The Influence of Visual Information on Word Predictability and Processing Effort

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultäten
der Universität des Saarlandes

vorlegt von

Christine Susanne Ankener

aus Rodalben

Saarbrücken, 2019

Dekan:	Univ.-Prof. Dr. Heinrich Schlange-Schöningen
Berichterstatter:	Dr. Maria Staudte
	Prof. Dr. Jutta Kray

Tag der letzten Prüfungsleistung: 24. Juni 2019

To my mother . . .

Day after day, day after day,
We stuck, nor breath nor motion;
As idle as a painted ship
Upon a painted ocean.
Water, water, every where,
And all the boards did shrink;
Water, water, every where,
Nor any drop to drink.
– *The Rime of the Ancient Mariner* –

Acknowledgements

I thank my dad and my brother. I love you.
– Rotlaust tre fell – (*Einar Selvik*)

Thank you Maria, Sébastien and Nikolina for the advice, patient guidance and encouragement.

Abstract

A word's predictability or surprisal in a linguistic context, as determined by cloze probabilities or language models (e.g., Frank, 2013a) is related to processing effort, in that less expected words take more effort to process (e.g., Hale, 2001). This shows how, in purely linguistic contexts, rational approaches have been proven valid to predict and formalise results from language processing studies. However, the surprisal (or predictability) of a word may also be influenced by extra-linguistic factors, such as visual context information, as given in situated language processing. While, in the case of linguistic contexts, it is known that the incrementally processed information affects the mental model (e.g., Zwaan and Radvansky, 1998) at each word in a probabilistic way, no such observations have been made so far in the case of visual context information. Although it has been shown that in the visual world paradigm (VWP), anticipatory eye movements suggest that listeners exploit the scene to predict what will be mentioned next (Altmann and Kamide, 1999), it is so far unclear how visual information actually affects expectations for and processing effort of target words. If visual context effects on word processing effort can be observed, we hypothesise that rational concepts can be extended in order to formalise these effects, hereby making them statistically accessible for language models. In a line of experiments, we hence observe how visual information – which is inherently different from linguistic context, for instance in its non-incremental - at once - accessibility – affects target words. Our findings are a clear and robust demonstration that the non-linguistic context can immediately influence both lexical expectations, and surprisal-based processing effort as assessed by two different on-line measures if effort (a pupillary and an EEG one). Finally, we use surprisal to formalise our measured results and propose an extended formula to take visual information into account.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

(Probabilistic) Prediction in *situated* Language Processing This thesis aims at observing the (statistical) effects of *visual* context information on target word expectations and language related processing effort, which, as opposed to purely *linguistic* context information, is still majorly unknown. Before going into detail with respect to the role and effects of visual information, we shortly outline what is already known about the effects of linguistic context information on predictive processing and the link between predictions, informed by linguistic context information, and actual processing effort.

Although some open questions about details of predictive processing still remain (such as, for instance, how necessary predictions are for language comprehension (for a discussion, see e.g. Huettig and Mani, 2016), or how specific comprehenders predict in different contexts (see e.g., Van Petten and Luca, 2012) (for a more detailed discussion about those questions, see the subsequent Background Chapter)), psycholinguistic research has in general collected ample evidence for the existence of predictive processing in purely linguistic contexts in natural language comprehension. Such evidence exists in the form of various data collected in psycholinguistic and psychological experiments that can only be explained by comprehenders predicting linguistic units before actually encountering them.

One very prominent example is the so-called *garden path* phenomenon in sentence comprehension, where comprehenders assign an initial interpretation to a temporarily ambiguous sentence part, which then needs to be revised as further bottom-up input is encountered. Revising means that an alternative, syntactically less frequent and, hence, less expected interpretation of the sentence has to be found. This causes increased processing difficulty, as reflected by, for instance, longer reading times (e.g., Ferreira and C. Clifton, 1986; Kuperberg and Jaeger, 2016a). The fact that an alternative interpretation needs more time to be processed suggests that comprehenders had actually predicted something a priori, namely the more frequent interpretation, so that increased effort is required for revision and processing of the

less predicted interpretation. It has further been discovered in very early research, that more, compared to less, predicted words are linked to shorter reaction times in behavioural tasks (Fischler and Bloom, 1979) and even that predictive processing in linguistic contexts can explain results from electrophysiological measures, such as the ERP component N400 (Kutas and Hillyard, 1980).

Not least due to this great explanatory power, first models of language processing soon came up in the early eighties in the field of cognitive psychology, which included predictive processing. Such models are, for instance, the so called (mental) situation models, as introduced by Johnson-Laird (1983) and extensively reviewed by Zwaan and Radvansky (1998). They were based on the idea that predictions in linguistic contexts are computed on-line and based on a comprehender's current model, or interpretation, of what is being communicated. More specifically, situation (or mental) models generally describe the complex mental representation in a comprehender's mind, simulating different aspects of the situation currently being communicated. In the case of linguistic context information – which is known to be processed incrementally – new information is permanently (i.e. at each linguistic unit) fed into the mental model. The new information affects the statistics of the mental model as it is integrated and can then result in predictions about what will be the next linguistic unit best fitting the previous context (Zwaan and Radvansky, 1998). At which time points in the input visual information – which is not processed as incrementally as language, but rather provides a lot of information at once – is fed into the model, and possibly affects the statistics, is still unclear.

A large body of work in the following decades observed further important details of predictive processing in (mainly) linguistic contexts, showing, for instance, how linguistic context information affects the mental model statistically: While early theories suggested that, based on previous linguistic context, listeners would settle on one specific prediction, having to re-analyse the sentence if this prediction was not confirmed by the actual input (for a discussion, see Kutas et al., 2011), more and more collective behavioural and neural evidence instead proposes that comprehenders rather evaluate linguistic context information to extract probabilities. In other words, instead of settling on one specific prediction, probabilistic evaluations of the linguistic context can result in adapted and respectively probabilistic predictions about upcoming input. The information encountered with each new word or linguistic unit in the input stream hence affects the statistics and alters probabilities of the mental model, which – in most cases – results in probabilistic predictions about upcoming words or units (for a detailed review of experimental evidence, see Kuperberg and Jaeger, 2016b). Predictive processing, according to this idea, is hence probabilistic and highly adaptive with respect to the given (linguistic) context. In this view, it becomes more apparent why predictive

processing can be beneficial for comprehenders: If predictions are probabilistic and highly adapted to the statistics of the current context, they are less likely to be wrong, that is, they are less likely to require revision. The encounter of predictions being probabilistic hence defuses one of the major arguments against the predictive nature of language comprehension (e.g., Delaney-Busch et al., 2017).

Probabilistic predictive processing in linguistic contexts has quickly proven to be of high psychological validity, as it ideally explains *graded* effects on psychological measures of processing effort, such as reading times (e.g., Hare et al., 2007). Interestingly, despite the fact that probabilistic prediction in linguistic context has been in the focus of many studies, again, it is still not clear whether the same rules hold for visual contexts. More precisely, it is neither known how visual information affects the mental model and the resulting predictions *statistically*. Nor is it transparent how interleaved the evaluation of visual and linguistic information is, although the influence of additional visual context is very relevant in situated language processing, as given in a lot of very common everyday situations.

With results like those found by Hare et al. (2007) – showing that the predictability of a word in a given linguistic context affects processing effort as assessed by reading times, in a graded way the actual connection – the correlation between predictability and processing effort becomes more apparent. The link here is that the amount of effort each word or linguistic unit in an input stream requires for processing is majorly determined by the word's probabilistic expectancy, as derived from the statistics of the current mental model: The more predictable a linguistic unit is in its context, the less new information it contributes to the recent interpretation of the situation. Less information then needs less effort to be processed and integrated into the congruent previous information structure of the mental model. An unexpected linguistic unit, however, contributes more information, and can hereby majorly alter the current interpretation, since it is less congruent with the previous information. The additional amount of information then requires more effort for processing and integrate into the current interpretation. As a result, processing effort increases as information conveyed by the linguistic units increases as a function of decreasing expectancy.

Here again, it is unknown whether the same linking hypothesis holds for the case of visual information. Although at first sight, it may sound intuitive to assume that it does because what someone sees in her direct environment could majorly influence the statistics of the mental model, it is indeed a highly appropriate question. Although visual information may be very likely to generally have an effect on word processing effort that the respective word alone cannot account for, it does not necessarily need to be identical to linguistic context effects. After all, it is possible that the non-incremental nature of visual information could require differently distributed effort to process. Non-incremental, visual information could,

for instance, have an overall increasing effect on processing effort, or, for instance, affect it only at very specific points during the input (i.e. whenever what is seen alters the model of what is being communicated). It is even possible that visual context affects a *different kind* of effort, such as effort with respect to saccade planning. A detailed observation of visual context effects is hence a vital consecutive step and contributes to a better understanding of processing difficulty in situated language processing as given not only in every day situations but also in the context of the widely used VWP. Current, lively debates about how "prediction encouraging" (Mani and Huettig, 2014) visually presented alternatives are, and whether or not visual context effects possibly disqualify the Visual World Paradigm as an objective tool to study predictions, show just how important further research and a possible quantification of visual context effects are at this time.

The main goal of this thesis is hence the observation of *whether* and *how* visual context information is evaluated for the sake of language comprehension, and *how* exactly it really affects the statistics of the comprehender's mental model, and subsequently, the target word predictability and linguistic processing effort.

In fact, we do not need to start from scratch, as there are already some well established findings in the literature, with respect to visual information. The most important one in the context of this thesis is an influential observation made by Kamide et al. (2003): The authors showed that comprehenders can generally use verbal constraints, mapped to visual information, in order to *anticipate* target words in the Visual World Paradigm (VWP). These findings are an appropriate starting point, as they tell us *that* visual information is used by comprehenders to anticipate words, but not *how* this information is evaluated statistically, and neither *how* it is integrated into the current interpretation of the situation being communicated, hereby possibly altering purely linguistic predictions.

Additional indication for a possible statistical relevance of visual context comes from very recent results, suggesting that comprehenders rapidly and rationally adapt local interpretations – and hence, their predictions about target words – to the statistical characteristics of their *broader* environment (e.g., Kleinschmidt and Jaeger, 2016). It is reasonable to assume that in the case of situated language processing, the "broader environment" includes the statistical characteristics of the visual context in which an utterance is processed. Further, Frank and Goodman (2012) proposed that – in the context of a referencing game in a multi-modal context – comprehenders deploy Bayesian inference based on information from *all* involved modalities to reconstruct a speaker's intended referent.

If all involved modalities can be used to draw conclusion about a referent in this setup, it is generally appropriate to assume that comprehenders evaluate multiple modalities in an

interleaved way, as long as it is beneficial for their current aim. The derived information should hence have an impact on the current interpretation and the resulting predictions.

Similar to the question raised earlier in the context of the discussion about the prediction encouraging nature of visual contexts, here it would again be very beneficial to observe at which point during the input the evaluation of multiple modalities is actually interleaved, and what the statistical effect of visual information is on the mental model. It is, for instance possible, that the effect of visual information is similarly fine grained and probabilistic as in the case of purely linguistic information. Visual information could then cause comprehenders to expect target words based on all possible alternative target references – as defined by the linguistic information – observed in the visual context. Alternatively, a probabilistic evaluation of visual information could be too effortful – at least when a task has to be solved. In this case, no further probabilistic evaluation could be performed with respect to the visual context, leaving it with no further effect on linguistic processing effort, as long as no information from the visual context is disruptive (i.e., in some way incoherent with the sentence).

The mentioned questions show that, although the effect of visual information on linguistic processing effort and predictions is relevant for a wide range of studies, more research is required in order to observe further important details. We therefore approach these questions by using a novel combination of different paradigms and measures of effort (i.e., behavioural, pupillary, and ERP measures), that enables us to observe *and* to additionally quantify effects of visual information on the statistics of the mental model, word expectations and word processing. We hereby establish an important link between multi-modal – that is, visually informed – predictability of a word and actual brain activity related to the processing of that word. The subsequent paragraphs will elaborate in further detail on how this will be done. Our results support most recent models of rational adaptation and bear important implications for the use of visual world setups to observe predictions in language comprehension, as well as for statistical models of language in situated language processing.

The linking hypotheses: Correlating multi-modal predictability and processing effort
A suitable linking hypothesis is essential when linking visually informed predictability and (linguistic) processing effort. As previously mentioned briefly, such a hypothesis already exists in the case of purely linguistic context effects: Namely, the more predictable a word is in its linguistic context, the less information it conveys, and the less effort is needed to process it. This correlation can be described by various rational approaches, which, unsurprisingly, have already proven to be a very powerful tool in language science, especially when it comes to objectively accounting for processing difficulty data collected in purely linguistic

experiments. Specifically two information theoretic concepts have increasingly been in the focus of psycholinguistic literature, namely surprisal and entropy (reduction). We will now use those two concepts to illustrate why rational approaches are inherently powerful for the linking of word predictability and processing difficulty: Surprisal and entropy both originate from the fields of statistical mechanics (Tolman, 1938) and physics, where surprisal describes the statistical surprise when a random variable is sampled, and entropy refers to the number of all possible micro states a system can take on at a certain point in time.

Shannon (1949) picked up the basic idea in his influential work on information theory, in which he aimed at finding a logarithmic (which is mathematically more suitable for engineering tasks) measure of the amount of information conveyed by a linguistic unit. Shannon, inspired by the concepts of predictability and entropy, subsequently formalised the information content of a unit as its statistical (conditional) probability to appear in the current context. The amount of information is then expressed in bits needed to describe the respective linguistic unit. Although he never explicitly used the term surprisal, Shannon (1949) defined the statistical structure of an utterance as a set of transition probabilities, that is, the probability of a letter or word x to be followed by letter or word y is calculated over the range of all possible continuations, which is later often termed surprisal (see, e.g. Bernstein and Levine, 1972). Closely linked to the predictability (surprisal) of a unit, he defined entropy as uncertainty (i.e. missing information) about upcoming input. He hereby formalised the previously only informally described idea that very predictable linguistic units convey less new information, resulting in higher uncertainty, compared to unpredictable ones.

Shannon's work is the basis for much work done in various research fields, such as, for instance, psychology (Attneave, 1959) and – most importantly in the present context – psycholinguistics (e.g., Hale, 2001; Smith and Levy, 2008), where the concepts were adapted to answer questions in the field and have had a significant impact ever since. Based on Shannon's notion of the predictability of information in a context, Hale (2001) finally proposed the versatile notion of (linguistic) surprisal to predict processing complexity in linguistic contexts. Hales work established an important, direct link between (incremental) language processing and comprehension difficulty: The surprisal of a word predicts the effort it requires to process. Smith and Levy (2008) subsequently presented a broad-coverage analysis of the functional relationship between probability and reading times, extended the empirical coverage of Hale's approach. In fact, surprisal is powerful enough as a predictor of processing effort to explain a wide range of data and phenomena in language processing, such as garden paths. Since its introduction to the fields of psycholinguistics by Hale, surprisal has subsequently proven capable of describing a wide range of psycholinguistic data: It has, for instance, been shown that readers take more time to read words with higher

surprisal (Demberg and Keller, 2008; Smith and Levy, 2013). Frank (2013b), for example, could further show that a specific word's (linguistic) surprisal in its linguistic context can be predictive of the ERP component N400's amplitude, hereby revealing that the N400 can be a reliable measure of information content. It is common practise in such experiments to derive surprisal (or entropy) values from language models (LM) or cloze probabilities in order to quantify the amount of information conveyed by specific words, before using these values in statistical models as a predictor of processing effort experienced by the listener upon encountering the respective word (e.g., DeLong et al., 2005; Demberg and Keller, 2008; Linzen and Jaeger, 2014).

Similar to surprisal, the closely related entropy was soon also adapted for the context of linguistic data by utilizing it to describe the uncertainty about either the next word or the entire rest of the sentence (Hale, 2003). In linguistic terms, entropy defines the uncertainty over possible continuations of a sentence, at a specific linguistic unit (note that entropy reduction defines the amount of uncertainty reduced by a unit). In other words, as opposed to surprisal, which depends on the previous actual context, entropy captures the subsequent possible context of a linguistic unit.

Hale (2003) employs the concept to suggest that a processor eagerly performs ambiguity resolution, based on incoming (incremental) information, in order to decide at each point of the input stream which one of competing, alternative interpretations is most suitable. The amount of work that needs to be done in order to reduce uncertainty, or entropy for that matter, is directly linked to processing effort, as in his case, reflected by reading times. In other words, the more information a word submits, the further the processing and integration of that word into the current mental representation of what is communicated reduces uncertainty (i.e., entropy) about the entire sentence structure (i.e., including upcoming parts) and the longer it takes to read that word (Linzen and Jaeger, 2014). Entropy (reduction) has further potential, and could, for instance, account for recent findings by Maess et al. (2016): The authors ran an MEG study to observe prediction signatures in the brain and found enhanced activity for highly predictive compared to less predictive verbs, as well as an inverse correlation between the verb constraint and the N400 on the subsequent noun. That is, highly predictive (i.e., constraining) verbs took more effort to process, while the processing of the subsequently very predictable nouns was facilitated. In terms of entropy (reduction), this result could be interpreted as more constraining verbs, being more informative and reducing more entropy about the referent, hence requiring more effort to process.

Due to their major impact and power with respect to correlating linguistically derived context probabilities and the resulting predictability of linguistic units with actual processing effort, we hypothesise that surprisal and entropy (reduction) can be extended to further

account for data from situated language processing. That is, the concepts could be adapted in order to cover a possibly statistical influence of additional visual information on word predictability and actual processing effort. The metrics can then provide a linking function between probabilistic word expectations (i.e., as derived from the combination of what is seen *and* heard) and the actual processing effort required for the target words. This way, the concepts can help to explain processing effort in situated communication, hereby allowing for insights into how and when multi-modal information is integrated and evaluated.

The paradigm: Observing the influence of visual information Now that rational concepts for a potential between visually influenced prediction and processing effort are identified, a suitable paradigm allowing for the observation of such effects is required. Especially in the context of visual, combined with linguistic information, the Visual World Paradigm (VWP) is very likely the first thing that comes into mind. This is mainly due to two prominent main characteristics of the paradigm: First and foremost, displaying objects provides the exceptional benefit of enabling the experimenter to exactly control and monitor over quantifiable referential uncertainty and predictability of a referent word, which is extremely complicated in a purely linguistic context. With no extensively strict contextual constraints conveyed by the linguistic context, it is nearly impossible to say whether and how many alternative target referents participants think of. Displaying objects allows for the assumption that comprehenders think of what they actually see.

Second, effects attributable to predictive language processing (and the associated cost of dis-confirmed predictions) have been found in some but not in other studies featuring linguistic context (for a summary, see Van Petten and Luca, 2012), the VWP has the rare advantage of reliably revealing signs of target word anticipation, as manifest in (usually verb driven) eye movements towards possible target objects displayed, prior to the actual target word (Altmann and Kamide, 1999). Anticipatory patterns, enabled by visual context, reveal any (also temporarily) activated interpretations and are interpreted as signs of visual context evaluation in expectancy of the target referent, which is especially beneficial in our context.

Despite its striking advantages, the paradigm does not come without pitfalls that need careful discussion as they bare implications for the interpretation of results, specifically with respect to target word predictions. We will briefly outline the inherent duality (which is basically the down side of the above mentioned advantages) of this powerful paradigm before deploying it in the actual experiments: Because presenting comprehenders with objects, that are congruent with the linguistic stimuli, causes them to consider what they see in the direct context, the VWP can never be an entirely neutral tool for the observation of linguistic surprisal. Instead it always adds substantial information, as the presence of visual referent

options alters purely linguistic predictions. This feature has caused vital debates within the literature. Some researchers even call VWP contexts "prediction-encouraging" (Huettig and Mani, 2016). The authors hereby suggest that it is not necessarily fair to visually propose target referents and still call what is going on "prediction". The discussion in Huettig and Mani (2016) is, at the same time, one of various examples showing how important it is to achieve a better understanding of whether and how visual information alters linguistic predictions and processing effort. Based on this discussion, we state that we do not consider surprisal in the VWP to be identical with purely linguistic context surprisal. However, we neither suggest that the VWP is overly "prediction encouraging", or even inappropriate for observing language related processing difficulty. First of all, because processing language in a visual context is a very common thing in everyday situations. Second, not every visual context is necessarily overly suggestive and transparent with respect to target word information.

Since the paradigm's strength make it an essential and contemporary irreplaceable tool for approaching open questions about effects of visual context information on word surprisal and effort, we argue that the paradigm is a valid context for observing effects of visual information. At the same time, it would be wrong to ignore the contra-argument and to not acknowledge the impact of visually presented target options on linguistic predictions. We hence distinguish between purely linguistic predictions and visually informed expectations for a target word. That is, similar to other authors before, we suggest a clear definition of the term prediction in the context of our experiments. Van Petten and Luca (2012), for instance, observe ERP signatures in reaction to confirmed versus dis-confirmed predictions, and find that it is most appropriate to think of their results as attributable to more general, rather than very specific predictions. In this context, a distinction between more general and lower level predictions about specific words was essential. In our context, however, we rather focus on the role of visual information for statistics of current linguistic interpretation and resulting predictions than on questions about the specificity of predictions. We therefore propose a similar distinction that suits our context: The term expectation is used to refer to visually influenced target word predictions (i.e., more or less specific predictions informed by not only one, but two modalities), while the term prediction means purely linguistically informed expectations in contexts where no additional information is presented visually (i.e., predictions derived from uni-modally conveyed information). With this distinction in mind, the VWP is deployed for the vital control of the exact number of referents considered by the comprehender, hereby enabling an (otherwise impossible) neat manipulation of target word's multi-modal predictability. A further advantage of the paradigm with respect to our initial research questions becomes apparent when it is combined with a novel, on-line pupillary measure of processing effort: Possible results can extend established findings concerning

anticipatory eye movements (Altmann and Kamide, 1999) as the combination of paradigm and measure allows us to observe any possible effects of anticipation on word processing effort. The respective measure will be described in more detail in the subsequent paragraph.

The measure: The pupillary system and the Index of Cognitive Activity Mental, or cognitive effort, especially with respect to language processing, is commonly assessed by behavioural measures such as, for instance, reading times (e.g., Demberg and Keller, 2008; Smith and Levy, 2013), as well as more direct, psychophysiological measures as, for instance, the ERP component N400 (Frank, 2013b). While the mentioned measures are reliable and well established in purely linguistic contexts, things become increasingly difficult with the addition of visual information: As comprehenders are required to move their eyes around a scene freely, ERP components can be very delicate with respect to ocular artifacts caused by those eye movements. Although there are mathematical and statistical methods for ocular correction, it is still very likely that significant data are contaminated. At the same time, reliable behavioural measures such as reading times drop out if comprehenders are a) busy looking at scenes presented with the linguistic stimuli, and, a fortiori, b) auditorily presented with sentences. In those cases, one specific, very direct (with short latencies, neurophysiological measure can be deployed to assess processing difficulty, namely pupillometry. It has been used for more than 50 years, based on the observation that (small) changes in the pupil diameter take place in reaction to task induced mental effort and is, most importantly, robustness with respect to eye movement artifacts when assessing of processing effort in the VWP.

When assessing effort with pupillometry, instead of causing contaminating artifacts, eye movements can be tracked and evaluated as a source of additional, highly valuable information. That is, overt attention (e.g., as reflected by anticipatory eye movements) can reveal even temporarily activated objects during incremental language processing, which is a feature unique to this measurement. The simultaneous assessment of eye movements and processing effort can be crucial when observing direct effect of encountered, coherent visual information on current interpretations and predictions on the linguistic level.

In fact, it has already been shown that pupillometry is sensitive specifically with respect to processing effort related to visual information during language comprehension: Scheepers and Crocker (2004), for instance, showed that participant's pupil size increased upon encountering visual information disambiguating a noun phrase as being the anticipated object of an O-V-S order, compared to when the visual context suggested this noun phrase to be the anticipated subject of a more usual S-V-O order. Further, Engelhardt et al. (2009) observed that pupil diameter, rather than task performance, reliably reflects processing effort. In

their studies, pupil diameter indicated a positive effect of visual context on the resolution of temporary syntactic ambiguities in sentences, despite the presence of conflicting prosodic structures.

Besides all its prominent advantages, however, pupillometry has one major pitfall: It can not only be influenced by the psycho-sensory reflex related to mental processes and cognitive effort, but also by a range of other factors. More specifically, while emotional arousal, near-reflex (leading the ocular motor system to control the depth of the eye's field), as well as other deviations of the optical system may affect the pupil's diameter, the main determinant is the light reflex (Beatty, 1982). This means that the eye largely reacts to changes in environmental illumination and light which, depending on the setup, can hardly be fully controled. This bears implications for the establishment of a reliable correlation between pupillary responses and (linguistic) processing effort. Namely, pupillary responses need a further, more detailed differentiation with respect to what causes them.

The vital basic observations about pupillary responses to mental effort, on which later differentiation approaches are based, were already made very early on: Hess and Polt (1964), for instance, reported pupillary responses in reaction to multiplication exercises, where pupil diameter increased with the difficulty of the problem participants were confronted with. This lead the authors to interpret the measured dilations in their experimental context as an indicator for reasoning in arithmetic tasks. A few years later, Kahneman and Beatty (1966) conducted a short-term memory task in which they presented strings of 3 – 7 digits. As a result, they observed that the overall pupillary diameter in participants increased with each additional digit, which suggests a correlation of overall pupil size and task difficulty in the context of short-term memory task. In this case, the hence pupil dilated in reaction to the amount of information actively being processed at a certain time.

Based on these findings, Beatty (1982) propose an early approach to actually differentiate between dilations caused by emotions, light or distance, and those caused by effort related to task difficulty. Based on their analysis of time-locked pupillary responses to critical events in the context of an attention task, the authors introduce so called task-evoked pupillary responses (TEPRs), which are distinct from responses induced by alternative factors, such as light. Measures such as mean pupil dilation, peak dilation and latency count as TERPs and are suitable for the assessment of cognitive effort. This vital distinction is majorly enabled by the idea that, although dilations are generally enabled by circular muscles contracting the pupil and radial muscles dilating it, the underlying activation and inhibition patterns differ, depending on the cause of the dilation. One result of different inhibition - activation patterns is dilations related to cognitive effort being shorter and more abrupt movements of less than 0.5 mm in extent, as compared to, for instance light induced dilations. Light induced dilations

are longer and, most importantly, of a larger magnitude (see also, e.g., Beatty, 1982; Beatty and Lucero-Wagoner, 2000).

Although the magnitude is a valid feature for the distinction of dilations caused effort, the nature of optical reflexes can cause issues: Dilations caused by light are comparatively large in extend and can therefore easily mask the more subtle TEPRs (Beatty and Lucero-Wagoner, 2000). This still requires a full control of the luminance level of both, the test environment and the stimuli, which is often more difficult than one might expect. Even under well controlled, constant light conditions, the pupil may dilate irregularly to a certain extent, for instance in reaction to fixating darker or lighter objects or parts of a scene (Demberg and Sayeed, 2016), or can even be artificially induced by the mere suggestion of differences in brightness (e.g., by presenting pictures of the sun and the moon Binda et al., 2013). Further, although pupillometry is in general comparatively robust with respect to eye movement artifacts, pupil size can sometimes falsely vary with the gaze position, especially in experiments involving a desktop tracker. More specifically, the pupil may, for instance, appear contracted when it is not, due to parts of the pupil not being detected, depending on what a participant looks at in relation to the ankle of the tracker lens (Hayes and Petrov, 2016).

Marshall (2000) provide a powerful approach also targeting these issues: They introduce an index that not only majorly improves the clarity of pupillary results by no longer confounding dilations caused by light and effort in the output, but is, due to the way of filtering, also more robust with respect to gaze position. More specifically, the so called index of cognitive activity (ICA) separates the different activation patterns using a wavelet analysis on the pupil dilation record to extract distinctively short- and abrupt contractions attributable to effort. The increased resistance against changes in lighting and errors caused by gaze position is based on the fact that (pseudo-) dilations caused by those factors cannot pass the threshold set for abruptness and extend of the relevant dilations. Effort related dilations are almost computed in real time and the filtered index returns the exact number of times per second that an abrupt discontinuity in the pupil signal is detected (with a resolution of only 100 ms). These events are then referred to as ICA events (for a method description, see Marshall, 2002). For analysis, the number of ICA events is then counted within the specified time periods of interest. The higher cognitive effort is in a task, the more ICA events are counted in the respective time windows. Generally, high ICA values (i.e., more ICA events in a given time window) reflect higher cognitive effort, while low ICA values suggest comparatively less effort. The ICA was originally introduced as a measure of mental effort while participants interacted with visual screens and has been tested in the context of math tasks (Marshall, 2002). Since then it has shown to be responsive to cognitive effort in

a variety of different tasks, such as, for instance, a driving task featuring increased mental demands (Schwalm et al., 2008).

Since cognitive effort is not per se synonymous with (linguistic) processing effort, it is important to note that the index has also already been tested and proven reliable in the context of language processing tasks: Demberg and Sayeed (2016), for instance, showed that the ICA is sensitive to linguistic processing effort in both, reading and auditory tasks, even when combined with driving. Sekicki and Staudte (2017), as well as Ankener and Staudte (2018), further observed that the ICA is responsive to cognitive effort induced by processing of language in varying (visual) contexts (Ankener and Staudte, 2018; Sekicki and Staudte, 2017). We hence deployed the ICA as an ideally suitable measure of (linguistic) processing effort, and, due to its tolerance, in combination with traditional eye movements in the VWP. This combination allows for unique insights into on-line processing effort during situated processing. In order to obtain ICA events, binocular eye-tracking was used at 250 Hz on an Eye-Link II tracker. The calculation of rapid small dilations from the tracker data was conducted in the EyeWorks Workload Module software (Version 3.12). We analysed the raw ICA workload, that is, the workload module's output containing information of the exact timing of each detected ICA event.

1.1 Hypothesis

Now that we determined the appropriate paradigm and measure, an overview of what is hypothesized with respect to the way visual information affects predictability and processing effort of a target word is provided. The respective hypotheses are based on previous research results from the literature observing predictions, surprisal and processing difficulty in purely linguistic contexts, as well as on first observations made in visual contexts. In the case of linguistic contexts, it is known that: a) information from a single modality, presented sequentially is evaluated in a probabilistic way, b) new information gained from the linguistic input is permanently fed into the mental model and affects its statistics (see, e.g. Nieuwenhuis, 2011), c) linguistic information affects predictions about upcoming input in a probabilistic way, and, finally, e) the resulting predictability of upcoming words is directly associated with processing difficulty affecting the actual mental effort needed to process the respective word.

So far, it has not been observed whether and when the same holds for additional visual information. Specifically, it is not only unclear how detailed listeners evaluate visual context with respect to language, but also whether the evaluation of visual and linguistic information is interleaved at any time, or rather at specific points in time, and how the extracted information

affects the statistics of the (probabilistic) mental representation of what is currently being communicated as well as actual word processing.

A useful guidance for the examination of these fundamental questions and some very first hints for the statistical relevance of visual context can be found in the recent literature. Frank and Goodman (2012) in their rational actor model, for example, propose that a listener engaged in a referencing game – featuring a linguistic and visual context – can use Bayesian inference to assess a speaker's intended referent. Comprehenders hereby evaluate an object's contextual salience and use features of the visual context in order to interpret what the speaker was most likely referring to. This highly suggests a general statistical influence of visual information, at least in context in which the extracted information in beneficial to the comprehenders aim. It now necessary to observe when this evaluation happens during the input and how this statistical influence can be described.

On the one hand, it is possible that listeners perceive the probabilistic details of their visual environment in which language is comprehended, and that the extracted information from both modalities heavily influences their current beliefs and probabilistic expectations. On the other hand, however, although it seems very natural to think about visual and linguistic information being evaluated in a similar manner, a closer look reveals that this is far from clear: After all, visual information is inherently different in nature in the sense that it is presented simultaneously, instead of sequentially, and naturally presented in a second modality. More specifically, statistics derived from visually conveyed information could fundamentally differ from linguistically derived statistics, due to the fact that visual context makes information accessible very suddenly, leaving comprehenders with less time for statistical evaluation. This could result in an extraction of more coarse grained information from the visual context, based on linguistic information. Such coarse grained information could, for instance, be whether the visual context contains any relevant and helpful information at all. It could further be too effortful to feed visual information into the mental model at any time during the input. Specifically, we hypothesise that, just like each encountered word influences the comprehender's mental model of what is currently communicated, which in turn permanently and dynamically changes with the context, visual information might be evaluated probabilistically, and hence affect the statistics of this model, either at each word, or, alternatively, only at specific words in the input. Specific words could for example be words enabling a grounding of visual and linguistic information. That is, comprehenders may perceive and evaluate statistical details of what they see in their visual environment with respect to what they simultaneously hear, while this information could possibly not always affect processing effort of each word but only for words of specific relevance.

A detailed, probabilistic influence of multi-modal information on word expectancy and processing effort generally has the advantage of making a maximum amount of information available for expectations about upcoming input, which can in many cases facilitate the processing of upcoming words and enable a quicker reaction and adaptation. Such a fine-grained, probabilistic influence of multi-modal information on word processing is in support of the idea of a dynamic interaction of cognitive processes, rather than their modularity, and could further contribute to the precision of future statistical models of situated language comprehension. A disadvantage, on the other hand, can be that a detailed evaluation of multi-modal context is computationally expensive and might not be the most efficient way of processing in every situation. The same holds for the time points at which the evaluation of multi-modal context is interleaved. It can enable a quicker adaptation and reaction if information is fed into the mental model at any time during the input, but could also be too effortful, especially when visual information is static and therefore remains more constant and predictable than sequential linguistic input with a higher entropy.

Alternatively, as previously mentioned, it is possible that listeners perceive more coarse-grained information in their visual environment while language is being processed, hence not treating extra-linguistic information in a probabilistic way. It is also possible that visual information is evaluated in a way that allows for a decision whether it is congruent with what is heard. In this case, linguistic information could be processed in the usual, probabilistic way, while visual information does not have a fine-grained influence on the statistics of the comprehender's mental model. This may provide less information for expectations about upcoming input, but can also be quicker and more efficient in some situations, because it very likely requires less cognitive resources. It additionally needs to be considered that both, the fine- and the coarse-grained evaluation method, are in principle applicable by the comprehender, possibly depending on the respective comprehension situation she finds herself in. The human brain may adaptively react to external factors such as noise, increased task demand or time constraints. The adaptation may result in either a quicker, computationally less expensive method being preferred over a more informative, but possibly more expensive probabilistic one.

Based on these hypotheses, we examine how what people see influences their current interpretation of what they hear, as well as what they expect to hear next and, finally, how this affects the effort needed to process what they hear. We additionally quantify the observed effects using information theoretic measures such as surprisal and entropy reduction. Based on the aforementioned line of computationally influenced psycholinguistic work using information theoretic concepts to explain and quantify experimental results for statistical models of language, we expected that the most relevant concepts, namely surprisal (e.g.,

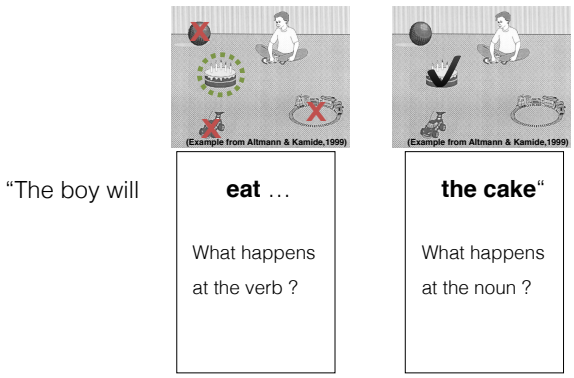


Figure 1.1 Graphic illustrating the research questions about how visual information can influence anticipation and how this could be linked to effects on actual linguistic processing effort for the critical words.

Demberg and Keller, 2008; Frank, 2009) and entropy (reduction) (Hale, 2003; Linzen and Jaeger, 2014) are suitable to describe visual context effects and hence suggest that both rational concepts can be valid predictors of processing effort related to multi-modally derived word predictability. Our approach can provide a next step in making situational aspects in language processing statistically assessable for probabilistic language models, which, for pragmatic reasons, so far only account for linguistic context information.

1.2 Overview

In this thesis, the (statistical) influence of visual context information is investigated using behavioural, psychological and electro-physiological measures. In addition to investigating effects of visual information, we aimed at describing and quantifying them using and extending information theoretic concepts.

While Chapter 2 provides an overview of this thesis' theoretical background, chapters 3 to 6 present a line of experiments, using different measures and paradigms in order to observe effects of linguistic and visual information on the statistics of the mental model and on target word expectations.

Specifically, Chapter 3 sets a baseline for effects of target word surprisal and predictability in the absence of any visual context, as reflected by well established, behavioural measures. This is especially important prior to deploying the comparatively novel ICA measure. The baseline consists of two reading studies on linguistic materials (of the type: “The man spills/orders soon the water/ice-cream/book”) that were conducted, while assessing processing effort via reading times (as a behavioural reference measure). Results reveal whether target word expectations affect different measures of processing effort in a setup that did

not include visual information. In this Chapter, we further review our measured result in the context of existing evidence for and against effects of predictive processing in the current literature. All experiments reported in this, as well as in the following Chapter 4 were designed, ran and evaluated in collaboration with M. Sekicki.

Chapter 4 subsequently reports on two experiments with a similar design and set of stimuli, presented in a different modality, namely auditory, processing effort is this time assessed via a pupillary measure. The first of those two listening experiments does not feature visual context, while the second one introduces additional visual information presented simultaneously along with the sentences. In the second setup, the potential of the additional pupillary measure of effort is deployed in combination with eye tracking measures assessing anticipation in the VWP. Results from the VWP setup are then compared to the data collected in the first experiment in this Chapter, in order to quantify the actual effect of the visual context information on word expectancy and processing effort (i.e. to see whether visual information affects the statistics of the mental model, as indicated by anticipatory eye movements in the VWP). Parts of Chapters 3 and 4 were published in Frontiers in Psychology (Ankener et al., 2018).

Chapter 5 reports on two studies testing not whether, but how and when visual context affects the statistics of the mental model in a neat manipulation merely the scenes displayed, while any linguistic variation is eliminated. Effort was assessed using two different measures, a pupillary and an EEG one. Results reveal the exact magnitude of visual information's influence on word expectancy and processing effort as related to surprisal in at least two different measures of effort. Deploying the EEG as a measure on the same experimental design further proves that our results can be replicated, or even extended in a (potentially more sensitive) electro-physiological measure. Possible implications of effects on the ERP component N400 are discussed in the context of the recent hypotheses in EEG literature.

Chapter 6 then outlines details about possible influences and limitations of predictive processing in multi-modal contexts. Experiments described in this chapter are setup to test a) whether context evaluation processes preceding visually informed word expectations can rightly be predicted by visually informed surprisal, or rather by a "one-many-nothing" evaluation, and b) whether this observation also holds in increased visual context complexity, as well as c) whether results hold in a context where anticipatory eye movements are prohibited, hereby observing the actual role of overt attention with respect to expectations. Conclusively, results are interpreted in the light of recent theories about the rational adaption of processing mechanisms and with respect to the role of the LC/NE system in effort related pupil contractions, as well as the correlation of the pupillary and ERP measures.

