen das *early Wednesday morning the*	**Vogelhaus** *bird*	in dem Keller. *in the basement.*
32	Die Tante *The aunt*	**bügelt** *irons*	am späten Donnerstagmittag das *late Thursday noon the*	**T-Shirt** *T-shirt*	in dem Schlafzimmer. *in the bedroom.*

Note. All item sentences of Experiment 2 are shown with their item number. Constraining verbs and target nouns are presented in bold, approximate English translations in italics.

Table A3

Filler Sentences of Experiment 2

Nr.	Sentence
1	Beim Toben am späten Montagabend macht der Junge seine Handschuhe schmutzig. *While romping late Monday evening, the boy gets his gloves dirty.*
2	Mit seinem Boot geht der alte Mann am späten Sonntagmorgen fischen. *With his boat, the old man goes fishing late Sunday morning.*
3	Für den Ausflug am frühen Samstagmorgen packt die Schwester den Rucksack. *For the excursion early Saturday morning, the sister packs her backpack.*
4	Im Winter schmückt der Junge den schönen Weihnachtsbaum in dem Wohnzimmer. *In winter the boy decorates the beautiful Christmas tree in the living room.*
5	Im Herbst bastelt der Junge einen bunten Flugdrachen in der Küche. *In autumn the boy makes a colorful flying kite in the kitchen.*
6	Mit seiner Schippe schaufelt der Junge den Sand in dem Garten. *With his shovel the boy shovels the sand in the garden.*
7	Die Mutter liest fast jeden Tag die Zeitung bei dem Frühstück. *The mother reads the newspaper almost every day at breakfast.*
8	Die Frau wählt den neuen Badeanzug für den Urlaub am Strand. *The woman chooses the new swimsuit for vacation on the beach.*
9	Der Junge wünscht sich seit Jahren ein neues Skateboard zu Weihnachten. *The boy has wanted a new skateboard for Christmas for years.*
10	Die Frau verziert die leckeren Törtchen mit Streuseln in der Küche. *The woman decorates the delicious cupcakes with sprinkles in the kitchen.*
11	Dass der wunderschöne Hut zu teuer ist, bedauert die Tante sehr. *The aunt is very unhappy that the beautiful hat is too expensive.*
12	Da der Junge alleine ist, stielt er ein Bonbon im Süßwarenladen. *Since the boy is alone, he steals a piece of candy from the candy store.*
13	Suppe ist bei der Großmutter immer lecker, da sie selbstgemacht ist. *Soup is always delicious at grandmother's house, because it is homemade.*
14	Am See trägt der Junge einen Rettungsring, damit ihm nichts passiert. *At the lake, the boy wears a life preserver so that nothing happens to him.*
15	Weil sie am schnellsten schwimmt, erhält die Schwester eine goldene Medaille. *Because she swims the fastest, the sister receives a gold medal.*
16	Um Plätzchen zu machen, nimmt die Oma Mehl und ein Ei. *To make cookies, grandma takes flour and an egg.*

Note. All filler sentences of Experiment 2 are shown with their filler number and approximate English translation in italics.

Appendix B — Visual Stimuli

Table B1

Semantic Categorization and Name Agreement Scores for all Pictures of the Final 32 Stimuli

Item	Picture/Noun	Verb	Type	Cat	NA	Test	*N*
1	Waffel	Verschlingen	Target	87	93	2	15
1	Pizza	Verschlingen	Competitor	87	87	2	15
1	Brezel	Verschlingen	Competitor	87	87	2	15
1	Wurst	Verschlingen	Competitor	87	87	2	15
2	Lupe	Putzen	Target	98	90	1	40
2	Badewanne	Putzen	Competitor	95	95	1	20
2	Brille	Putzen	Competitor	90	100	1	20
2	Toilette	Putzen	Competitor	87	87	2	15
3	Banane	Schneiden	Target	100	100	1	40
3	Karotte	Schneiden	Competitor	100	100	1	20
3	Gurke	Schneiden	Competitor	100	80	1	20
3	Zwiebel	Schneiden	Competitor	100	85	1	20
4	Lampe	Reparieren	Target	98	95	1	40
4	Brille	Reparieren	Competitor	80	100	1	20
4	Waschmaschine	Reparieren	Competitor	90	90	1	20
4	Uhr	Reparieren	Competitor	90	90	1	20
5	Erdbeere	Probieren	Target	98	95	1	40
5	Karotte	Probieren	Competitor	100	100	1	20
5	Birne	Probieren	Competitor	100	95	1	20
5	Wurst	Probieren	Competitor	95	100	1	20
6	Gießkanne	Befüllen	Target	95	100	1	40
6	Badewanne	Befüllen	Competitor	90	90	1	20
6	Flasche	Befüllen	Competitor	95	90	1	20
6	Tasse	Befüllen	Competitor	100	90	1	20
7	Himbeere	Naschen	Target	100	70	1	40
7	Gurke	Naschen	Competitor	95	95	1	20
7	Orange	Naschen	Competitor	100	90	1	20
7	Paprika	Naschen	Competitor	90	100	1	20
8	Rakete	Entzünden	Target	95	70	1	40
8	Kerze	Entzünden	Competitor	100	100	1	20
8	Laterne	Entzünden	Competitor	90	90	1	20
8	Fackel	Entzünden	Competitor	95	35	1	20
9	Tomate	Gießen	Target	100	100	1	40
9	Blume	Gießen	Competitor	100	100	1	20
9	Zitrone	Gießen	Competitor	100	80	1	20
9	Erdbeere	Gießen	Competitor	100	95	1	20

Note. Table continued on the next page.

Table B1 (Continued)

Semantic Categorization and Name Agreement Scores for all Pictures of the Final 32 Stimuli

Item	Picture/Noun	Verb	Type	Cat	NA	Test	*N*
10	Trompete	Spielen	Target	95	85	1	40
10	Flöte	Spielen	Competitor	95	100	1	20
10	Gitarre	Spielen	Competitor	95	100	1	20
10	Rassel	Spielen	Competitor	100	100	1	20
11	Gitarre	Hören	Target	100	95	1	40
11	Trommel	Hören	Competitor	100	95	1	20
11	Flöte	Hören	Competitor	100	75	1	20
11	Mundharmonika	Hören	Competitor	100	15	1	20
12	Pizza	Kauen	Target	100	98	1	40
12	Schokolade	Kauen	Competitor	100	100	1	20
12	Nuss	Kauen	Competitor	95	90	1	20
12	Brezel	Kauen	Competitor	100	100	1	20
13	Tasche	Schließen	Target	100	98	1	40
13	Kasse	Schließen	Competitor	90	70	1	20
13	Truhe	Schließen	Competitor	100	90	1	20
13	Tür	Schließen	Competitor	100	100	1	20
14	Kirsche	Pflücken	Target	95	93	1	40
14	Brombeere	Pflücken	Competitor	93	93	2	15
14	Blume	Pflücken	Competitor	85	100	1	20
14	Birne	Pflücken	Competitor	95	95	1	20
15	Sprudel	Verschütten	Target	100	50	1	40
15	Kaffee	Verschütten	Competitor	100	90	1	20
15	Saft	Verschütten	Competitor	100	20	1	20
15	Tee	Verschütten	Competitor	100	25	1	20
16	Computer	Starten	Target	98	95	1	40
16	Fernseher	Starten	Competitor	95	85	1	20
16	Zug	Starten	Competitor	100	95	1	20
16	LKW	Starten	Competitor	100	95	1	20
17	Kuchen	Backen	Target	100	95	1	40
17	Keks	Backen	Competitor	85	70	1	20
17	Donut	Backen	Competitor	90	55	1	20
17	Muffin	Backen	Competitor	95	100	1	20
18	Bus	Bremsen	Competitor	95	100	1	40
18	Roller	Bremsen	Target	98	100	1	40
18	Bagger	Bremsen	Competitor	87	87	2	15
18	Traktor	Bremsen	Competitor	90	95	1	20
19	Salat	Ernten	Target	90	93	1	40
19	Apfel	Ernten	Competitor	90	100	1	20
19	Mais	Ernten	Competitor	90	80	1	20
19	Kürbis	Ernten	Competitor	95	85	1	20

Note. Table continued on the next page.

Table B1 (Continued)

Semantic Categorization and Name Agreement Scores for all Pictures of the Final 32 Stimuli

Item	Picture/Noun	Verb	Type	Cat	NA	Test	*N*
20	Handschuh	Nähen	Target	78	98	1	40
20	Bademantel	Nähen	Competitor	95	65	1	20
20	Rock	Nähen	Competitor	95	75	1	20
20	Pullover	Nähen	Competitor	95	70	1	20
21	Tannenzapfen	Sammeln	Target	100	50	1	40
21	Knochen	Sammeln	Competitor	70	95	1	20
21	Pilz	Sammeln	Competitor	90	100	1	20
21	Stein	Sammeln	Competitor	80	95	1	20
22	Kakao	Trinken	Target	100	50	1	40
22	Saft	Trinken	Competitor	100	50	1	20
22	Tee	Trinken	Competitor	100	55	1	20
22	Wein	Trinken	Competitor	100	35	1	20
23	Pokal	Zerbrechen	Target	88	88	1	40
23	Spiegel	Zerbrechen	Competitor	95	80	1	20
23	Teller	Zerbrechen	Competitor	100	100	1	20
23	Fernseher	Zerbrechen	Competitor	85	55	1	20
24	Schal	Stricken	Target	100	100	1	40
24	Handschuh	Stricken	Competitor	95	100	1	20
24	Rock	Stricken	Competitor	85	85	1	20
24	Pullover	Stricken	Competitor	95	70	1	20
25	Ball	Werfen	Target	100	98	1	40
25	Pfeil	Werfen	Competitor	90	70	1	20
25	Ast	Werfen	Competitor	80	90	1	20
25	Würfel	Werfen	Competitor	80	100	1	20
26	Bus	Parken	Target	98	93	1	20
26	Bagger	Parken	Competitor	100	55	1	20
26	Traktor	Parken	Competitor	100	100	1	20
26	LKW	Parken	Competitor	100	95	1	20
27	Ei	Genießen	Target	95	98	1	40
27	Brot	Genießen	Competitor	95	90	1	20
27	Croissant	Genießen	Competitor	100	90	1	20
27	Fleisch	Genießen	Competitor	90	50	1	20
28	Glas	Waschen	Target	93	90	1	40
28	Kleid	Waschen	Competitor	100	85	1	20
28	Messer	Waschen	Competitor	95	100	1	20
28	Handtuch	Waschen	Competitor	100	95	1	20
29	Bonbon	Futtern	Target	100	93	1	40
29	Croissant	Futtern	Competitor	100	90	1	20
29	Gummibärchen	Futtern	Competitor	100	90	1	20
29	Eis	Futtern	Competitor	95	100	1	20