Additionally, Chapter 7 outlines the mathematical formalisation of our results by means of rational concepts from information theory. The chapter hence contains an extended surprisal formula that is able to account for the measured influence of visual context information on actual linguistic processing effort. Finally, a general discussion of the collected data from all experiments and their contribution to recent research is provided. Collective evidence for the major impact of visual information on word expectations and processing is presented, bearing implications for recent (language-centric) models of language processing as well as for the use of the VWP, before concluding the thesis.

Chapter 2

Background

The previous chapter outlined the main research questions of this thesis, the linking hypotheses, as well as the paradigms and measures used. The following chapter introduces the reader to important ongoing debates in the field, in the light of which our results will necessarily be interpreted later and subsequently outlines the theoretic background of the present work in more detail. That is, the general role and specificity of prediction in language processing will be discussed, and the importance of rational approaches in linguistic research, as well as the relation between (processing) effort and pupil dilations as culled by the ICA measure will be outlined in more detail.

2.1 On the general role of predictions in language processing

As already briefly mentioned, although many important details about effects of linguistic information on processing effort and predictions have been observed and well described in the literature, some open questions about the general nature of predictive processing remain. It is, for instance, still widely debated how necessary and specific (linguistically informed) predictions in language comprehension are.

How necessary are predictions for language processing? As of today, the predictive nature of language processing is generally accepted and established. Evidence exists in the form of results on the neural and the algorithmic level: A range of psycholinguistic work shows that results from various experimental setups can only be explained by language processing being predictive – at least to a minimal extent (see e.g., Kuperberg and Jaeger, 2016a). Early work in the field, for instance, by Morton (1964), could already show that

predictive processing (i.e. the predictability of a word) could explain variations in behavioural measures, in this case, visual duration thresholds for a word. Fischler and Bloom (1979) further concluded that incomplete sentence contexts only facilitated subsequent lexical decisions reflected by shorter reaction times, when the target word was very predictable from the previous truncated sentence context. Only a few years later, eye-tracking studies additionally showed that differences in reading (Ferreira and C. Clifton, 1986) and fixation times (Balota et al., 1985), could also be predicted by the target word's predictability in a given context. One of the most influential findings with respect to predictability effects on neurophysiological measures is Kutas and Hillyard (1984)'s discovery that the ERP component N400's amplitude is an inverse function of word expectancy (as assessed via cloze probabilities). This suggests that the N400 is a reliable index of semantic priming or activation.

On the algorithmic level, the concept of predictive processing was implemented via so called predictive feature frameworks (e.g., incremental belief updating), for instance, in recurrent connectionist networks. A very early example are simple recurrent networks (SRN), as proposed by Elman (1990). Those networks learned to predict which sequence of words would be likely to follow next, given the previous (linguistic) context. Interestingly, whenever predictions by these networks are disconfirmed, the formalisation of errors is already very close to the definition of Bayesian surprise (see also Kuperberg and Jaeger, 2016a). However, even though it is evident from a very large body of work, that language processing can be predictive, it is not automatically clear how necessary predictive processing actually is for language comprehension:

For instance, (Federmeier, 2007) observed prediction related processing cost, but also signs of possible age-related deterioration in predicting, hereby suggesting that prediction is not necessarily vital for the comprehension process. The authors hence propose that the human brain, although being generally able to use sentence context information for predictive processing, does not do so under any circumstances. In a later paper (Wlotko and Federmeier, 2015), the authors elaborate on this idea. That is, while they still call predictive processing a "core component of normal language comprehension", they also identify specific influential factors (in this case, timing) affecting the degree to which the human brain engages in predictive processing. The conclusion here is that prediction can be a component of language processing, but very likely is not fundamentally vital to it. There is even evidence for the possible influence of additional factors on prediction: Mishra et al. (2012) propose that observed differences between adult low and high literates in predictive language processing can be best explained by reading acquisition and practice, hereby suggesting that both influence the degree to which comprehenders can or do predict.

In the case of children, even further factors could be influential: Mani and Huettig (2012) suggests vocabulary size significantly influences prediction. Specifically, the authors found that children with larger production vocabularies predicted upcoming linguistic input, while children with less production skills did not. In fact, Huettig and Mani (2016) developed an especially critical view on the role of prediction. They even propose that signs of predictions are mostly observed in what they call "prediction encouraging" contexts, meaning that tasks and design used in many studies observing prediction provide a highly suggestive and overly optimised background for predictive processing, which may not be given in real world situations.

These examples show how various evidence shapes opinions with respect to the necessity of predictions along a scale. The two opposite ends of the scale are marked by the hypotheses that predicting is highly important and the best method of dealing with speed, imperfection and noise variance of mainly spoken input (as described via the ideal adapter framework proposed by Kleinschmidt and Jaeger, 2015) on the one hand, and the hypothesis that prediction is not vital for language processing at all (Huettig and Mani, 2016) on the other hand.

Since the the main focus of this thesis is the observation of how and when information from two modalities is evaluated and integrated in order to expect upcoming linguistic input, the presented research is not directly targeted at answering questions about the necessity of prediction in language comprehension. However, results can be interpreted with respect to this debate. Especially since the set-ups in the presented experiments feature an innovative way of presenting visual context while not being highly suggestive, since they are mostly ambiguous or even of high visual complexity. Those contexts hence reveal whether and how strongly participants predict (or, expect target words, for that matter) if the evaluation of contextual information is comparatively effortful and they are not confronted with easy and unambiguous target referent suggestions. Results can hence be reviewed with respect to whether they provide evidence for the importance or irrelevance of predictive processing, or, whether they support a moderate view in between the two ends of this scale.

How specific are predictions in language processing? The question how specific comprehenders predict in certain contexts is closely linked to the "necessity" debate. That is, the type and granularity of context information used to predict upcoming input can significantly influence how useful, reliable and necessary predictions are in a given context. The idea here is that overly specific, low level predictions, or predictions made based on unreliable context are very likely error prone beyond efficiency. This means that risky predictions are likely to result in a revision, which is suggested to be linked to an increased demand in resources (see

also Federmeier (2007), Van Petten and Luca (2012), or Xiang and Kuperberg (2015), who found that detection of failed event predictions caused a later posterior positivity and/or an anterior negativity effect in the ERP). High level contextual influences such as plausibility or coherence, on the other hand, allow for more reliable but mostly also less specific and possibly less beneficial predictions. Indeed, one argument brought up in the literature against predictive processing, especially on the lower level, is the idea that information gathered from the context is often insufficient for reliable predictions that are any more specific than rough syntactic categories. In this view, the efficiency of computing predictions is hence questionable in the first place, given the fact that processing resources are limited (Jackendoff, 2002). On the other hand, some researchers propose that predictions in linguistic context can even be done to a very specific extent (DeLong et al., 2005).

To a certain extent, evidence exists for both, high level semantic categories (Kamide et al., 2003), as well as more fine-grained information about semantic characteristics of the possible target referent can be predicted (Altmann and Kamide, 2007; Xiang and Kuperberg, 2015). Extremely low level prediction, such as on the phonological level is, however, controversial in the sentence processing literature. That is, although the basic hypothesis is not unusual: In speech recognition, for example, prediction on the phoneme level is not an unusual concept (see, e.g. Dahan and Magnuson, 2006; Kuperberg and Jaeger, 2016a). The controversy in sentence processing, however, continues to exist and has only recently been fuelled by the finding that results – the so far best and only evidence for prediction of phonological forms of exact words – found by DeLong et al. (2005) were not reliably replicable. The authors used phonological regularities of English indefinite articles ('an' vs. 'a' preceding nouns) in an EEG and found a graded modulation of the ERP component N400, elicited by the articles and the nouns in their stimuli, in reaction to the probability that those target nouns were the continuation of the previous sentence fragment. Those results could not be confirmed in a direct replication attempt by nine different labs Nieuwland et al. (2017).

As a result, it was proposed that very specific predictions play, if at all, a very minor role in language comprehension. Van Petten and Luca (2012) even suggest that prediction is generally not a good description of what comprehenders actually do. Based on the finding that contextual benefits on processing are measurable, while little evidence for direct costs of failed predictions (i.e., failed predictions on the word level) was found so far, the authors suggest that instead, collective results from ERP and behavioural data strongly support general expectations about upcoming semantic content. Most recent research mainly proposes a more flexible interpretation of the granularity of predictions: As opposed to the idea that comprehenders make fine-grained low levels predictions regardless of any contextual factors, Federmeier (2007), in the context of their dynamic framework, suggest that the human brain

generally uses predictive processing to deal with the speed and complexity of input, while granularity and likelihood with which predictions are made may depend on factors such as costs, benefits, availability of resources (as e.g. influenced by the experimental task) and ageing, which the brain compensates for by also implementing bottom-up, integrative processing strategies. Kuperberg and Jaeger (2016a) further propose that prediction can happen on various levels, while the actual level can depend on the expected utility of the prediction in the current context, including the comprehenders' intrinsic aims as well as the reliability of their knowledge in combination with the recent input. Very recently, Delaney-Busch et al. (2017) found that participants in their EEG studies used global factors such as the predictive validity of the general experimental environment in order to rationally adapt the strength of their semantic predictions to the statistical structure of this environment. In other words, if global factors affect the strength (and possible the granularity) of predictions, they can be made on different levels, but the brain adapts them to the broader context. These more flexible concepts are promising with respect to their psychological validity, not least given the fact that their basic assumptions are coherent with the general, adaptive nature of the human brain.

Experiments in this thesis feature various contexts, some of which provide additional, simultaneously presented visual information conveying more information, possibly making expectations more reliable and beneficial, while others only provide a mildly restrictive linguistic context, making specific predictions less efficient and more likely to be dis-confirmed. By keeping the linguistic stimuli as similar as possible across all experiments, results from differently informative and reliable contexts can be compared with respect to the adaption of strength and granularity with which comprehenders predict or expect upcoming linguistic input. Results will then be evaluated in the light of the previously named recent findings in the literature.

2.2 Anticipation and possible effects on probabilistic prediction

Besides various debates, the probabilistic nature of predictions derived from *linguistic* context is nowadays widely accepted: Early concepts thought of prediction as either being absent or very specific, settling irreversibly on one interpretation (see, e.g. Ferreira and C. Clifton, 1986). Modern concepts acknowledge that contextual information alters the current state of the comprehender's mental model at any time, hereby probabilistically predicting upcoming input, which eventually results in facilitated processing. Since it has

been found that (linguistic) context information can affect processing effort in a graded fashion – depending on the strength of the established bias for or against an interpretation (see, e.g. Hare et al., 2007) – a range of studies could elaborate on the gradedness by providing evidence for *probabilistic* prediction. Evidence was collected by correlating a word's *surprisal* in its linguistic context with measures of word processing effort, such as reading or reaction times (see, e.g. Frank et al., 2015; Hale, 2001; Smith and Levy, 2008). This way, it was perfectly shown how contextual information conveyed by a single modality affects the statistics of a comprehender's mental model of what is communicated in a probabilistic way, resulting in a respective processing facilitation for more or less expected upcoming input.

As briefly mentioned, it is so far unclear whether and how contextual information from a second modality is being integrated and how it affects statistics of the mental model. Especially in the case of visual information, which is fundamentally different in its nature because it allows for an immediate assessment of all information at once, as opposed to the sequentially conveyed information in linguistic context.

Still, when observing the effect of multi-modal information on the mental representation of statistical characteristics of the context, one does not have to start naively from scratch: The basic mechanisms involved in the evaluation of linguistic context are known, as well as the fact that participants can use visual information to anticipate potential target words, as reflected by anticipatory eye movements toward target options prior to hearing the actual word (Altmann and Kamide, 1999). This shows how incremental eye-tracking measures can not only be very helpful in combination with written stimuli, where they deliver continuous data (as opposed to button presses or reaction times) and reveal a words predictability depending on how long it is fixated (Balota et al., 1985; Demberg et al., 2012). They also further reveal important processing and integration steps, and even partially activated concepts that might be revised in the course of incremental interpretation when visual information is given. This makes their implementation specifically interesting in the context of the visual world paradigm (VWP), where aspects of situated language processing – as given in everyday situations in the real world – can be introduced, showing that extra-linguistic information may be considered by a comprehender.

The link between eye movements and linguistic information was first drawn by Cooper (1974), who observed that listeners move their eyes to those objects presented in a visual scene, that match the referring expressions in the simultaneously presented sentence.

Some years later, Allopenna et al. (1998) tested a paradigm in which participants manipulate real world objects or their pictures on a screen according to spoken instructions while their eyes were tracked. They found that eye movements to objects that could be referred to

provided a very sensitive measure of language processing since the movements were closely time-locked to the expressions that elicited them. Finally, Altmann and Kamide (1999) could specifically prove that verbal constraint information can be used to immediately narrow down the domain of possible target referents displayed in order to anticipate the target object prior to hearing it. If this is linked to actual differences in processing effort, entropy reduction is very likely a good predictor. Kamide et al. (2003) extent this observation by showing that it does not only hold for semantic, but also syntactic information being used to anticipate target words based on visually presented referents. These verb-driven anticipatory eye-movements not only provide rare evidence for expectations about target words, prior to the actual word (i.e., as opposed to, for instance, N400 effects that are typically found on the target itself), but further hint at when and how multi-modal information is evaluated in an interleaved way.

It is now necessary and important to observe whether and how this affects the comprehender's mental model and, possibly processing effort for the critical words. In the experiments presented in this thesis, a combination of incremental eye-tracking measures, the VWP and an additional pupillary measure of on-line processing effort were deployed in order to do so.

More specifically, based on the previous findings by Altmann and Kamide (1999), this novel combination allowed us to observe whether, when and how the interleaved evaluation of linguistic and visual information, as indicated by the anticipatory patterns in the eye movements, affects statistics of the mental model as well as actual processing effort related to the, now multi-modal, predictability of the linguistic input.

Finally, we outline existing, considerable criticism concerning the observation of expectations in the VWP. Huettig and Mani (2016), for instance, suggest that presenting participants with a few selected images of thematically highly appropriate, as well as obviously inappropriate referents prior to the auditory target words is a setup that encourages prediction beyond what could be expected in real world situations where targets are not clearly displayed in front of comprehenders. A relevant contra argument is that prediction of upcoming target words in such setups also happens when no ideal target is presented visually, as shown by Rommers et al. (2013), who found that participants would fixate objects that shared visual features such a shape with the actual preferred target that was not displayed. Specifically in our setups featuring ambiguous visual context, it is similarly questionable whether the provided context is highly suggestive since it takes comprehenders effort to evaluate the different possibilities, rather than being confronted with a clear option. Although visual displays in an experimental setup are usually less complex, compared to real-world situations (for a discussion, see Altmann and Mirkovic, 2009), we do not consider our designs to be overly suggestive and propose that comprehenders could be confronted with similar situations when having to quickly chose from various pre-activated referents in the real world (e.g., in the context of

assembly tasks, cooking). We hence consider our setup appropriate for the observation of possible effects of visual context evaluation, as reflected by anticipatory patterns, on the statistics of mental models on comprehension, as well as on linguistic processing effort.

2.3 Information Theory: The importance of rational approaches in psycholinguistics

In order to account for possible results with regard to how and when visual information affects the statistics of the comprehender's mental model, we chose rational concepts from information theory, due to their evident power of describing results from purely linguistic contexts. This chapter elaborates on the background of the most important rational concepts for psycholinguistics, namely surprisal and entropy (reduction) and their correlation with linguistic processing data in more detail. For many years, work within the field of functional linguistics aimed to achieve a deeper understanding of language systems (for both, processing and production) via what was interpreted as the underlying general cognitive principles, described by the ideas of *complexity* and *utility* (for an overview, see Jaeger and Tily, 2011). The idea behind those models was to understand principles of grammar through the way language is used (Hawkins, 2004). In other words, cognitive principles of processing and production complexity were suggested to determine grammatical forms of language use. For instance, it was hypothesised that the distribution of grammatical complexity can be explained by their processing complexity. However, those concepts were often criticised as being rather intuitive, while not finding much empirical support (Jaeger and Tily, 2011).

The fact that such frameworks also lack a mathematical notion of information and processing difficulty makes it even harder to collect empirical evidence from experimental data. Many modern approaches in the field hence offer formal notions and develop better approaches with higher validity. In those cases, strong empirical evidence for an actual correlation between processing difficulty, as assessed by different measures of effort, and the *predictability* of linguistic units (on different levels of granularity) in their wider context could be gathered quickly (e.g., Bell et al., 2009; Frank, 2013b; Kutas and Federmeier, 2011; Rayner et al., 2004a; Rayner and Well, 1996; Smith and Levy, 2013; Van Petten and Luca, 2012).

Additional evidence comes from results showing that comprehenders can actively expect (e.g., DeLong et al., 2005) or, in the case of the VWP, anticipate (Kamide et al., 2003), what is a likely continuation of a recent sentence, hereby explaining and supporting the ease of processing and integrating linguistic input depending on how well it matches with the

listener's expectations made in advance. The intention to formalise this correlation resulted in the hypothesis that processing difficulty in language comprehension can be suitably expressed by quantifying the *information* conveyed by a word via the *predictability* of a linguistic unit in its context (Shannon, 1949). This is exactly the idea behind the formal notion of surprisal, as proposed by Hale (2001) and later refined by (Smith and Levy, 2008).

More specifically, surprisal in information theory quantifies the amount of information conveyed by a *unit* in terms of *bits*, based on the *predictability* of that (linguistic) unit in its context. Shannon (1949) defined the predictability of a (linguistic) $unit_i$ as:

Pred $(unit_i) = \mathrm{P}(unit_i|\mathrm{Context})$,

Based on the formalisation of predictability, the actual amount of information conveyed by a unit can be formalised as:

Surprisal $(unit_i) = \log \frac{1}{P(unit_i|Context)}$.

Surprisal hence formalises the general idea that less probable linguistic units convey more information, while the information content of the unit is majorly determined by the context in which it occurs. This way, the information content of linguistic units can be objectively quantified, which, in turn, allows researchers to reason about rational cognitive processes underlying language comprehension. Especially surprisal – as a quantification of information content – has been proven be of greater psychological validity and ability to account for experimental data. That is, compared to more traditional models of sentence processing such as, for instance, symbolic logic models. Such models usually featured much simpler assumptions such as that meaning can be accounted for solely by compositionality, or that low-level features such as word frequency alone defined predictability (as proposed by Zipf's law, for a critical review see Piantadosi, 2014), or simply that syntactic expressions correlated with semantic form (Blackburn and Bos, 2005). As opposed to most older concepts, surprisal values can explain and formalise the finding that less expected words take longer to read (Demberg and Keller, 2008; Smith and Levy, 2013), they can be correlated with the ERP component N400 (Frank, 2013b) – hereby allowing for the conclusion that the N400 can indeed index information content –, and can even account for the use of shorter utterance forms for more predictable words in language production, suggesting the rational adaption of human language use to this principle.

Although surprisal generally has shown to be a valid predictor of processing effort, it is important to note that, while within (psycho-)linguistic literature, surprisal is usually defined via the formula introduced by (Hale, 2001), the term unit can refer to different levels of (linguistic) granularity, such as a specific word, a phoneme, a sentence, or a part of speech. Hence, different interpretations of surprisal occur: Unlexicalised surprisal for example, has been shown to be a good predictor for measures such as word reading times (e.g., Demberg

and Keller, 2008; Frank, 2009). This form of surprisal takes structural probabilities into account, meaning Surprisal is calculated via the same formula, but considering the probability of a word‘ s part-of-speech to come up, given all previous parts of-speech in the sentence. Lexicalised surprisal, on the other hand, considers the exact words of a string to calculate structural and lexical probabilities of a specific target word. It is hence necessary to provide a clear definition of units and surprisal in our context as well: Based on the main regularities of information theoretic formalisations (i.e., low predictable units convey more information which is greatly determined and influenced by the context), as well as on existing evidence for the influence of simultaneously presented visual information on target word expectations (e.g., Altmann and Kamide, 1999), we propose that relevant information in situated communication is distributed across and gathered from different modalities.

It is hence reasonable to hypothesise that surprisal and possibly other information theoretic notions can be used to account for the effect of multi-modal information. Results from our studies will hence be quantified using information theoretic notions and made assessable for implication in statistical models of language processing. Due to the studies' design largely featuring homogenous sentence structures, it is in our case appropriate to define surprisal with respect to the lexical, rather than to the structural properties of an upcoming linguistic unit. Surprisal is therefore calculated as the negative log probability of a target unit (which in the present case means on the lexical level), given its visually enriched context.
Thus, surprisal at word w_i is:

$S(w_i) = -\ log_2\ p(w_i|w_1, w_2 ... w_{i-1})$ [1].

Further, entropy (reduction) (Shannon, 1949), has shown to be capable of predicting processing effort in psycholinguistic experiments: It can, for instance, perfectly account for phenomena such as the well-known garden-path effect, potentially even better than surprisal (Hale, 2003) in the sense that the disambiguating word causes confusion as conveys enough information to reduce uncertainty to basically zero. Much like surprisal, entropy is clearly defined by a formula, while being applicable on different levels of granularity. That is, although it describes the uncertainty about the outcome of a random variable (Frank et al., 2015), or, in other words, the next step in language processing, this next step can, for instance, be a specific word, an entire parse, or a structural part of a sentence. When Hale (2003) introduced the entropy reduction hypothesis, he formalised the ambiguity resolution work via the information conveyed by a word (in PCFG-generated sentences) and quantified entropy

[1] Since this type of surprisal is naturally highly affected by frequency effects, as well as possibly to additional factors on the visual level, such as salience, these factors had to be controlled for in the present work.

as the uncertainty (i.e., the missing information) over parses of a sentence. The entropy at a given word in a sentence is hence defined as the remaining uncertainty about possible parses after the information of the word has been integrated.

Entropy at word w_i is therefore

$$H(w_i) = - \sum_{parse \in SetOfPotentialParses} P(parse) log^2 P(parse)$$

As a consequence, entropy *reduction* is caused by the information conveyed by word w_i and is suggested to relate to the processing effort for that respective word.

Linzen and Jaeger (2014), on the other hand, who also suggest that entropy (reduction) effects can explain their linguistic data, employed slightly different realisations of the concept when observing different ways in which uncertainty can affect processing difficulty (as assessed by reading times). The authors differentiate between single-step entropy (meaning the uncertainty over the next step in a syntactic derivation) and total entropy (defined as the uncertainty over entire parses of the sentence as derived by a probabilistic context-free grammar (PCFG)). They found that, in addition to surprisal, only total entropy, but not single step entropy, is a reliable and significant predictors of reading times. In other words, the more information a word submits, the further the processing of that word reduces uncertainty about the entire sentence structure, and the longer it takes to read that word. Although entropy is not always easily distinguishable from surprisal, it is in the respective case, which leads the authors to suggest that models of sentence processing should consider both, surprisal and entropy. In the present work, the concept of entropy reduction is hence considered as a predictor of processing effort related to a possible reduction of referential uncertainty. That is, we test whether entropy reduction can predict processing effort resulting from the interleaved evaluation of visual and linguistic information, which could possibly lead comprehenders to concentrate on objects matching what they hear (as indicated by anticipatory eye movements), while ignoring distractors, hereby reducing uncertainty about target referents.

2.4 The role of the LC/NE System in effort related pupil dilations

A significant component of our setup is the pupillary index ICA, which was briefly introduced in the introduction. This measure enables a continuous on-line assessment of (linguistic)

processing effort, by filtering short and abrupt pupil contractions and returning a list of those, so called, ICA events along with precise time stamps. When aligned to the experimental audio files, the number of events within the critical time windows (in our case the verb and the noun) can be interpreted as increased (more ICA events) or decreased (less ICA events) effort attributable to the processing of the respective (more or less informative/expected) word.

Pupillometry has been around for many decades, functioning as a handy index of effort, and has been increasingly deployed lately, not least due to recent technical advancements in eye-trackers. Ever since (Beatty, 1982) observed that the pupil diameter can be influenced not only by cognitive effort but also by a range of other factors such as the light- or the near-reflex, different approaches to filter the causes of dilations have been made. The ICA wavelet analysis introduced by Marshall (2000) is one of the most recent ones. The index resulting from this wavelet analysis has been shown to overcome some traditional problems in pupillometry: That is, it is relatively robust with respect to gaze position (in relation to the tracker lens) and fluctuations in lighting (Marshall, 2000). While the measure can deal well with those well known problems in pupillometry, which makes it a very attractive measure, especially for the VWP, it is still important to take care of some factors. First and foremost, although it culls dilations related to mental effort, it is not clear what effort in those cases means. In other words, it needs to be carefully investigated with respect to what we can interpret effort indexed by the ICA. Since effort itself is not a fixed term, it can relate to many different factors causing it, not all of them being connected with linguistic processing. So far, the ICA has been used in the context of various tasks:

While Marshall (2002) assessed effort caused by the participant's interaction with visual screens as well as related to math tasks, Schwalm et al. (2008), for instance used the ICA to assess mental effort related to driving and strategic shifts of attention by the driver in a simulated driving tasks with varying additional demands such as changing lanes. Demberg and Sayeed (2016) were the first to approach the question whether and how the ICA was also sensitive with respect to effort induced by language processing. The authors proved that the index can be an appropriate measure of linguistically induced processing effort in the context of three reading, three dual-task (driving combined with language comprehension) and one VWP experiment. Further support for the ICA's sensitivity to linguistic processing effort comes from recent studies using the measure in the context of (VWP) language comprehension tasks in varying (visual) contexts (Ankener and Staudte, 2018; Sekicki and Staudte, 2017). Especially due to the measure's proven sensitivity with respect to language processing but also effort induced by other factors such as increased memory load or attention, it is important to further observe when the ICA indexes what effort and how

this can be explained by the neural mechanisms underlying the effort induced dilation of the pupil. Indeed, although pupillary responses to task difficulty and processing effort have been investigated for a while (e.g., Beatty, 1982; Hess and Polt, 1964), it is not entirely clear *why* the pupil dilates in reaction to increased mental workload (see also Demberg and Sayeed, 2016). A recent body of work observes the question of how (distinctively short and abrupt) contractions of the dilator pupillae (iris dilator) muscle are connected with mental (processing) effort in order to explain how it is possible for the eye to reflect, different kinds of mental effort. Demberg and Sayeed (2016) proposed a connection between pupil dilation as culled by the ICA and activity in a brain stem nucleus called the locus coeruleus (LC), as proposed by Aston-Jones and Cohen (2005), who, after conducting a series of experiments on primates, found a strong correlation between the animals' pupil sizes and increased activity in the small, bilateral brain stem region located under the fourth ventricle. The LC region is one of a number of small neuromodulatory brainstem nuclei and supplies, or, in medical terms, innervates the forebrain, the brainstem and brain regions associated with higher cognitive and affective processes (i.e., it widely projects into the hippocampus and the neocortex) with the neuromodulatory neurotransmitter norepinephrine (NE, sometimes also called noradrenaline) (Berridge and Waterhouse, 2003). In early theories, the locus coeruleus-norepinephrine (LC-NE) system was thought to simply be involved in arousal. More recent theories, however, connect it to a more complex modulation of behaviour, for instance with respect to alert waking versus sleep, as well as the alerting effects of salient stimuli and events and the modulation of processes underlying the acquisition and processing of salient information (Aston-Jones and Cohen, 2005). More specifically, the neurons clustered in the LC area emit NE, which facilitates the synchronization of neurons, and hence, the functional integration of different brain regions (Aston-Jones and Cohen, 2005). In general, neuromodulators – such as NE – and activity in neuromodulatory systems of the brain – such as le LC – are nowadays thought to play a major role in the regulation and mediation of psychological states and behavioural processes such as attention and motivation, as well as of sensory processes and memory retrieval, hereby influencing an organisms behaviour in a given context (Sara, 2009). Malfunctions of neuromodulatory systems in general are even increasingly connected to disordered cognition, as well as to various psychiatric and neuro-degenerative issues, such as post traumatic stress disorder (Aston-Jones and Cohen, 2005; Sara, 2009). Probably the most relevant of the various influences of the LC noradrenaline system for the present case is the observation that LC/NE neurons show increased activity in response to unexpected, that is, surprising changes (Sara and Segal, 1991) in order to modulate the alerting effect of this information (Aston-Jones and Cohen, 2005), and to facilitate attentional and cognitive shifts in order to (ideally) adaptat to changes in the given environmental context (Sara, 2009). Aston-

Jones and Cohen (2005) define this as a facilitation of "behavioural responses to the outcome of task-specific decision processes", which makes it reasonable – in the present context – to hypothesis that the LC/NE system might be involved in the modulation of behavioural responses of subjects in our experimental setups. More specifically, it could be relevant in the modulation of attention shifts, as reflected by eye-movements when mapping linguistic to visual information and for the reaction to more or less surprising input in the case of the more or less predicted target words, which could be reflected by pupil contractions. Indeed, the connection between LC/NE activity and pupillary responses has not only been established in the case of primates: Among others, Gilzenrat et al. (2010) clearly suggest that a correlation between pupil dilations and diameter and the different modes of LC/NE activity is also observable in humans. The direct causal connection here is that NE, as a neurotransmitter, affects target neurons via receptors. Hereby, the kind of receptor determines and defines the specific effect of NE. That is, activation of α1-adrenergic receptors is associated with excitation, while α2-adrenergic receptor activity is correlated with inhibition (Berridge and Waterhouse, 2003). The relevant receptors for NE are also located in cells of the ocular tissue (Woldemussie et al., 2007), and, specifically in the sphincter pupillae, contracting the pupil (Alphen, 1976). Although, so far, no specific direct connection of activity in the brain's LC area and language processing has been observed, it is still reasonable to believe LC/NE activity can be relevant here as well, given its role in behaviour modulation and resource management. In other words, according to the recent state of research concerning language and information processing, there is no specific reason to believe that the mechanisms involved in language processing are fundamentally different from those involved in other the processing of and adaption to other, non-linguistic information. This assumption is further in line with recent frameworks including a dynamic embodied view of mental activity, which suggests that language and other, non-language related cognition should be treated as dynamically interacting with, rather than being independent of perception and action (Spivey et al., 2009).

Indeed, some studies suggest a correlation between stimuli and language related EEG components such as the P3 (assessed by picking up currents caused by postsynaptic potentials related to the release of neurotransmitters), and increased LC/NE activity Nieuwenhuis et al. (2005). Additionally, Demberg and Sayeed (2016) specifically propose that activity in the LC area, which is very wide-spread in the brain, very likely *can* also project into brain regions associated with language and language processing, and hence, that NE can affect rapid pupil dilations as picked up by the ICA index in tasks involving language comprehension.

Data collected in the presented work using the ICA index, culling pupil contractions related to mental effort, as well as other measures of effort, will hence be evaluated and

interpreted based on the background of the hypothesised link between increased LC/NE activity and (effort related) abrupt pupil contractions in reaction to visual and linguistic stimuli and, most importantly, their expectancy in a given context.

Chapter 3

Pre-test: Processing effort in behavioural measures

3.1 Experiment 1: Linguistic context - Reading

In the course of this thesis, we use different measures of effort to assess the effects of visual information. It is hence important to initially set a baseline for expectation-related processing effort in the absence of visual information, that is, in a purely linguistic context. This way, we are later able to assure that measured effects are indeed attributable to the presence of visual context. Additionally, we set a baseline using an already well established behavioural measure, namely reading times, in order to have a reference before introducing further measures of effort. It is, for example, possible that the different measures show different sensitivities with respect to effort induced by various factors (possibly also non-language related effort). In this case, we can possibly draw interesting comparisons.

The following chapter hence presents an initial baseline reading study using the same set of stimuli that is later used in the following experiments in order to keep results maximally comparable. In the later studies, visual context was simply added to the linguistic stimuli. Here, possible effects of specific (linguistic) predictions or more general expectations about upcoming words on processing effort were assessed via reading times. Results set a baseline for effects of target word predictability and processing effort in the absence of any visual information as reflected by the well established behavioural measure. We further give a detailed explanation on how stimuli sentences were designed and validated prior to using them in the actual experiments.

3.2 Linguistic materials: Design and validation

Linguistic Stimuli Design The set of linguistic stimuli used in the experiments presented in this thesis was held similar to the maximum possible degree (see Table 3.1 for an example) in order to keep results comparable. All experimental items were German independent main clauses, uniform in their syntactic structure (NP-V-ADV-NP) and designed so that sentences' subject NPs did not contain any helpful information with respect to the predictability of the verb or the object noun phrases.

Experiment 1, **Experiment 2**, and **Experiment 3** featured two different verb categories, that is, highly constraining[2] (*spill*), and unconstraining (*order*) verbs. These differences in verb constraints were used to manipulate the target noun's predictability and surprisal.

The adverb following the verb in all stimuli (*The man spills **soon** the water*) served as a padding region in order to give time for linguistic predictions or, in the later experiments, for visually informed expectations concerning the object noun.

In order to observe possible differences in predictability and surprisal of the target nouns in the first three studies, both verb categories were paired with two different object nouns (see Table 3.1, (1) and (2)), one of which is more plausible in the verb context (see Table 3.1, column 3). If the differences in verb constraint information – either alone as in **Experiment 1** and **Experiment 2**, or in combination with visual context as in **Experiment 3** – were enough to have more or less specific expectations about the target noun, surprisal and processing effort for the less plausible nouns were expected to increase.

All auditorily presented stimuli were recorded using *Audacity* (Version 2.0.6). Factors with potentially substantial influence on a word's processing effort and (linguistic) surprisal, such as word length and frequencies (Schilling et al., 1998) were controlled for. In particular, noun frequencies were derived a priori from the word lists DeReWo[3] of the German research corpus (DeReKo) and held approximately constant within an item. Differences in word length were integrated in the analyses (by either including the length factor as predictor in the model, or using mid-word time windows, see *"Analysis"* subsections for details). All nouns were concrete and inanimate, in order to make them easily depictable and, especially in the ERP study, to prevent differences in the N400 amplitude caused by abstract words (West and Holcomb, 2000). Fillers were always plausible sentences with differing length and of differing syntactic structure in order to prevent fatigue effects. Half of the fillers were followed by yes/no comprehension questions (such as "Did the man spill the lemonade?") to keep participants focused.

[2] for validation of the categories, see 3.2

[3] (DeReKo corpus size $>$ 28 billion words. Source: Korpusbasierte Wortgrundformenliste DeReWo,2012, v-ww-bll-320000g-2012-12-31-1.0, http://www.ids-mannheim.de/derewo)

For the sake of observing the effect manipulations on merely the visual context level on word processing, the linguistic stimuli were altered for studies **Experiment 4**, **Experiment 5**, **Experiment 6**, and **Experiment 7**. That is, only highly contraining verbs were used, paired with only the more plausible noun continuations in order to exclude *any* linguistic variation and hence its effect on target word predictability and processing effort.

Linguistic Stimuli Validation Both linguistic manipulations used in the first three studies, i.e. strength of verb constraint and the nouns' plausibility in their contexts were validated offline prior to using the stimuli in the actual online experiments.

Verb constraint The more constraining a verb is, the fewer plausible continuations it allows. A classical sentence completion task for cloze probability hence assessed to what extent the verbal constraint (i.e. highly constraining (*spill*) or unconstraining (*order*)) increased the predictability of the target noun (3.1, column 4)[3]. 17 German native speakers participated voluntarily in this online questionnaire. All items were truncated prior to the target noun and presented in one list, containing 50 % fillers, shown in randomised order. Participants were asked to spontaneously complete the sentences with a noun best fitting the sentence context. Unique participation of each webform user was controlled for. Results from the Cloze task showed differences between nouns in the highly constraining verb context only (see Table 3.1, column 4, 1a) vs. 1b)). Cloze probabilities ranged from 4% to 55% for plausible nouns in high constraining verb contexts (*spill water*) and from 0% and 4% for the less plausible nouns in the same contexts (*spill ice cream*). Unconstraining verbs produced Cloze probabilities $<.01$ for all critical nouns. In sum, higher constraint led to higher cloze probabilities of the two subsequent nouns (see Table 3.1, column 4).

Verb-noun plausibility A plausibility rating on a seven-point Likert scale assessed how plausible participants would rate a target noun to be in its sentence context (3.1, column 3). 14 German native speakers participated voluntarily in the online questionnaire. Participants read the stimuli sentences in a webform and were asked to spontaneously judge the plausibility of each sentence combination, resulting from the Verb - Object manipulation, ranging from 1 (very plausible) to 7 (not plausible at all). Items were presented in randomised lists, containing 50 % filler sentences. Each participant had only a single access to the webform. Results plotted in Fig.3.1 show that plausibility – or thematic fit – of the nouns used in the stimuli differed in the context of high constraining verbs (see Table 3.1, column 3, 1a) vs.

[3]Note that the Cloze task was only used to assess the strength of verb constraints. All object nouns were filtered from DeReWo, rather than being picked according to the Cloze results.

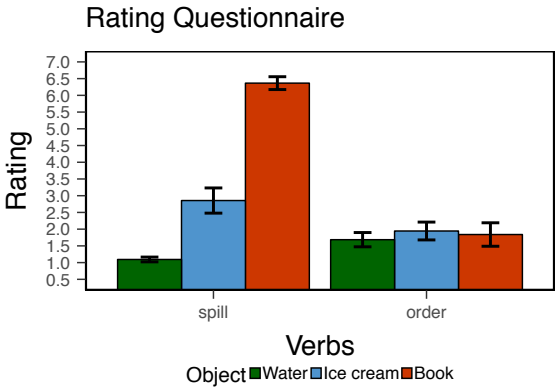


Figure 3.1 **Pre-test results** (verb-noun plausibility). Participants rated how plausible a noun was in the context of the previous verb, using a 7-point Likert scale where 7 is not plausible at all and 1 is perfectly plausible. Error bars reflect 95% confidence intervals (CI).

Table 3.1 **Sample items** and corresponding pretest results for the two nouns in each verb condition: highly constraining (1) and unconstraining (2).

Item	noun	plausibility *M (SD)*	cloze% *M (SD)*
(1) *The man spills soon the*	a) *water*	1.12 (0.68)	13.67 (18.06) [1]
	b) *ice cream*	2.76 (2.17)	0.16 (0.54)
(2) *The man orders soon the*	a) *water*	1.65 (1.50)	$<$ 0.01 (0.0)
	b) *ice cream*	1.90 (1.80)	$<$ 0.01 (0.0)

1b)), while both nouns were equally plausible in the unconstraining verb context (see Table 3.1, column 3, 2a) vs. 2b)), reflecting the intended manipulation.

3.3 Method

The first on-line experiment assessed effects of *purely linguistic* context on processing effort of the critical target nouns, in terms of reading times.

An "implausible" condition as well as a spill-over region following the verb's argument in all conditions were added only for this experiment (e.g. *the man soon spills the book at the restaurant*). The implausible condition served as a reference measure for effects of

[1] Note that the relevant nouns were pre-selected from the DeReKo corpus (in order to control for frequency) and not collected from the cloze task answers. Hence the relatively high SD for these particular nouns in the cloze task.

implausibility on reading times, and as sanity check for the design. This condition was expected to clearly elicit longer reading times for the – in the context of the verb *spill* – highly unpredicted and hence highly surprising noun (*the book*).