Note. Table continued on the next page.

Table B1 (Continued)

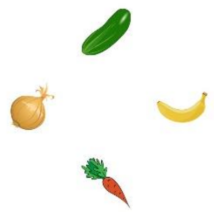
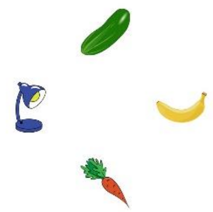
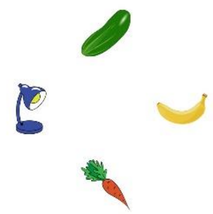
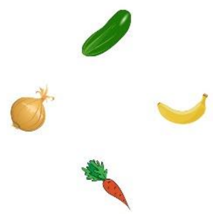
Semantic Categorization and Name Agreement Scores for all Pictures of the Final 32 Stimuli

Item	Picture/Noun	Verb	Type	Cat	NA	Test	*N*
30	Skateboard	Fahren	Target	100	88	1	40
30	Fahrrad	Fahren	Competitor	100	100	1	20
30	Auto	Fahren	Competitor	100	100	1	20
30	Motorrad	Fahren	Competitor	100	95	1	20
31	Vogelhaus	Bauen	Target	100	98	1	40
31	Iglu	Bauen	Competitor	90	40	1	20
31	Papierboot	Bauen	Competitor	90	100	1	20
31	Baumhaus	Bauen	Competitor	100	100	1	20
32	T-Shirt	Bügeln	Target	98	75	1	40
32	Kleid	Bügeln	Competitor	95	100	1	20
32	Hemd	Bügeln	Competitor	95	20	1	20
32	Unterhemd	Bügeln	Competitor	95	90	1	20

Note. The column “Item” shows the number of the item to which each picture and its ratings belong. “Picture/Noun” indicates which object was visualized by the picture and thus, which target noun was represented. “Verb” is the semantically constraining verb for which the semantic categorization scores were collected (for reasons of space, only the German nouns and verbs could be presented). “Type” indicates whether the object was a target or competitor object. “Cat” shows the semantic categorization scores. “NA” shows the name agreement scores. “Test” indicates whether the ratings were collected in the first or second pretest. “N” is the number of children based on which the scores were calculated. Note that this table is most easily to comprehend if it is interpreted in combination with the object pictures of the visual scenes presented on the next page in Table B2.

Table B2

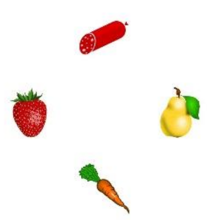
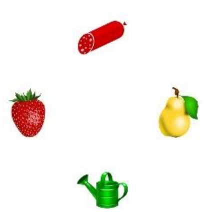
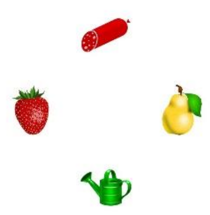
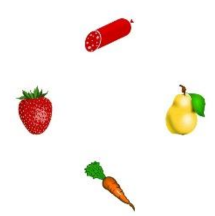
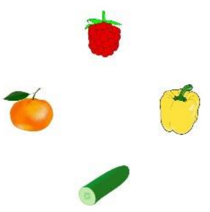
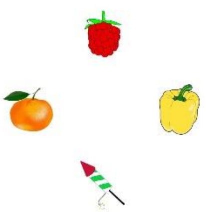
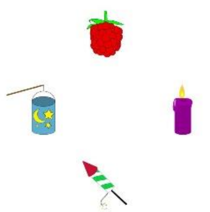
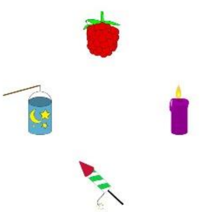
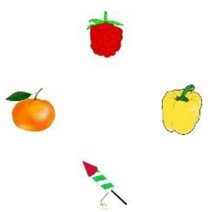
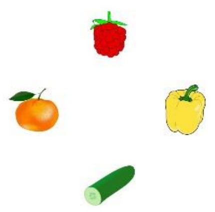
Visual Scenes of the 32 Items of Experiments 1 and 2

Item 1 (**verschlingen**, *eat*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 2 (**putzen**, *clean*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 3 (**schneiden**, *cut*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 4 (**reparieren**, *repair*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. Table continued on the next page.