The added adverbial phrases served as post target spill-over regions for the time-dependent measure, which requires longer time windows (as compared to ERPs or the ICA).

The verb-noun manipulation resulted in a 2x3 design in which constraining (*spill*) and unconstraining (*order*) verbs were paired with objects that were more plausible in the constraining verb context (*water*, i.e., most predictable and least surprising as compared to the other objects used), less plausible (*ice cream*, i.e., mid predictable and surprising) and implausible (*book*, i.e., unpredictable and most surprising). As reflected by the plausibility rating ran prior to the experiment, all three objects were equally plausible in the unconstraining verb context while the target noun's plausibility differed only in the constraining verb context.

36 experimental and 36 filler items were distributed across six lists, using the Latin square design in such a way that each participant would see each item in only one condition. 24 native speakers of German (students of Saarland University) gave informed consent before participating in this study for monetary reimbursement. Their age ranged from 18 to 32 years ($M = 22.71$).

Sentences were presented as a whole, in the centre of the screen (Times New Roman, 20 pt), with a drift correct fixation point, shown at the top left corner in order to avoid initial fixations at the sentence. Participants were instructed to read for comprehension, at their own pace.

Predictions In the recent psycholinguistic literature, it is well established that in human sentence processing, the probability of a word to appear in its previous (linguistic) context largely affects the time it takes to read this word. This correlation has further often been quantified using information theoretic surprisal. Van Berkum et al. (2005), for example, measured significantly increased reading times after expectation-inconsistent adjectives in self-paced reading. Smith and Levy (2013) further even suggest that the quantitative form of the relationship between a word's reading time and its predictability, that is, the strength of a comprehender's prediction for a word, is indeed logarithmic. Only recently, Goodkind and Bicknell (2018) demonstrated that the degree of predictive power of word surprisal, as derived from different language models for reading times can even be a reliable, linear function of the respective language model's quality.

Based on this corpus of studies, longer reading times reflecting increased processing effort were expected on or after the most surprising and least predictable *implausible* target nouns, *only* when appearing in a context allowing for predictions, namely when following

the constraining verbs. If, however, the verbal constraint alone was not enough to elicit (lexical) predictions about the target nouns, no differences between the object conditions in the constraining verb context were expected. According to many studies suggesting a strong context dependence of predictions, no differences in processing effort were expected on the verb, although the verbal constraint could cause participants to make more detailed assumptions about the target noun (e.g., something "spillable"). However, compared to studies using bigger linguistic context to strengthen expectations about target words prior to them, the stimuli used in this experiment are not embedded in a wider context. The only information driving possible expectations or even specific predictions about the target noun hence come from the verb's constraint.

Analysis If not stated differently statistical analyses of the data collected in this and all following experiments were conducted using the R statistical programming environment (RCoreTeam, 2013) and the *lme4* package (Bates et al., 2015). The dependent measures is all experiments were always analysed within two non-overlapping time windows on the critical words. Since the verb's information was hypothesized to possibly drive a reduction of mainly visual objects not matching its constraints in the later VWP experiments (Experiments **3** to **7**), the time window critical for analysis was on the verb. In order to not bias results towards the visual context having an influence, a comparison for processing effort on the verb without visual context was needed. The verb window was hence also analysed in the first two experiments not featuring visual context, although the purely linguistic context was not expected to provide strong enough constraints for participants to decide against certain nouns. In the second critical time window, namely the target noun itself, differences in processing effort and surprisal were expected in all experiments as soon as expectations about the noun can be informed by either linguistic or visual information, or even by the combination of both.

In **Experiment 1**, reading times were hence measured and analysed within the two non-overlapping time windows on the verb and on the target noun. Since reading time differences are often measured with a slight delay, the spill-over regions following the respective critical words were additionally analysed. Time measures were log-transformed due to the natural skewness of reaction time data, and entered as dependent variables into linear mixed-effects models. The contrast-coded Object and Verb conditions as well as the scaled length of the target word (measured in characters) were entered as fixed factors. Following Barr et al. (2013), the models were run with the maximal converging random effects structure, including intercepts and slopes for Subject and Item.

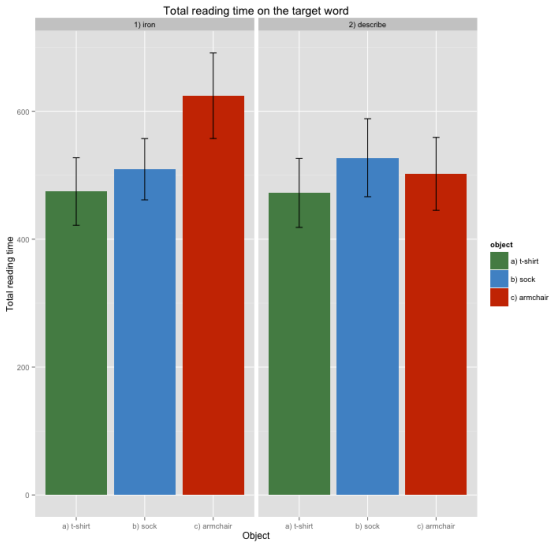


Figure 3.2 **Total dwell time results** in all levels of the conditions in **Experiment 1**. Error bars reflect 95% confidence intervals (CI).

3.4 Results and Discussion

Total dwell-time is shown in Fig.3.2. Data reveal a significant difference in processing effort as assessed by the measure on the critical region (the noun) only for the implausible condition, in comparison to the less plausible condition (*spill the ice cream* (M = 6.09, SD = 0.55) vs.*spill the book* (M = 6.25 , SD = 0.62), $p < .005$) if the verb was highly constraining (*spill*; significant interaction with Verb, $p < 0.05$). No differences in reading times were found between the *more* and the *less plausible* conditions (*spill the water* vs. *spill the ice cream*). Further, analysis of the first-pass measurement did not yield significant results. Regressions to the pre-target region showed the same pattern as total dwell-time. Analyses of the spill-over region showed no significant effects.

Discussion In line with the expectations, results show that verbs with different strengths of constraint did not differ in processing effort in the absence of visual context.

This is opposed to Maess et al. (2016), who in an MEG study found increased neural magnitudes on highly predictive compared to less predictive verbs. However, besides the fact that reading times may provide a less direct measurement of processing effort, the verbs used in their study (e.g. "conduct") had stronger constraints with respect to the subsequent noun, compared to the ones used in this study (e.g. "spill") which could explain the lack of an effect on the verb and spill-over region's reading times.

Moreover, the noun's processing effort was only significantly affected when the respective noun was implausible in its linguistic context and hence unexpected, i.e. highly surprising.

This may appear surprising, given that several previous studies found effects of predictability and plausibility in reading or listening (e.g., Kutas and Federmeier, 1999). Rommers et al. (2013) even found a graded N400 effect in response to more or less expected target words.

Again, however, the stimuli used in the presented study were not embedded in wider contexts; hence no additional information, apart from the comparatively low verb constraint, was given. That is, no further information was available for the listener to form (lexical) expectations about the target noun.

We hence suggest that in this case, verb constraints alone did not elicit concrete lexical expectations about the target nouns (beyond a semantic category).

This is mainly in line with Rayner et al. (2004b), who found an immediate effect of implausible words on eye movements in a reading task, but only very small and delayed effects for less severe violations of plausibility. Whether these findings would be replicated within a different measure and presentation mode was tested in the subsequent **Experiment 2**.

Chapter 4

Processing effort in the pupillary measure

4.1 Experiment 2: Linguistic context - Listening

In our initial experiment, using reading times to assess processing effort of target words in purely linguistic context, significant differences in effort were only measured in the implausible condition ("The man spills soon the book"), while no differences were observed between the more of less plausible continuations ("The man spills soon the ice cream" vs. "The man spills soon the water"). We interpreted the results as to reflect no specific (lexical) expectations about the target nouns, beyond a rough semantic category, relatable to the fact that the verbal constraints alone did not convey enough information for comprehenders to expect more than a category. Processing effort as reflected by reading times differed only when something clearly not spillable was presented as target noun. An alternative explanation, however, could be that either the presentation mode (written) or the sensitivity of the time-dependent measurement (reading times) were not sensitive enough for our manipulation and – at least partially – caused the null effect. The second experiment was hence designed in a different modality (auditory) and therefore used a different measure of effort (a pupillary measure). We again aimed to observe whether effects of target word expectations beyond the rough semantic category could be measured on the same set of linguistic stimuli (minus the implausible condition and the spill-over regions since both were no longer needed). The auditory experiment further marks an additional baseline for processing effort of the critical words in the absence of visual context in the respective pupillary measure which was also used in the later experiments featuring visual context. After all, establishing a baseline in the same measure allows for valid comparisons.

The third experiment in the presented series was designed to introduce additional visual context to the same linguistic stimuli. This time, processing effort was assessed using the pupillary measure ICA, in addition to traditional eye movements, which were hypothesised to simultaneously reflect patterns of anticipation in the presence of visual information. Especially the combination of those measurement allowed us to not only capture anticipation in replication of previous results by Altmann and Kamide (1999), sowing that verbal constraints cause a mapping of visual and linguistic context, hereby driving attention shifts to objects considered possible targets. It further enabled us to simultaneously assess whether such anticipatory patterns had an effect on linguistic processing effort on either the driving verb, or the target object, which has never been done before.

4.2 Method

All 36 native speakers of German, who participated in this study, gave informed consent and had not participated in the previous study. They were, again, monetarily reimbursed for their contribution. All were students of Saarland University and their age ranged from 19 to 46 years (M = 24.72). 20 experimental items and 26 filler items in four conditions (*constraining* or *unconstraining* verbs, crossed with *more plausible or less plausible* objects (see Table 3.1)) were used in each of the four lists for this experiment. The analysis was done using the identical models and time windows on the verb and the noun, minus the additional spill over regions. Even though no visual stimulus was presented, participants' eyes were tracked in order to extract the ICA values from the pupil jitter.

Predictions It was previously suggested that the verbal constraints alone did not contain enough information to cause participants to have lexical expectations about target nouns, resulting in highly similar processing effort for more or less constraining verbs and more or less plausible object nouns following those verbs. If this result was, however, due to the presentation mode (written) or the sensitivity of the time-dependent measurement (reading times), differences in processing effort were expected, as assessed by the ICA, for the same stimuli when presented auditorily. In that case, a lack of effects between the verbs and the plausible and implausible nouns following the constraining verbs in **Experiment 1** could not be accounted for by the nature of expectations in the purely linguistic context and surprisal-based effort itself.

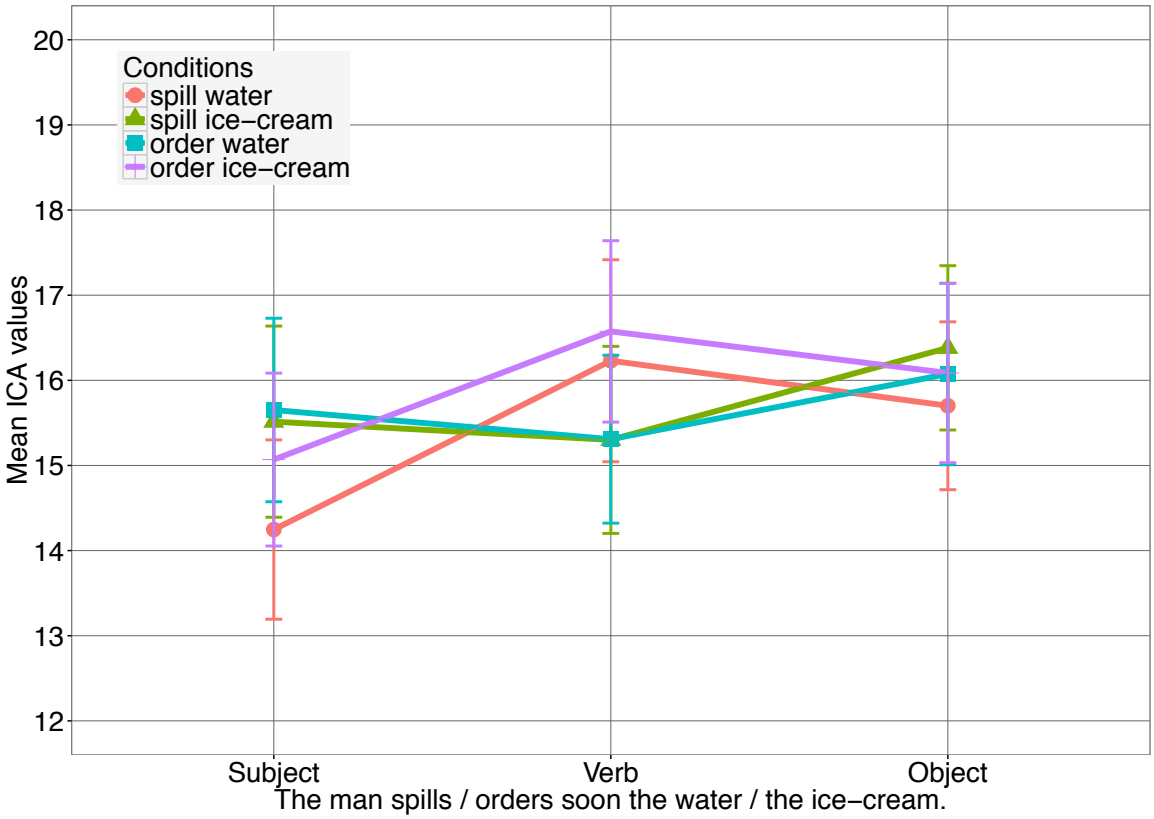


Figure 4.1 **ICA Results** for **Experiment 2** in all levels of the conditions. Error bars reflect 95% confidence intervals (CI).

4.3 Results and Discussion

Demberg and Sayeed (2016) analysed ICA events within a time window taken 600-1200 ms from the onset of the critical word. The critical words used in this and the following studies, however, differed in length across items. In order to compensate for these differences, the appropriate time window length of 600 ms was taken starting from the middle of the duration of the critical words (verb and noun). Word length was hence not additionally included as a covariate in the statistical models used to analyse the data from the ICA studies [4] .

ICA event counts were obtained for both eyes separately, but were summed (i.e., the respective 100 ms time bins from the left and the right eye were summed up) for analysis since no theoretical reason is given as to why differences should be expected for the two eyes (Demberg and Sayeed, 2016). Data from both eyes (per 100 ms) were then summed for the entire 600 ms time windows. ICA events during the two critical time windows (verb & noun) were the basic dependent variable. Since those events were treated as a count variable, generalised mixed effects models with Poisson distribution were used. All independent variables were contrast coded for the analysis.

Again, the verb and object manipulations (i.e., *order water / ice cream* and *spill water / ice cream*) did not result in differences in effort on the noun, as assessed by the ICA. Fig.4.1 shows how both verb conditions cause almost identical average processing effort on the noun (*spill water/ice cream*, M = 16,05, SD = 6.51 vs. *order water/ice cream*, M = 16.07, SD = 7.09). That is, no effect of verb constraint was observed in the ICA values of either critical time window in any of the conditions.

Discussion Reading time results from **Experiment 1** were replicated in the pupillary measure, showing no differences in processing effort for the different verbs and for more or less plausible nouns following constraining or unconstraining verbs. It is hence suggested that the lack of differences between conditions was due neither to the measure, nor to the presentation mode. Rather, it can be explained by the verb not containing enough information for the listener to concretely expect the target noun on the lexical level without any further (linguistic or visual) information.

Results from **Experiment 2** further serve as a reference for processing effort in the absence of any visual context in the pupillary measure used in most of the following experiments and show how both verb and noun conditions require equal processing effort in a purely linguistic context.

[4]Neither were noun frequencies included as covariate, since head noun frequencies were controlled for within each item. Frequencies were assessed using the word lists DeReWo, http://www.ids-mannheim.de/derewo.

Although no effects of processing facilitation related to expectations or even predictions were found, we interpret this outcome in accordance with Wlotko and Federmeier (2015). That is, although the authors found such effects in the N400 if 500 ms stimulus onset asynchrony were used, demonstrating that timing is a major influence on prediction. Namely, in the sense that Wlotko and Federmeier (2015) interpret their findings as to show that prediction is not an invariant process but rather results from the brain flexibly using its resources in order to most efficient and effectively achieve comprehension. In other words, effects caused by prediction are measurable in some contexts and may not be visible in other setups, depending on how beneficial possible predictions are for the comprehender.

In case of **Experiment 2**, the weakly constraining context – as opposed to, for example wider multi-sentence contexts as used by Wlotko and Federmeier (2015), or even visually enriched contexts as used in VWP studies in which signs of anticipation have been measured (Altmann and Kamide, 1999) – did very likely not bear enough information to motivate and justify detailed (lexical) expectations as they would very likely to be wrong in this context and, hence, not beneficial for the aim of comprehension.

Whether the addition of visual context contributes enough information that is used by the comprehender to make expectations – hereby causing effects of processing facilitation for less surprising target nouns and probably even effects of an exclusion of non matching objects during the verb – will be tested in the following experiments where additional visual context information in introduced while the same stimuli were presented.

4.4 Experiment 3: Linguistic and visual context

Both previous experiments used purely linguistic stimuli to measure the effect of verb constraints on not only the processing effort for the verb itself but also on target word expectations and processing effort for the reference noun.

Results defined a reference for expectation based processing effort in the absence of visual information and only revealed effects on the target noun if the respective word was an implausible continuation following a highly constraining verb. No effects were found on the verb itself (i.e., attributable to a reduction of entropy, or possible referents, in case of the highly constraining, compared to the unconstraining verb) or between the more or less plausible nouns following the restrictive verb, suggesting that the information conveyed by the verbal constraint was not enough for participants to have expectations about the target noun (at least at the level of our manipulation). Visual context reducing the number of possible referents to the displayed ones can add crucial information which is likely to make more specific expectations about the target word less error-prone, and hence more beneficial

for comprehension if the visual information has proven to be reliable. Indeed, the most reliable effects attributable to anticipation were found in experimental setups featuring visual contexts. Kamide et al. (2003), for example, found that verb information drives anticipatory eye movements towards objects in the display that match the verb and can possibly depict the target word. That is, hearing "The boy will eat..." lead comprehenders to direct more gazes toward the depicted edible object, while hearing "The boy will move..." did not cause such patterns in the eye movements, according to Altmann and Kamide (1999) referred to as *anticipatory* eye movements.

Although it is known that visual information can inform target word anticipation, not much is known about the actual effect of the visual context on linguistic processing effort. Measuring and quantifying a possible effect can not only answer open questions about the prediction encouraging nature of VWP setups, but can further explain how non-incremental information is being integrated and how it can affect statistics of the mental model and the resulting target word expectations. Especially when formalising the results, importance and power of rational approaches for the explanation of psycholinguistic data can be extended to situated language comprehension.

All following studies hence added visual context to the linguistic stimuli sentences in order to test whether and how the additional visual information really affects (lexical) expectations and actual processing effort of the critical words.

By deploying the pupillary measure of effort (ICA), it was possible – for the first time – to observe whether anticipatory patterns, which were expected in the setup, are related to actual effects on processing effort for the driving word, or solely cause effort that so far has not been reported to be reflected by the measure, such as effort with respect to motor decisions or actual actions. At the same time, it is possible to measure potential effects on the processing effort for the noun, caused by more specific expectations made prior to it due to the visual context.

4.5 Visual Materials: Design and Validation

Design All scenes presented in the following three VWP studies were composed in the same way: They always consisted of four simple pieces of clip art, arranged around the screen center, while no agent or background was given in order to prevent influences from face recognition or background integration.

In *Experiment 3*, one of the four objects corresponded to the target mentioned in the sentence (e.g., spill *water*). In the case of the constraining verb, a second object was a competitor, matching the constraint information only to some extent (less plausible, e.g. spill

Figure 4.2 **Example of visual stimuli** as used in **Experiment 3**. For more items see Appendices A and B

ice cream), while the remainder were non-matching distractors (e.g., spill *suitcase* or *coat* as shown in the sample display in Fig.4.2). Distractor objects always belonged to a different class, compared to suitable, competing objects (e.g. *drinks* vs. *clothes*), and were clearly distinguishable in every display.

In the case of the unconstraining verb, all four objects in a display matched the category introduced by the word (e.g., order *water, ice cream, suitcase, coat*).

None of the objects in any of the displays corresponded to the sentences' highest-cloze nouns, in order to not bias comprehenders into clearly preferring one object after hearing the verb due to purely linguistic probabilities.

Clip art items within one visual display were of similar complexity, uniformly salient in terms of colours, and depicted concrete and inanimate objects to keep visual processing as similar as possible. The scenes were always counterbalanced between two items in such a way that the display for one condition of one sentence also served as display for the other condition of a second item sentence. Additionally, the positions of targets, competitors and distractors were rotated in order to prevent possible effects of habituation. Filler trials introduced further variation in terms of the number of categories displayed (i.e., edible, drinkable, or wearable objects, but also ridable or ironable ones, etc.). This way, a possible grouping of objects based on features other than the compatibility with the verb was aggravated.

Validation All Clip art used in the following studies was pre-tested for naming to make sure all of the objects were recognisable and well distinguishable. A naming test was performed on-line, where the unique participation of each user was controlled for. All clip art was presented in two randomised lists. Participants were asked to spontaneously write down the name of the object they saw in the picture. This way, it was ensured that participants in the actual experiment would not only recognize the objects well, but also would chose the exact

same word to relate to them. 24 people participated in this naming task. Pictures were used in the experimental items only if they were recognised reliably and named correctly ($> 90\%$ of participants recognised each object correctly).

4.6 Method

By adding visual context to the same stimuli used in the previous experiments, the immediate effects of visual context on the statistics of the mental model, and, hence, expectations and the related processing effort for a spoken target word were assessed. It was further observed whether anticipatory eye movements are relatable to differences in actual processing effort required for the verb itself, e.g. due to reduction of visual entropy upon encountering the verbal constraint.

The linguistic stimulus set, manipulation and design were identical to **Experiment 2**. The simultaneously presented *visual* stimuli were arranged as shown in Fig.4.2 and functioned as an enhancement of both manipulations on the linguistic level (plausibility and verb constraint). That is, by decreasing the number of potential target object options from a non-assessable number of nouns matching the verb in ***Experiment 2***, to a countable number of options shown in the display.

Visual displays were presented from 1000 ms before sentence onset and during the whole sentence (see Fig.5.3 for an example trial). Participants were asked to interact naturally with the scenes, not forcing themselves to look at or away from items while their eyes were tracked in order to obtain eye movement data and extract the ICA values from the same data set. 36 native speakers of German (all students of Saarland University) between 19 and 38 years of age ($M = 23.25$), all of whom had not participated in any of the previous lab experiments, were tested. All Participants gave informed consent and were monetarily reimbursed for their participation. Data from two participants had to be excluded from the analysis due to technical problems.

Predictions Based on results typically found in similar VWP setups (e.g., Kamide et al., 2003), a replication of verb-driven anticipatory eye movements towards depicted target options matching the verb was expected as listeners exploit visual context information to expect the target word.

That is, in case of the highly constraining verb, increased anticipatory eye movements to the possible target object and also to the competitor were expected, compared to looks towards the two distractor objects not matching the verb constraint.

If anticipatory eye movements at the verb are related to actual differences in processing effort for the respective word informing and driving them – possibly due to the listener reducing entropy, that is, narrowing down the domain of subsequent target reference more in the case if the highly constraining verb, as proposed by Kamide et al. (2003) – the ICA on this region was expected to differ between the two verb conditions. Namely, processing effort for the highly constraining verb was expected to increase, as it allows for a reduction of more references. This would be in line with Hale (2001), suggesting that words reducing more entropy take more effort to process.

In accordance with Demberg and Keller (2008); Smith and Levy (2013), who found that more surprising and less expected words take more effort to process and also with Maess et al. (2016), who specifically report a correlation between the effort of reducing referent options in advance and facilitated processing of the actual target referent, a processing facilitation for the more predictable target nouns was expected as a result of the anticipations.

That is, less ICA events (i.e., lower processing effort) were expected in the case of the more plausible noun (*water*), when following the constraining verb (*spill*) if visual context information significantly influenced predictability and surprisal, and hence processing effort for the target nouns. In other words, the more predictable (i.e., the less surprising) a noun becomes in its multi-modal context, the easier it should be to process it.

Since the ICA index is thought to be sensitive to effort related to (linguistic and visual) information processing (e.g., Demberg and Sayeed, 2016; Marshall, 2002), we additionally expected an overall increase in processing effort (i.e., more ICA events in all time windows) throughout an entire trial, compared to the previous experiment in which no additional visual context had to be processed simultaneously, as opposed to incrementally as in bigger linguistic contexts.

4.7 Results and Discussion

Eye movement Data For presentation purposes, fixation plots of the following studies show the overall fixation distribution across an averaged trial length in all conditions. Dashed lines mark the area of interest for eye movements to potential target options in anticipation of the noun, that is, the verb onset on the left and the noun onset on the right. Fig.4.3 shows increased anticipatory eye movements towards the object best matching the verb in the constraining verb condition (*spill*). At the same time, no such preference for any of the displayed objects was found in the unconstraining verb condition (*order water/ice cream*: see c & d of Fig.4.3), suggesting that, based on the context information, none of the objects could be specifically expected.

Statistical significance of differences in anticipatory eye movements in the different verb conditions was assessed by analysing the new inspections (i.e., the first in a series of inspections towards a region during the time periods of interest, interpreted as an active attention shift towards the respective object) between both conditions.

That is, probabilities of verb-driven attention shifts towards objects matching the verb constraint were analysed, based on the idea that the highly constraining verb was more likely to enable active, anticipatory attention shifts to a potential target object.

The time windows for the analysis spanned from verb start to article start and from article start to noun end. New inspections of the different objects in a visual scene were encoded as a binary dependent variable (as being either present or absent) and were analysed using generalised mixed effects models with a Poisson distribution.

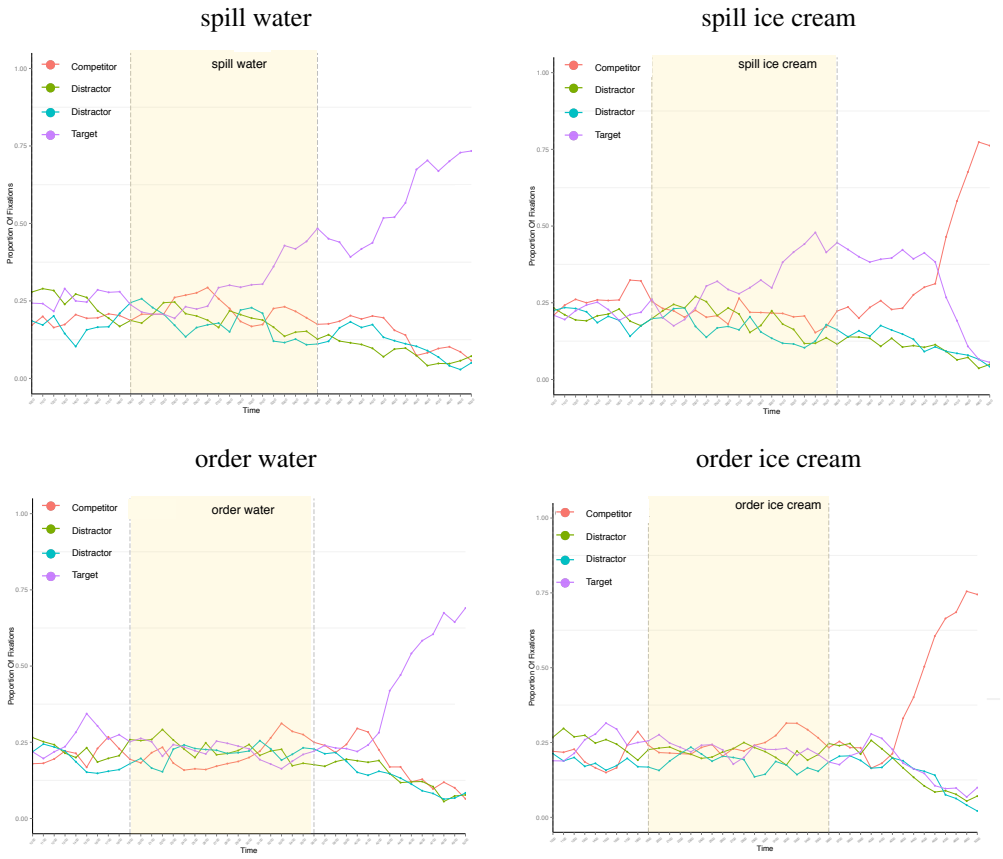
In the verb window, looking toward the target object (*water*) was significantly less probable if the verb was unconstraining compared to when the verb was (highly) constraining: *order* (M = 0.11, SD = 0.32) vs. *spill* (M = 0.16, SD = 0.37), $\beta = -.378$, $SE = .093$, $z = -4.038$, $p < .001$. Further, inspections towards objects *not* corresponding to the target noun were significantly more likely if the verb was unconstraining: *order* (M = 0.34, SD = 0.47) vs. *spill* (M = 0.3 , SD = 0.46), $\beta = 0.141$, $SE = 0.071$, $z = 1.985$, $p < .05$.

For the noun region, new inspections towards the actual target were compared to inspections towards any other object displayed. By this, it was assessed if participants stayed focused and still mapped linguistic to visual information, as well as how quickly listeners identify the object corresponding to the noun they are hearing, while more expected words should be discovered earlier. In line with many former VWP studies (see e.g., Cooper (1974)), data revealed that participants looked more towards the mentioned object than towards any other object in the display, while no difference in timing was measured.

ICA To assess effects of visual context on expectations and surprisal-based processing effort, ICA event counts were again analysed within the same non-overlapping 600 ms time windows on the verb and the noun, starting from the middle of the critical word's duration, as done in the previous ICA study. ICA events obtained within the two critical time windows were used as the basic dependent variable in generalised mixed effects models (see 4.1 for the full model). Subjects and items were included as completely crossed random factors.

No significant effects were found in the verb window, where eye-movement data showed clear patterns of anticipation in the presence of visual context information. However, a non-significant trend towards higher ICA values in the case of the highly constraining verb, allowing for a higher reduction of visual uncertainty (entropy in this case) was observed.

Figure 4.3 **Proportion of fixations** across trial length in all conditions of **Experiment 3**



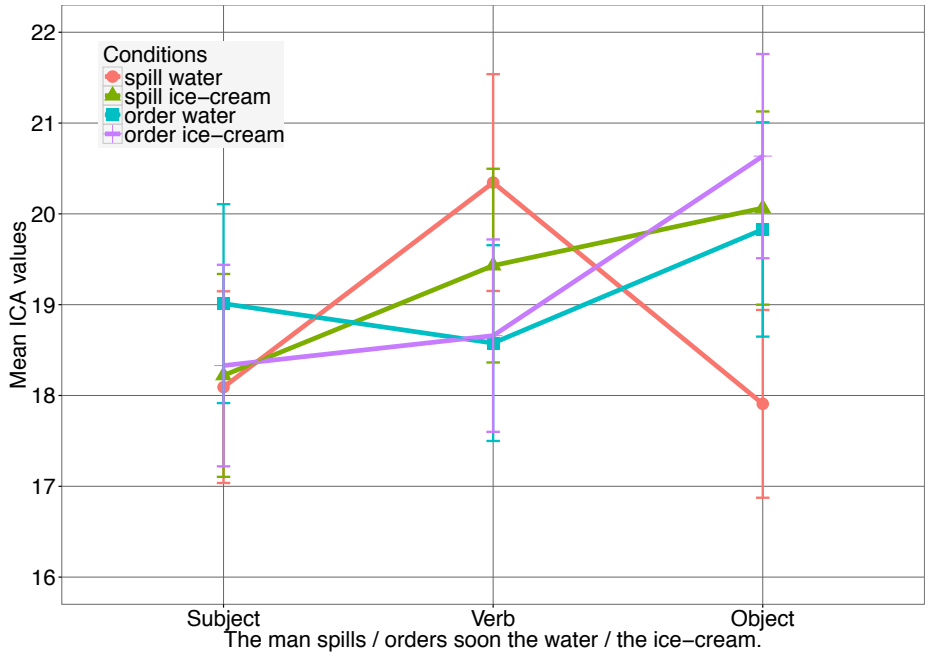


Figure 4.4 **ICA Results** for **Experiment 3** in all conditions. Error bars reflect 95% confidence intervals (CI).

Most importantly, this time, values from the noun window showed a significant interaction of *Verb* and *Object* ($\beta = -.071$, $SE = .035$, $z = -2.05$, $p < .05$), as well as a significant main effect of *Verb* (*spill* M = 18.99 , SD = 6.99 vs. *order* M = 20.24, SD = 7.7, $\beta = .063$, $SE = .032$, $z = 1.98$, $p < .05$). This suggests that *water* and *ice cream* affect processing effort to a different degree when succeeding the constraining verb *spill* than they do when following the unconstraining verb *order*.

Planned pairwise comparisons revealed a significant effect of *Noun* in the case of the constraining verb *spill* (*water* M = 17.91 , SD = 6.91, *ice cream* M = 20.06, SD = 7.07, $\beta = .113$, $SE = .045$, $z = 2.51$, $p < .05$), but not for the unconstraining verb *order* (*water* M = 19.83 , SD = 7.91, *ice cream* M = 20.64, SD = 7.49, $p = .46$), implying that *water* was easier to process than *ice cream* only when following *spill*. In line with this, a significant effect of *Verb* was found, in the case of the slightly preferred object *water* ($\beta = .094$, $SE = .039$, $z = 2.41$, $p < .05$), but not for *ice cream* ($p = .675$).

Experiment Comparison ICA values collected within the critical time windows in **Experiment 2** and **Experiment 3** were compared in order to assess whether processing effort rises as more information is presented in combination and has to be processed *simultaneously*, as opposed to bigger linguistic contexts, where information is given incrementally.

Table 4.1 **Differences in ICA** for **Experiment 3**,
Model1: ICA values on verb/noun ∼ Verb-Object Interaction + Verb + Object + (1 + Verb-Object Interaction + Verb + Object || Subject)+ (1 + Verb-Object Interaction + Verb + Object || Item), family=poisson (link = "log")
Model2 (Main effect): ICA values on noun ∼ Object + (1 + Object || Subject)+ (1+ Object || Item), family=poisson (link = "log")

Time window: Verb	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model1)	(Intercept)	2.9286	.0360	81.24	<0.001
	Verb - Object Interaction	.0395	.0655	.60	.547
	Constr. vs. unconst. verb	-.0538	.0313	-1.72	.086
	More vs. less plausible noun	-.0069	.0264	-.26	.795
Time window: Noun	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model1)	(Intercept)	2.948	.0346	85.16	<0.001
	Verb - Object Interaction	-.0722	.0610	-1.18	.2365
	Constr. vs. unconst. verb	.0619	.0319	1.94	.0529
	More vs. less plausible noun	.0782	.0422	1.85	.0636
Time window: Noun	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model2)	(Intercept)	2.9185	.0377	77.35	<0.001
	More vs. less plausible noun	.1129	.0450	2.5	<0.05

Each experiment's ID (*2* vs. *3*), as well as both Verb and Object conditions were entered into the generalised mixed effects models as contrast coded fixed factors.

Fig.4.5 shows overall higher processing effort if additional visual context had to be processed along with the utterance.

It further shows a significant interaction of *Verb* and *Experiment* in the verb ($\beta = -.08$, $SE = .029$, $z = -2.82$, $p < .005$) and noun window ($\beta = .066$, $SE = .032$, $z = 2.05$, $p < .05$). This implies that target nouns were only more predictable in the context of the constraining verb, as compared to the unconstraining one, if visual context information was available.

In accordance with the graph, models revealed a highly significant main effect of *Experiment* in the verb window (**Experiment 2**: M = 15.80 , SD = 7.25 vs. **Experiment 3**: M = 19.19 , SD = 7.25, $\beta = .22$, $SE = .020$, $z = 10.7$, $p < 0.001$) and noun window (**Experiment 2**: M = 16.06 , SD = 6.79 vs. **Experiment 3**: M = 19.76 , SD = 7.41, $\beta = .262$, $SE = .020$, $z = 13.05$, $p < 0.001$). This highly suggests that processing effort – as assessed by the ICA in this setup – indeed increases exponentially with the amount of information presented simultaneously.

We further found a significant interaction of *Verb* (constraint) and *Experiment* in the verb ($\beta = -.008$, $SE = .029$, $z = -2.82$, $p < .005$) window. The same interaction was also found in the noun window ($\beta = .066$, $SE = .032$, $z = 2.05$, $p < .05$). In both cases, follow up comparisons revealed however, that the interaction was carried by the opposite direction of the non-significant trend between the two verbs in the two studies. That is, compared to the unconstraining verb (*order*), the constraining verb (*spill*) tended to require more effort to process in **Experiment 3**, where visual context was given, while requiring slightly less effort in Experiment 2, in the absence of visual context.

In sum, these results show not only that visual context, in combination with constraining verbs, results in differences in expectation based processing effort for the target noun, but also that the ICA increases with the amount of information presented simultaneously, supporting the idea that the index indeed culls effort related to information processing.

Discussion The first VWP experiment presented in this thesis shows the immediate significant effect if visual information on word expectation and processing effort.

Initially, eye-movement data revealed listeners' use of verbal constraints to narrow down visually present referent options to expect the noun, as in line with the expectations based on results found by Kamide et al. (2003). That is, in replication of their results, participants in this study were more likely to fixate objects matching the highly constraining verb in the time window prior to the noun, suggesting that listeners exploited the visual information in combination with the verb in order to anticipate the noun with more or less certainty. No such

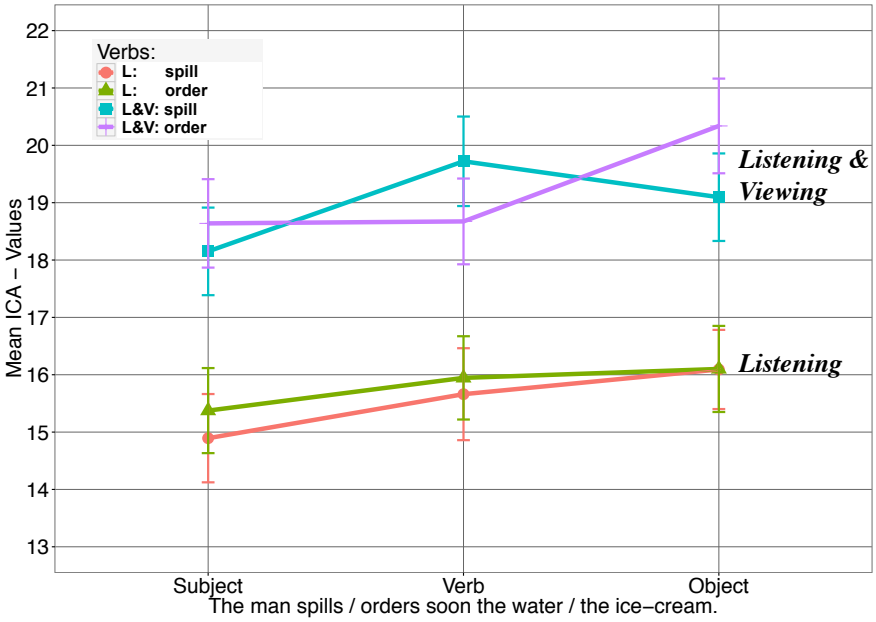


Figure 4.5 **Comparison of ICA values** in **Experiment 2** and **Experiment 3**. Both object nouns of a verb condition are considered together for the plot. Graphs show the significant main effect of *Experiment* and a significant interaction of *Verb* and *Experiment*. Error bars reflect 95% confidence intervals (CI).

Table 4.2 **Comparison of ICA values** in **Experiment 2** and **Experiment 3**, i.e., with and without visual context.
Model 1: ICA values on verb ∼ Study (2 vs. 3) Study + Verb + Verb - Study Interaction (1 + Verb | Subject)+ (1 + Verb Interaction | Item), family=poisson (link = "log")
Model 2: ICA values on noun ∼ Study (2 vs. 3) Study + Verb-Object Interaction + Verb + Object + Verb-Study Interaction + Object-Study Interaction (1 + Verb + Object + Verb-Object Interaction | Subject)+ (1 + Verb + Object + Verb-Object Interaction | Item), family=poisson (link = "log")

Time window: Verb	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model 1)	(Intercept)	2.8403	.0309	92.07	<0.001
	Study	.2196	.0205	10.70	<0.001
	Constr. vs. unconst. verb	-.0108	.0192	-.56	.5744
	Study-Verb Interaction	-.0832	.0295	-2.82	<0.005
Time window: Noun	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model 2)	(Intercept)	2.8583	.0269	106.11	<0.001
s	Study	.2624	.0201	13.05	<0.001
	Constr. vs. unconst. verb	.0329	.0229	1.43	.1518
	More vs. less plausible noun	.0418	.0314	1.33	.1835
	Verb-Object Interaction	.0782	.0422	1.85	.0636
	Study-Verb Interaction	.0659	.0322	2.05	<0.05
	Study-Object Interaction	.0265	.0337	.79	.4316

anticipatory patterns were found in the context of the unconstraining verb. In other words: Like the verb "eat" in Kamide et al. (2003), that caused comprehenders to look towards the only edible object shown, the verb "spill" caused participants to anticipate the noun "water", while the verb "order" did not.

The processing effort for the respective word, however, was not significantly affected by anticipation (as reflected by the anticipatory eye movements). Only a non-significant trend for differences in processing effort for the highly restrictive, compared to the non-restrictive verb was found in the data. This trend showed that the unconstraining verbs, carrying less information to exclude potential visually present referents, were slightly easier to process. Very likely, the observed trend is attributable to differences in the verbs' nature in the visual context. That is, the lack of a constraint of verbs such as "order" could cause the listener to put less effort into mapping linguistic to visual information upon encountering the verb, as less information can be gained from this process (see also e.g., Maess et al., 2016). However, since the direction of the trend shows a pattern that would also be in line with entropy reduction, it could also be attributable to the listener's reduction of visual uncertainty, that is, the exclusion of objects from the display not matching the constraining verb. The subsequent experiment will hence test whether the non-significant trend for processing differences on the verb could rightly be attributed to the reduction of visual uncertainty, rather than to the nature of the verb itself.

As opposed to the previous (purely linguistic) experiment, the pupillary measure of processing effort deployed simultaneously this time revealed differences in expectation based processing effort on the noun.

More specifically, processing effort differed between both noun conditions (*water* vs. *ice cream*) *only* when the verb was highly constraining (*spill*). No such difference was found after the unconstraining verbs (*order*), which conveyed less information supporting the narrowing down of potential referents given in the visual context in order to expect the target noun.

The same linguistic stimuli, which resulted in no significant differences in the previous study without visual context, now caused significant differences in effort for processing the noun between the two verb types. This does not only add to the existing evidence that listeners match linguistic with visual information to make the target nouns more predictable (i.e., less surprising) but most importantly also shows that the resulting anticipations actually lead to expectations affecting word processing (i.e., noun processing is facilitated is expectations can be made in advance). This strongly suggests a significant influence of non-linguistic context on lexical expectations and expectation-based processing. Whether this influence is describable by (multi-modal) surprisal remains to be tested.

As a by-product, results from this experiment reveal that the noun *"ice cream"* was equally hard to process in both verb conditions, although a clearly favoured competitor (*water*) was present in the constraining verb condition. The equal amount of processing effort measured at *"ice cream"*, despite having a favoured competitor in only one of the conditions is in line with a the recent hypothesis that predictions, or in this case, visually enriched expectations are not necessarily competitive. This in accordance with Gambi and Pickering (2017), who found that participants were equally able to recognize target words, independent of whether a present alternative continuation was more probable.

Finally, a comparison of the two studies deploying the pupillary measure with and without visual context (**Experiment 2** vs. **Experiment 3**) revealed a significant overall increase of processing effort in the presence of additional visual information which had to be processed simultaneously. This is interpreted as further evidence for the direct link between processing effort, as assessed by the pupillary measure, and the amount of information presented.

The following experiment a) tested whether the trend for differences in verb processing effort are attributable to a reduction of visual uncertainty, and b) observed the important impact of visual information on word processing.

That is, it was designed to exclude any variation in linguistic surprisal to specifically quantify the effect of visual context seen on the noun in the present experiment, in order to observe if multi-modal surprisal of a word – as modulated by the visual referential context – can predict pupillometric measures of on-line processing effort.

Chapter 5

Number of competing (potential) referents

The data from our previous studies suggested that, while only effects of coarse grained expectations for the target word were measured in purely linguistic contexts of the baseline experiments, the addition of visual context resulted in information mapping and integration, followed by more fine grained target word expectations. More specific, effects caused by manipulations of word surprisal on actual processing effort only came up as visual context was added to the sentences, revealing a potential trend which could be described by comprehenders reducing entropy (i.e., visual uncertainty) about upcoming target words already on the verb, which then result in clear effects of expectancy on the noun.

While already suggesting a strong effect of visual information on linguistic processing, possibly describable by surprisal also in the case of extra-linguistic information, a subsequent step is required in order to test how strongly visual context information actually can contribute to the measured effects. Since at least two predictor variables were included on the linguistic level in the previous experiments (i.e., manipulations of *predictability* and *plausibility* of target words in their context) in order to observe possible effects of this manipulation on surprisal-based processing effort, it is not entirely clear whether the trend for differences describable by entropy reduction were caused by the difference in verb types or by an actual differing reduction of (visual) uncertainty. As a subsequent step, this influence needs to be fully exposed (i.e., any measured effects need to be clearly attributable to visual information being integrated with linguistic context) before and in order to quantify it. We consecutively designed an experiment that eliminated any possible variation on the linguistic level, while only the visual context, that is, its statistical regularities when combined with the verb, was manipulated. The following chapter describes two experiments using the same design and, again, the same set of stimuli (plus additional ones for the EEG experiment in order to achieve

average power in such a setup) and task to observe the effect of different target word surprisal, manipulated merely via variations in the visual context. The first experiment again deploys the ICA as a pupillary measure of effort, while the second one uses a more established on-line measure of processing effort, namely the ERP component N400. As a consequence of the design, it is not only better possible to directly observe how comprehenders make use of the different visual contexts, but also to relate possible differences to the rational concepts of surprisal and entropy reduction.

5.1 Experiment 4: Number of competing (potential) referents in the ICA

Both experiments presented in this chapter use the same design and task (adapted to the measures) while using not only a pupillary, but also an ERP measure of effort. Both were designed to include only *one* predictor describing a manipulation *only* of the visual context, while any linguistic variation was excluded. This was done by using the exact same linguistic stimuli across all conditions, featuring only constraining verbs (*spill*) and nouns with high thematic fit (*water*), each presented in four visual contexts where the number of displayed objects matching the verb constraint was manipulated (e.g. **0**, **1**, **3**, or **4** "spillable" objects).

This way, linguistic surprisal was held identical across conditions, while the same verb reduced visual uncertainty to different degrees, resulting in different expectancy of the target word *only* due to the variant visual contexts each sentence was presented in. This manipulation results in a well controlled context to observe whether the processing effort for a word – as expected based on situated surprisal – can also predict the deployed pupillometric measure when *only* the visual referential context alone modulates the target word's actual expectancy and surprisal.

In this case, results could provide important first evidence for a possible link between *visually informed* surprisal (as opposed to purely linguistic surprisal) and actual processing effort. This would strongly suggest that surprisal is also appropriate for the description of visual context effects on word expectancy in the VWP, and can bear further implications about the way participants evaluate visual context information in order to expect target words.

In addition, the design also allowed to test whether the previously observed trend for differences in processing effort on the verb was indeed attributable to a reduction of (referential) uncertainty, rather than being caused by possible conceptual and linguistic differences between the verbs in the previous study. In this case, the design of the present study would reinforce the effect as the same verb reduces uncertainty to different degrees in different

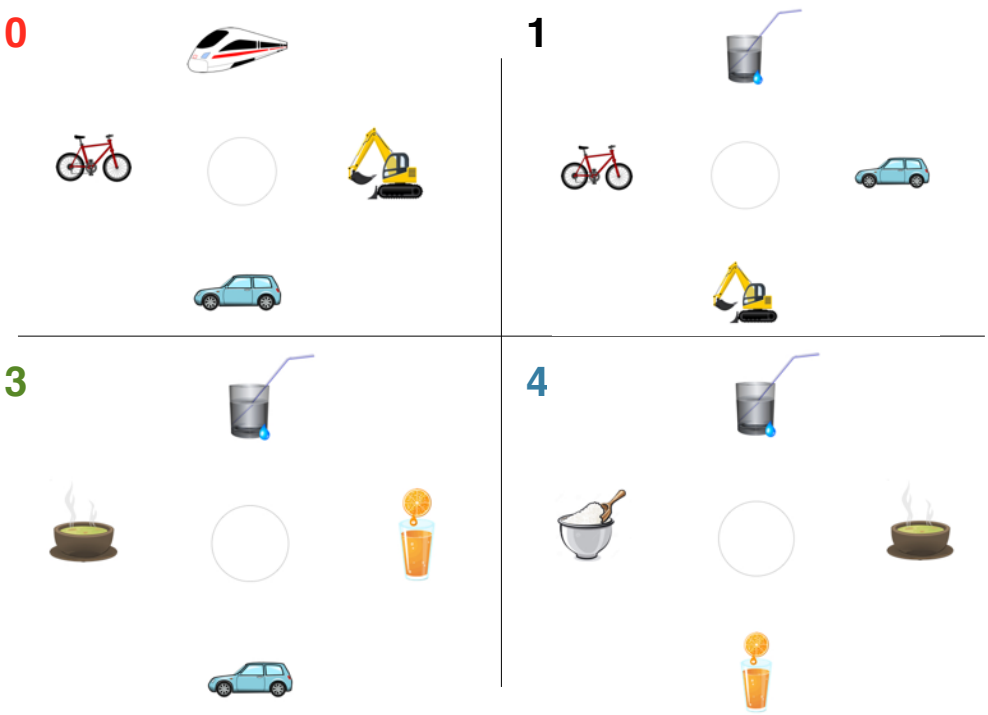


Figure 5.1 **Example Stimuli** from **Experiment 4**. From left to right and top to bottom: **0**, **1**, **3**, and **4** possible targets, given the sentence *"The man spills soon the water."* (Numbers were not depicted in the experiment). For more items see Appendices A and B.

visual contexts. Differences in processing effort for the verb were then expected to increase between conditions in line with the entropy reduction hypothesis introduced by Shannon (1949).

Visual Stimuli Validation A pre-test assessed whether the pieces of clip art used in this experiment were indeed interpreted as related to the verb they were intended to be matched with in the actual stimuli combinations. That is, the entire clip art was presented in two randomized lists in an online web-form that was filled in by 40 voluntary participants. They were asked to spontaneously decide whether or not an *object* was *"verb-able"*, by ticking a box stating either "yes" or "no". All experimental items were only used in the actual experiment presented here if they were well relatable to the verb they were presented with, that is if the pictures where related to the verb with > 90% certainty (correct answers per item).

5.2 Method

All 20 constraining verbs from the previous studies (followed by the *plausible* object noun), plus 20 additional new sentences of the same type (in order to increase power), were used in this experiment. In sum, 40 item and 40 filler sentences were combined with the visual displays (for all items see Appendix) in such a way that all four conditions of each display (i.e., 4 variations of each of the 40 displays, summing up to 160 displays in total) shared one sentence. The visual scenes were adapted in the sense that the number of instantiations of a category selected by the verb, i.e., the potential referents, or competitors for that matter, differed. That is, either none, one, three, or all four of the pieces of clip art shown in a scene could be target referents matching a verb (see Fig.5.1, from left to right and top to bottom). The displays were counterbalanced between two items in such a way that, for example, a *0* target condition picture for one item served as a *4* target condition in another item. In order to prevent habituation to a specific position, positions of targets, competitors and distractors in the displays were rotated. The 40 additional filler trials were followed by a task (identical to the one used in **Experiment 3**) to keep participants focused and introduced variation in terms of the number of categories displayed (i.e., edible, wearable, or driveable objects, etc.) to prevent possible grouping effects based on the categories shown rather than on the verb information. This way it was not easily possible to expect a certain verb based on visual priming or grouping of pieces of clip art based on obvious features. All sentences in the experiment were presented auditorily and again always simultaneously with the visual displays in order to prevent memory effects at the points of reference resolution.

32 native speakers of German (all students of Saarland University), aged between 18 to 32 years (M = 24.56), who had not participated in any of the previous experiments, were tested under informed consent and were monetarily reimbursed for their contribution. For the entire experiment, a total of 160 visual displays were paired with 40 sentences and split into 4 lists using a Latin square design. Each participant went through one of the lists and therefore heard each sentence in only one visual condition.

Predictions In line with the Entropy Reduction Hypothesis (Hale, 2003, 2006; Linzen and Jaeger, 2014), suggesting that the further the information submitted by a word reduces uncertainty about subsequent input, the harder it is to process that word, it is expected that the same verb is harder to process in conditions where it selects fewer of the four objects (i.e, **1** and **3**, as compared to **4**). That is, given the previously observed trend for processing effort differences on the verb is related to entropy reduction for the sake of target word expectations.

According to this hypothesis, the verb is further expected to be easiest to process in condition **4**, where referential entropy can not be further reduced based on it's constraints. In the special case of condition **0**, it also could be easier to process, compared to conditions 3 and 1. However, an alternative hypothesis for this exceptional case is to expect increased processing effort caused by the mismatch between word and picture information and the subsequently obstructed mapping process.

For the noun window, a facilitation of processing is expected as fewer potential target options are shown (**1** and **3**), as compared to **4**. That is given that only the number of depicted competitors in the visual context in the absence of any linguistic variation has a strong enough influence on target word expectations to affect the effort of processing that word.

This is in line with Shannon (1949) and also in extension of the results previously reported for example by Demberg and Keller (2008), who found shorter reading times for words with lower surprisal, hereby linking purely linguistic surprisal to differences in processing effort as assessed by reading times. In case of the present experiment differences on the noun would not only support these results and demonstrate their reliability even in a different (online) measure, but also further suggest that surprisal highly influenced by visual information – as opposed to purely linguistic surprisal derived from LMs – can also be directly linked to processing effort.

Again, increased processing difficulty can be expected for the noun of condition **0**, as compared to any other condition as none of the object matches the noun, meaning that the mapping of picture and noun is obstructed as well as none of the objects prepared the participants in order to expect the actual target noun.

Analyses

Eye movement Data New inspections to the actual target objects during the verb (i.e., in anticipation of the noun) were compared between conditions, expecting less anticipatory looks to the actual target as more competitors were considered, as done and described previously. New inspections to the actual target mentioned were compared to new inspections to any other object in the display during the target noun (i.e., as a the matching object is identified), additionally indexing whether participants paid attention to the stimulus.

ICA ICA data was analysed as before, using the identical time windows (600 ms from the middle of the verb and the object noun). Differences between the conditions (*Number* e.g., zero versus one possible target objects displayed) were contrast coded and entered into the models as fixed factors.

5.3 Results and Discussion

Eye movement Data The overall distribution of fixations across an averaged trial is plotted for all conditions in Fig. 5.5 for presentation purposes only. They show an increase in fixations towards objects matching the verb from the onset of the corresponding word onwards (left dashed line) when either one or three objects matched the verb constraint. That is, anticipatory patterns occur in cases where the visual scene allowed for more specific expectation about the upcoming target noun, even when more than one possible target objects are shown. This indicates a discrimination between those objects that matched the verb and those that did not, while all matching objects were equally considered. No increase in fixations was found when the context did not allow for any further specific anticipations (i.e., the mismatch case in condition **0** and when all objects matched in condition **4**).

In order to assess statistical significance of anticipatory glances during the verb, inferential statistics were again ran on new inspections between conditions to assess the probabilities of verb-driven attention shifts towards objects matching the verb (see Fig. 5.2). It was expected that anticipatory inspections to the actual target object decrease if other competitors are taken into consideration.

New inspection data on the verb was in line with the patterns seen in the fixation distribution plots: They revealed a significant increase in attention shifts towards the object corresponding to the target noun upon hearing the verb as fewer competitors are shown, i.e. in condition **1** (M = 0.21, SD = 0.41), compared to **3** (M = 0.17, SD = 0.38) ($\beta = -.221, SE = .099, z = -2.21, p < .05$) and to **4** ($M$ = 0.16, SD = 0.36) ($\beta = -.293, SE = .099, z = -2.97, p < .01$).

For the noun region, new inspections towards the actual target were again compared to inspections towards any other object displayed to assess if and how quickly listeners identify the object corresponding to the noun they are hearing. Data showed that participants always directed more new inspections towards the mentioned objects than towards any other object in the display while no difference in timing was detected.

Again, as previously, verb-driven anticipatory eye movements were replicated in the present experiment, even in conditions where more than one possible target object was displayed. This hints at more (when one object matched the verb) or less (when three objects matched the verb) specific anticipation of the target noun and shows that listeners equally considered all competitors rather than deciding for one of the possible targets. Whether these anticipations alter surprisal and processing effort either on the verb or on the target noun was assessed by the simultaneously obtained ICA values.

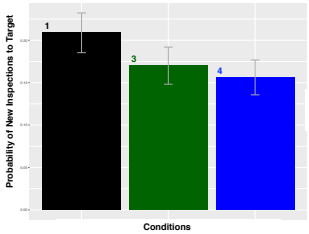


Figure 5.2 **Probability of verb-driven new inspections** of target object before target word onset in all possible conditions of **Experiment 4**.

ICA Fig.5.4 shows how the number of effort related pupil changes as indexed by the ICA is very similar on the verb is similar, while differences between conditions appear on the noun and in reaction to the manipulation. More specifically, noun processing was facilitated as fewer competitors were displayed, making the noun more predictable and less surprising.

The obtained ICA values within the pre-defined time windows of interest were treated as count variable and used as the basic dependent variable in generalised mixed effects models of poisson type. Both time windows for ICA analysis were non-overlapping and 600 ms in length, starting from the middle of the critical word's duration, as previously established for this measure (see e.g., Sekicki and Staudte (2017), Demberg and Sayeed (2016)). Differences between the conditions (e.g., **0** versus **4** competitors, i.e., "spillable" objects displayed) were contrast coded and entered into the model as fixed factors.

In line with the plot showing the ICA values, analysis of the verb window did not reveal significant differences between the conditions. This even holds for the linguistic–visual mismatch condition **0**, where the mapping of the verb information and the visual display was obstructed. The overall lack of differences in processing effort on the verb suggests that anticipatory eye movements, although verb driven, do not elicit measurable ICA differences as related to surprisal and processing effort on the word itself. Consequently, the trend observed on the verb in the previous experiment was not confirmed.

In order to examine whether listeners possibly grouped the displayed objects according to any visual or conceptual features prior to hearing the verb, namely immediately as the display appears on the screen, an additional time window of 600 ms length starting from trial onset was analysed. No significant differences were found in this region.

In the noun window, however, comparisons of ICA events between conditions revealed a significant processing facilitation (i.e., lower ICA values) if three competitors were shown (M = 19.37, SD = 8.17), compared to the unhelpful condition **0**, where none of the objects shown were potential referents (M = 20.90, SD = 8.12) ($\beta = .08, SE = .03, z = -2.32, p < .05$). Further, the same noun was significantly easier to process when the target object was most predictable, that is, in the presence of only one potential target object (M = 17.40, SD = 7.79),

Table 5.1 **Differences in ICA** for **Experiment 4**, Model: ICA values on verb/noun ~ Nr. of possible Targets + (1 + Nr. of possible Targets | Subject)+ (1 + Nr. of possible Targets | Item), family= poisson (link = "log")

Time window: Verb	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model)	(Intercept)	2.9522	.0388	76.17	<0.001
	Zero vs. Four poss. Targets	.0044	.0329	.13	.895
	Zero vs. Three poss. Targets	.0017	.0345	.05	.961
	Three vs. One poss. Targets	.0124	.0362	.34	.732
Time window: Noun	Predictor	Coefficient	*SE*	Wald Z	*p*
(Model)	(Intercept)	2.929	.0485	60.36	<0.001
	Zero vs. Four poss. Targets	.0470	.0409	1.15	.2503
	Zero vs. Three poss Targets	.0776	.0335	2.32	<0.05
	Three vs. One poss. Targets	.1186	.0466	2.54	<0.05

compared to when three competitors were displayed ($\beta = -.12, SE = .05, z = 2.54, p < .05$). Differences in processing effort between condition **3** and **4** (M = 20.13, SD = 8.45) did not reach significance. These results clearly demonstrate how different visual information directly affects processing effort for and surprisal of the same target word.

Correlating anticipatory eye movements with ICA ICA values on the noun were interpreted as to vary with respect to the manipulation of the visual domain, that is the number of depicted competitors. An alternative explanation, however, can be that each listener simply tries to find one object in the display matching the verbal constraint and then sticks to it, regardless of possible other options. In this case, the expectation would be correct with 100% certainty in case of condition **1**, with 33% certainty in condition 3, with 25% in condition 4 and simply not possible to compute in condition **0**. If this distribution is assumed to be relatively equal across participants, increased ICA values on the noun in conditions with more competitors can also be caused by the accumulated effects of wrong assumptions about the target word.

In order to test whether the first interpretation is supported, it is necessary to reliably relate facilitated noun processing as indexed by the ICA measure to actual anticipation taking place prior to hearing it, namely at the verb, as indexed by the eye movements. Hence, the statistical significance of the correlation between both measures was additionally assessed by using anticipatory glances measured during the verb of a trial as a predictor of the noun ICA of this trial. In order to do that, ICA values obtained in the noun window of each trial

were entered as dependent variable into a linear model of poisson type. At the same time, anticipatory glances towards the mentioned target measured during the verb (again, for each trial) were encoded as a binary predictor and entered as fixed factor into the model in order to assess the correlation between the two measures, on a trial-by-trial basis.

The mean number of ICA events differed between trials in which participants directed anticipatory glances towards the target referents as they heard the verb (M= 16.8) and trials in which they did not (M= 17.6), hinting at a possible correlation between anticipatory looking and processing effort on the noun (although fewer trials were recorded in which no anticipatory eye movements towards the target object were found). Including the binary variable of target inspections (present or absent) during the verb window in the model resulted in a significantly higher model fit ($\chi^2(1) = 7,25, p < .05$). The model revealed a main effect of anticipatory target glances on ICA in the noun window ($p < .05$), suggesting that their presence during the verb window can predict processing effort as assessed by the ICA values on the subsequent noun. This result supports the hypothesis that differences in the noun's processing effort are indeed attributable to the anticipation and expectancy of the word in its multi-modal context, rather than to effort related to the correction of falsified predictions.

Discussion The present experiment's manipulation in the VWP context allowed us not only to replicate verb-driven anticipatory eye movements towards matching objects (as previously found in many other studies, e.g. by Altmann and Kamide (1999)), but also to extend those results to scenes in which entire groups of objects could be considered as upcoming target referents. This strongly suggests that listeners exploited the visual information in combination with the verb to anticipate the noun with more or less certainty, thereby considering all objects matching the verb constraint. In order to support the interpretation that indeed all objects were considered by each participant to expect the target word, rather than different objects being considered by different listeners, an additional analysis tested the actual correlation between the measured anticipatory patterns during the verb and the processing effort needed for the subsequent noun. The analysis revealed that anticipatory target glances during the verb were a significant predictor of the processing effort for the noun, as reflected by the ICA events ($p < .05$), demonstrating that more specific anticipation indeed facilitated subsequent word processing.

Although eye movement data showed clear patterns of anticipation, no impact was found on the actual processing cost of the verb as assessed by the pupillary measure. With respect to the initial research question about possible mechanisms of referential entropy reduction affecting the verb's processing effort, as inspired by the observed trend in the previous study, experiment **4** clearly demonstrated that no such effects could be measured in this setup.

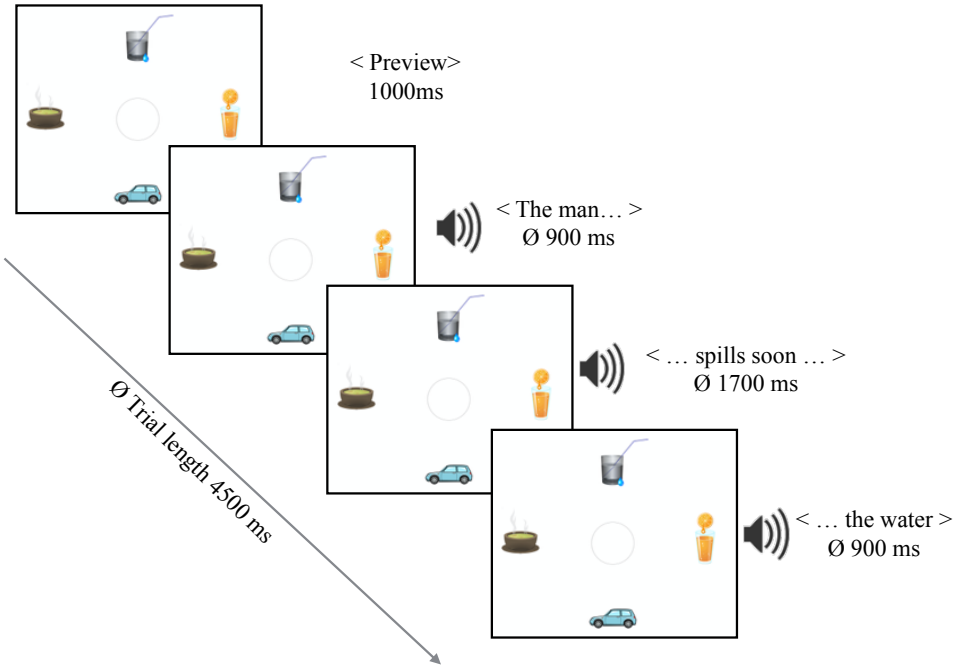


Figure 5.3 **A trial timeline example** for **Experiment 4**. The example scene shows three possible target referents.

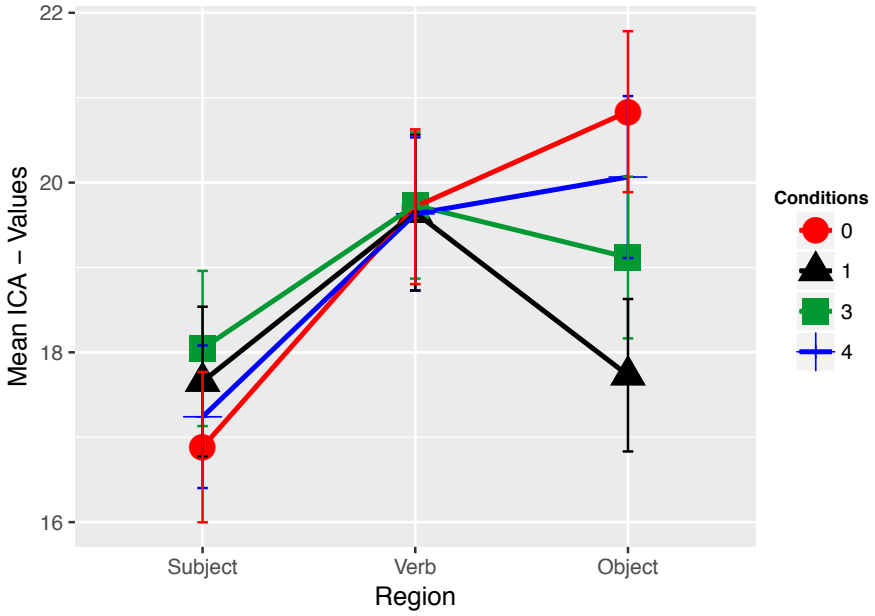
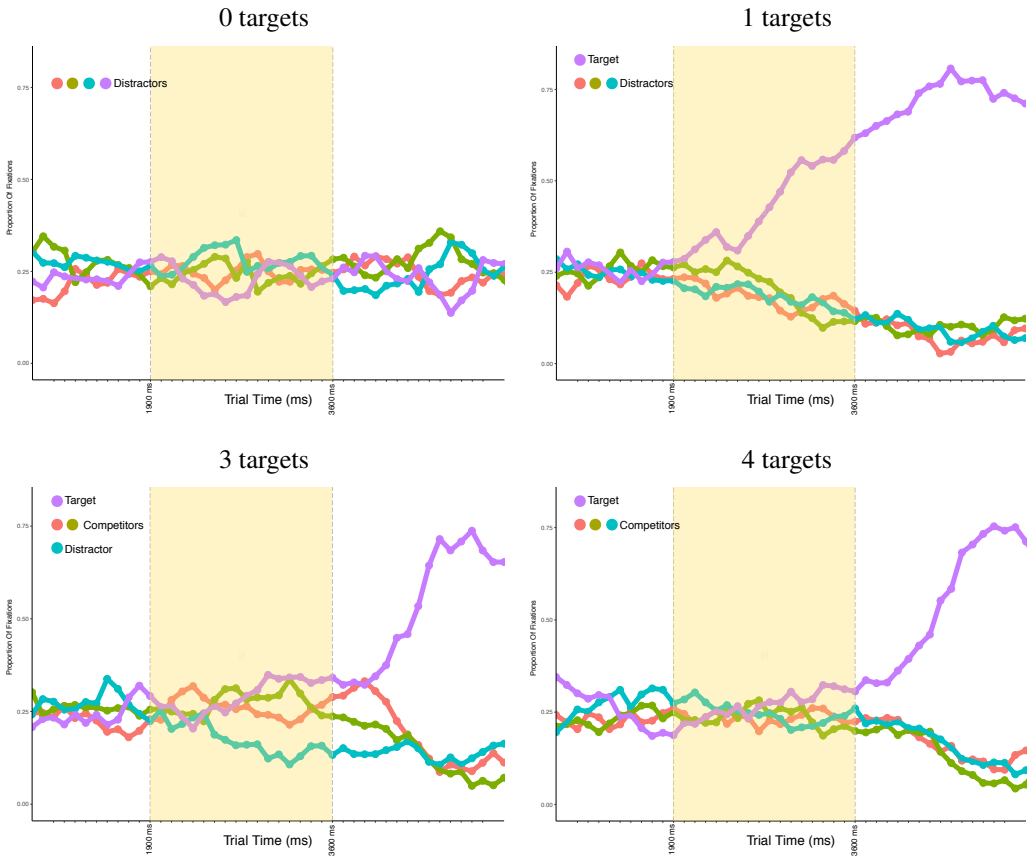


Figure 5.4 **ICA Results** for **Experiment 4** in all conditions. Error bars reflect 95% confidence intervals (CI).

That is, although participants anticipated the target noun, the verb (e.g., *spill*) took equal effort to process, no matter how many competitors ("spillable" objects) were displayed. This was even the case in condition **0**, where the eye striking incongruence between the visual and linguistic information (nothing "spillable" was displayed) did not cause significantly increased processing cost for neither the verb, nor the noun, although no visual information could be used to prepare the listener for the words. This shows that the detection of a modality-mismatch possibly elicits effort of a different quality, that is, effort with respect to something distinct from information processing which is reflected in the ICA in case of the presented experiments. If this is the case, it is possible that the mismatch and the resulting obstruction in mapping visual and linguistic information is visible in a different measure, for instance, in the ERP component N400, which is well known to be sensitive to the detection of incongruence.

Overall, the combined results from eye movement and pupillary data is interpreted as reflecting listeners shifting attention towards possible target objects based on the verb information (as clearly indicated by the eye movements) but possibly refraining from deciding on an ultimate exclusion of distractors without any further evidence or hints.

Figure 5.5 **Proportion of Fixations** across trial length in all conditions of **Experiment 4**. Each line corresponds to one of the four pieces of clip art in the visual displays



Since not only the anticipatory patterns in the eye movement, showing that listeners focused more on competitors, but also the differences in ICA on the noun in reaction to the number of competitors can be interpreted in support of the hypothesis that listeners do in fact reduce entropy at some point between the verb and the actual noun, an alternative explanation can be that such a reduction does either not induce any additional effort, or that the ICA measure is not sensitive towards this sort of effort.

Results from Maess et al. (2016), revealing enhanced activity for highly predictive compared to less predictive verbs in a different measure, namely in the MEG, along with a matching correlation between verb constraints and the noun's N400 amplitude might support the latter. Whether possible effects on processing effort attributable to the reduction of uncertainty in order to expect target words can be assessed by other measures needs further testing.

With respect to the second initial question of whether visual information alone could (systematically) affect word surprisal and processing, ICA values measured on the noun clearly showed positive results.

More specifically, ICA values differed between conditions **0**, **3** and **1**, which strongly suggests that visual context information clearly affects the statistics of the mental model and hence expectations and processing effort for the target word, as describable by multi-modal surprisal.

These effects can be interpreted as being analogous to the number of competitors, that is, the probability of a target object in the visual display to correspond to the actual target word coming up: If only one possible target was shown, the noun itself was least surprising and easiest to process, as the displayed object would correspond to the noun with 100% certainty (condition **1**). The same noun was more surprising when three competitors were shown and this correspondence was only 33% certain (condition **3**).

At the same time, the very close conditions **3** and **4** did not differ significantly from each other. This lack of a significant difference may be attributable to a lack of power, ant to condition three and four being too similar to result in measurable differences. The interpretation would be perfectly in line with recent literature suggesting that expectations or even specific predictions in language processing are probabilistic and at multiple levels (see e.g. Kuperberg and Jaeger (2016b)), by showing that patterns of probabilistic expectation can also be found if visual context has to be evaluated. Results would then also provide additional proof that surprisal is an appropriate predictor of processing effort, as suggested by Frank (2013b), even in multi-modal contexts (i.e. situated language processing), hereby emphasising the overall broad importance of rational approaches such as information theoretical measures in language science (especially in sentence processing).

Alternatively, the lack of differences between **3** and **4** can be a symptom of the overall results rather being interpretable as to reflect a way less fine-grained decision of whether one, many, or none of the objects matched the verb, while no further evaluation of the context is performed by the listeners.

In sum, findings from this experiment are interpreted as robust evidence for the significant effect of visual context information on (linguistically identical) processing effort for the target word, which (up to our best knowledge) has never been measured and quantified before. However, it leaves several questions open which will be closer examined in the subsequent experiment, using a different, more established measure. Specifically, an electrophysiological measure might help to answer remaining questions with respect to processing effort as reflected by the ICA. It is further suitable for testing whether the ICA results can be validated or even extended in a more established – and possibly more sensitive – measure.

5.4 Experiment 5: Number of competing (potential) referents in the EEG

The previous study revealed significant differences in processing effort, as assessed by the deployed pupillary index. Results suggested a (more or less detailed) correlation between processing effort needed for a word and the probability characteristics of the word's multimodal context. While eye movements on the verb clearly revealed mapping of linguistic and visual information, followed by expectations about the target word, it remained unclear whether listeners evaluated the multi-modal context in a detailed, probabilistic, or rather in a less detailed "one-many-nothing" way, without caring further about probabilities.

While in both cases, results provide clear evidence for the direct and significant influence of visual information on word expectation, which has not been shown before, the case of detailed probabilistic evaluation would additionally demonstrate how detailed surprisal can be extended to describe data from situated language comprehension in visual contexts.

The following experiment is an important step in validating and extending previous eye movement and pupillary results. Event-related potentials (ERPs) were used as a dependent measure that has already been well established in the literature, while experiment design and task were identical to experiment **4**.

ERPs are not only more frequently used as a measure, but also allow for deeper, electrophysiological insights into different cognitive domains involved in meaning construction as they are a direct measure of activity in the neocortex, continuously reflecting different brain states during cognitive processing. In the present setup, they were hence used to examine

whether previous results can be replicated, or even extended in a different measure, as well as whether any additional differences in processing effort on the verb would occur in the potentially more sensitive method.

Results can additionally contribute to the extension of current (usually language-centric) ERP finding regarding surprisal (e.g., Frank, 2013b) to also take visual information into account, showing the direct effect of co-present visual information on the expectancy and the related processing difficulty of linguistic material.

5.5 The N400 and multi-modal information

Many language comprehension studies deploy ERP measures as an index of the brain's electrical activity during processing of usually linguistic input. While recent literature suggests possible correlations between the ICA measure and the ERP components P3 (more specifically, its subcomponent P3b) and P6 (Demberg and Sayeed, 2016), which are also thought to be dependent on the neuromodulator norepinephrine (Sassenhagen et al., 2014), a different component was used here, namely the N400. This decision was based on the fact that a general possible correlation between the component and the ICA index was not the focus of the study. Instead it should be tested whether an additional and more established measure for word surprisal is also sensitive to the influence of visual information on word expectancy, thereby validating and reinforcing the relevance of the measured ICA results. It is, however, *not* argued that the N400 and the ICA index are in general similar or equivalent.

Here the ERP component N400 was specifically chosen as dependent measure, as it is known to reliably show a reduced neural signal – or amplitude – in reaction to semantically predictable as opposed to unpredictable words in linguistic contexts, hereby indexing the difficulty of processing those words (Kutas and Hillyard, 1980). In general, the N400 is a negative-going deflection in the average ERP, peaking at approximately 400 ms after stimulus onset. It distributes globally across the scalp and reaches maximal amplitudes at midline centro-parietal sites if it is caused by linguistic stimuli.

Originally, the component was discovered and interpreted as a reaction to very specific stimuli, namely as the electrophysiological response to anomalous or violating information, mediated by semantically incongruent *sentence-final* words in reading tasks (Kutas and Hillyard, 1980). However, the use of the N400 quickly evolved, revealing additional sensitivity, for example to the congruency of single word pairs, the frequency of words, and even to semantic context effects in different stimulus types, including mathematical symbols, signed language and visualised words (Kutas and Federmeier, 2011).

Especially the latter is interesting in the context of the present study. Results from Ganis et al. (1996), who presented participants with written sentences either ending with normal written words, or with a line drawing instead of the word: The authors found that the N400 elicited by the drawings was very similar – although more frontally-distributed – to the one elicited by the actual words (see also Kutas and Federmeier, 2011). This clearly suggests that N400 effects of cognitive processing can indeed be generalized across different input modalities.