Table B2 (Continued)

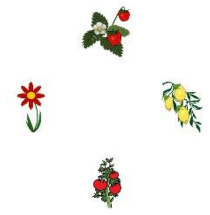
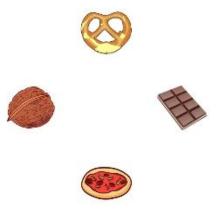
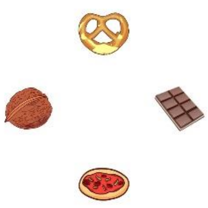
Visual Scenes of the 32 Items of Experiments 1 and 2

Item 5 (**probieren**, *taste*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 6 (befüllen, *fill*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 7 (naschen, *nibble*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 8 (entzünden, *ignite*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. Table continued on the next page.

Table B2 (Continued)

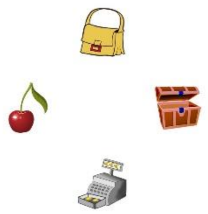
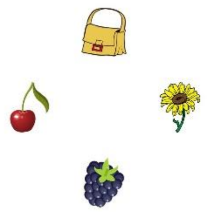
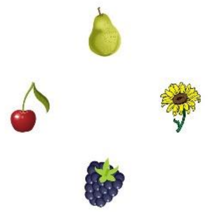
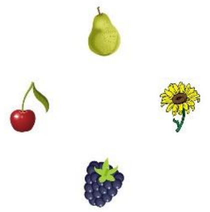
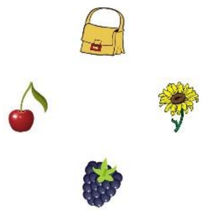
Visual Scenes of the 32 Items of Experiments 1 and 2

Item 9 (**gießen**, *water*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 10 (**spielen**, *play*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 11 (**hören**, *hear*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 12 (**kauen**, *chew*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. Table continued on the next page.

Table B2 (Continued)

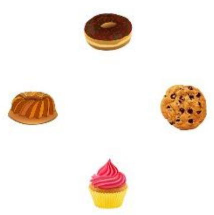
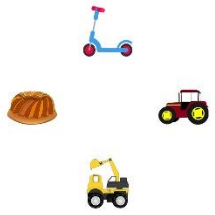
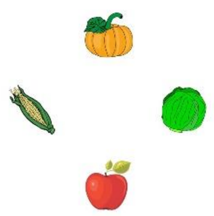
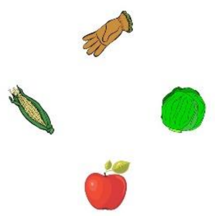
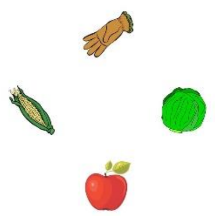
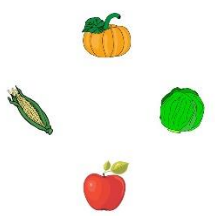
Visual Scenes of the 32 Items of Experiments 1 and 2

Item 13 (**schließen**, *close*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 14 (**pflücken**, *grab*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 15 (**verschütten**, *spill*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 16 (**starten**, *start*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. Table continued on the next page.

Table B2 (Continued)

Visual Scenes of the 32 Items of Experiments 1 and 2

Item 17 (**backen**, *bake*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 18 (**bremsen**, *brake*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 19 (**ernten**, *harvest*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 20 (**nähen**, *sew*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. Table continued on the next page.

Table B2 (Continued)

Visual Scenes of the 32 Items of Experiments 1 and 2

Item 21 (**sammeln**, *pick*)

4-consistent	3-consistent	1-consistent	0-consistent

Item 22 (**trinken**, *drink*)

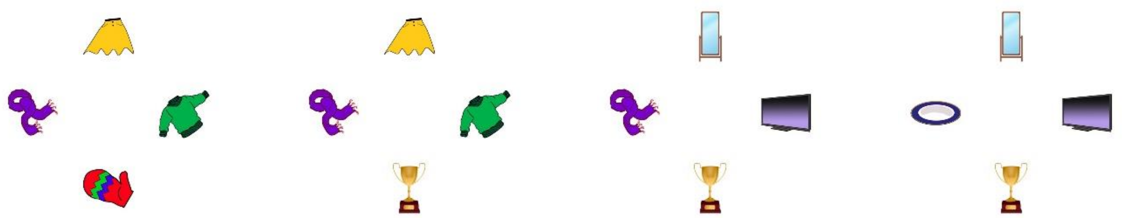
4-consistent	3-consistent	1-consistent	0-consistent

Item 23 (**zerbrechen**, *break*)

4-consistent	3-consistent	1-consistent	0-consistent

Item 24 (**stricken**, *knit*)

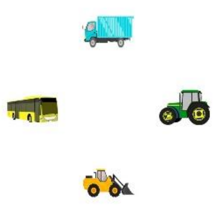
4-consistent	3-consistent	1-consistent	0-consistent