Further, the finding that the N400 is sensitive to the cloze probability of a word in its linguistic context (i.e., irrespective of contextual constraint) already hinted at a correlation of the measure with probabilities and expectancy (Kutas and Federmeier, 2011). In addition to these results, it has also been shown that the component can indeed be affected by surprisal manipulations on the word level. For example, Frank (2013b) examined possible correlations of surprisal estimates, as derived from Markov models, probabilistic phrase-structure grammars, and recurrent neural networks with four different ERP components during sentence reading. The authors found that surprisal values derived by any of the tested model types were a significant predictor of *only* the N400 component's amplitude (this correlation appeared to be strongest for content words), hereby showing that word surprisal is not only predictive of word-reading time (Smith and Levy, 2013) but also affects EEG components, making surprisal a reliable index of processing difficulty.

It is hence argued here that surprisal, as derived online from integrating language and additional (as opposed to *substituting*, as in Ganis et al., 1996) visual context, can also predict the N400's amplitude. That is, surprisal estimates considering combined linguistic and visual context, and already shown to affect processing effort as indexed by the pupillary measure are hence expected to affect the nouns' N400 amplitudes.

If differences in the nouns' processing effort can also predict the N400's amplitude, findings would validate previous ICA results and strengthen the correlation between (linguistic) processing effort and (multi-modal) probabilities, thereby supporting the idea that language and perception are parts of a dynamic interaction, rather than being independent modules (Spivey et al., 2009).

5.6 Method

In order to keep results comparable, this study used the same manipulation and type of stimuli (plus additional Stimuli to increase power) as in the previous experiment, while deploying ERPs as a well-established measure in (predictive) language processing. This time, 96 linguistic stimuli were used, each again paired with four different visual displays.

An equal amount of fillers was added, before all sentences were parted into four lists and randomized as before. Conditions and task were identical to the ones used in **Experiment 4**. The only differences were the written presentation of the sentences, and the now quadrangular arrangement of the four pieces of clip art around the center of the screen with an added sepia filter, in order to exclude disadvantaged positions on the vertical axis and to tone down salient colours.

36 right-handed native speakers of German (*M* age: 25, 22; Age range: [19, 34]; *SD*: 3.47; Female: 29) with normal or corrected-to-normal vision took part in the ERP experiment. 8 participants were removed from the analysis due to more than 20% of their data being influenced by eye artefacts. All Participants gave informed consent and were monetarily reimbursed for their participation.

Visual displays were again presented with a 1000 ms preview time, in which participants were allowed to move their eyes around in order to identify and inspect the clip art items. As soon as the fixation cross appeared for a jittered duration (to prevent any effects of habituation with respect to the length of the cross' display time) in the middle of the display, participants were asked to keep their eyes focused on the cross and the phrases presented subsequently. They were further instructed to prevent blinking throughout the sentence. Sentences were presented in phrases and in the centre of the screen with a presentation period of 400 ms and a 100 ms ISI. The visual displays stayed on the screen for the entire trial time. Subsequent to the sentences, the visual displays disappeared and a question appeared on the screen. Questions were presented after each trial and either concerned the visual (e.g.*Was the milk on the right?*), or the linguistic content (*Did the man spill the milk?*). Subjects were asked to answer on a button press. The correspondence of the left and right button to Yes and No was alternated. Answers were recorded using a Cedrus Response Pad RB-834 (Cedrus Corporation). All stimuli were presented using the E-prime software (Version 2.0.10.353 Psychology Software Tools, Inc.). Participants were seated in a quiet environment in front of a 19" Dell 1908FP TFT UltraSharp monitor (resolution of 1280x1024 with a refresh rate of 75 Hz). The distance between the participant and the screen was always 103 cm in order to keep all of the objects in a 5° visual angle from the center of the screen in order to minimize eye movements throughout the experiment. Prior to the actual experiment, participants were presented with written instructions and completed five practice trials. The experiment lasted approximately 55min.

Predictions If the N400 is sensitive to visually influenced surprisal and processing effort, similar differences in the N400's amplitude were expected as previously found in the ICA measure. Additionally, differences in amplitude were expected on the verb if the N400 is

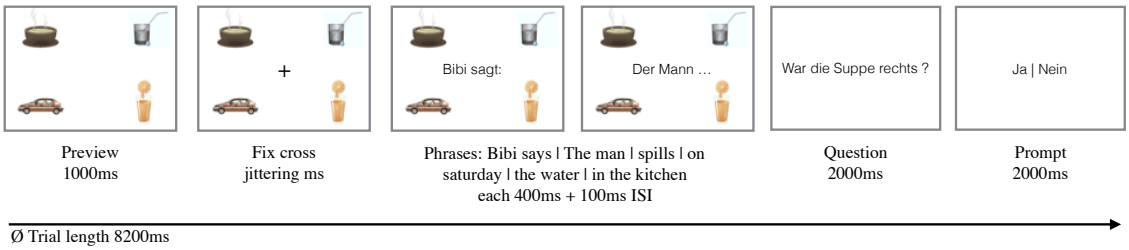


Figure 5.6 **A trial time line example**. The example scene shows three possible target referents.

more sensitive to effort induced by anticipation, or, reduction of referential uncertainty for that matter, compared to the pupillary measure. Specifically, a higher N400 amplitude was expected as the verb constraint excludes more distractors from the display as possible targets.

Electroencephalographic recording and processing parameters The EEG was recorded by 24 Ag/AgCl scalp electrodes embedded in a cap (acti-CAP, BrainProducts) and amplified with a BrainAmp (Brain- Vision) amplifier. Electrodes were placed according to the 10-20 system (Sharbrough et al., 1995). The cap was positioned by placing the Midline Central electrode (corresponding to Cz) at 50% of the nasion-inion distance, and at 50% of the distance between the mastoid processes. The electrodes were referenced online to the reference electrode (FCz) and later re-referenced offline to the average of both mastoid electrodes. The ground electrode was located at AFz. Horizontal eye movements were monitored via bipolar recording of the electrooculogram (EOG) with an electrode on the outer canthus of the left and right eye. Blinks (Vertical EOG) were monitored with electrodes over the supraorbital and the infraorbital ridge of the left eye, referenced to the left mastoid. Electrode impedances were kept below 5 kΩ. The signal was sampled at 500 Hz, using an anti-aliasing low-pass filter of 250 Hz online during recording. Data were later band pass filtered offline at 0.01-40 Hz (Luck, 2014). All records were semi-automatically examined and marked offline for EOG and other artefactual contamination such as electrode drifts, amplifier blocking and excessive muscle activity. Artefactual trials were excluded with a rejection threshold of 20% per condition for participant rejection, resulting in the exclusion of 8 participants from the analysis. No additional participant was removed due to the behavioural data (with a threshold for participant removal of 15% wrong answers).

5.7 Results and Discussion

Single-participant data were averaged for each of the four experimental conditions within 800 ms windows from the onset of the verb and the target noun. The segments were aligned to a 200 ms pre-critical baseline and data sets for both critical time windows were then exported from the averaged ERP data, using BrainVision Analyzer's (Version 2.1) Area-Information export function. The grand average of all participants was then analysed, time-locked to the onset of the critical words. All analyses were conducted using the *ez* package for R, to perform repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser corrected p-values. In addition, F-values, as well as η^2 (generalised eta-squared, see Bakeman, 2005) values are reported as a measure of effect size. ANOVAs were performed on data sets including the Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO9, PO10, O1 and O2 electrodes, including ROIs for frontal (F3, Fz, F4), central (C3, Cz, C4) and posterior (P3, Pz, P4) distributions. We analysed a typical N400 time window between 300 and 500 ms after onset of the verb and noun. Main effects were assessed by running omnibus ANOVAs with electrode site (frontal/central/parietal) and experimental condition (number of competitors matching the verb) as within factors.

Fig. 5.7 shows how only the mismatch condition **0** elicited an increased negativity at 400 ms after verb onset.

In the noun region, all four conditions elicited a modulated ERP response to the more or less predictable target word (see Fig. 5.8). That is, the N400, peaking at 400 ms after onset of the critical word, differed in amplitude between conditions, although the linguistic context never changed and only the visual display varied.

The model assessed the statistical significance of these effects, revealing a main effect for condition ($F(3.81) = 8.18, p < 0.05, \eta^2 = 0.06$) on the verb. Follow-up pairwise comparisons showed that significantly larger negativity was elicited by condition **0** ($-1.34\,\mu V$), compared to the baseline condition **1** ($-0.69\,\mu V$) ($F(1.27) = 8.49, p < 0.05, \eta^2 = 0.06$). Negativity was widespread across frontal, central and parietal regions, while being largest in the latter. However, conditions **3** ($-0.75\,\mu V$) and **4** ($-0.4\,\mu V$) did not yield significant differences in the N400 component, compared to **1**, suggesting a binary evaluation of whether the visual display matched the verb, rather then more detailed distinctions of the displayed options.

In the noun window, a more detailed effect was found. Namely, further analysis of the significant main effect of condition ($F(3.81) = 7.74, p < 0.05, \eta^2 = 0.13$) revealed that condition **1**, in which the noun was most predictable, resulted in the lowest N400 amplitude ($0.07\,\mu V$). Conditions **3** ($-0.8\,\mu V$) and **4** ($-0.79\,\mu V$), where the target noun

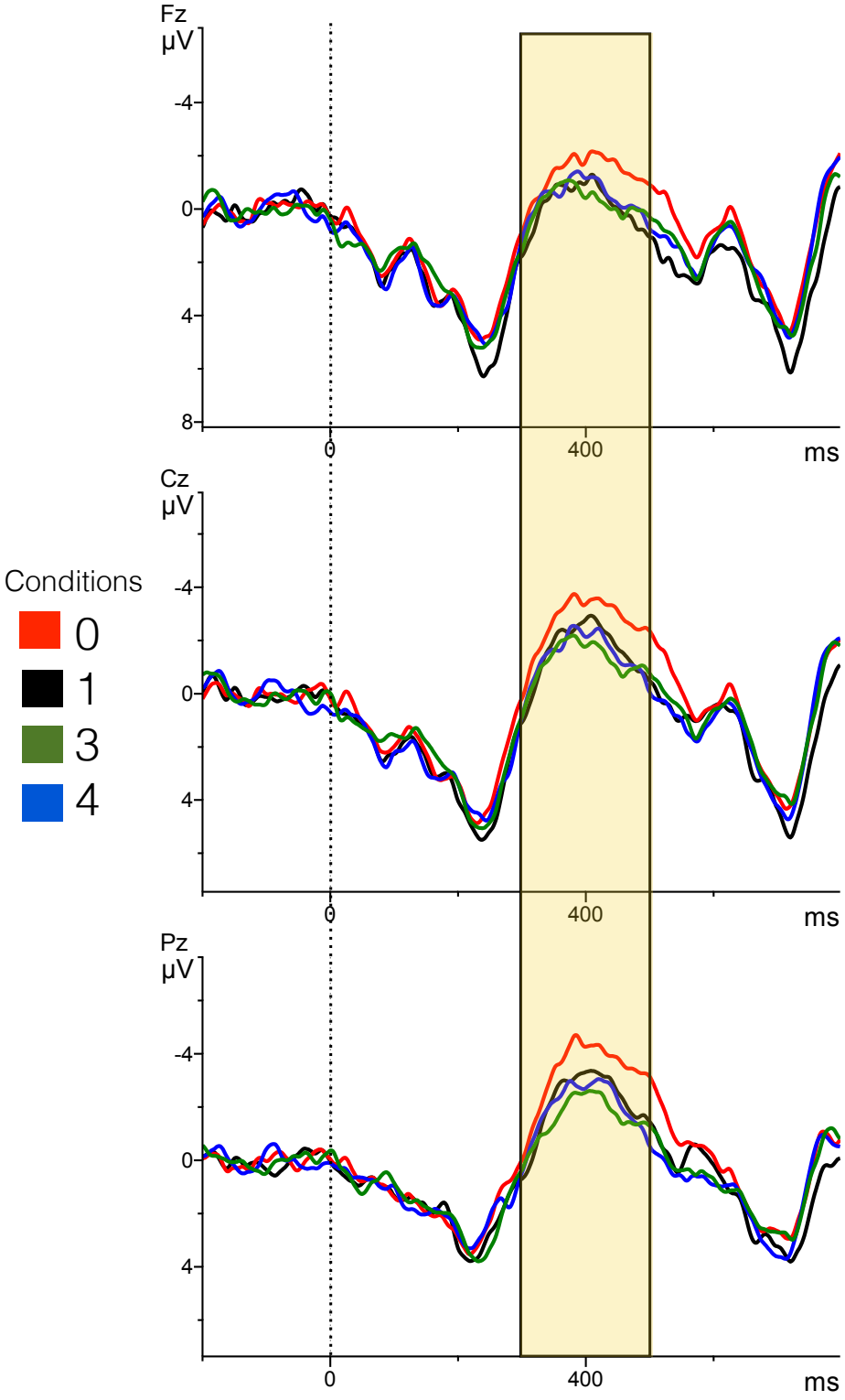


Figure 5.7 **ERP time-locked to the onset of the verb** (dotted line) and separated by the experimental conditions. The reported region is highlighted. The data shows the electrode subset Fz, Cz and Pz (unfiltered) for presentation purposes only.

Table 5.2 **N400 amplitude differences** for **Experiment 5**, Model: *ezANOVA (dv = N400 value in each time window, wid = Subject, within = Targets, region)*

time window: Verb	*Predictor*	*F-value* (DFn, DFd)	eta^2	*p* (GG corrected for overall)
Overall	Targets	8,18 (3,81)	.06	< .05
Follow up	One vs. Zero poss. Targets	8.49 (1,27)	.056	< .05
Follow up	One vs. Three poss. Targets	.001 (1,27)	4.00	> .05
Follow up	One vs. Four poss. Targets	1.78 (1,27)	.007	> .05
time window: Noun	*Predictor*	*F-value* (DFn, DFd)	eta^2	*p* (GG corrected for overall)
Overall	Targets	7.74 (3,81)	.13	< .05
Follow up	One vs. Zero poss. Targets	20.68 (1,27)	.215	< .05
Follow up	One vs. Three poss. Targets	9.92 (1,27)	.11	< .05
Follow up	One vs. Four poss. Targets	7.93 (1,27)	.096	< .05

could be expected with 33% and 25% certainty, resulted in a significantly higher amplitude (three: $F(1.27) = 9.92, p < 0.05, \eta^2 = 0.11$, four: $F(1.27) = 7.93, p < 0.05, \eta^2 = 0.096$). Condition **0** ($-1.3\,\mu V$), where none of the clip art items in the visual display could be used to predict the target noun, yielded the highest difference in the N400 amplitude, compared to **1** ($F(1.27) = 20.68, p < 0.05, \eta^2 = 0.215$).

Discussion This experiment was designed to investigate whether the N400, which is known to be sensitive to linguistic probabilities (and hence surprisal) as well as to meaningful stimuli in different modalities, is also sensitive towards target word expectancy and surprisal, when probabilities have to be derived by combining linguistic and visual information, hereby validating and possibly even extending previous ICA results. That is, it was further observed whether the component is equally or even more sensitive with respect to surprisal-based processing effort or to effort induced by the reduction of visual uncertainty (as reflected by eye movements in the previous experiments) on the verb.

Results from the EEG experiment indeed revealed a clear, globally distributed ERP response in reaction to the same words, presented in different visual contexts: On the verb, the mismatch condition **0** (where non of the objects matched the verb) elicited a significantly increased N400 amplitude, as compared to the other three conditions. Interestingly, this effect had not been reflected by the ICA in the previous experiment **4**. In other words, the combination of these two results highly suggests that the ICA does not reflect a mismatch

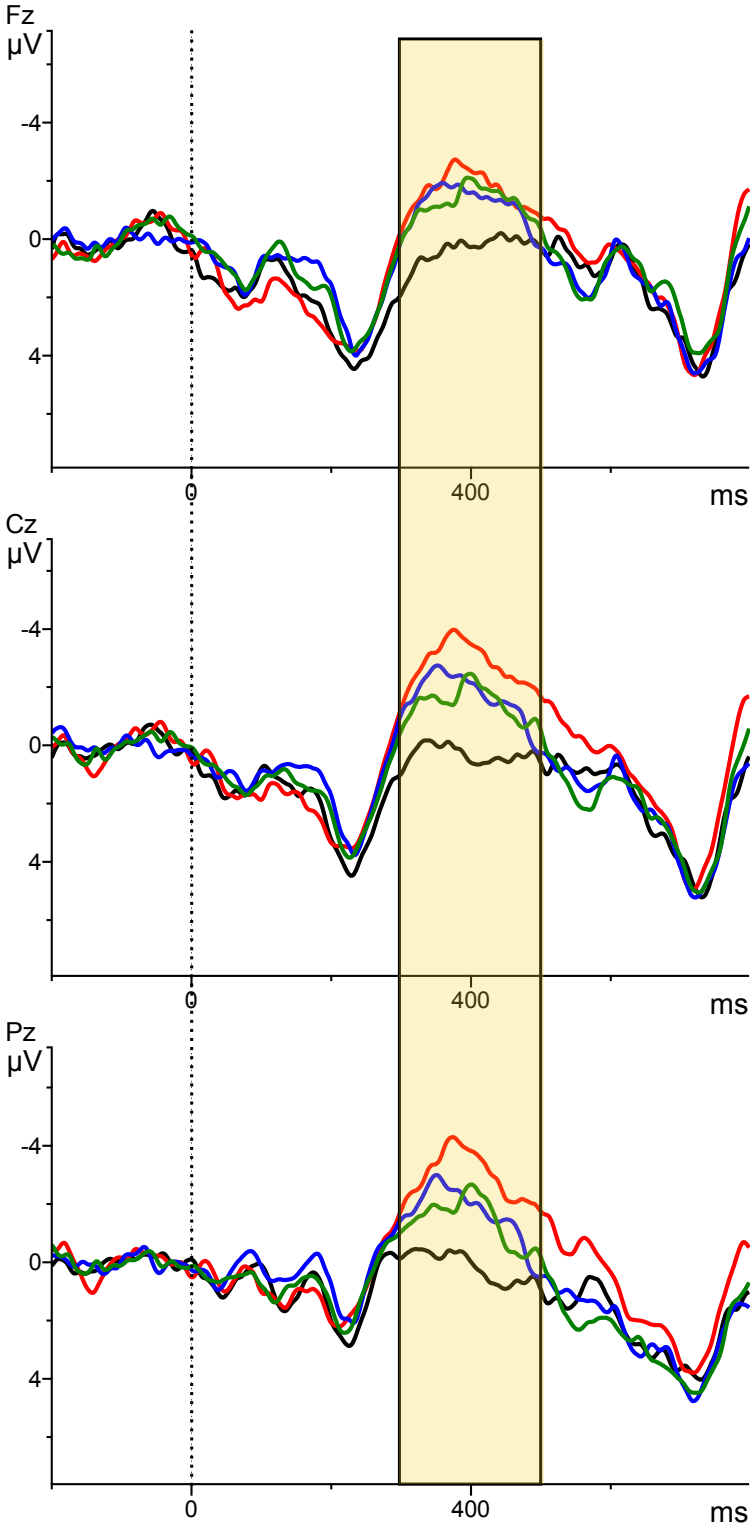


Figure 5.8 **ERP time-locked to the onset of the noun** (dotted line) and separated by the experimental conditions. The reported region is highlighted. The data shows the electrode subset Fz, Cz and Pz (unfiltered) for presentation purposes only.

between linguistic *and* visual information, which shows that a very careful interpretation of effort is required with respect to its source and that the direct comparison of the two measures is intricate. At the same time, the verb in the remaining three conditions, that is, the same verb, although being more or less informative in the visual contexts showing either **1**, **3** or **4** competitors, was equally difficult to process, eliciting highly similar N400 amplitudes. First of all, this pattern on the verb again shows no effects attributable to the reduction of (visual) uncertainty, as also previously measured in the pupillary index. We hence conclude that the lack of entropy reduction effects was not simply due to the ICA measures insensitivity. Instead, it is possible that either the verbal constraint information was not sufficient for participants to exclude target options, possibly because exclusion would be too effortful or risky, or that the exclusion process does not elicit effort visible in *either* of the two measures, namely the ICA *or* the N400.

What we found instead of an entropy reduction effect was a binary pattern (*something* matches or *nothing* matches), which clearly suggests an instant integration of visual and linguistic information, as previously reflected by the eye movements. This inherently means that no additional effect on (linguistic) processing effort appears, as long as the integration of visual and linguistic information is not obstructed by an obvious mismatch. This result is indeed surprising, since again, results from the subsequent noun window showed differences in processing effort relatable to target word expectations based on a previous context evaluation. More specifically, noun window results validated the differences in processing effort for the target noun in reaction to the manipulation, as previously found in the ICA index. An additional analysis, testing whether the verb N400 is a valid predictor of the noun N400 showed no significant results on a trial by trial basis, but for the averaged data in all conditions. Again, conditions **3** and **4** did, however, not differ from each other. Instead, their amplitudes were even closer together than the respective ICA values from the previous experiment. This could be explained by the nature of the N400, which, as mentioned previously, is *not* thought of as being directly comparable, or even equivalent, to the pupillary index. The similarity between conditions **3** and **4** could also – at least partly – be caused by the lack of eye movements in the EEG experiment, making a detailed discrimination between 3 and 4 objects more difficult. Most likely, however, the null result with respect to differences between both conditions is simply caused by the low numeric difference between the presented competitors, that is, by a lack of power. Based on the EEG results, it is hence still not possible to clearly decide whether comprehenders evaluate the multi-modal context in a probabilistic way, or simply decide whether one, more than one or no objects in the display match the verb. Therefore, two main questions result from this experiment: 1) are expectations in situated comprehension probabilistic or rather coarse grained (i.e., is the

lack of a difference between conditions 3 and 4 attributable to both conditions being too similar, or, in other words, numerically too close together) and 2) What is the actual role of overt attention, as reflected by anticipatory eye movements in context evaluation for target word expectations (i.e., are they necessary for a fine grained, probabilistic evaluation of the multi-modal context)? The subsequent chapter hence describes follow up ICA experiments exploring these questions. More specifically, we first increased the numerical difference in competitors between conditions in order to test whether expectations would show to be probabilistic in those contexts. Secondly, we ran the identical experiment again, asking participants to keep their eyes fixated during the trials in order to mimic the circumstances given in the EEG experiment in order to closely observe the effect of overt attention on probabilistic expectations and context evaluation.

In sum, the EEG results presented in this chapter clearly emphasize not only the reliability but also the importance of the visual context effects on word processing by showing that those effects are measurable in at least two independent measures, affecting not only pupil dilations but also the N400. This also bears implications for the use of visual stimuli experiments observing language comprehension by showing how strongly pictures can influence and shift linguistic expectations. Specifically, our EEG results additionally imply that the N400 component is affected not only by a word's expectancy in purely linguistic context, but also by *combined* multi-modally derived expectations in visual contexts. These findings extend the suggestions of Frank (2013a), by showing that surprisal can not only be used as an appropriate predictor of the N400's amplitude in reaction to purely linguistic information, but it can also account for data from situated language comprehension, bearing implications for surprisal's extended possibilities to quantify hypotheses as well as for its explanatory potential.

Chapter 6

Fine grained expectations: The role of visual complexity and eye movements

The two experiments presented in the previous chapter replicated and extended findings by Altmann and Kamide (1999), showing that comprehenders shifted attention towards objects matching the verbal constraints upon hearing the verb, in anticipation of the target noun, even when more than one possible target option was shown. At the same time, no effects on processing effort, relatable to a reduction of (visual) uncertainty, were found during the verb. On the target noun, however, differences in processing effort for the *same* noun were measured, attributable to the only manipulation, namely, the different numbers of displayed competitors matching the verb, and hence, the statistical properties of the multi-modal context. While this demonstrated the important direct effect of visual context on actual processing effort for the target word, it cannot entirely clarify whether participants had fine grained *probabilistic*, or rather rough "one-many-nothing" expectations about the target nouns after mapping visual and linguistic information. More specifically, although processing effort – as reflected by a pupillary and an ERP measure – differed, depending on whether none, one, or more than one pieces of clip art matched the verb, no fine grained differences between three and four competitors were found in either measure.

Specifically, anticipatory patterns in the eye movement data recorded in *Experiment 4* rather suggest that comprehenders shifted attention towards *all* possible competitors shown, instead of finding one or more than one matching object without caring for further evaluation. It is therefore reasonable to assume that the difference between three and four competitors (i.e., 33% and 25% certainty with respect to the actual target word) was indeed too small to significantly affect processing effort for the noun. In case of the ERP experiment, the lack of eye movements could additionally aggravate a fine grained context evaluation, making probabilistic expectations too effortful. We hence concluded that two main questions

arose from the results, namely, 1) whether target word expectations would prove to be probabilistic, rather than coarse grained when the numerical difference in competitors was increased between conditions, and, 2) whether overt attention as reflected by anticipatory eye movements was essential for the fine grained evaluation of the multi-modal context and, hence, for probabilistic target word expectations. Here, we hence describe a total of three experiments, designed to target the two questions resulting from the previous chapter. Again, the ICA was deployed in order to be able to best observe the role of free eye movements while assessing processing effort on-line (which is more complicated in the EEG), while similar noun results could be expected as shown previously. All following studies featured conditions with an increased numeric difference in competitors, in order to observe whether the numeric similarity between the previous conditions three and four can account for the result. Here again, the same linguistic stimuli, design and task as before were used in order to keep results comparable.

6.1 Experiment 6 : Fine grained expectations in increased visual complexity

This study uses displays with increased visual complexity in order to achieve a larger numerical differences between the numbers of competitors in the different conditions, while still counterbalancing the experimental items. Data resulting from this design can hence contribute to answer the question whether visual information really affects the statistics of the mental model in a probabilistic way, or, in other words, whether comprehenders had fine- or rather coarse-grained expectations about the target word in the presence of visual information.

Importantly, however, the increased complexity of the visual scenes (from four to eight objects) alternatively could have a different effect: Namely, it can possibly lead to an overall decrease in expectations for the target word due to it not being feasible, or not being the most efficient strategy anymore. Complexity (as closely related to time) could therefore be another factor majorly affecting expectations in the VWP.

In fact, visual displays in VWP studies are generally of varying complexity, spanning from single or few very simple objects Smith et al. (2013) to complex scenes, even depicting events Coco and Keller (2015). While it has already been shown that preview time can play an important role in the generation of predictions, studies rarely observed the role of visual complexity for (detailed) expectations or predictions in language processing.

Especially in the light of the idea that (detailed) expectations might highly depend on their computational efficiency and utility in a given context (see e.g., Kuperberg and Jaeger (2016b)), it is, however, highly appropriate to consider the complexity of the visual information as a possibly influential factor on how detailed expectations are. Indeed, Ferreira et al. (2013) conducted a line of VWP studies, considering the timing and complexity of visual information. Results showed that classic garden-path effects in the VWP only occur when the scene was not too demanding. That is, given an appropriate preview time, and, more importantly, not too many objects shown in one visual scene. This entails that the use of visual context in order to expect linguistic content (and to narrow down possible interpretations computed on-line) was significantly aggravated or even inhibited by (even moderately) increased context complexity. At the same time, the authors found no correlating decrease in accuracy of overt attention, compared to the contexts with fewer objects. This suggests that useful information was still obtained from the complex displays. The authors interpret these results as adding to a growing body of evidence in favour of flexible and adaptive, as opposed to fixed, language processing strategies (e.g., Kleinschmidt and Jaeger (2016), Kuperberg and Jaeger (2016b)).

With this background in mind, we again deployed the pupillary measure of effort (ICA) in the VWP in order to additionally observe if increased visual complexity revealed clear signs of a probabilistic influence of visual information on the mental model, or even at one point caused a decrease in target word expectations. Results can hence answer open questions resulting from our previous data and can further extend insights into flexible mechanisms of language processing by revealing the actual effect of (increased) visual complexity on processing effort.

Visual Stimuli Validation All visual stimuli used in the experiment were pretested for naming and verb relatability beforehand. The complete displays were presented in four randomized lists and in the same way as they were planned to appear in the experiment, using a web form. 30 people participated voluntarily and were asked to spontaneously decide whether or not an *object* was *"verb-able"*, filling the matching object's names into pre-defined blanks. Unique participation was controlled for. All experimental items used in the actual experiment were well relatable to the verb they were presented with (> 90% correct answers per item).

6.2 Method

This study features increased visual complexity in order to test whether a greater numerical difference in competitors between conditions could reveal clear effects of probabilistic of coarse-grained statistical effects of visual information. Alternatively, the increased number of clip art pieces in each visual scene could lead an overall decrease in expectations, and possibly, to listeners changing language comprehension strategies, due to hampered mapping of visual and linguistic information.

As previously done, the same sentence was paired with four different visual displays in a 1 x 4 design. This time, however, visual complexity was increased by showing 8 objects in each display, of which either **1**, **2**, **4**, or **7** were potential target referents (see Fig. 6.1, from left to right and top to bottom).

The items were counterbalanced. That is, condition **1** was condition **7** for another sentence, as well as **2** at the same time served as condition **4** for another sentence (in those cases, we had to add two unrelated distractor objects from a different category in order to have eight objects per display). The pieces of clip art were now arranged quadrangular around the screen center. Sentences were again presented auditorily and always simultaneously with the visual displays. In compensation of the increased number of clip arts in the visual displays, comprehenders were given an increased preview time of 2500 ms before onset of the audio.

The same 40 item sentences were used for four conditions as previously, combined with 156 new visual displays. The stimuli were again mixed with 40 filler sentences, combined with 40 displays and parted into four lists, using Latin square design.

Again, each item was followed by yes/no comprehension questions, concerning either the position of a clip art in the visual display (e.g.: “Was the water on the right?”), or the utterance (e.g.: “Did the man spill the lemonade?”) to keep participants focused. Questions could be answered by button press on a keyboard (Model: “Cherry G230"). 32 students of Saarland University, that were all native speakers of German (*M* age: 23.7; Age range: [19, 40]; *SD*: 3.86; Female: 27) were tested and monetarily reimbursed for their contribution.

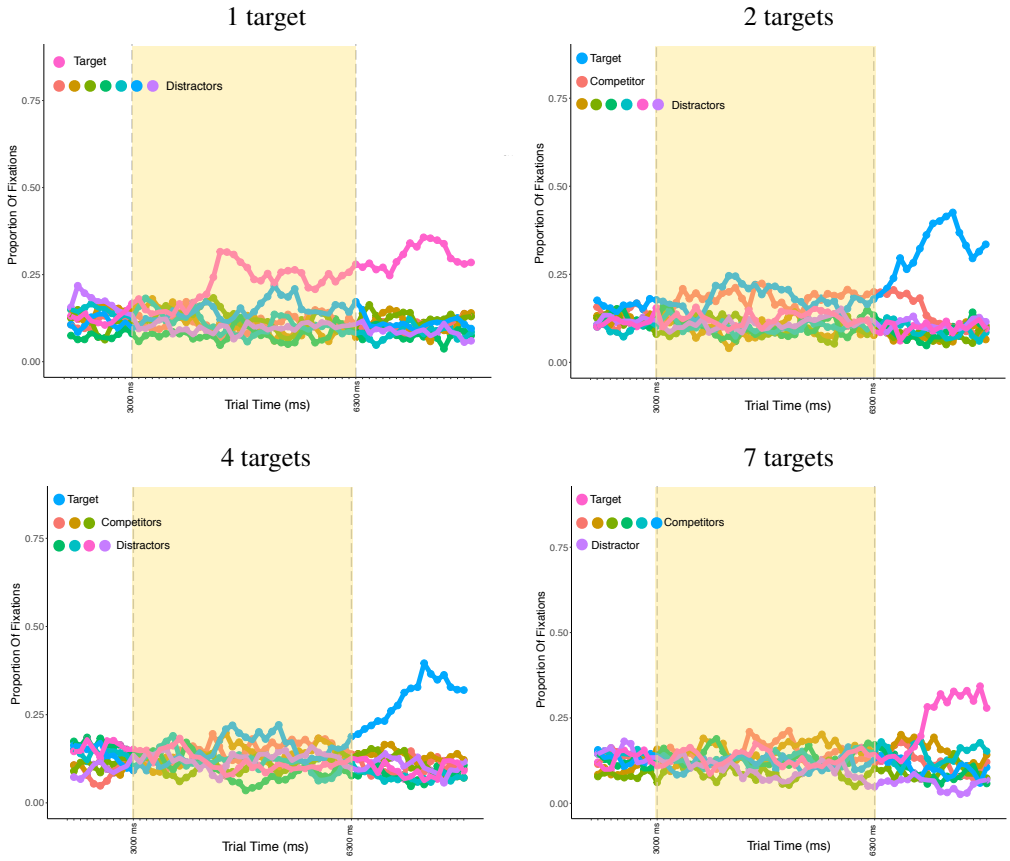
Predictions If visual information indeed has a probabilistic effect on the statistics of the mental model, and hence the predictability-dependent processing effort for a word, the design of the present experiment should reveal clear differences between conditions due to the increased numeric difference in competitors matching the verb. That is, while previously no effects on processing effort were measured between conditions showing three versus four competitors, the following study increases differences between conditions to two, four, and finally seven competitors from a total of eight objects in each display. If the lack of an effect

Figure 6.1 **Example Stimuli** for **Experiment 6**. From left to right and top to bottom: **1**, **2**, **4**, and **7** possible targets, given the sentence *"The man spills soon the water."* (Numbers were not depicted in the experiment). For more items see Appendices A and B

between three and four competitors was due to the small numeric difference between the conditions simply being too weak to be measured, the present study should increase the differences in effort between conditions, therefore possibly revealing a clear, probabilistic effect of visual information on the processing effort for the noun. At the same time, increased gazes towards target competitors were again expected during the verb. Again, as in line with previous results, as well as with data from Ferreira et al. (2013), more eye movements to the target than to any other object were expected as the actual target noun is heard, if substantial information can be gained from the visual displays, despite their complexity.

However, due to the increased differences between the competitors in the different conditions of this experiment while still counterbalancing the items, the visual complexity of the displays increased very strongly, namely by 100 %. Seeing eight objects at once could significantly aggravate expectations and task solving. It hence alternatively needs to be considered that expectations could also not be the most efficient strategy in the given context any more and could therefore possibly decrease overall. In this case, anticipatory eye movements during the verb region, as well as surprisal-based differences in processing effort for the noun were expected to decrease as the number of competitors increases if (detailed) expectations gradually waned as a result of visual complexity and aggravated mapping of

Figure 6.2 **Proportion of Fixations** across trial length in all conditions of **Experiment 6**. Each line corresponds to one objects in the visual displays.



information – possibly as a result of not being the computationally most efficient strategy any more.

Analysis

Eye-movement Data New inspections on relevant interest areas (targets, competitors, distractors) were analysed in the verb and noun time window, as done previously.

ICA ICA data was analysed as before, using the identical time windows. Differences between the conditions (*Number*) were orthogonally contrast coded and entered into the model as fixed factors.

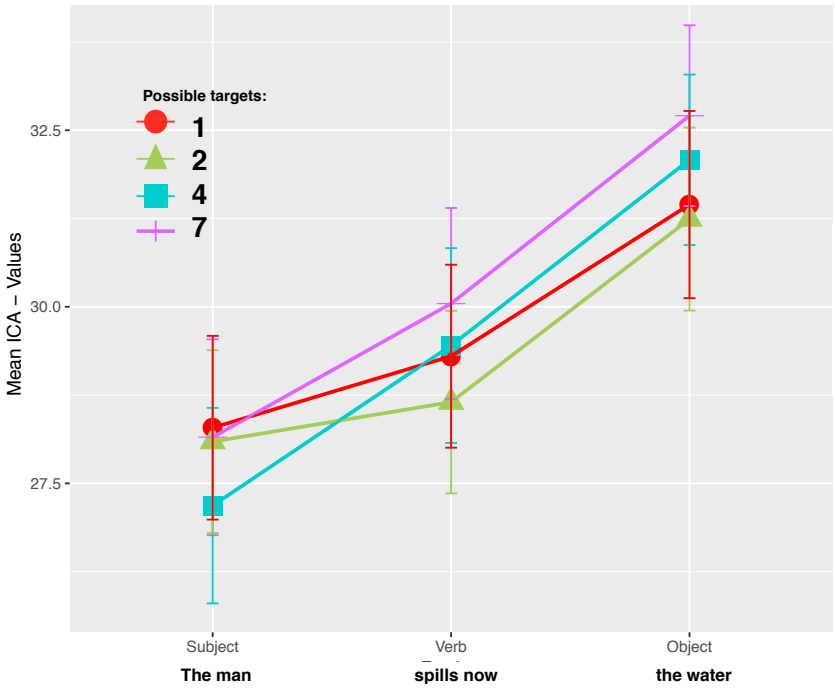


Figure 6.3 **ICA Results** for **Experiment 6** in all conditions. Error bars reflect 95% confidence interval (CI).

6.3 Results and Discussion

Eye-movement Data Fixation distribution across an averaged trial length is shown in Fig.6.2 for all conditions. An increased number of eye movements towards the target and competitor object, that is, towards the object(s) matching the verb, is clearly visible for condition **1** and partly for condition **2**. In the latter condition, data shows that, despite the fact that target and competitor have both been identified, listeners do not as clearly decide for those two objects as seen in previous experiment with less objects shown per display. This can be interpreted as reflecting increased difficulties in integrating visual and linguistic information in the complex visual context, where it is takes longer to decide whether each of the many objects shown is a possible target.

Moreover, anticipatory patterns in the verb region gradually further decrease as the number of competitors increases in conditions **4** and **7**.

In line with this, parallel analyses of new inspections in the verb window revealed significantly more new inspections toward the target object upon hearing the verb if only the target object was displayed **1** $M = 0.12$, $SD = 0.33$) , compared to seven options (**7** $M = 0.09$, $SD = 0.29$) ($\beta = -.298, SE = .13, z = 2.21, p < .05$). Marginally more looks were directed

Table 6.1 **Fixation data on the Verb:** Anticipatory first Inspections to target object *between* conditions for **Experiment 6**, Model: *First Inspections on Target Object ~ Nr. of possible Targets + (1 | Subject)+ (1 | Item), family="binomial"*

Predictor	*Coefficient*	*SE*	*Wald Z*	*p*
(Intercept)	-2.136	.051	-41.16	< .001
One vs. Seven poss. Targets	.298	.134	2.21	< .05
Two vs. Seven poss. Target	.233	.135	1.72	.085
Four vs. Seven poss. Target	.182	.136	1.34	.18

to the actual target when two objects were displayed (**2** M = 0.12, SD = 0.32), compared to the context featuring seven competitors ($\beta = -.233, SE = .135, z = 1.72, p = .085$). No significant differences in new inspections of the target were found between conditions **1** vs. **2** and **2** vs. **4**.

Interestingly, the fixation distribution plots (6.2) further show a slight delay in eye movements towards the actual target object upon hearing the corresponding noun for condition **4**, and a clearly visible delay for condition **7,** compared to all other conditions. In other words, it took listeners more time to actually *find* the object corresponding to the target noun in the complex displays as the number of competitors increases.

Along with this, first inspections in the noun window (i.e., between article onset and noun offset) reflected the expected mapping of visual and linguistic information, that is, increased inspections towards the mentioned object, for conditions **1** ($\beta = .862, SE = .408, z = 2.11, p < .05$) and **2** ($\beta = .752, SE = .382, z = 1.97, p < .05$). At the same time, increased looks to the actual target (reflecting it's identification) in conditions **4** ($\beta = .976, SE = .363, z = 2.69, p < .05$) and **7** ($\beta = 1.487, SE = .399, z = 3.730, p < .005$) were only found as the time window was moved to span noun onset until 200ms after noun offset.

ICA Again, no significant differences were found in the ICA values on the verb between conditions. Fig. 6.3 shows how the same verb again requires almost identical processing effort in the different visual contexts (**1**, M = 29.75, SD = 11.13 vs. **2**, M = 28.90, SD = 11.38 vs. **4**, M = 29.57, SD = 12.07 vs. **7**, M = 30.33, SD = 11.56). Moreover, in the presence of complex visual displays containing eight pieces of clip art, no differences were found in the ICA values on the target noun (**1**, M = 31.45, SD = 11.78 vs. **2**, M = 31.24, SD = 11.51 vs. **4**, M = 32.08, SD = 10.77 vs. **7**, M = 32.71, SD = 11.38). That is, no effect of visual context on

Table 6.2 **Differences in ICA** for **Experiment 6**, Model: *ICA values on noun ~ Nr. of possible Targets + (1 + Nr. of possible Targets| Subject)+ (1 + Nr. of possible Targets| Item), family=poisson (link = "log")*

Time window: Verb	*Predictor*	*Coefficient*	*SE*	*Wald Z*	*p*
	(Intercept)	3.360	.042	79.89	< .001
	One vs. Two poss. Targets	.030	.028	1.08	.280
	Two vs. Four poss. Target	-.014	.036	-0.38	.702
	Four vs. Seven poss. Target	-.034	.035	-.97	.331
Time window: Noun	*Predictor*	*Coefficient*	*SE*	*Wald Z*	*p*
	(Intercept)	3.4334	.042	82.53	< .001
	One vs. Two poss. Targets	-.003	.032	-.09	.926
	Two vs. Four poss. Target	-.035	.028	-1.24	.216
	Four vs. Seven poss. Target	-.019	.027	-.72	.468

processing effort, or, target word surprisal and predictability was observed in the ICA as the complexity of visual contexts increased.

Discussion This study used the ICA index to assess processing effort in contexts featuring increased numerical difference in competitors between conditions. It was designed in order to test whether this increase would lead to a clear result with respect to the question whether participants evaluate the multi-modal context in a fine grained way, as indicated by the eye movements in previous studies, or rather decided whether one, many, or none of the items shown are possible targets.

However, based on findings in the recent literature, suggesting that the complexity (and timing) of visual information can cause a change in language processing strategy (see, e.g. Ferreira et al., 2013), we additionally had to consider that the increased complexity of visual displays could lead to expectations being computed to a reduced extend. More specifically, anticipation, as reflected in the eye movements gradually could decrease as visual contexts get more complex. Indeed, eye movement data showed a *gradual decrease* of anticipatory patterns in the context of eight pieces of clip art as the number of competitors among those objects increased. More specifically, anticipatory patterns were found as either one or two of the eight depicted objects matched the verb, but not if four or even seven objects were actual competitors. Fixation distribution data plotted in Fig.6.2 showed a decreasingly clear distinction between competitors and distractors from the verb on in condition **2** (i.e., as

compared to fixation data from previous experiments), while anticipation is finally not visible at all anymore in conditions **4** and **7**.

In line with results reported by Ferreira et al. (2013), this is interpreted as to show how complex displays are still used for information extraction. However, it gets more difficult for listeners to use visual context to constrain and specify their interpretations as the contexts become more complex. Importantly, fixation distribution on the noun also reflected a delay in mapping linguistic to visual information in addition to this. That is, it took participants longer to direct gaze towards the target objects upon hearing the actual noun if more competitors are among the eight objects in the display. This is especially important because it reflects an increased visual search followed by a delayed identification of the target object, most likely attributable to expectations about the target noun not being made in advance. Since the mapping of linguistic information onto eight different visual objects is more difficult, it possibly takes too much time to additionally compute expectations about the target noun in advance to hearing it. In this experiment, this would hence mark the limit of complexity that can still be handled and hence the point at which expectations disappeared all together.

Despite the patterns of decreasing target word expectations, the data is in line with Ferreira et al. (2013), who report no increase in overt errors in displays featuring 12 objects. Participants in this experiment did correctly identify the target without increased errors. This was even true for condition **7**, however, it took them more time to do so, as compared to studies with less complex displays and even as compared to equally complex displays featuring less competitors. Along with Spivey et al. (2001), we suggest that the lack of expectations for the target word results in hampered incremental interpretation of the actual word as it is encountered, which is reflected by increased processing time required for visual search. In sum, we propose to interpret these findings as to reflect a gradual change in comprehension strategy, very likely as a result of feedback with respect to the efficiency of the current strategy in the given context.

Simultaneously assessed ICA values reflecting processing effort this time did not show any differences between conditions in either of the observed time windows. Instead, values in the different conditions equally increased over the lengths of the trial. That is, although task and linguistic stimuli stayed consistent, compared to the previous experiments, the increased complexity of the visual scenes seemingly caused equally high processing effort for the same target word in all different visual contexts. Although at the first sight these results could be interpreted as to reflect the absence of any expectations about the target word, it is unlikely that this caused the identical ICA values. Interestingly, the noun is even hard to process in condition **1**, where anticipatory patterns were measured prior to the noun and clearly only one objects is the possible target. This time, not even a distracting or mismatch condition

(as condition **0** in the previous studies) was presented, making any possible expectation very reliable in this context.

If the measured ICA results were solely attributable to the fact that listeners decreasingly computed expectations in advance to the noun in the complex visual context as the number of competitors increased, facilitated processing would still have been expected in condition **1**, where anticipatory eye movements were measured during the verb and retrieval of the actual target object when hearing noun was immediate.

Hence, alternatively or even additionally, other – possibly not language related – factors may have contributed to these results. The effort reflected by the ICA data again might not entirely be attributable to processing effort of the target word, but rather also reflect non-linguistic factors such as increased alertness in participants caused by the higher task difficulty, as previously suggested in experiment *7 B* with the only difference that this time, attentiveness increased in reaction to the complex visual scenes and the restricted time period given by each audio. Even if different factors caused the higher task demand, it may either way result in a reaction of the LC-NE system in order to optimize performance. This would be in line with Marshall (2000) as well as with Demberg and Sayeed (2016), who report an increased amount of rapid pupil dilations as picked up by the ICA measure in task that required focused concentration. It can also be interpreted as being in line with Aston-Jones and Cohen (2005), stating that the LC system lapses into a tonic activation mode whenever a the utility of a current behaviour for task performace decreases. This mode is proposed to facilitate disengagement from the current task and the search for alternative behaviours. Berridge and Waterhouse (2003) further propose that increased tonic discharge activity, as for example caused by increased stress, reduce phasic discharge, which according to the authors, is associated with overt attention. In this case, patterns of overt attention would simply not be visible any more. In consequence, the ICA and the eye movement data in this experiment can not be seen as complementary, as it was the case for example in experiment **4**, where a correlation between anticipatory looks and facilitated noun processing was found. Instead it is suggested that results from the eye movement data of this experiment reflect aggravated search and difficulties in mapping of information due to the fact that now features of not five, but eight objects need to be considered, while the ICA is at least partially caused by factors not related to language processing.

6.4 Experiment 7: Fine grained expectations & the role of eye movements

Although data from the previous experiment did reveal an interesting decrease in expectations in complex visual contexts, and even hinted at different sources of effort reflected by our pupillary measure, it could not answer the question whether visual information has a probabilistic effect on the mental model. Additionally, we are still left with a question that already came up in the context of the EEG experiment. Namely, whether the absence of overt attention affects such a possible probabilistic influence of visual information. As a next step, we hence ran two versions of the identical experiment, now featuring less complex visual contexts while still increasing the numerical difference in competitors between conditions (at the cost of counterbalancing the items). The following two versions of the same experiment reduced visual complexity form eight to five pieces of clip art while keeping an increased numeric difference in competitors between conditions. The first version (7 A) was designed to observe whether participants had fine grained probabilistic, or rather rough "one-many-nothing" expectations about the target nouns, while the second version (7 B) tested whether overt attention, as reflected by anticipatory eye movements, was a crucial factor in fine grained context evaluation. In this version, free eye movements were hence prohibited, hereby mimicing the eye movement conditions of the previous EEG study while deploying the pupillary measure.

6.4.1 Experiment 7 A: Fine grained vs. "one-many-nothing" expectations

Visual Stimuli Validation The visual stimuli used in the following two studies were subsets of the previously used displays. Since the number of clip arts in each display was reduced from eight to five, no additional pretests were required for this reduced version. All experimental items used in the experiment had yielded > 90% correct answers per item in the previously pretested versions, that is, the pieces of clip art used were actually matched with the verb.

6.4.2 Method

Both studies featured a within-subject 1 x 3 design in which, for the sake of comparability, again the same kind of manipulation and type of stimuli as in the previous experiments were used. 39 linguistic stimuli (we dropped one in order to get a multiple of the three conditions)

Figure 6.4 **Example Stimuli** for **Experiment 7**. From left to right: **1**, **2**, and **5** possible targets, given the sentence *"The man spills soon the water."* (Numbers were not depicted in the experiment). For more items see Appendices A and B

were paired with 117 different visual displays (three for each sentence). In order to test whether the effect of visual context information on the statistics of the mental model, the resulting expectations and finally the linguistic processing effort was indeed probabilistic, the number of clip art objects in each display was set to five, arranged in a star shape around the screen center. This way, the numeric differences between conditions could be enlarged (as compared to **Experiment 4**), resulting in three conditions: **1**, **2**, or all **5** objects were now potential target referents (see Fig. 6.4, from left to right), while avoiding too complex contexts as in **Experiment 6**. Due to the design of the conditions in the experiment (**1**,**2** or **5** competitors) and the number of objects in each display (five), displays could not be counterbalanced. Positions of targets, competitors and distractors were again rotated.

An equal amount of fillers introducing variation in terms of the number of categories displayed (i.e., edible, wearable, driveable objects etc.) was added, before the sentences were parted into three lists using latin square design and randomized as before. As in the previous ICA study, all sentences were presented auditorily and always simultaneously with the visual displays.

Visual displays were shown with an adapted preview time of 1500 ms due to the increased amount of objects in the displays. As in the EEG Study, each item was followed by yes/no comprehension questions, concerning either the position of a clip art in the visual display (e.g.: "Was the water on the right?"), or the utterance (e.g.: "Did the man spill the lemonade?") to keep participants focused. Questions could be answered by button press on a keyboard (Model: "Cherry G230").

24 students of Saarland University, all of them being native speakers of German (*M* age: 23.8; Age range: [19, 35]; *SD*: 3.7; Female: 19) gave informed consent before being tested and were monetarily reimbursed for their contribution.

Items were presented on a Samsung S27D390 monitor with a 1920x1080 resolution, a 60 Hz refresh rate, and 32 bit colour depth. ICA values were extracted from the eye movement data, using the same procedure as previously described.

Predictions If visual information has a probabilistic effect on the mental model and expectations, the increased (as compared to the differences between conditions in **Experiment 4** and **5**) numerical difference between the number of competitors is hypothesised to result in significant differences in processing effort for the target noun, as reflected by the ICA.

Conditions **1** and **2** are especially interesting, as we saw from previous experiments, that a small numeric difference between displayed competitors (as previously in conditions **3** and **4** in **Experiment 4** and **Experiment 5**) possibly causes null results with respect to differences in processing effort. On the one hand, a replication of the previous results could be interpreted as to confirm this hypothesis. On the other hand, however, if the evaluation of visual context in probabilistic, the difference in percent would be much larger this time. More specifically, the difference in percentages between one and two competitors is as much as 50% (not 8% as it was between conditions **3** and **4** in **Experiment 4** and **Experiment 5**) which, at the end, could be enough to cause probabilistic differences in processing effort.

If, however, comprehenders only roughly decide whether one or many objects matching the verb, we expected no difference between conditions **2** and **5**, since both feature "many" (as opposed to one or zero) competitors. In line with previous findings, we expected verb driven, anticipatory eye movements towards competitors prior to hearing the actual target noun, as well as more looks towards the actual target objects once the target noun is encountered. No delay in retrieving the target object was expected if expectations were made prior to the target noun and the mapping of visual and linguistic context is not obstructed by visual complexity.

Analysis

Eye-movement Data Again, linguistically driven new inspections (attention shifts) towards possible targets were analysed in the respective verb and noun time window, as done previously.

ICA ICA data was analysed as before, using the identical time windows. Conditions were contrast coded, entered into the model as fixed factors and orthogonally compared.

Table 6.3 **Fixation data on the Verb:** Anticipatory first Inspections to target object *between* conditions for **Experiment 7 A**, Model: *First Inspections on Target Object ~ Nr. of possible Targets + (0+ Nr. of possible Targets | Subject)+ (0 + Nr. of possible Targets | Item), family="binomial"*

Predictor	*Coefficient*	*SE*	*WaldZ*	*p*
(Intercept)	-1.8519	.0318	-58.31	< .001
One vs. Two poss. Targets	-.0151	.0755	-.20	.8415
Two vs. Five poss. Target	.2554	.0787	3.25	< .005

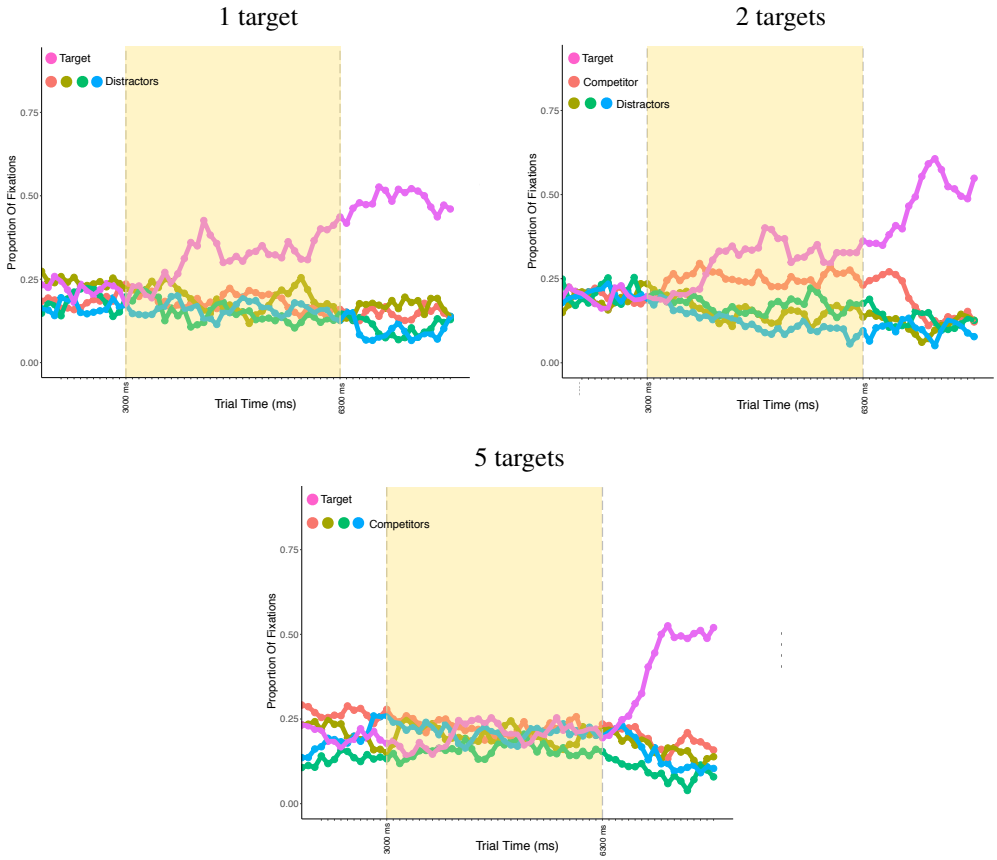
6.4.3 Results and Discussion

Eye-movement Data The overall proportion, that is, the distribution of fixations across an averaged trial length in percent is again plotted for presentation purposes in Figure 6.5 for all conditions. Data shows a replication of the previously found anticipatory patterns towards objects matching the verb, from the verb onset (left dashed line). That is, listeners increasingly inspected competitors upon mapping the linguistic information to the visual context, in expectancy of the target. Again, no increase in competitor fixations was found when the context did not allow for a discrimination between matching and mismatching objects (condition **5**).

Inferential statistics were ran on new inspections (attention shifts) to the target object between conditions, which were expected to decrease as the number of competitors increases if all matching objects are considered (see Fig. 5.2). In line with the previous data and along with the fixation distribution, first fixation models revealed a significant decrease in glances to the target object between condition **2** (M = 0.15, SD = 0.35) and **5** (M = 0.11, SD = 0.32) ($\beta = .255, SE = .08, z = 3.25, p < .005$) At the same time, anticipatory glances at the verb decreased only non-significantly between the two conditions which were closer to each other in terms of the number of competitors, that is, between **1** (M = 0.14, SD = 0.35) and **2** ($\beta = -.02, SE = .07, z = -1.69, p = .84$). The noun region revealed, as in the previous experiments, that participants were significantly more likely to inspect the mentioned object compared to any other object in the display, upon hearing the target word.

ICA As in all previous studies, no significant differences in effort related, abrupt contractions of the pupil, as culled by the ICA, were measured on the verb. That is, no effects attributable to an exclusion of mismatching objects were found. From Fig. 6.7, it can be seen how again the same verb requires similar processing effort in each of the conditions, independent of the number of competitors, or, respectively distractors that could be excluded

Figure 6.5 **Proportion of Fixations** across trial length in all conditions of **Experiment 7 A**. Each line corresponds to one objects in the visual displays. The Plots show a clear discrimination between target word competitors and unrelated distractors (the difference between actual target object and competitor in condition **2** is not significant).



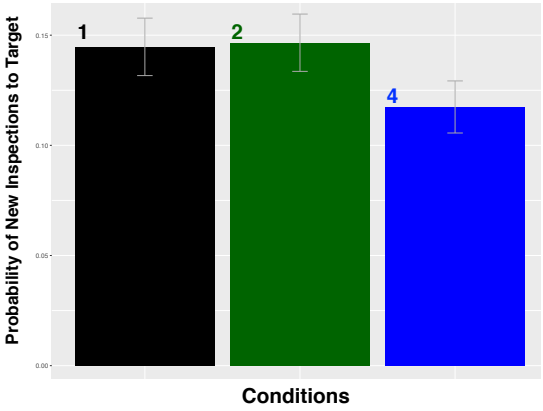


Figure 6.6 **Probability of verb-driven new inspections** of target object during the verb (prior to the target word onset) in all possible conditions of **Experiment 7 A**.

Table 6.4 **Differences in ICA** for *Experiment 7 A*, Model: *ICA values on noun ~ Nr. of possible Targets + (1 + Nr. of possible Targets| Subject)+ (1+ Nr. of possible Targets| Item), family=poisson (link = "log")*

time window: Verb	Predictor	Coefficient	*SE*	WaldZ	*p*
	(Intercept)	3.4601	.0402	86.07	< .001
	One vs. Two poss. Targets	-.0278	0.0369	-.75	.451
	Two vs. Five poss. Target	-.0003	.0313	-.01	.993
time window: Noun	Predictor	Coefficient	*SE*	WaldZ	*p*
	(Intercept)	3.4933	.0397	88.09	< .001
	One vs. Two poss. Targets	-.0232	.0315	-.74	.4595
	Two vs. Five poss. Target	-.0794	.0296	-2.68	< .01

as possible targets (**1**, *M* = 32.00, *SD* = 12.01 vs. **2**, *M* = 32.64, *SD* = 11.48 vs. **5**, *M* = 32.33, *SD* = 12.29).

Most importantly, in addition to the replication of the null results on the verb, the ICA graph shows a graded difference in processing effort for the same target noun in the different visual contexts, in support of a rather probabilistic influence of visual information on expectations and processing effort. That is, processing effort for the same noun was clearly higher when five competitors were shown, compared to just one or two potential targets. The generalised mixed effects models of poisson type with the ICA values as (contrast coded) basic dependent measure revealed a significant processing facilitation, that is, lower ICA values, if two possible targets were shown (*M* = 33.51, *SD* = 11.82), compared to five possible target objects (*M* = 35.48, *SD* = 10.80) ($\beta = -.08, SE = .03, z = -2.68, p < .001$). ICA differences between conditions **1** (*M* = 32.28, *SD* = 12.09) and **2** did not reach significance ($\beta = -.02, SE = .03, z = -.74, p = .46$). Again, analysis of an additional time window of 600 ms length starting from trial onset, where possible effects of object grouping may appear showed no significant differences in processing effort.

Apart from replicating the previous results showing the direct impact of visual context on word processing and surprisal, the data this time clearly suggest that the observed differences in processing effort are indeed attributable to a probabilistic evaluation and expectation about the actual target word.

Discussion The present experiment featured visual contexts with reduced complexity (as compared to the previous study **Experiment 6**) and either **1**, **2** or **5** competitors, in order to

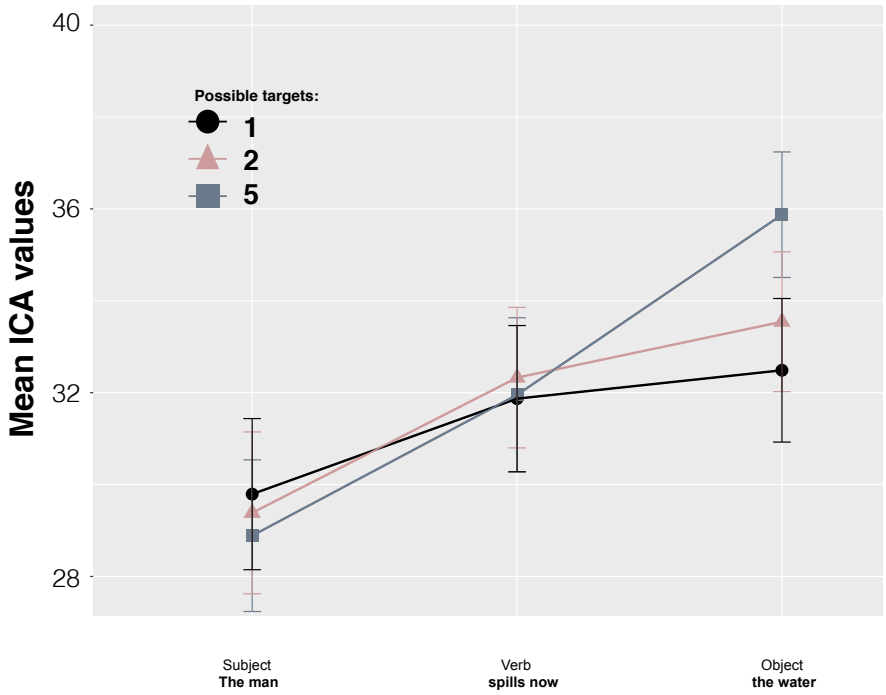


Figure 6.7 **ICA Results** for **Experiment 7 A** in all conditions. Error bars reflect 95% confidence intervals (CI).

answer the question whether visual information has a similarly probabilistic effect on the statistics of the mental model as linguistic context information. This question was the result of **Experiment 4** and **5**, where significant differences were found between conditions **1**, **0** and **4**, but not between **3** and **4**. It could hence have been possible that comprehenders decided between one, more than one, or no object matching the sentence. In **Experiment 6**, increased visual complexity resulted in effects related to language processing not being visible any more. We subsequently argued that – based on on previous findings about probabilistic expectations in linguistic context (see, e.g. Hare et al., 2007), as well as on anticipatory patterns observed in the simultaneously obtained eye movements in our experiments – this pattern might be the result of the small numeric difference in competitors shown in conditions **3** and **4** in question, namely only one more competitor displayed, resulting in a probabilistic difference of not more than 8 %. The three conditions in this study were hence designed to increase numeric (and hence the probabilistic) difference between the number of displayed competitors(**1**, **2** or **5** competitors), while stimuli type, setup and task remained identical to the previous experiments, and visual complexity was reduced to only five objects in each display. This way, it was not only possible to observe whether expectations were probabilistic,

but also whether the lack of a difference between conditions was caused by the too small differences in competitor numbers between them.

Results from the present experiment's eye movement data reliably replicated verb-driven anticipatory eye movements towards matching objects, indicating a detailed context evaluation, that is, a fine grained discrimination between possible targets and distractors, while again, no differences were found in the pupillary measure of processing effort in the same time window (i.e., during the verb).

Most importantly, ICA results from the target noun not only replicated previous results by generally showing a robust and significant effect of visual context on target noun processing, but were further revealing with respect to the initial question of how this influence can be described: More specifically, this time processing effort differed significantly between conditions **2** and **5** while interestingly, the difference between **1** and **2** did not reach significance. Based on the significant difference between conditions **2** and **5**, as well as in the light of eye movement data suggesting a detailed evaluation of visual context, we still interpret these results as evidence in support of probabilistic expectations (as opposed to rough "one-many-nothing" decisions), enabled by the mapping of linguistic to visual context and the fine grained evaluation of the multi-modal context. We hence suggest that visual information is evaluated at each possible point, that is, whenever language can be mapped to the visual context, and has a similarly probabilistic influence on the statistics of the mental model and expectations-based, linguistic processing effort. We propose that the null result between conditions **1** and **2** could be attributable to a ceiling effect in the sense that the scenes were similarly transparent in both conditions (i.e., it is comparatively easy to identify one or two competitors in a display of five objects). In other words, while the lack of a difference between the previous conditions **3** and **4** may be attributed to the small numerical (and percental) difference between competitors, this time it may be caused by the conditions both being so easily grasped by the comprehenders.

The interpretation of our results in favour of a probabilistic effect on the mental model is not only in accordance with studies reporting graded effects of (linguistic) context on word processing (see, e.g. Hare et al., 2007), but further shows that these findings can be extended to non-linguistic context information, possibly hinting at predictive mechanisms not necessarily being language specific in their nature. Now that we showed results in support of the hypothesis that visual information has a probabilistic effect on expectations and processing effort, we are still left with the question what the actual role of eye movements and overt attention is in the evaluation of visual context underlying this effect. The following version of the same experiment was setup to observe exactly this role of (anticipatory)

eye movements in the evaluation of the visual context and the calculation of probabilistic expectations about the target word.

6.4.4 Experiment 7 B: Overt vs. covert attention: The role of eye movements in fine grained expectations about target words

In line with previous research in purely linguistic paradigms, the results from the preceding study show that expectations about the target noun are also fine grained and probabilistic (as understandable within a Bayesian computational framework) in multi-modal contexts. However, we also saw from **Experiment 4** and **5** that no effects were found between conditions with a small numeric difference in competitors (at least not with the the tested amount of participants) in either the pupillary or the ERP measure of effort.

In fact, the difference between conditions **3** and **4** were even smaller in the EEG (**Experiment 5**), which left us with the question whether this is solely attributable to the N400 not being directly comparable to the ICA: After all, like most EEG studies, we deployed a paradigm of covert attention, due to the fact that eye movements linked to overt attention shifts cause severe artifacts. This form of attention could additionally aggravate a fine grained evaluation of visual context, as it might be more difficult to shift attention using working memory, compared to visual shifts.

Eye movements in general have long been used as an unbiased and informative measure of (perceptual) cognitive processing, being able to reflect partially active concept taken into consideration while an interpretation is being formed (Richardson and Spivey, 2004). Previous research has shown that comprehenders look towards things they anticipate to come up next (Altmann and Kamide, 1999). More recently, it has also been explicitly proposed to think of motor movements, especially hand and eye movements (as opposed to only the mental representations of them) as being symptoms, or even indeed components, of linguistic and cognitive processes (Spivey et al., 2009). This is based on the idea that language, perception and action are not independent modules, but rather parts of a complex, dynamic interaction with the rest of the brain and body. In this framework, cognition is thought of as quick, dynamic procedures such as anticipated perceptions and actions prepared accordingly.

It has further been shown that, although cognitive control needed to shift overt, perceptual attention, as indicated by eye movements, and covert attention shifts in working memory share collective parts of the brain's attentional control network, both ways of shifting attention towards relevant stimuli exhibit distinct patterns of neural activity (Tamber-Rosenau et al., 2011). In other words, overt and covert attention share control processes, however covert attention further involves inhibition of eye movements (Kulke et al., 2016). It can hence be

assumed that covert attention (and possibly the inhibition of eye movements) aggravates the detailed evaluation of visual context, possibly resulting in less detailed effects on expectation based processing effort. Alternatively, it is possible that probabilistic effects of visual information on processing effort are also visible when overt attention is suppressed. This would hint at a different role of (anticipatory) eye movements in the evaluation of visual context, such as, for example, an "outsourcing" of memory load as the information is still present. Although many studies have worked with (anticipatory) eye movements, their specific role with respect to context evaluation and predictive processing has so far not been researched. Hence, the present study exploits the advantages of the pupillary on-line measure of effort to investigate on the role of overt attention in the probabilistic evaluation of visual context. Results are interpreted in the light of our previous findings, especially in **Experiment 7 A**, where in the very same experiment, eye movements were not inhibited. The gathered data will provide deeper insights into the actual role of eye movements in the VWP.

6.4.5 Method

Apparatus, stimuli, design and task were identical to the ones used in **Experiment 7 A** in order to have a direct comparison where the only difference is the fact that no overt eye movements were allowed this time. . That is, the same three conditions were deployed, resulting a design featuring 39 item sentences, combined with 156 different visual displays and 39 filler sentences, combined with 39 visual displays. Three lists in Latin square design were used as before. The only difference was a fixation cross appearing with the onset of the audio stimuli and disappearing with the audio offset. All visual stimuli were again shown throughout the entire trial with an additional 1500 ms preview during which eye movements were allowed in order to enable the identification of the pieces of clip art in the display prior to the onset of the audio. After sentence offset, visual displays stayed on screen for an additional 1000 ms in order to prevent participants from spending the time during the actual audio encoding positions to solve the task.

Participants were asked to keep their eyes fixated on the cross from sound onset throughout the whole trial. As in the EEG experiment, the distance between each participant and the display screen was always 103 cm in order to keep all of the objects in a 5° visual angle from the center of the screen to minimize eye movements throughout the experiment.

24 students of Saarland University, all of them being native speakers of German (*M* age: 24.8; Age range: [18, 31]; *SD*: 3.8; Female: 17) with normal or corrected to normal vision took part and gave informed consent before being tested. All participants were monetarily reimbursed for their contribution.

Table 6.5 **Differences in ICA** for **Experiment 7 B**, Model: *ICA values on noun ~ Nr. of possible Targets + (1 + Nr. of possible Targets | Subject)+ (1 + Nr. of possible Targets | Item), family=poisson (link = "log")*

time window: Verb	Predictor	Coefficient	*SE*	Wald Z	*p*
	(Intercept)	3.5513	.0311	113.86	< .001
	One vs. Two poss. Targets	.0023	.0234	0.10	.920
	Two vs. Five poss. Targets	.0087	.0211	.41	.682
time window: Noun	Predictor	Coefficient	*SE*	Wald Z	*p*
	(Intercept)	3.5943	.0317	113.39	< .001
	One vs. Two poss. Targets	.0186	.0169	1.10	.272
	Two vs. Five poss. Targets	-.0134	.0204	-.66	.510

Predictions If anticipatory eye movements are crucial for the detailed context evaluation underlying probabilistic effects of visual information on mental statistics and target word expectations, less graded effects were expected in the ICA on the noun as saccades were inhibited. That is, if context evaluation and mapping of linguistic and visual information is aggravated in the absence of overt eye movements (despite an appropriate preview time given in order to identify the pieces of clip art), no fine grained expectations might be formed prior to the noun. If, on the other hand, expectations are formed as attention is shifted covertly, similar ICA results were expected as found in experiment **7 A**. In that case, overt eye movements may serve a different purpose such as a relief of memory load in the sense that looking at an object is easier than remembering its attributes and position.

Analysis

ICA Since no eye movements were allowed in this experiment, only the ICA data was analysed. The identical time windows and analysis method was used as before. Conditions were contrast coded, entered into the model as fixed factors and compared. Trials containing saccades away from the fixation cross during the critical time period (i.e., during the sentence) were excluded from the analysis.

6.4.6 Results and Discussion

ICA This time, apart from again not measuring a significant differences at the verb (although possibly for other reasons compared to the previous experiments), there were also

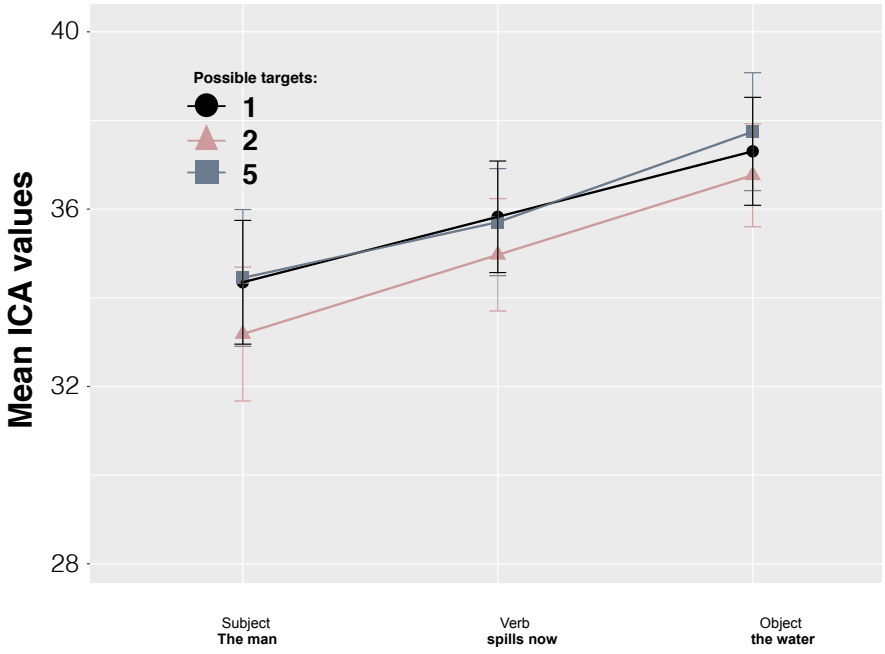


Figure 6.8 **ICA Results** for **Experiment 7 B** in all conditions. Error bars reflect 95% confidence interval (CI).

no significant differences found on the noun. At the same time, processing effort linearly increased throughout the trials and overall processing effort was higher than observed in any of the previous experiments. Fig. 6.8 shows, how ICA on the noun is highly similar between conditions **1** (M = 37.30, SD = 9.97), **2** (M = 36.76, SD = 9.47), and **5** (M = 37.74, SD = 10.57). That is, no effects of visual context on surprisal-based processing effort was found in a context where overt eye movements were inhibited.

Discussion While the previous **Experiment 7 A** confirmed a probabilistic pattern in surprisal-based processing effort on target nouns, that is, depending on the probability profile of the multi-modal context, **Experiment 7 B** only yielded null results for the same region, although the only difference between both versions of the same experiment was the inhibition of overt eye movements in the latter study.

Two possible explanation for the null result in **7 B** are close at hand: Either the lack of overt attention caused a major aggravation of probabilistic context evaluation resulting in no probabilistic expectation made, or probabilistic expectations are still made but are not detectable by the ICA any more. In the former case, the lack of a significant effect on the noun could be attributed to fine grained expectations no longer being the computationally

most efficient strategy as eye movements are inhibited. This could be accounted for by the idea of listener's rational adaption to the task demands, as for example brought up by Kuperberg and Jaeger (2016b), who propose that the estimated effort and utility of predictions in their context, as related to the intrinsic goal of the comprehender are decisive of how fine grained predictions are made. According to this hypothesis, the listener would then adapt to the demands and take on a more efficient strategy. In the case of the results from the presented study, this would however mean, that the *sweet spot* at which (fine grained) expectations are ideally efficient, would be left in a single, discrete step of only one more piece of clip art displayed in this experiment, compared to four clip art objects shown in the EEG, where expectations were still evident in the data. Note that it is nor argued here that a rational adaption, that is, a change in strategy due to (especially fine grained) expectations not being the most efficient strategy in a certain context, is not possible *per se*. Indeed there is increasing evidence against fixed and pre-determined language processing strategies. Ferreira et al. (2013), for example, argued for the role of prediction as an adaptive comprehension strategy, showing that comprehension strategies can change if, for example and amongst other factors, scene complexity increases. However, a change in language processing strategy, for example in response to (increased) task demands, would rather be expected to occur gradually according to feedback and evaluation of the current context, which would also potentially be of higher psychological validity, given the fact that an adaption of behaviour should be based on incremental feedback, rather a fixed threshold. The hypothesis of a gradual adaption is further backed by our findings from **Experiment 6**, featuring complex visual contexts, where similar ICA patterns were found, while eye movements still showed patterns of anticipation in the conditions with **1** or **2** possible targets among the eight objects. Only in the conditions **4** and **7**, those patterns gradually disappeared.

Importantly, the eye movement results found in the previous study **6** can support the hypothesis that (detailed) expectations about target words do not disappear in a discrete manner (i.e., as one more object is added to the display) but may rather be continuously reduced as they become more expensive to maintain and less efficient for task performance, as for example, in the context of increased visual complexity within a limited time period.

An alternative interpretation would be that fine grained expectations were still present in **7 B**, this time, however they are not reflected by the pupillary measure. This could be explained in the context of the initially discussed Norepinephrine-Hypothesis, suggesting that the neuro-modulator NE plays a significant role in pupil contractions and the ICA index. In other words, listeners might still have computed fine grained expectations about the target word, as they did in case of 4 objects while inhibiting eye movements in the EEG, but the linear increase in effort related pupil dilations may be caused by an overall higher LC-activity.

A possible explanation here can be that the increased task demand (due to the inhibition of eye movements) causes an overall increased attentiveness – as it was the case in **Experiment 6**, where no differences were found in the ICA, although eye movements revealed anticipatory patterns at least for the conditions with less competitors. In that case, although **Experiment 7** and **7 B** resulted in similar ICA patterns, the cause for this increased attentiveness is very likely different: Namely, high visual complexity in **Experiment 6**, and the inhibition of eye movements in **Experiment 7 B**, resulting in the same task being more difficult to solve. This hints at a facilitating role of overt attention with respect to the task that participants had to solve. The hypothesis that increased attentiveness could be reflected by overall increasing ICA values can be backed by Aston-Jones and Cohen (2005)'s adaptive gain theory: The authors suggest that LC-neurons have two distinct modes of activity, namely a so called *tonic*, and a *phasic* mode. While the latter is suggested to facilitate current behaviours and to help optimize task performance, the *tonic* activity mode is proposed to be elicited as disengagement from the current task and the search for alternative behaviours increases. Berridge and Waterhouse (2003) suggest that activity in the **tonic** mode causes less robust effects of overt attention (which would be visible in *phasic* mode). In line with this, the results from **7 B** can be interpreted as being caused by increased attention and probably the search for an alternative strategy, meaning not all the measured effort is language related.