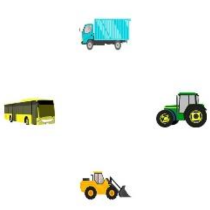


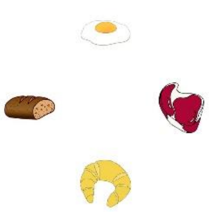
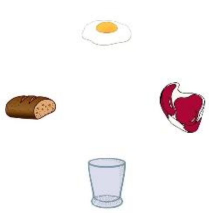
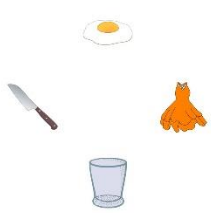
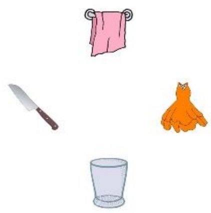
Note. Table continued on the next page.

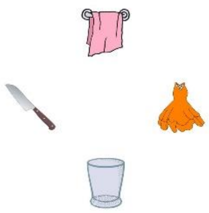
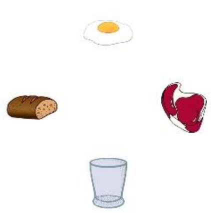
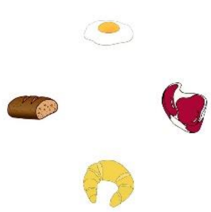
Table B2 (Continued)

Visual Scenes of the 32 Items of Experiments 1 and 2

Item 25 (**werfen**, *throw*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Item 26 (**parken**, *park*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

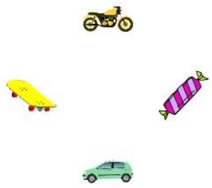
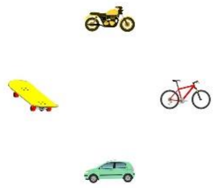
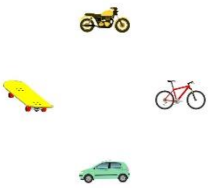
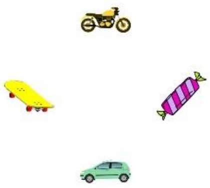
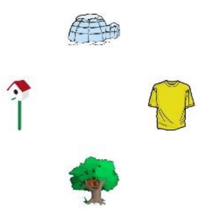
Item 27 (**genießen**, *enjoy*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Item 28 (**waschen**, *wash*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. Table continued on the next page.

Table B2 (Continued)

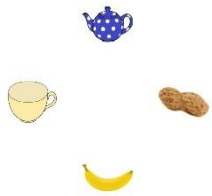
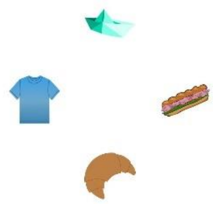
Visual Scenes of the 32 Items of Experiments 1 and 2

Item 29 (**futtern**, *guzzle*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 30 (**fahren**, *drive*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 31 (**bauen**, *build*)			
4-consistent	3-consistent	1-consistent	0-consistent
			
Item 32 (**bügeln**, *iron*)			
4-consistent	3-consistent	1-consistent	0-consistent
			

Note. For each of the 32 items we first show the constraining verb followed by the scenes in the four visual conditions (4, 3, 1, or 0 objects are consistent with the verb constraints).

Table B3

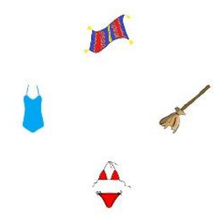
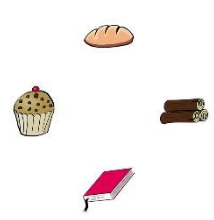
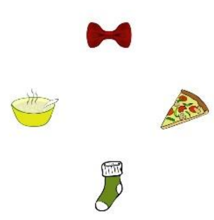
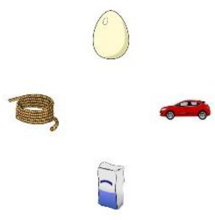
Filler Scenes of Experiment 1

Filler 1 (kleben, *stick*)	Filler 2 (schlucken, *swallow*)
	
Filler 3 (reinigen, *tidy*)	Filler 4 (bemalen, *paint*)
	
Filler 5 (löffeln, *spoon*)	Filler 6 (lackieren, *varnish*)
	
Filler 7 (falten, *fold*)	Filler 8 (kosten, *taste*)
	

Note. Visual scenes of the eight filler trials of Experiment 1. Each with the verb of the filler sentence first and the visual scene below.

Table B4

Filler Scenes of Experiment 2

Filler 1	Filler 2	Filler 3	Filler 4
			
Filler 5	Filler 6	Filler 7	Filler 8
			
Filler 9	Filler 10	Filler 11	Filler 12
			
Filler 13	Filler 14	Filler 15	Filler 16
			

Note. Visual scenes of the 16 filler trials of Experiment 2. Each with the filler number first and the visual scene below.