Further evidence for increased task demand are the overall higher ICA values measured in *Experiment 7 B*, although the study itself – including the task – was identical to **Experiment 7 A**. In other words, the inhibition of eye movements does not necessarily result in an absence of expectations about the target word, but rather increases task difficulty, reflected in not only the overall increased ICA values, but also in a lack of differences between conditions in processing effort on the noun.

The linear increase in ICA values throughout trial length could also be explained by a possible increased effort related to the actual inhibition of eye movements. Effort related to the suppression of saccades has previously been found in EEG experiments. Kulke et al. (2016), for example, reported an increased frontal positivity during covert attention shifts and suggested this to possibly reflect inhibition of saccadic eye movements and maintained fixation. In this case, however it would be reasonable to expect lower effort during the preview time, during which saccades were allowed, while effort should increase as the fixation cross appeared and listeners were asked to keep their eyes still. The preview region in **7 B**, however, did not show such a facilitation. Instead, ICA values from the preview region were slightly higher, compared to the following subject region (condition **1**: M = 33.97, SD = 12.04, condition **2**: M = 33.17, SD = 12.81, condition **5**: M = 33.50, SD = 13,19). We hence propose that the null results from **7 B** reflect an increase of task difficulty

in absence of overt attention, and that this can be interpreted as being in line with the idea that LC/NE activity plays a role in effort related pupil contractions, as culled by the ICA's algorithm and supports the hypothesis that eye movements have a facilitating role, rather than being crucial for probabilistic expectations..

6.5 The ICA and the N400

It has recently been discussed how norepinephrine-release (and respective measures reflecting it) might correlate with different ERP components. While originally no direct connection between the LC area (the sole source of NE for brain regions involved in higher cognition, as well as affective processes) and language processing has been established in the first place, by now a line of research proposes that LC activity might spread into brain regions associated with language processing.

This also seems plausible, as the tiny, neuro-modulatory nucleus in the brain stem plays an important role in regulating cortical function and enabling cognitive reorientation (e.g., whether a response is being executed or inhibited): Although it was initially believed that the system was only involved in arousal, more recent data suggest that its role in the modulation of behaviour might be much more complex. It is proposed to be crucial for task performance optimisation, as well as for the search for alternative behaviours and strategies if the recently adapted one is not ideal (Aston-Jones and Cohen, 2005). These functions are relevant in non-verbal behaviour, but may as well also project onto language comprehension and underlying strategies.

Indeed, Nieuwenhuis et al. (2005) suggest that the ERP component P300 (often referred to as P3), usually following salient and task-relevant input, reflects phasic activity of the LC-NE system, based on the idea that the LC-NE's phasic response plays a crucial role in information processing. In other words: the P 300 is proposed to reflect LC-NE activity which is the result of internal decision-making, and, as a consequence, affects, or, reinforces information processing. Further, Coulson et al. (1998) and later Sassenhagen et al. (2014) suggest that the P300 (b) may be functionally equivalent to the P600, a late positivity component with a posterior scalp distribution and a similar morphology to the P300, elicited upon encountering errors or anomalies in reading and listening. Based on this "P600-as-LC/NE-P3" theory, Demberg and Sayeed (2016) argue that potential similarities may arise between the (possibly LC-NE related) ICA index and the P600 component. That is, as something unpredicted, or surprising (on the structural or on the word level) is encountered by the listener, the ICA possibly reflects signals from higher cortical areas (involved in language processing) to the LC-NE system, indicating the need for further processing resources. The

suggested correlations further fit well into the modern view of language, perception and action dynamically interacting (Spivey et al., 2009) and especially the correlation between the ICA and a more established ERP measure would open up new insights and possibilities for research involving overt attention.

We, however, decided to focus on the N400 in our experiment, since we hypothesise that the pupillary measure corresponded to the surprisal of the respective word, and it has already been shown that the N400 can also predict word surprisal (Frank, 2013b). It was hence the most promising component with respect to our initial hypothesis. Indeed, results from the presented studies **Experiment 4** and **Experiment 5** showed similarities in sensitivity of the pupillary ICA index and the ERP component N400, with respect to visually influenced surprisal and effort of processing a target word.

This, however, does not imply that both measures are directly comparable, nor that such a correlation would hold in any other experimental setup. Instead, we propose that similarities between the index and the N400, as assessed in the presented experiments, may be caused by the fact that (especially in the ICA **Experiment 4**, where eye movements were allowed) no alternative factor caused increased “alertness“ (LC activity). In other words, listeners may have settled on a strategy to ideally solve the task and process the sentences and this strategy proved efficient. Hence, no additional resources were needed and no alternative strategy had to be found. In a different experimental setup, where other factors may cause increased LC activity, or a different paradigm, better optimised for the observation of the P600, we possibly would have found similarities between the ICA and the P300, or, respectively the P600.

In our **Experiment 6**, for instance, where visual complexity was increased, and **Experiment 7 B**, where participants were asked to keep their eyes fixated, a different factor may have increased "alertness". In those cases, the search for a more efficient strategy and the associated increase in alertness, or, attentiveness for that matter, could cause the increased ICA values and the lack of differences between conditions, which would then be invisible. As a result, possibly no similarity to the N400 would have been measurable. The (*tonic*) NE-activity reflected in the ICA may hence not exclusively and directly be related to language processing – at least not in those cases.

This would imply that the ICA, especially when related to NE-activity, can possibly not be projected to a single (language related) component, but rather show similarities with different components in different experimental setups. In the case of **Experiment 6**, and **Experiment 7 B** the ICA very likely reflects overall attentiveness, potentially due to rational adaption of the listener to the given task demands and no similarities. Here, possibly no similarity to the N400 or any other component would have been found, while in other setups, the ICA may be very similar to the P600 or the P300. It is an interesting question for future

research how exactly neuromodulators are involved in the pupil contraction culled by the ICA and in the elicitation of the various ERP components.

Chapter 7

Formalisation of visually inspired surprisal

In addition to the initial questions about how non-sequentially presented, visual information is behaving statistically when being integrated and evaluated with linguistic information (i.e., how and when visual information affects the comprehender's mental model statistically, and, possibly hereby the target word expectations and actual (linguistic) processing effort), we aimed at quantifying the measured effects in order to make them assessable for statistical models of language processing.

We initially hypothesised that participants could either perceive the *probabilistic details* of their visual environment, hereby extracting information from both modalities that can significantly influence their current believes and probabilistic expectations. Alternatively, we suggested that they could extract rather coarse grained information, such as, for instance, whether or not the visual scene is congruent with what they hear. Although verb-driven anticipatory eye movements in our experiments did not directly affect the verb's processing effort (i.e., as measured by the ICA) they did reveal that comprehenders considered each object matching the constraints introduced by the verb as possible target referent. In line with the eye movement patterns showing that listeners concentrated on the objects matching the verb, we found that linguistic processing effort for the target noun differed in a way that could only be explained by a fine grained, *probabilistic evaluation* of the multi-modal context prior to actual word. That is, the noun was more or less predicted, depending on how many target competitors were considered after evaluating the visual context with respect to the only linguistic constraint given, namely the verb. In extension of Frank and Goodman (2012), who showed that listeners can exploit features such as visual salience to assess a speaker's intended referent in the context of a referencing game, as well as in line with (e.g., Hare et al., 2007), who proposed that (linguistic) context can have graded effects on target word

processing, we hence suggested the following: The evaluation of combined extra-linguistic and linguistic information is interleaved whenever it is possible, and visual information can affect the mental model, as well as resulting expectations and the associated processing effort for target words statistically. That is, namely in a *probabilistic* way (i.e., as long as no factors such as a time constraint keeps comprehenders from doing so). Because the influence of visual information on word expectations and linguistic processing effort showed to be probabilistic, it is very likely that the measured effects on processing effort can be described by surprisal, in a similar way as the effect of linguistic context has shown to be predictable by surprisal (see, e.g. Frank, 2013b). That is, of course, given the formula can be adapted in order for it to account for both, linguistic *and* visual information (since multi-modal context evaluation is interleaved, linguistic probabilities are still needed, but are majorly influenced by visual context probabilities).

We approach this challenge in a consecutive step by proposing an adaption of Hale (2001)'s surprisal formula to enable the additional consideration of information extracted from co-present visual context, as given in situated language processing. We further provide a first prove of concept by applying the new, adapted formula to LM (language model) derived probabilities in order to test whether and how accurately it can predict the results measured in our experiments.

If the formula is potent in the context of our LM derived probabilities and experiment results, we propose that this extended formula can be used to approximate processing difficulties in future VWP set ups similar in design. Results could further extend the psychological validity of the concept of surprisal, even in situated comprehension, hereby adding to a body of current evidence for the importance of rational approaches in describing processing difficulties in language comprehension.

7.1 The formula

This paragraph elaborated the adaption of Hale's surprisal formula to account for visual context. That is, we use the potential of the rational, mathematical framework to quantify the influence of visual information in order to provide a description of linguistic processing effort in situated communication, and hereby a tool for estimating the influence of VWP designs on processing effort. The presented extension of surprisal is based on various experiments proving that surprisal can predict effort related to language processing in *linguistic contexts*, as assessed by different measures (see, e.g. Demberg and Keller, 2008; Frank, 2013b; Smith and Levy, 2013) on the one hand, and on our finding that the information theoretic concept

of surprisal is specifically promising to be able to account for our gathered data on the other hand.

Extending the formula (to make it account for the effect of what listeners see on what they expect to be the referent) specifically means to implement a possibility to restrict linguistic probabilities to only the visually presented objects, since our data showed that comprehenders specifically considered those objects in the display that matched the verb. In other words, in order to formalise visual context effects, we assume the basic linguistic probabilities for the target word options, given the previous linguistic context (i.e. the verbal constraint), but *only* relative probabilities corresponding to the visually presented options. This way, the exact number of visually presented competitors (as opposed to an undefined amount of competitors in a purely linguistic context) is important, as well as the relative probabilities of the corresponding words in the linguistic context.

That is, given Hale's definition of surprisal as a concept for quantifying the amount of information conveyed by a *unit* via its *predictability* (Shannon, 1949):

Predictability of a $unit_i$ is:

Pred $(unit_i)$ = P$(unit_i|$Context)

Hence, the amount of information conveyed by the unit is quantified as:

Surprisal $(unit_i) = log\frac{1}{P(unit_i|Context)}$

Based on the observed probabilistic influence of visual context information on linguistic processing effort, we now suggest the following approach to determine surprisal from the relative probabilities of *only* the co-present objects:

Initially, the number of target possibilities, as considered by the comprehender, must be defined, since this information is essential in order to get the accurate probabilities later in the process. We hence define λ as the sum of all linguistically-derived probabilities of all potentially upcoming words (i.e., all possible words a comprehender possibly have thought of after hearing the verb in purely linguistic context), resulting in 1. That is, assuming the contextual information is reliable, and given a certain constraint introduced by the verb, λ is the sum of all known objects matching the verb (e.g., all *spillable* things).

Now we implement the restriction of the visual context as suggested by our results, namely that, in the case of visually co-present objects, comprehenders will first and foremost consider the objects in sight to be possible references. Hence, "all potentially upcoming words" in this case means all potential target options among the depicted objects that match the verb, since the comprehender evaluates verbal information with respect to the visual context, causing a preference for the depicted objects over other possible non-present continuations. The depicted objects matching the verb (all *spillable* objects shown) therefore compose the new, well describable set of target reference alternatives.

Thus, as opposed to purely linguistic contexts, where the amount of target word possibilities thought of by the comprehender is hardly definable, λ in visual contexts reduces to the sum over visually presented, relevant nouns:

$$\lambda = \sum_{w_i \in PotentialTargetObjects} p(w_i|w_1,...,w_{i-1})$$

Now that the sum of all objects considered is defined, a further definition is needed, namely of how the probability for the actual target word to come up in a multi-modal context (i.e. relative to the words corresponding to other possible depicted target options) can be calculated. In a multi-modal context, especially when information is presented simultaneously, input from all different modalities needs to be considered. In our case, this means that not only information from the visual modality, namely the sheer presence of competitors, needs to be considered when calculating the probability of the actual target to come up, but also information from the linguistic level, namely the linguistic probabilities of all target options is important (i.e. it can cause preferences for some over other objects) and should not be neglected. In other words, since the evaluation of both modalities has shown to be interleaved, linguistic level information such as cloze probabilities is still relevant amongst the displayed target options. That is, by dividing the probability of a single target noun w_i by the amount of visually present alternatives (λ), the probability of w_i to come up as the actual target object of the sentence will be scaled to the available set of alternatives while preserving the relative (linguistic, LM derived) probabilities. The formalisation of the influence of visual information hence consists of the basic, linguistic surprisal of only the visually presented target word options. Visually-informed surprisal is thus proposed to be:

$$S(w_i) = -\log_2 \frac{p(w_i|w_1,w_2,...w_{i-1})}{\lambda}$$

Notice also that the proposed extension is not adverse to the validity of the formula in purely linguistic context, since the sum of probabilities for all possible nouns – if no visual context is given – again would be 1, resulting in the original linguistic surprisal value.

7.2 Applying extended surprisal - a first proof of concept

As previously mentioned, one advantage of the formalisation of visual context effects on (language related) processing effort is that the formula can be used on language model data in order to predict processing difficulty data in the experimental set up. Here, we hence test the reliability of the proposed formula extension by applying it to surprisal values that are

calculated based on LM derived probabilities. In other words, we used a language model to assess the linguistic probabilities of only the objects that matched the verb *and* were represented by the different pieces of clip art in a display and calculated situated surprisal based on these probabilities. Specifically, we calculated the probability distributions over objects, given the prior verbal constraints (translated into English), using the language model described in Tilk et al. (2016). This neural network model is especially suitable for our purposes, as it was specifically trained to represent event-relevant context and predict human thematic fit ratings, making it reasonable to calculate event-level surprisal from the model's probabilities.

We applied the extended surprisal formula proposed in the previous paragraph to the model-derived values for our items in the different experimental conditions used in *Experiment 4* and *Experiment 5*. That is, **1**, **3** and **4** (note that **0** is not calculated, due to the more complex processing of the mismatch between visual and verb information which needs further experimental examination before being formalised) possible competitors among the 4 pieces of clip art presented in each display, while the sentence remained the same in each condition (i.e. "the man spills soon the water" in three different visual contexts, either showing one, three, or four spillable things). We report the following results:

Surprisal on the target noun (plausible)[6], given the constraining verb (e.g. *spill the water*) in a purely linguistic context:
$S(noun)_{ling.only} = -\log_2(0.008)$
$S(noun)_{ling.only} = 6.96$

Surprisal on the target noun (plausible), given the constraining verb (e.g. *spill the water*) in a visual context featuring one possible target object (condition **1** in *Experiment 4/5*):
$S(noun)_{cond.1} = -\log_2 \frac{0.008}{0.008}$
$S(noun)_{cond.1} = 0$

Surprisal on the target noun (plausible), given the constraining verb (e.g. *spill the water*) in a visual context featuring three possible target objects (condition **3** in *Experiment 4/5*):
$S(noun)_{cond.3} = -\log_2 \frac{0.008}{(0.008)+(0.005)+(0.01)}$
$S(noun)_{cond.3} = 1.51$

[6]Probabilities of all nouns corresponding to depicted competitors, given the preceding verb (e.g. SPILL_WATER, SPILL_JUICE, etc.) were calculated using the Language Model described in Tilk et al. (2016).

Table 7.1 **Representative example Item** with average values for classical linguistic surprisal, visually informed surprisal and the corresponding ICA values from *Experiment 3*.

Item	Condition target options	Surprisal [7] (LM derived prob.)	Surprisal *visually-informed*	ICA
	1	6.96	0.0	17.6
(1) *The man spills soon the water*	3	6.96	1.51	19.2
	4	6.96	1.61	20.1

Surprisal on the target noun (plausible), given the constraining verb (e.g. *spill the water*) in a visual context featuring four possible target objects (condition **4** in *Experiment 4/5*):
$S(noun)_{cond.4} = -\log_2 \frac{0.008}{(0.008)+(0.005)+(0.01)+(0.002)}$
$S(noun)_{cond.4} = 1.61$

Note, however, that the absolute ICA values we measured in the visual studies were overall significantly higher, compared to the Experiments featuring purely linguistic contexts. We proposed that this effect is most likely attributable to the increased effort needed to simultaneously process information from multiple modalities. Hence, visually-informed surprisal alone cannot account for *all* kinds of cognitive effort involved in the processing of multi-modal information. It is, however, as mentioned previously, an appropriate predictor of differences in surprisal-based processing effort in VWP setttings, where the usual surprisal formula can not account for the data. In table 7.1, we show how the adapted surprisal formula can accurately predict the ICA data (averaged across condition, as measured in *Experiment 4*)collected in our experiment, while surprisal without the situated component fails to accurately predict the differences between conditions.

Column five shows the averaged ICA values in each conditon, while column four shows (visually informed) surprisal values achieved by applying the proposed formula, and column three contains the linguistic surprisal values as calculated from LM derived probabilities, along with a representative example item from the stimuli set. Note how the purely linguistic surprisal values cannot describe the changes in ICA values in response to the different probability profiles of the varying visual scenes, "visually-informed" surprisal can.

By applying our proposed surprisal formula extension on LM derived probabilities, we showed that is could reliable predict our measured ICA data from our VWP set up. We hence

[7]Condition 0 is not shown in the table as the complex processing of the mismatch between visual and verb information can not be formalised without further experimental examination of this process.

propose that the extended, or situated, surprisal formula is a suitable tool for the estimation of effort in future experiments.

Chapter 8

General Discussion

Summarising the results presented in this thesis, a series of experiments using behavioural, pupillary and electro-physiological measures in different paradigms have provided important evidence for the major role and effect of visual information on the statistics of a comprehender's mental model and the resulting word expectations, as well as actual, surprisal-based processing effort.

8.1 The effect of visual information on the mental model, word expectations and processing

Although it is known that, in purely linguistic, sequentially encountered contexts, comprehenders extract probabilistic information, and based on this information, upcoming linguistic units can be expected, it was so far unclear whether the same holds for extra-linguistic information that is not presented sequentially but rather allows for an immediate and continuous assessment of the respective information.

The importance of increased knowledge about the influence of visual information is not least reflected in the discussions about whether or not the VWP is an appropriate paradigm to observe predictive processing. As mentioned previously, considerable criticism with respect to the potential effect of visually presented target references on target word expectation in language processing comes, for instance, from Huettig and Mani (2016). The authors suggest that prediction-based mechanisms are not a necessary component of language processing or acquisition and mainly occur in encouraging context, such as VWP setups, highly suggesting a target option to the comprehender and hereby creating an unnaturally ideal condition, artificially enforcing target word expectations (Huettig and Mani, 2016). Altmann and Mirkovic (2009), on the other hand, argue that visual scenes can basically reflect realistic

complexities and circumstances as given in real-world situations. Visual contexts in the VWP are not overly encouraging per se and results are (to some extend) transferable to natural situated language processing.

We hence examined when and how interleaved comprehenders evaluate multi-modal contexts and measured the precise impact of visual information on the statistics of the comprehender's mental model, as well as on the resulting target word expectations, and finally, the related processing effort.

We initially compared setups with and without visual information. Data revealed differences in surprisal-based processing effort only when visual information was given. No such effects were measured on the same sentences in the absence of visual context. In other words, very high level predictions in purely linguistic context (predictions about target words were not made beyond rough semantic categories) were significantly altered by information extracted from the additionally and simultaneously presented visual context. Results showed that the parallel in situ provision of additional context information via the VWP caused an overall increase in processing effort throughout each trial, attributable to the extra information that had to be processed at the same time. Most importantly, however, the data revealed that visual information is evaluated in combination with language and can affect at least processing of the target word. Interestingly, although the patterns found on the target object were predictable by surprisal extracted from the multi-modal context, which speaks for a detailed evaluation of this context in advance, we did not find any significant effects directly on the verb prior to the target noun. This is especially surprising because the verb's constraints drove anticipatory eye movements in all our VWP experiments (i.e., as previously found by Altmann and Kamide (1999)).

We interpret the eye movement and ICA data measured at the verb as to reflect listener's mapping of linguistic information to the visual context as soon as the grounding of presented information was possible, in order to evaluate the multi-modal context and to inform *probabilistic* expectations about the target noun (which resulted in facilitated processing for those nouns), while participants – despite looking at objects matching the verb – refrained from excluding distractors from a possible set of considerable options.

We will elaborate more on the role of eye movements and the possible reduction of visual uncertainty (entropy) in the subsequent paragraphs. For now, it is important to note that the evaluation of visual and linguistic information is interleaved, while visual information does not affect the statistics of the comprehender's mental model at *each* word (i.e., as linguistic information would), but rather at *certain* time points during the input. In this case, the verb's information was used to evaluate the visual context with respect to which objects matched the verb and could hence be the sentence's referent. By deploying a design in which only the

number of possible target referents displayed was manipulated, while no variation between conditions was featured on the linguistic level, we could further show that the evaluation of the visual context, as driven by the verbal constraints, was not restricted to a rough decision of whether both are somehow coherent, but was indeed probabilistic. In other words, with respect to the initial questions, the gathered data showed that a) the immediate – as opposed to subsequent – presentation of information due to the visual context caused overall increased processing effort, b) the interleaved evaluation of two context modalities happened as soon as reliable context information could be extracted at the verb, and, finally, c) the statistical influence of visual information is probabilistic.

In a sense, these results are in line with Huettig and Mani (2016), who proposed that predictive processing "provides a unified theoretical principle of the human mind" [line 3], which means that, whenever an informative context is provided (i.e., even when this information is provided on a non-linguistic level), information from this context can be analysed with respect to predictive cues concerning upcoming information that is compliant with the achievement of the current intrinsic aims of the comprehender. However, according to the authors, this also shows that visual context always adds substantial information and hereby inevitable alters purely linguistic predictions. In other words, visually informed target word expectations are most likely *often not identical* to purely linguistically inspired predictions. Although this is an important aspect of the VWP that needs to be acknowledged when using the paradigm for the observation of predictive processing, we do not think it necessarily entails that the VWP is inherently overly suggestive, hereby unnaturally "encouraging" expectations. We neither propose that predictive processing mainly (or even exclusively) comes up in what Huettig and Mani (2016) called "prediction encouraging"contexts.

Indeed, the multi-modal setups used in this thesis were often ambiguous with respect to the target referent. Visual contexts were in some cases even designed to feature a very high complexity, hereby not providing an ideal or specifically encouraging context for predictive processing. Along with Altmann and Mirkovic (2009), we hence propose that VWP contexts can in general share enough overlapping features with real-world situations in which situated language processing occurs in order for researchers to draw conclusions about predictive mechanisms active in such situations. As a consequence, we would like to propose that predictive processing – as a very likely general, rather than language specific, mechanisms – does not necessarily require encouragement but can occur whenever it is not inhibited or impossible.

Our interpretation is hence in line with both, the hypothesis that predictive processing happens as long as it is not inhibited (as opposed to only when it is encouraged) and the idea that it is not vital for language comprehension: We argue that various factors may affect

expectations, as well as that high, or even low level predictions and expectations are made as soon as it is possible and not aggravated. The latter conclusion is not only based on the observation that predictive processing occurred in our suboptimal, often ambiguous, setups but also on the observation that any additional information given was eagerly used by the comprehenders to anticipate and expect target words.

Specifically Huettig and Mani (2016) argue, that the evaluation of a context's statistical regularities does not necessarily prove that there is prediction going on. What is unclear at this point is whether this is rather an issue of terminology. Especially when assuming predictive processing might be a general cognitive mechanism (in fact, physiological evidence for predictive processing exists on the field of neuroscience, while it can further explain various cognitive phenomena (for a review, see Euler, 2018)) used by organisms in everyday situations in order to prepare for action and reaction to more or less foreseen things, it is questionable what this evaluation of statistical regularities would be good for, other than predictive processing. Whether predictive mechanisms are a side product in the case of language comprehension cannot be answered in this thesis. However, even if they are, it is reasonable to believe that evaluation of statistical context features is a part of it.

With respect to the widely discussed levels of granularity in predictive processing (e.g., Fruchter et al., 2015; Van Petten and Luca, 2012), and the closely related question about the efficiency of more fine grained predictions (e.g., Federmeier, 2007), our data supports the view that predictive processing is highly flexible and adaptive, which in turn explains why various factors may have an impact. More specifically, we understand that the purely linguistic context featured in the initial experiments did not provide any information that justified more fine grained predictions (i.e., beyond broad semantic category, which is why only the object in "the man soon spills the book" needed more effort to process), since any prediction based on more than the presented context information would have been extremely error prone (note that no highest cloze items or idioms were used). The additional information provided by the visual context in the subsequent experiments, however, did allow for more reliable expectations. Possibly not only due to the extra information provided by the depicted objects, but maybe also because of their reliablity (i.e., whenever target options were depicted, one of them was always mentioned in the sentences, see also Delaney-Busch et al. (2017)). Here, one could again argue that these are ideal lab conditions, rather than a proxi of realworld situations. However, one would assume that no misleading information would be given in a situated conversation situation either. In sum, this hints at a consideration of all available information (possibly even global context information such as the reliability oft he context) in order to adapt the specificity of expectations about upcoming words – that is, with respect to a certain risk-benefit equation. For instance, participants in our experiments did not risk a

decision for one out of four possible target options without having any informational evidence for this choice (e.g. in the form of an additional adjective hinting at one oft he objects). Instead, expectations were just as specific as evidence from direct contextual information in the experiment allowed them to be. We specifically use the term risk, here rather than cost, as up to now, little evidence has been found for cost of dis-confirmed predictions or even expectations. Here, risk differs from cost, especially when hypothesising that prediction is a general, rather than a language specific mechanism: Risk may simply involve factors, other than increased processing cost. If prediction is used on various levels of cognitive processing, it may underlie general rules of minimising risks in real-world situations, rather than being bound to specific processing costs. Expectation of upcoming input is hence seen as an adaptive mechanism, which – based on the evaluation of the wider context (i.e., the experiment, rather than just one sentence) – can be more fine or coarse grained but is involved whenever contextual circumstances generally allow for it. This is in line with Wlotko and Federmeier (2015), who suggest that predictive processing results from the brain's flexible use of resources to most efficiently achieve a primary goal (e.g., comprehension).

Given these suggestions, it is necessary to consider what it takes for a context to allow for predictions, or in our case, expectations. As already pointed out, no fine grained expectations were made in the purely linguistic context, while the same sentences resulted in differences in surprisal-based processing effort for the target noun.

The data from **Experiment 6** additionally showed that increased visual complexity can lead to a decrease of a priori expectations of upcoming linguistic information. Here, patterns of anticipation in the eye movement data waned as more competitors were presented among the eight objects in the displays, suggesting that, although no ICA effects were found on the noun, expectations were more difficult as more competitors were shown and hence slowly decreased in adaptation to the context. In the condition where all but one object were possible target referents, it took participants even slightly longer to look at the representation of the target object when it was already mentioned, showing that participants had no prior expectations and started retrieving the matching object name as they heard the word. In line with Spivey et al. (2001), we interpreted the increased processing time required for visual search as related to the more difficult incremental interpretation of the referring expression due to it not being expected. This does not only hint at an adaptation of the comprehension mechanisms to the recent context, which causes multiple factors such as time or complexity, to affect the granularity of expectations. Interestingly, this also suggests something very important: Namely, that the pupillary measure could cull different sorts of effort, some of which might not directly be related to language processing but, for example, to the increase of attentiveness and possible, physical reactions. This is due to the fact that contractions

culled by the measure may be a result or symptom of an increase in certain neuromodulators, being very difficult to separate with respect to what caused their activity. Further evidence was found in **Experiment 7 B**, where the oppression of overt attention as reflected by eye movements caused an increase in task difficulty which again resulted in a similar, very specific ICA pattern (for a more detailed discussion, see 8.1). In both cases, different factors could have caused increased attentiveness in adaption to the context, resulting in similar ICA patterns.

In line with Huettig and Mani (2016), these results suggest that predictive processing is not necessary for language processing and comprehension. It seems to be the case, however, that predictive processing only wanes as it must, as opposed to only occurring when it is encouraged, which is in accordance with Delaney-Busch et al. (2017), who suggest that comprehenders adapt their predictions to the predictive validity of a current experimental environment, as well as with Wlotko and Federmeier (2015), who propose predictive processing depends on the brain's flexible use of available resources for achieving a primary goal.

8.2 Quantifying the effect of visual information using Information-theoretic concepts

In addition to measuring and observing the (statistical) effect of visual information on word expectation and processing effort, we aimed at formalising our results in order to make them statistically assessable. We hence chose the currently most potent and promising mathematical framework available for the quantification of information content, predictability and processing effort in psycholinguistic literature: Information theory. From this framework, we specifically chose the concepts of surprisal and entropy (reduction) (Hale, 2001; Linzen and Jaeger, 2014; Shannon, 1949; Smith and Levy, 2008) to observe whether they would be suitable to explain and formalise our findings (i.e., results from our pupillary and ERP measures assessing processing effort).

We hypothesised that verb driven anticipatory eye-movements could be related to a reduction of (visual) uncertainty. That is, they could be attributable to an actual decision by the comprehender to concentrate on the matching objects, while possibly excluding the unmatching ones from the domain of possible targets. More information contributed by the verb would hence lead to a further reduction of visual uncertainty (or entropy) and was thought to result in increased processing effort as more objects can be excluded as possible targets. In other words, if processing effort for the verbs differed with the amount of uncertainty

they reduced when their respective constraint information is mapped to the visual context, those patterns were thought to be describable by entropy reduction (Hale, 2003; Shannon, 1949). Our hypothesis was based on evidence from recent literature: Linzen and Jaeger (2015) found that uncertainty about a probabilistic outcome would be best quantified using Shannon's notion of entropy, and Linzen and Jaeger (2014) proposed that *total* entropy (over sentence parses, as opposed to single step entropy) was a significant predictors of reading times, further leading them to conclude that any complete model of sentence processing should imply entropy *and* surprisal. Moreover, Maess et al. (2016) found increased activity in an MEG study for highly constraining (i.e., predictive), relative to less constraining *verbs*, which was proposed to reflect the pre-activation of the expected nouns. Most interestingly, this suggestions was fortified by the author's finding the inverse activity pattern in the evoked responses of the *nouns* (i.e., increased activity for less predicted nouns). Although Maess et al. (2016) do not explicitly mention it, the patterns in their results would be describable by entropy reduction (on the verb) and surprisal or entropy reduction (on the noun).

Especially by deploying the pupillary measure of effort while allowing for free eye-movements, we approached the challenge of assessing the impact of visual information on the processing effort for not only the noun, but also the verb driving anticipation. Interestingly, although patterns of anticipation were repeatedly measured in the eye-movement data, and processing effort for the nouns differed with respect to their multi-modally derived surprisal, no effects on effort relatable to an active reduction of visual uncertainty were found at that point. It hence had to be considered that eye movements were not related to an active reduction of visual uncertainty (entropy) that affected the statistics of the mental model directly at the verb. Alternatively, the ICA could be insensitive to this sort of effort (as opposed to e.g., linguistic processing effort). Therefore, the same design, stimuli and task were used in a follow-up experiment, where effort was assessed in the EEG.

Here, we could mainly replicate the surprising null results on the verb (i.e., interestingly, apart from a mismatch effects in condition 0, where nothing *spillable* was depicted), again showing no effects attributable to entropy reduction, while processing effort for the nouns again differed with respect to the target word's multi-modally derived surprisal. Especially the interesting mismatch effect on the verb in the EEG (while still no effects of entropy reduction were found), as well as eye movements patterns from the previous experiments suggest that the verb information is being integrated with information from the visual context, and the statistical regularities of the multi-modal context are detected (as visible from the effects found on the noun). However, the verb information did not cause an effect attributable to a definite exclusion of objects and this null result could not be (merely) explained by the ICA's potential insensitivity.

Instead, a reasonable interpretation is that eye movements in anticipation of the noun reflect comprehender's shift in attention to possible targets considered but do not lead to a decision for an exclusion of distractors based on the information conveyed by the verb constraint, which is simply not informative enough. Alternatively, it is possible that anything beyond a mere attention shift, such as the active exclusion of visually presented objects does not happen because "ignoring" a present object – even though it might not match the verbal constraints – might be difficult and effortful, as well as incommensurate with the aim of comprehending in order to solve the task and is hence not beneficial but rather hindering. When interpreted with respect to the idea that predictive processing is a general mechanism, the latter seems reasonable because actively ignoring an information (even if this information is thought to not be important for the current interpretation) may simply be risky in a real world situation, as the classification of the information as being rather unimportant might need to be revised quickly if anything unexpected occurs.

The second hypothesis we had initially was that, as a result of mapping verbal constraint to visual information and evaluating the multi-modal context, differences in processing effort for the more or less predictable nouns would occur which were thought to be describable by the concept of surprisal. We found that – as soon as visual context was presented in addition to the sentences – processing cost for the same target noun in the same sentential context differed with respect to the statistical regularities of the multi-modal context (i.e., the verbal mapped to the visual information) and that these differences are perfectly describable by surprisal. More specifically, the fewer competitors were displayed, the more predicted and the less surprising was the actual target noun and the more competitors were among the displayed objects, the more surprising was the actual target noun. Now, one could argue here whether the measured results, rather than being attributable to a general underlying mechanism involved in evaluating visual context in order to expect words, may be caused by noun processing being different from verb processing. In that case, even in the context of a different syntactical construction, effects presumably related to visual context and predictive processing would not show up on any other word. We hence supervised a Master's Thesis by Christine Muljadi that investigated on this question. Interestingly, results from this thesis showed that similar patterns to those that we measured on the nouns (i.e., relatable to visual context effects on the mental model as well as on the resulting expectations) can also be measured in the ICA on verbs if their positions in the stimuli sentences are turned around (e.g., "*Sag mir, ob* ***das Wasser****, das der Mann* ***verschüttet****, links ist*"). That is, graded surprisal effects were found on the sentence - final verb and no effects (of uncertainty reduction) were found on the preceding noun (Muljadi et al., 2019). This highly supports the interpretation of our noun results as being attributable to the same noun being more or less predictable in the

same sentential context – simply due to differences in the visual context which is evaluated by the comprehender in advance to hearing the target word.

We hence proposed that the incremental evaluation of statistical regularities in the context takes place prior to the noun (even when eye movements are inhibited as in the EEG) and results in effects on the target word in either measure. That is, visual and linguistic information are mapped, integrated, and evaluated in combination. This is done in a way such that visual information can affect (linguistic) processing effort in a similarly probabilistic way as purely linguistic context information does. The important difference between visual and linguistic context is, however, that visual context is immediately simultaneously assessable, which is likely to result in visual information not affecting the mental model at each word of the input but only at certain times (for instance, at words that allow a referencing to what is seen via the information they convey, such as verbs or nouns). Hence, multi-modal surprisal of a word reliably predicts on-line processing effort as assessed by at least a pupillometric (ICA) and an ERP (N400) measure, showing that surprisal, and its associated processing effort, is not determined by the linguistic signal alone but rather reflects expectations derived online (at least) from the relevant visual environment in which language is used. This is not only in support of current findings showing that surprisal can predict processing effort as assessed by reading times (Demberg and Keller, 2008; Smith and Levy, 2013) or the ERP component N400's amplitude (Frank et al., 2015), but it further extends those results by showing that surprisal can even be adapted to describe data from situated language processing, where influences of non-linguistic information occur. We provided an extended surprisal formula in *Chapter 7*, followed by an initial validation, showing that language model derived information quantities (i.e., surprisal values calculated based on LM derived probabilities) in combination with the formula were a significant predictor of our ICA values. In sum, this shows that surprisal is indeed capable of formalising the effect of visual information to make it assessable for statistical models of language, hereby showing the immense power of rational approaches in explaining linguistic and psychological data. The actual quantification of visual context effects on language processing and expectations can further provide an orientation when designing VWP stimuli or when setting up and analysing processing data from VWP studies. Our results additionally connect not only linguistic, but multi-modal surprisal to its possible neural implementation in the form of the ERP component N400, which has shown to be correlated in amplitude not only to linguistic *or* visual surprisal (Frank et al., 2015; Kutas and Federmeier, 2011) of a word, but also to surprisal that has to be derived by combining visual and linguistic information. In sum, while (multi-modal) surprisal was efficient in describing our results, entropy reduction could not predict any of our measured effects in the given setups.

8.3 The role of eye movements and overt attention

Based on findings by Altmann and Kamide (1999), we initially stated that it is still unknown whether and how anticipation may affect actual word processing effort. It could have been possible, for instance, that verb driven anticipatory eye movements are linked to a focusing on only specific objects, namely those that are considered to be possible target referents (i.e. matching the verbal restrictions in our contexts). While no effect relatable to a possible (visual) uncertainty reduction during during the verb was found in the different on-line measures of effort, we reliably measured patterns of anticipation in the eye movements. In other words, eye movements as a sign of overt attention did not cause any *immediate* effect on processing effort (this is especially interestingly because they should cause effort, for instance, with respect to motor decisions and action, which is then obviously not reflected by the measures used). From these results, we concluded that anticipatory eye movements are possibly related to an active attention shift towards objects matching the verb, but not necessarily to a decision to exclude non-matching objects, although the differences in processing effort measured at the noun would suggest this. By running the identical VWP experiment twice, with and without allowing for free eye movements, we hence observed what the actual role of eye movements was in situated communication and the evaluation of the context's statistical regularities. Based on the finding that the difference between conditions in which more than one competitor was presented decreased in the EEG, where no eye movements were allowed during the critical region, we hypothesised that overt attention may be crucial for the detailed evaluation of the context. Alternatively, eye movements could generally function as a relieving factor with respect to memory load, but not be crucial for detailed expectations. The lack of overt eye movements could then lead to a less detailed evaluation, since drawing fine grained details from memory, rather than looking at the respective objects, may result in more effort, making it less efficient to evaluate details from the context. In *Experiment 7 A*, we observed differences in surprisal-based processing effort (attributable to probabilistic expectations) on the noun in presence of overt attention, while, interestingly, such differences did not simply *decrease* but were entirely *absent* if overt eye movements were prohibited. What at first sight could be interpreted as a significant role of (anticipatory) eye movements in context evaluation and in forming expectations about the target word, should, however, be interpreted with care. More specifically, the observed null result in the pupillary measure might not be simply attributable to the lack of eye movements. It is important to note, that, although in the previous EEG Experiment, results could have been interpreted as rather coarse grained (i.e., not probabilistic), however differences between one, more than one and no competitors were measured even in the absence of eye movements. At least a rough evaluation of the context is hence possible without eye movements during

the sentence, since participants were given an appropriate preview time and the objects were always in sight. Although we do not claim the ICA and the ERP component N400 are directly comparable, we still suggest that participants' behaviour in both experiments can be compared. Indeed it is very likely that the ICA and the N400 show similarities only in very specific cases (for a detailed discussion, see next paragraph). At the same time it is not certain what sort of effort the ICA reflects in which setups. Hence, an alternative interpretation of the results could have been that the ICA reflects increasing effort of suppression eye movements. Based on the finding that similar patterns resulted from the previous *Experiment 6*, where visual complexity increased while eye movements were allowed, as well as the fact that no lower ICA values were measured during the preview time (where participants were allowed to move their eyes in order to identify the objects in the display), we suggest, however, that this might not be the correct interpretation. At this point, it is reasonable to ask what kind of effort is culled by the pupillary measure and whether it can in fact be entirely attributed to language and information processing. Although effort related to motor decisions was not indexed by the ICA, it is still possible that, for instance, increased attention, which could indeed be related to the neuromodulator NE and activity in the LC/NE system (see, e.g. Aston-Jones and Cohen (2005), who propose that the LC/NE system might be involved in the modulation of behavioural responses), is reflected by the ICA as well. That is, instead of indexing the reaction to more or less surprising target words, the ICA could in this case reflect rising attentiveness related to higher task difficulty under the given circumstances. In other words, we propose that the null result is not directly caused by the suppression of eye movements, but is rather, at least partly, a consequence of increased attentiveness in reaction to higher task difficulty due to the lack of eye movements. This is in line with Spivey et al. (2009), who suggests that ample evidence exists in support of the idea that cognitive processes highly depend and rely on perceptual mechanisms. If eye movements are suppressed, other cognitive mechanisms involved in task solving may be significantly aggravated. In fact, a sign rather attributable to the waning of expectations about the target word was found in condition **7** of **Experiment 6**, where not only no anticipatory patterns were measured although free eye movements were allowed, but also the identification of the object referred to by the target noun was belated, suggesting that it was not anticipated. In sum, we hence suggest that overt attention and anticipatory eye movements are not crucial for probabilistic expectations in the VWP, but do facilitate task solving, probably by preventing and "outsourcing" increased strain of the short time memory.