Appendix C — Psychometric Tests

Overview

In both of our Experiments additional cognitive and verbal tests were applied to control for the cognitive and verbal status of the sample. In Experiment 1 participants worked on the Semantic Verbal Fluency Task and the Color Naming Task. Here, task instructions each were presented verbally by the experimenter. In Experiment 2 participants worked on the Semantic Verbal Fluency Task, the Phonemic Verbal Fluency Task, and the Digit Symbol Substitution Task. Since Experiment 2 was an online study, task instructions each were presented as text on the screen.

Semantic Verbal Fluency Task

This task was applied in Experiments 1 and 2. It assesses semantic memory and executive functioning (Bialystok & Poarch, 2014; Brandeker, 2017; Friesen et al., 2015; Nielsen & Waldemar, 2016; Rosselli et al., 2002; Troyer et al., 1997), thus can control the verbal and cognitive capacities of a sample. Participants were instructed to name all the animals coming to their mind within 60 seconds while avoiding repetitions. The verbal answers were recorded via Audacity (version 2.3.2) for Experiment 1 and via the microphone of participants' computer for Experiment 2. Audio files were annotated in Praat (version 3.7.3) by a German native speaker expert who transcribed correct answers (verified by https://www.duden.de, last access: December 9, 2021) and filtered out errors (e.g., "sun") and disfluencies (e.g., "hm"). We automatically extracted the number of correct responses as the dependent variable using a custom Python script (version 6.0.37).

Phonemic Verbal Fluency Task

This task was only applied in Experiment 2. It is a measure of specific aspects of executive functioning such as memory, planning of strategies, maintaining and inhibiting information (Martins et al., 2007; Moura et al., 2014). Besides, it has been shown to be a proxy of literacy (Kavé, 2006; Kavé & Sapir-Yogev, 2023). Participants were instructed to name all words with the initial letter "S" coming to their mind within 60 seconds while avoiding repetitions and naming of countries, cities, or names. The verbal answers were recorded with the microphone of participants' computers. Audio files were annotated in Praat (version 3.7.3) by the same German native speaking expert who transcribed correct answers (verified by https://www.duden.de, last access: December 9, 2021) and filtered out errors (e.g., cities) and disfluencies (e.g., "hm"). The number of correct responses as the dependent variable was extracted automatically with a custom Python script (version 6.0.37).

Digit Symbol Substitution Task

This task was only applied in Experiment 2. It measures perceptual speed of processing (Hoyer et al., 2004; Huettig & Janse, 2016; Salthouse, 2000) and can be applied with adults (e.g., Huettig & Janse, 2016; Kray et al., 2008) and children (e.g., Karbach & Kray, 2007; Kray et al., 2008). While the original version of this task is a paper-pencil test (Salthouse, 1992; Wechsler, 1955), we applied an adapted online version (cf. Haeuser & Kray, 2022; Häuser et al., 2018, 2019). Here, participants were presented with a matching key on top of the screen that consisted of the digits from one to nine. A symbol was displayed under each digit (e.g., a bracket). Participants were first allowed to familiarize with the key. Then, they were presented with combinations of digits and symbols while the key remained on the screen. They should distinguish as quickly and as exactly as possible via button press whether the digit-symbol combinations were correct (button "S") versus incorrect (button "L") given the matching key. Each digit-symbol combination remained on the screen until an answer was given. Then, the next combination appeared. Only for 10 practice trials at the beginning participants received feedback. Then, 90 experimental trials followed, 45 each correct versus incorrect given the key. The mean reaction time for correct responses was calculated as dependent variable (e.g., Häuser et al., 2019).

Color Naming Task

This task was only applied in Experiment 1. It measures perceptual speed of processing and is an adapted version of the Digit Symbol Substitution Task that is suitable for the usage with young children (Karbach et al., 2011; Kray et al., 2006; Salthouse, 1992; Vergilova et al., 2021). Participants were presented with three sheets of paper on top of which they saw the same matching key consisting of four symbols in four colors (blue cross, green square, red triangle, yellow circle). Below them there were seven rows of these symbols, each containing all four objects in a different order, but colorless. Thus, there were 28 symbols on each sheet, for a total of 84 blank symbols. The matching key was presented on each sheet throughout the task. Participants were asked to name the color corresponding to each symbol referring to the key. They started with the first symbol on the first sheet and worked their way through line by line. Participants were asked to respond as quickly and as accurately as possible within 60 seconds and were told that they might not reach the last symbol. They received no feedback (aside from four practice trials on an extra sheet with a different key). Responses were noted, and participants could correct errors. The number of correct responses minus the number of errors was calculated as the dependent variable (cf. Salthouse, 1992).

Appendix D — Object Fixations

Table D1

Averaged Target Advantage Scores

Condition	Age group	Window	Object		
			Distractor 1	Distractor 2	Distractor 3
0-consistent	Adults	Baseline	-.01 (.42)	.00 (.41)	.01 (.42)
		Verb	.04 (.36)	.01 (.36)	.04 (.36)
		Noun	-.04 (.42)	.01 (.39)	-.02 (.43)
	Children	Baseline	.03 (.39)	-.01 (.43)	.00 (.42)
		Verb	.01 (.33)	-.05 (.38)	-.02 (.35)
		Noun	.04 (.40)	-.04 (.45)	.00 (.42)
			Distractor 1	Distractor 2	Distractor 3
1-consistent	Adults	Baseline	-.01 (.42)	.00 (.41)	.00 (.44)
		Verb	.25 (.42)	.22 (.45)	.24 (.42)
		Noun	.43 (.52)	.43 (.53)	.46 (.49)
	Children	Baseline	.00 (.45)	-.02 (.44)	.01 (.42)
		Verb	.26 (.41)	.28 (.40)	.24 (.44)
		Noun	.58 (.49)	.56 (.51)	.59 (.47)
			Competitor 1	Competitor 2	Distractor
3-consistent	Adults	Baseline	.01 (.40)	.00 (.39)	-.01 (.42)
		Verb	.02 (.36)	-.04 (.41)	.10 (.35)
		Noun	.47 (.45)	.47 (.45)	.51 (.43)
	Children	Baseline	-.02 (.43)	-.01 (.39)	-.05 (.43)
		Verb	.00 (.42)	-.01 (.39)	.13 (.33)
		Noun	.45 (.44)	.42 (.48)	.52 (.36)
			Competitor 1	Competitor 2	Competitor 3
4-consistent	Adults	Baseline	.00 (.38)	-.02 (.39)	-.01 (.39)
		Verb	.02 (.37)	-.02 (.4)	.01 (.39)
		Noun	.43 (.42)	.43 (.44)	.42 (.46)
	Children	Baseline	-.06 (.43)	-.07 (.42)	-.02 (.40)
		Verb	-.03 (.40)	.00 (.38)	-.01 (.40)
		Noun	.45 (.41)	.43 (.42)	.44 (.42)

Note. The averaged target advantage scores for each condition, time window, and age group are shown. The column "Object" shows which object was referenced to the (pseudo-)target. Standard deviations are presented in parentheses.

Appendix E — Index of Cognitive Activity

Table E1

Results of the Models on the ICA Values in the Baseline and Postview Region

Region	Comparison	β	*SE*	*z*	*p*	95% *CI*
Baseline	Intercept	3.52	0.02	150.42	.000	3.08, 3.95
	v1	0.01	0.03	0.41	.681	-0.41, 0.46
	v2	-0.03	0.03	-1.16	.245	-0.47, 0.40
	v3	-0.04	0.03	-1.29	.197	-0.49, 0.38
	a	-0.06	0.05	-1.33	.182	-0.68, 0.57
	v1:a	-0.12	0.07	-1.78	.075	-0.77, 0.50
	v2:a	0.09	0.05	1.59	.113	-0.55, 0.72
	v3:a	0.04	0.06	0.60	.549	-0.57, 0.70
Postview	Intercept	3.45	0.03	117.98	.000	3.39, 3.50
	v1	-0.07	0.04	-1.83	.068	-0.14, 0.01
	v2	0.00	0.04	-0.06	.949	-0.08, 0.07
	v3	0.00	0.04	-0.04	.970	-0.07, 0.07
	a	-0.04	0.06	-0.70	.485	-0.15, 0.07
	v1:a	0.03	0.08	0.40	.692	-0.12, 0.18
	v2:a	0.02	0.07	0.34	.737	-0.12, 0.17
	v3:a	-0.08	0.06	-1.26	.208	-0.21, 0.05

Note. The models on the ICA values (ica) in the baseline and postview region covered the factors condition and age group. v1, v2, and v3 are the first, second, and third condition contrasts. a is the age group contrast. The converged models:

Baseline: ica~(v1+v2+v3)*a+(1+(v1+v2+v3)||subject)+(1+(v1+v2+v3)*a||item)
Postview: ica~(v1+v2+v3)*a+(1+(v1+v2+v3)||subject)+(1+(v1+v2+v3)*a||item)

Table E2

Averaged ICA Values

Age group	Region	Condition			
		0-consistent	1-consistent	3-consistent	4-consistent
Adults	Baseline	34.25 (11.77)	32.95 (12.70)	33.15 (13.45)	33.59 (12.41)
	Verb	32.37 (11.81)	31.57 (13.22)	31.69 (13.03)	30.64 (13.30)
	Spill-over	32.77 (11.36)	30.47 (12.29)	31.69 (12.81)	31.61 (12.53)
	Noun	34.85 (10.50)	34.36 (11.07)	36.72 (10.61)	36.39 (10.32)
Children	Baseline	34.61 (13.27)	35.97 (12.16)	36.68 (11.19)	38.25 (10.83)
	Verb	32.86 (14.18)	33.62 (14.32)	34.64 (12.43)	32.91 (12.85)
	Spill-over	33.18 (12.18)	32.67 (11.56)	36.69 (9.45)	32.71 (11.61)
	Noun	35.57 (12.13)	35.28 (11.96)	36.45 (11.63)	35.21 (13.08)

Note. The averaged ICA values for each condition, region, and age group are shown.

Standard deviations are presented in parentheses.

Appendix F — Pupil Sizes

Table F1

Results of the Model on the Pupil Sizes in the Baseline Region

Comparison	β	*SE*	*df*	*t*	*p*	95% *CI*
Intercept	3000.79	95.03	60.27	31.58	.000	2814.67, 3186.86
v1	31.42	25.60	70.70	1.23	.224	-18.58, 81.62
v2	-7.68	18.04	1480.40	-0.43	.671	-43.10, 27.59
v3	4.25	23.57	63.91	0.18	.858	-42.02, 50.30
a	-518.60	189.63	59.75	-2.74	.008	-889.91, -147.15
v1:a	22.90	51.21	70.78	0.45	.656	-77.26, 123.15
v2:a	14.48	36.06	1480.69	0.40	.688	-56.13, 85.21
v3:a	-20.85	47.15	63.91	-0.44	.660	-112.87, 71.96

Note. The model on the raw pupil size values (p) in the baseline region covered the factors condition and age group. v1, v2, and v3 are the first, second, and third condition contrasts. a is the age group contrast. The converged model:

p~(v1+v2+v3)*a+(1+(v1+v3)||subject)+(1+a||item))

Table F2

Averaged Proportional Pupil Sizes

Age group	Condition	Region				
		Baseline	Verb	Spill-over	Noun	Postview
Adults	0-consistent	2716 (897)	1.95 (7.74)	1.49 (10.74)	2.46 (10.87)	4.20 (13.46)
	1-consistent	2775 (852)	1.14 (7.56)	-0.02 (10.81)	1.09 (12.73)	0.81 (13.34)
	3-consistent	2749 (851)	1.61 (7.88)	0.78 (10.08)	3.35 (14.31)	2.93 (15.80)
	4-consistent	2749 (861)	1.16 (7.07)	1.18 (10.64)	3.37 (13.26)	2.77 (13.27)
Children	0-consistent	3259 (562)	2.17 (7.86)	2.35 (10.58)	4.16 (13.21)	5.96 (12.54)
	1-consistent	3320 (578)	2.35 (6.82)	2.64 (8.44)	4.22 (10.33)	3.63 (10.94)
	3-consistent	3293 (655)	1.64 (6.54)	1.70 (8.96)	3.02 (10.82)	2.51 (11.36)
	4-consistent	3266 (557)	1.82 (7.15)	2.57 (8.80)	5.21 (11.50)	3.94 (11.94)

Note. The averaged values for each condition, region, and age group are shown. The column "Baseline" refers to the *raw* pupil size values in the baseline region (without decimals for ease of comprehension). All other regions refer to the *proportional* pupil sizes relative to the baseline. Standard deviations are presented in parentheses.

Appendix G — Processing Times

Table G1

Averaged Raw and Log-Transformed Processing Times

Region	Condition	Processing times			
		Raw		Log-transformed	
		Children	Adults	Children	Adults
Baseline	0-consistent	900 (500)	543 (293)	6.67 (0.52)	6.17 (0.51)
	1-consistent	925 (502)	541 (322)	6.70 (0.52)	6.14 (0.54)
	3-consistent	926 (494)	546 (302)	6.70 (0.50)	6.17 (0.51)
	4-consistent	901 (493)	532 (286)	6.67 (0.52)	6.14 (0.52)
Verb	0-consistent	1018 (597)	527 (290)	6.76 (0.59)	6.13 (0.52)
	1-consistent	1071 (644)	529 (304)	6.80 (0.61)	6.13 (0.53)
	3-consistent	1018 (565)	518 (248)	6.78 (0.55)	6.13 (0.49)
	4-consistent	966 (523)	495 (242)	6.73 (0.54)	6.08 (0.50)
Post-Verb	0-consistent	3230 (1618)	1807 (938)	7.95 (0.52)	7.37 (0.53)
	1-consistent	3068 (1501)	1707 (869)	7.91 (0.50)	7.31 (0.53)
	3-consistent	3197 (1668)	1821 (990)	7.94 (0.51)	7.36 (0.55)
	4-consistent	3147 (1561)	1739 (907)	7.93 (0.50)	7.32 (0.54)
Noun	0-consistent	1107 (571)	667 (399)	6.88 (0.51)	6.35 (0.54)
	1-consistent	1075 (547)	659 (359)	6.85 (0.52)	6.36 (0.52)
	3-consistent	1126 (570)	646 (349)	6.90 (0.51)	6.33 (0.53)
	4-consistent	1149 (580)	659 (338)	6.92 (0.52)	6.36 (0.51)
Post-Verb	0-consistent	3035 (1529)	1989 (1119)	7.90 (0.48)	7.46 (0.51)
	1-consistent	2938 (1521)	1942 (1024)	7.87 (0.48)	7.45 (0.50)
	3-consistent	2946 (1346)	1932 (906)	7.89 (0.44)	7.47 (0.44)
	4-consistent	2983 (1449)	2020 (1069)	7.89 (0.47)	7.49 (0.49)

Note. The averaged raw and log-transformed processing times for each condition, region of interest, and age group are shown. Raw values in the post-verb and post-noun region are substantially higher than in the other regions because they summarize the processing times of three words each. Standard deviations are presented in parentheses.