8.4 The connection between ICA, ERP Components and the LC/NE System

As previously mentioned, our results can be interpreted as to suggest that effort with respect to not only language or information processing, but also with respect to increased attentiveness might be indexed by the same pupillary measure. Although we compared participant's behaviour from an eye tracking and an EEG experiment, it is important to note that we do not suggest that both measures are identical or mappable per se. Instead, we suggest that both measures were reliable indicators of surprisal-based processing effort in our setups. So far, however, we did not state why it is difficult to directly compare the both measurements of effort. There is an ongoing debate in the recent literature concerning the role of the brain's LC/NE system in pupil dilations and stimuli related ERP components. Nieuwenhuis et al. (2005), for instance, suggest that the ERP component P300, which is hypothesised to have a functional role in information processing, indexes phasic activity of the LC-NE system. Phasic activity is thought to facilitate behavioural responses in reaction to a current task and decisions related to solving it, has been proposed to be one of two possible modes of activity of LC neurons, the second mode being tonic activity (Aston-Jones and Cohen, 2005). In other words, based on the idea that the LC/NE system has a modulatory effect on information processing in the sense that it regulates responses to the outcome of internal decisions and behavioural actions with respect to the current task demands and motivation, Nieuwenhuis et al. (2005) propose in their "LC-P300 hypothesis" that those modulatory effects (i.e., noradrenergic potentiation of information processing) of the system are reflected by the component P300, hereby linking measured brain activity to actual behaviour. This connection between the P300 and activity in the LC/NE system make it an interesting component for a comparison with the ICA measurement, which as well has been hypothesised to reflect LC/NE activity Demberg and Sayeed (2016). Coulson et al. (1998) already stated that, instead of focusing on a clear distinction of syntactic and semantic processing (i.e., the corresponding differences between the components P600, following syntactically deviant stimuli and though to index linguistic re-analysis, and N400, elicited in reaction to semantic expectancy), it makes sense to believe that neural implementations might be not strictly language specific. The authors hence propose that specifically the late positivity component P600 resembles – and might be functionally equivalent to – the P300(b), which is traditionally elicited in reaction to non-linguistic, rare categorical events, so called odd balls. Further evidence for the P600-as-P300 hypothesis comes from Sassenhagen et al. (2014), who analysed aligned single-trials, showing that the P600, rather than the P300(b) is aligned with response times. The authors emphasise the relevance of the biological processes

underlying ERP effects (attributable to language) in the sense that interpreting the P600 in sentence processing as an LC/NE-P300, it can be thought of as to index the point in time where a linguistic unit was identified as important and results in adaption in some way. Based on this, finally Demberg and Sayeed (2016) analyse reaction times data from their single task driving setup with respect to alignment with the ICA, and, as a consequence, propose a possible relation between the ICA (thought to be related to pupil contractions attributable to LC-NE activity) and the P600. Based on the results yielded in our experimental setups, on the other hand, we can maximally propose a possible relation between the ICA index and the late component N400, while no P600/P300 similarities were found. We hypothesise, however, that these similarities are not necessarily reliable, as, for instance, in some contexts (e.g., in the increased visual complexity given in *Experiment 6*, as well as in *Experiment 7 B*, where eye movements were prohibited) the ICA values show patterns that are very likely attributable to effort not directly related to language processing. In those contexts, differences in surprisal-based processing effort might not be visible in the pupillary index any more as overall attentiveness increases with task difficulty. In other words, it is likely that additional – possibly not directly language related – factors influenced the pupillary measure, helping to adapt to the overall increasingly demanding current situational contexts. An overall increased attentiveness, as opposed to a stimulus resulting in increased attention at a certain point in time (which should then be visible in the P600), might not be aligned to critical words in the sentences, but might still be related to LC/NE activity. Results may still indicate that the ICA is related to the neurotransmitter NE, however, it might not be mappable to a single ERP component in general. Instead, we propose it is possible that the ICA may show similarities to different stimuli related ERP components, depending on the current situation and, especially, task. That is, as long as no additional factor causes increased attentiveness or even stress, as in **Experiment 6** and **Experiment 7 B**, the ICA may be similar to several components, while increased overall attentiveness (e.g., in reaction to task difficulty and as opposed to an increased amount of information that has to be processed, which we hypothesised to be reflected by overall higher ICA values) possibly overwrites stimulus related responses, making it, at least at this point, impossible to separate the potentially different sorts of effort in the pupillary measure. This interpretation is in support of the basic idea that neural implementations might not be language specific, as stated by Coulson et al. (1998). It further supports the approach that language, perception and action dynamically interact (Spivey et al., 2009) and most importantly suggests, that it might not be possible to always clearly distinguish language related from language unrelated factors influencing pupil contractions as reflected by the ICA. This bears possible implications for future experimental setups deploying the ICA or similar pupillary measures of processing effort with respect to what

sort of effort the experimental conditions and especially tasks could induce. It could become increasingly important to interpret effort with respect to what causes it in order to learn more about how brain activity is linked to behaviour.

8.5 Conclusion

In sum, this thesis aimed at answering vital questions concerning the evaluation and (statistical) influence of context information, inherently different from linguistic information. As a result, we have demonstrated the substantial influence of visual context on word expectancy and actual processing effort associated with surprisal.

We initially hypothesised that, either, similar to linguistic information, visual information could be evaluated probabilistically, or more coarse-grained (for reasons of efficiency), while the combined evaluation of information form both modalities could be more or less interleaved.

Results from our experiments showed that the direct at once presentation of information via a visual context caused increased processing effort in general, as well as that multi-modal surprisal of a word even when modulated merely by the visual referential context predicts both, pupillometric (ICA) and ERP (N400) measures on on-line processing effort. We hence show for the first time (up to our knowledge), that visual information, although being inherently different from linguistic information, is evaluated together with and similar to the latter. This is in line with modern frameworks of language processing, such as the dynamic embodied view of mental activity proposed by Spivey et al. (2009), as well as with the suggestion that comprehenders likely adapt their expectations about upcoming input to the statistical characteristics of their recent (linguistic and extra-linguistic) environment (Delaney-Busch et al., 2017). Similar to the latter, we propose that listeners rationally adapt their expectations (e.g., in their granularity) to the current context. This adaptation is, according to our data, based on the evaluation of integrated information from different modalities. We hence conclude that predictive processing as a possibly not language specific mechanism allows comprehenders to eagerly integrate every bit of information given in any modality, in order to precisely evaluate the context with respect to its statistical regularities, resulting in probabilistic expectations about the target noun, describable by surprisal. Expectations further showed to not have to be encouraged but rather to be computed as long as no factor (such as time) keeps comprehenders from doing so.

In addition to observing how and when visual information affected expectations and word processing, we aimed at quantifying our results, using rational approaches that have already been proven valid in describing psycholinguistic data, namely entropy reduction and surprisal.

More specifically, we suggested that if visual information affected verb processing, possible effects could be described by entropy reduction, while surprisal should be suitable to predict processing effort for the nouns if the statistical effect of visual information is probabilistic. Interestingly, although we could replicate and extend former findings concerning anticipatory eye movements (Altmann and Kamide, 1999), we did not measure any effect related to entropy reduction on the verbs that inspired anticipation in either measure of processing effort. We suggested that, although the verb constraint was exploited to direct looks towards possible target referents displayed (even considering more than one target option), this information was not sufficient for the participant to definitely decide for an exclusion of distractors, possibly because it would be effortful to ignore objects in sight. Most importantly, however, we found that, based on the probabilistic nature of the effects measured on the target nouns, effects of visual information on the processing effort for these words were predictable by an adapted version of surprisal. This lead us to propose an adapted formula to calculate multi-modal surprisal of target words, based on linguistic probabilities of only visually presented objects. An initial prove of concept was delivered by demonstrating that surprisal values achieved by applying this formula were a significant predictor of processing effort in linear mixed effects models. This does not only further support the role of rational approaches in explaining linguistic and psychological data, but also makes aspects of situated language processing assessable for statistical models of language, and, finally, provides a proxi for processing effort when designing VWP stimuli.

Our data additionally allowed us to closer observe the role of eye-movements in visual context evaluation an expectation computing, and hinted at some interesting questions with regard to what kinds of effort are actually reflected by pupil contractions as culled by the ICA:

We found that eye movements are not vital for a probabilistic evaluation of visual (or multi-modal) context, but seem to contribute to facilitated task solving. In experiments where eye movements were inhibited, an interesting overall increase in effort was measured, while no differences in surprisal-based effort were visible. We suggested this overall increase to reflect enhanced attentiveness caused by aggravated task solving due to the lack of eye movements. Actual signs of waning expectations, however, were found in the most complex conditions in increased visual complexity, where time can become a crucial factor. These results may show that effort with respect to different cognitive mechanisms (directly, and possibly only indirectly related to language processing) could be confounded in the pupillary data, making it important as well as challenging to differentiate between them. The question of what sorts of effort are possibly reflected in pupillary data in different experimental setups,

and given different task difficulties, is a fruitful and promising question for the future and definitely needs further research.

Chapter 9

Ethics & Funding

Ethics

The studies involving human subjects were carried out in accordance with the recommendations of the American Psychological Association, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee by the Deutsche Gesellschaft für Sprache (DGfS).

9.1 Funding

The reported research is funded by the Deutsche Forschungsgemeinschaft (DFG) under grant SFB1102. I gratefully acknowledge support by the Cluster of Excellence Multimodal Computing and Interaction (MMCI).

Bibliography

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

Alphen, G. W. V. (1976). The adrenergic receptors of the intraocular muscles of the human eye. *Invest Ophthalmol*, 15(6):502–505.

Altmann, G. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

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

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

Ankener, C. and Staudte, M. (2018). Multimodal surprisal in the N400 and the index of cognitive activity. In *Proceedings of the 40th annual conference of the cognitive science society*.

Ankener, C. S., Sekicki, M., and Staudte, M. (2018). The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in Psychology*, 9:2387.

Aston-Jones, G. and Cohen, J. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28:403–450.

Attneave, F. (1959). *Applications of Information Theory to Psychology: A summary of basic concepts, methods and results.* Holt, Rinehart and Winston.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3):379–384.

Balota, D. A., Pollatsek, A., and Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3):364–390.

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

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1–48.

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

Beatty, J. and Lucero-Wagoner, B. (2000). The pupillary system. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology (2nd ed.)*, pages 142–162. Cambridge University Press.

Bell, A., Brenier, M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, (60):92–111.

Bernstein, R. B. and Levine, R. D. (1972). Entropy and chemical change. i. characterization of product (and reactant) energy distributions in reactive molecular collisions: Information and entropy deficiency. *The Journal of Chemical Physics*, (57):434–449.

Berridge, C. W. and Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, (42):33–84.

Binda, P., Pereverzeva, M., and SO., M. (2013). Attention to bright surfaces enhances the pupillary light reflex. *The Journal of Neuroscience*, (33):2199–2204.

Blackburn, P. and Bos, J. (2005). *Representation and inference for natural language: A first course in computational semantics.* Stanford: CSLI Press.

Coco, M. and Keller, F. (2015). Integrating mechanisms of visual guidance in naturalistic language production. *Cognitive Processing*, 6(2):131–150.

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

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, 13(1):21–58.

Dahan, D. and Magnuson, J. S. (2006). Spoken word recognition. In Traxler, M. J. and Gernsbacher, M. A., editors, *Handbook of psycholinguistics (Vol. 2)*, pages 249–284. Academic Press.

Delaney-Busch, N., Morgan, E., Lau, E., and Kuperberg, G. (2017). Comprehenders rationally adapt semantic predictions to the statistics of the local environment: a bayesian model of trial-by-trial N400 amplitudes. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 283–288.

DeLong, K., Urbach, P., and Kutas, M. (2005). Probabilistic word pre-activation during comprehension inferred from electrical brain activity. *Nature neuroscience*, 8:1117–21.

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

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

Demberg, V., Sayeed, A. B., Gorinski, P. J., and Engonopoulos, N. (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 356–367.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Engelhardt, P. E., Ferreira, F., and Patsenko, E. (2009). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, 63:639–645.

Euler, M. J. (2018). Intelligence and uncertainty: Implications of hierarchical predictive processing for the neuroscience of cognitive ability. *Neuroscience Biobehavioral Reviews*, 94:93–112.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4):491–505.

Ferreira, F. and C. Clifton, J. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25(3):348–368.

Ferreira, F., Foucart, A., and Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, 69:165–182.

Fischler, I. and Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18(1):1–20.

Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(25):998.

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

Frank, S. (2013b). Word surprisal predicts N400 amplitude during reading.

Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing.

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

Fruchter, J., Linzen, T., Westerlund, M., and Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *Journal of Cognitive Neuroscience*, 27(10):1912–1935.

Gambi, C. and Pickering, M. (2017). Linguistic prediction is a non-competitive process: Evidence from the processing of spoken sentences. In *30th CUNY Conference on Human Sentence Processing.*

Ganis, G., Kutas, M., and Sereno, M. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 2(8):89–106.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and 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*, 10:252–269.

Goodkind, A. and Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, pages 10–18. Association for Computational Linguistics.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies.*, pages 1–8.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2):101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4):643–672.

Hare, M., Tanenhaus, M. K., and McRae, K. (2007). Understanding and producing the reduced relative construction: Evidence from ratings, editing and corpora. *Journal of Memory and Language*, 56(3):410–435.

Hawkins, J. (2004). *Efficiency and Complexity in Grammar.* Oxford University Press.

Hayes, T. and Petrov, A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, 2(48):510–527.

Hess, E. H. and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 13.

Huettig, F. and Mani, N. (2016). Is prediction necessary to understand language? probably not. *Language, Cognition and Neuroscience*, 31(1):19–31.

Jackendoff, R. (2002). *Foundations of language.* University Press.

Jaeger, T. and Tily, H. (2011). On language 'utility': processing complexity and communicative efficiency. *Cognitive science*.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language,inference and consciousness.* Harvard University Press.

Kahneman, D. and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 3756(154):1583–5.

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

Kleinschmidt, D. F. and Jaeger, F. T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2):148–203.

Kleinschmidt, D. F. and Jaeger, F. T. (2016). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin & Review*, 23(3):678–691.

Kulke, L. V., Atkinson, J., and Braddick, O. (2016). Neural differences between covert and overt attention studied using eeg with simultaneous remote eye tracking. *Frontiers in Human Neuroscience*, 10:Article 592.

Kuperberg, G. R. and Jaeger, T. (2016a). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience. In press.*

Kuperberg, G. R. and Jaeger, T. (2016b). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1):32–59.

Kutas, M., DeLong, K., and Smith, N. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In Bar, M., editor, *Predictions in the brain: Using our past to generate a future*, pages 190–207. Oxford University Press.

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

Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62:621–647.

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

Kutas, M. and Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163.

Linzen, T. and Jaeger, F. (2014). Investigating the role of entropy in sentence processing.

Linzen, T. and Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: evidence from subcategorization distributions. *Cognitive science*.

Luck, S. J. (2014). *An introduction to the event-related potential technique.* The MIT Press.

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

Mani, N. and 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.

Mani, N. and 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.

Marshall, S. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. *U.S. patent no. 6,090,051*.

Marshall, S. (2002). The index of cognitive activity: Measuring cognitive workload. *Proceedings of the 7th conference on Human Factors and Power Plants*, IEEE:75–79.

Mishra, R. K., Singh, N., Pandey, A., and 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.

Morton, J. (1964). The effect of context on the visual duration threshold for words. *British Journal of Psychology*, 55:165–180.

Muljadi, C., Ankener, C., Sikos, L., and Staudte, M. (2019). Verb surprisal in the visual world. In *32nd CUNY Conference on Human Sentence Processing.*

Nieuwenhuis, S. (2011). Learning, the p3, and the locus coeruleus-norepinephrine system. In Mars, R., Sallet, J., Rushworth, M., and Yeung, N., editors, *Neural Basis of Motivational and Cognitive Control*, pages 209–222. Oxford University Press.

Nieuwenhuis, S., Cohen, J., and Aston-Jones, G. (2005). Decision making, the p3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131:510–532.

Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Wolfsthurn, S. V. G. Z., Bartolozzi, F., Kogan, V., Ito, A., Meziere, D., Barr, D. J., Rousselet, G., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., Donaldson, D. L., Kohut, Z., Rueschemeyer, S., and Huettig, F. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *bioRxiv*.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin and Review*, 21:1112–1130.

Rayner, K., Ashby, J., Pollatsek, A., and Reichle, E. D. (2004a). The effects of frequency and predictability on eye fixations in reading: Implications for the e-z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4):720–732.

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

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

RCoreTeam (2013). R: A language and environment for statistical computing. *Computer software manual*, pages Retrieved from http://www.R–project.org/.

Richardson, D. and Spivey, M. (2004). Eye tracking: Research areas and applications. In Wnek, G. and Bowlin, G., editors, *Encyclopedia of Biomaterials and Biomedical Engineering*, pages 573–582. Marcel Dekker, Inc.

Rommers, J., Meyer, A., Praamstra, P., and Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 3(51):437–447.

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

Sara, S. J. and Segal, M. (1991). Plasticity of sensory responses of locus coeruleus neurons in the behaving rat: Implications for cognition. *Prog. Brain Research*, (88):571–585.

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The p600-as-p3 hypothesis revisited: Single-trial analyses reveal that the late eeg positivity following linguistically deviant material is reaction time aligned. *Brain and language*, 137:29–39.

Scheepers, C. and Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. In Carreiras, M. and Clifton, C. J., editors, *The On-line Study of Sentence Comprehension: Eyetracking, ERP and Beyond.* Psychology Press.

Schilling, H., Rayner, K., and Chumbley, J. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition*, 26(6):1270–1281.

Schwalm, M., Keinath, A., and Zimmer, H. D. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. In De Waard, D., Flemisch, F., Lorenz, B., Oberheid, H., and Brookhuis, K., editors, *Human Factors for Assistance and Automation.*, pages 75–88. Shaker.

Sekicki, M. and Staudte, M. (2017). The facilitatory effect of referent gaze on cognitive load in language processing. In *Proceedings of the 39th annual conference of the cognitive science society*, pages 3107–3112.

Shannon, C. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

Sharbrough, F., Chartrian, G. E., Lesser, R. P., Luders, H., Nuwer, M., and Picton, T. W. (1995). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, (8):200–202.

Smith, A., Monaghan, P., and Huettig, F. (2013). The multimodal nature of spoken word processing in the visual world: Testing the predictions of alternative models of multimodal integration. *Journal of Memory and Language*, 93:276–303.

Smith, N. and Levy, R. (2008). Optimal processing times in reading: a formal model and empirical investigation. *Psychological Science*, 14(4):328–333.

Smith, N. and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128:302–319.

Spivey, M., Richardson, D., and Dale, R. (2009). The movement of eye and hand as a window into language and cognition. In Morsella, E., Bargh, J. A., and Gollwitzer, P. M., editors, *Oxford Handbook of Human Action*, pages 225–249. Oxford University Press.

Spivey, M., Tyler, M., Eberhard, K., and Tanenhaus, M. (2001). Linguistically mediated visual search. *Psychological Science*, 12(4):282–286.

Tamber-Rosenau, B. J., M. Esterman, M., Chiu, Y.-C., and Yantis, S. (2011). Cortical mechanisms of cognitive control for shifting attention in vision and working memory. *Journal of Cognitive Neuroscience*, 23(10):2905–2919.

Tilk, O., Demberg, V., Sayeed, A., Klakow, D., and Thater, S. (2016). Event participant modelling with neural networks. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 91(2):171–182.

Tolman, R. C. (1938). *Principles of Statistical Mechanics.* Clarendon.

Van Berkum, J., Brown, C., Zwitserlood, P., Kooijman, V., and 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.

Van Petten, C. and Luca, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2):176–190.

West, W. and Holcomb, P. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, 12:1024–1037.

Wlotko, E. W. and Federmeier, K. D. (2015). Time for prediction? the effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68:20–32.

Woldemussie, E., Wijono, M., and Pow, D. (2007). Localization of alpha 2 receptors in ocular tissues. *Visual neuroscience*, 24(5):745–756.

Xiang, M. and Kuperberg, G. R. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6):648–672.

Zwaan, R. and Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123:162–85.

Appendix A

Linguistic Stimuli

Experiment 1

Table A.1 **Linguistic stimuli** for **Experiment 2** and **Experiment 3**.
The Table shows all linguistic items in each condition: Each item had a constraining and an unconstraining verb condition (see columns 3& 4), each paired with two different objects (see columns 6&7). Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentence in the VWP studies. The complete set of all visual stimuli is listed in Section B

Item Nr.		Verb constraining	Verb unconstraining		Object 1 plausible	Object2 possible
1	Die Mutter *The mother*	löffelt *sups*	bewertet *rates*	gleich *soon*	die Suppe *the soup*	den Kaffee *the coffee*
2	Der Grossvater *The grandfather*	verschüttet *spills*	bestellt *orders*	gleich *soon*	das Wasser *the water*	das Eis *the ice cream*
3	Die Schwester *The sister*	schmilzt *melts*	kontrolliert *checks*	gleich *soon*	die Butter *the butter*	den Honig *the honey*
4	Die Cousine *The cousin*	montiert *assembles*	ersetzt *replaces*	gleich *soon*	die Antenne *the antenna*	das Motorrad *the motorbike*
5	Der Grossvater *The grandfather*	kocht *cooks*	verpackt *wraps*	gleich *soon*	die Kartoffel *the potato*	die Banane *the banana*
6	Die Grossmutter *The grandmother*	trinkt *drinks*	testet *tests*	gleich *soon*	den Kaffee *the coffee*	den Joghurt *the yoghurt*
7	Der Grossvater *The grandfather*	serviert *serves*	fotografiert *photographs*	gleich *soon*	das Eis *the ice cream*	die Zitrone *the lemon*

Table A.1 **Linguistic stimuli** for **Experiment 2** and **Experiment 3**. .
The Table shows all linguistic items in each condition: Each item had a constraining and an unconstraining verb condition (see columns 3& 4), each paired with two different objects (see columns 6&7). Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentence in the VWP studies. The complete set of all visual stimuli is listed in Section B

Item Nr.		Verb constraining	Verb unconstraining		Object 1 plausible	Object2 possible
8	Der Grossvater *The grandfather*	isst *eats*	nimmt *grabs*	gleich *soon*	das Brot *the bread*	den Pfeffer *the pepper*
9	Der Cousin *The cousin*	poliert *polishes*	erblickt *beholds*	gleich *soon*	das Auto *the car*	den Zug *the train*
10	Der Vater *The father*	kühlt *cools*	reklamiert *exchanges*	gleich *soon*	den Wein *the wine*	die Waffel *the waffle*
11	Die Schwester *The sister*	bestickt *embroiders*	behält *keeps*	gleich *soon*	das Kissen *the pillow*	den Stiefel *the boot*
12	Die Grossmutter *The grandmother*	zuckert *sweetens*	prüft *checks*	gleich *soon*	den Tee *the tea*	die Zitrone *the lemon*
13	Der Mann *The man*	fährt *drives*	sieht *sees*	gleich *soon*	das Auto *the car*	das Schiff *the ship*
14	Der Cousin *The cousin*	näht *sews*	berührt *touches*	gleich *soon*	die Jacke *the jacket*	das Sofa *the sofa*
15	Die Frau *The woman*	schneidet *cuts*	holt *gets*	gleich *soon*	das Brot *the bread*	die Pommes *the fries*
16	Die Mutter *The mother*	flickt *repairs*	vergisst *forgets*	gleich *soon*	die Jeans *the Jeans*	den Besen *the broom*
17	Die Frau *The woman*	bügelt *irons*	beschreibt *describes*	gleich *soon*	das T-Shirt *the t-shirt*	die Socke *the sock*
18	Die Mutter *The mother*	strickt *knits*	bekommt *receives*	gleich *soon*	den Schal *the scarf*	die Decke *the blanket*
19	Die Frau *The woman*	erntet *harvests*	wäscht *washes*	gleich *soon*	den Apfel *the apple*	den Reis *the rice*
20	Der Bruder *The brother*	repariert *repairs*	zeichnet *draws*	gleich *soon*	den Laptop *the laptop*	das Raumschiff *the space ship*

Table A.2 **Linguistic stimuli** for **Experiment 4**, **Experiment 5**, **Experiment 6**, and **Experiment 7**. **Experiment 5** featured 60 additional items shown in the subsequent table.
Each item had four different corresponding scenes in **Experiment 4**, **Experiment 5** and **Experiment 6**, and three corresponding scenes in **Experiment 7**. All visual items are shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences. The asterisk indicates which items have also been used in the previous experiments.

Item Nr.		Verb		Object
1*	Die Frau	löffelt	gleich	die Suppe
	The woman	*sups*	*soon*	*the soup*
2*	Die Frau	verschüttet	gleich	das Wasser
	The woman	*spills*	*soon*	*the water*
3*	Die Frau	packt	gleich	den Koffer
	The woman	*packs*	*soon*	*the suitcase*
4*	Die Frau	fährt	gleich	das Auto
	The woman	*drives*	*soon*	*the car*
5	Die Frau	entsaftet	gleich	die Orange
	The woman	*juices*	*soon*	*the orange*
6*	Der Mann	montiert	gleich	die Antenne
	The man	*assembles*	*soon*	*the antenna*
7*	Die Frau	bügelt	gleich	das T-Shirt
	The woman	*irons*	*soon*	*the t-shirt*
8*	Der Mann	trinkt	gleich	den Kaffee
	The man	*drinks*	*soon*	*the coffee*
9	Die Frau	verbiegt	gleich	die Büroklammer
	The woman	*bends*	*soon*	*the paper clip*
10	Der Mann	würzt	gleich	den Salat
	The man	*seasons*	*soon*	*the salad*
11*	Die Frau	erntet	gleich	den Apfel
	The woman	*harvests*	*soon*	*the apple*
12*	Der Mann	poliert	gleich	das Auto
	The man	*polishes*	*soon*	*the car*
13	Der Mann	pflanzt	gleich	die Blume
	The man	*plants*	*soon*	*the flower*
14*	Die Frau	bestickt	gleich	das Kissen
	The woman	*embroiders*	*soon*	*the pillow*

Table A.2 **Linguistic stimuli** for **Experiment 4**, **Experiment 5**, **Experiment 6**, and **Experiment 7**. **Experiment 5** featured 60 additional items shown in the subsequent table.
Each item had four different corresponding scenes in **Experiment 4**, **Experiment 5** and **Experiment 6**, and three corresponding scenes in **Experiment 7**. All visual items are shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences. The asterisk indicates which items have also been used in the previous experiments.

Item Nr.		Verb		Object
15	Die Frau	bäckt	gleich	den Keks
	The woman	*bakes*	*soon*	*the cookie*
16*	Die Frau	isst	gleich	den Burrito
	The woman	*eats*	*soon*	*the burrito*
17	Der Mann	knackt	gleich	den Tresor
	The man	*cracks*	*soon*	*the safe*
18*	Die Frau	näht	gleich	die Jacke
	The woman	*sews*	*soon*	*the jacket*
19	Der Mann	spült	gleich	den Topf
	The man	*washes*	*soon*	*the pot*
20	Der Mann	schält	gleich	die Zwiebel
	The man	*peels*	*soon*	*the onion*
21	Der Mann	raucht	gleich	die Zigarre
	The man	*smokes*	*soon*	*the cigar*
22	Die Frau	grillt	gleich	die Wurst
	The woman	*barbecues*	*soon*	*the sausage*
23	Die Frau	lenkt	gleich	den Hubschrauber
	The woman	*steers*	*soon*	*the helicopter*
24	Die Frau	stimmt	gleich	das Klavier
	The woman	*tunes*	*soon*	*the piano*
25	Die Frau	belegt	gleich	die Pizza
	The woman	*tops*	*soon*	*the pizza*
26	Die Frau	bindet	gleich	den Schuh
	The woman	*ties*	*soon*	*the shoe*
27	Die Frau	gießt	gleich	die Rose
	The woman	*waters*	*soon*	*the rose*
28*	Der Mann	kocht	gleich	die Kartoffel
	The man	*cooks*	*soon*	*the potato*

Table A.2 **Linguistic stimuli** for **Experiment 4**, **Experiment 5**, **Experiment 6**, and **Experiment 7**. **Experiment 5** featured 60 additional items shown in the subsequent table.
Each item had four different corresponding scenes in **Experiment 4**, **Experiment 5** and **Experiment 6**, and three corresponding scenes in **Experiment 7**. All visual items are shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences. The asterisk indicates which items have also been used in the previous experiments.

Item Nr.		Verb		Object
29*	Die Frau	schneidet	gleich	das Brot
	The woman	*cuts*	*soon*	*the bread*
30*	Die Frau	flickt	gleich	die Jeans
	The woman	*repairs*	*soon*	*the jeans*
31	Der Mann	liest	gleich	die Zeitung
	The man	*reads*	*soon*	*the newspaper*
32*	Der Mann	zuckert	gleich	den Tee
	The man	*sweetens*	*soon*	*the tea*
33	Der Mann	brät	gleich	das Fleisch
	The man	*cooks*	*soon*	*the meat*
34*	Der Mann	repariert	gleich	den Laptop
	The man	*repairs*	*soon*	*the laptop*
35	Der Mann	streicht	gleich	den Stuhl
	The man	*paints*	*soon*	*the chair*
36*	Der Mann	serviert	gleich	das Eis
	The man	*serves*	*soon*	*the ice cream*
37*	Der Mann	rührt	gleich	den Joghurt
	The man	*stirs*	*soon*	*the yoghurt*
38	Der Mann	spitzt	gleich	den Buntstift
	The man	*sharpens*	*soon*	*the crayon*
39*	Die Frau	strickt	gleich	den Schal
	The woman	*knits*	*soon*	*the scarf*
40	Die Frau	buttert	gleich	die Brezel
	The woman	*butters*	*soon*	*the pretzel*

Table A.3 **Additional linguistic stimuli** for **Experiment 5**. All items in the EEG Experiment had extended spill over regions following the verb and the object noun (adverbials, local after the verb and temporal after the noun). Those regions never added substantial information and were never mismatching the sentence content.
Each item had four different corresponding scenes shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences.

Item Nr.		Verb	Object
41	Die Frau	düngt	die Rose
	The woman	*fertilises*	*the rose*
42	Der Mann	sät	die Erbsen
	The man	*seeds*	*the peas*
43	Die Frau	lackiert	den Zaun
	The woman	*paints*	*the fence*
44	Der Mann	justiert	die Uhr
	The man	*adjusts*	*the watch*
45	Die Frau	kompostiert	die Bananenschale
	The woman	*composts*	*the banana peel*
46	Der Mann	verfüttert	das Heu
	The man	*feeds*	*the hay*
47	Die Frau	rollt	das Nudelholz
	The woman	*rolls*	*the rolling pin*
48	Der Mann	baut	das Puppenhaus
	The man	*builds*	*the doll house*
49	Die Frau	pflückt	die Birne
	The woman	*picks*	*the pear*
50	Die Frau	zermahlt	die Kaffeebohne
	The woman	*grinds*	*the coffee bean*
51	Der Mann	graviert	den Ring
	The man	*engraves*	*the ring*
52	Die Frau	leert	den Eimer
	The woman	*empties*	*the bucket*
53	Die Frau	entsteint	die Avocado
	The woman	*pits*	*the avocado*
54	Die Frau	zerreisst	die Urkunde
	The woman	*rips*	*the certificate*
55	Der Mann	beizt	den Schrank

Table A.3 **Additional linguistic stimuli** for **Experiment 5**. All items in the EEG Experiment had extended spill over regions following the verb and the object noun (adverbials, local after the verb and temporal after the noun). Those regions never added substantial information and were never mismatching the sentence content.
Each item had four different corresponding scenes shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences.

Item Nr.		Verb	Object
	The man	*stains*	*the cabinet*
56	Der Mann	schärft	das Messer
	The man	*sharpens*	*the knife*
57	Der Mann	dünstet	den Spargel
	The man	*stews*	*the asparagus*
58	Die Frau	frittiert	die Kartoffeln
	The woman	*fries*	*the potatoe*
59	Der Mann	zerknittert	das Papier
	The man	*crinkles*	*the paper*
60	Die Frau	kürzt	den Rock
	The woman	*shortens*	*the skirt*
61	Der Mann	versprüht	das Parfum
	The man	*sprays*	*the perfume*
62	Der Mann	zerteilt	das Hähnchen
	The man	*dissects*	*the chicken*
63	Die Frau	bastelt	die Laterne
	The woman	*crafts*	*the lantern*
64	Der Mann	röstet	die Haselnüsse
	The man	*roasts*	*the hazelnuts*
65	Die Frau	verstreicht	die Erdnussbutter
	The woman	*spreads*	*the peanut butter*
66	Der Mann	verdünnt	die Farbe
	The man	*dilutes*	*the paint*
67	Die Frau	kopiert	den Pass
	The woman	*copies*	*the passport*
68	Der Mann	salzt	die Pommes
	The man	*salts*	*the fries*
69	Der Mann	kaut	die Möhre
	The man	*chews*	*the carrot*

Table A.3 **Additional linguistic stimuli** for **Experiment 5**. All items in the EEG Experiment had extended spill over regions following the verb and the object noun (adverbials, local after the verb and temporal after the noun). Those regions never added substantial information and were never mismatching the sentence content.
Each item had four different corresponding scenes shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences.

Item Nr.		Verb	Object
70	Die Frau	pellt	die Kartoffel
	The woman	*peels*	*the potatoe*
71	Die Frau	schiebt	den Einkaufswagen
	The woman	*pushes*	*the cart*
72	Der Mann	schlürft	den Cappuccino
	The man	*sips*	*the cappuccino*
73	Die Frau	gart	die Paprika
	The woman	*cooks*	*the bell pepper*
74	Die Frau	schmilzt	den Zucker
	The woman	*melts*	*the sugar*
75	Die Frau	schneidert	den Rock
	The woman	*tailors*	*the skirt*
76	Der Mann	betoniert	die Treppe
	The man	*concretes*	*the stairs*
77	Die Frau	frühstückt	die Waffel
	The woman	*eats (for breakfast)*	*the waffle*
78	Der Mann	püriert	den Lauch
	The man	*mashes*	*the leek*
79	Die Frau	beschriftet	das Heft
	The woman	*labels*	*the note book*
80	Der Mann	archiviert	den Ordner
	The man	*archives*	*the folder*
81	Die Frau	verschluckt	den Kaugummi
	The woman	*swallows*	*the chewing gum*
82	Die Frau	mischt	die Karten
	The woman	*shuffles*	*the cards*
83	Der Mann	renoviert	das Haus
	The man	*renovates*	*the house*
84	Der Mann	zersägt	die Palette

Table A.3 **Additional linguistic stimuli** for **Experiment 5**. All items in the EEG Experiment had extended spill over regions following the verb and the object noun (adverbials, local after the verb and temporal after the noun). Those regions never added substantial information and were never mismatching the sentence content.
Each item had four different corresponding scenes shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences.

Item Nr.		Verb	Object
	The man	*saws*	*the pallet*
85	Die Frau	wirft	den Würfel
	The woman	*throws*	*the die*
86	Der Mann	verzehrt	das Croissant
	The man	*eats*	*the croissant*
87	Die Frau	parkt	das Wohnmobil
	The woman	*parks*	*the camper van*
88	Der Mann	schmiedet	das Schwert
	The man	*forges*	*the sword*
89	Die Frau	mietet	die Schlittschuhe
	The woman	*rents*	*the ice skates*
90	Der Mann	töpfert	die Schüssel
	The man	*potters*	*the bowl*
91	Die Frau	nascht	das Gummibärchen
	The woman	*snacks on*	*the gummy bear*
92	Der Mann	schokoliert	die Himbeere
	The man	*chocolate coats*	*the raspberry*
93	Der Mann	verschrottet	das Motorrad
	The man	*scraps*	*the motor bike*
94	Der Mann	belädt	den LKW
	The man	*loads*	*the truck*
95	Die Frau	faltet	die Serviette
	The woman	*folds*	*the napkin*
96	Die Frau	züchtet	die Orchidee
	The woman	*breeds*	*the orchid*
97	Der Mann	versichert	das Fahrrad
	The man	*insures*	*the bike*
98	Die Frau	konstruiert	die Brücke
	The woman	*constructs*	*the bridge*

Table A.3 **Additional linguistic stimuli** for **Experiment 5**. All items in the EEG Experiment had extended spill over regions following the verb and the object noun (adverbials, local after the verb and temporal after the noun). Those regions never added substantial information and were never mismatching the sentence content.
Each item had four different corresponding scenes shown in B. Column 1 indicates the ID of the item itself and the scenes which were shown together with the sentences.

Item Nr.		Verb	Object
99	Die Frau	verkocht	die Nudeln
	The woman	*overcooks*	*the noodles*
100	Der Mann	raspelt	die Mandeln
	The man	*shreds*	*the almonds*

Appendix B

Visual Stimuli

Figure B.1 *Visual stimuli for* ***Experiment 3***[1]. Tables in A indicate the matching linguistic stimuli presented simultaneously with the pictures.

Item Nr.1

Item Nr.2

Item Nr.3

Item Nr.4

Item Nr.5

Item Nr.6

Item Nr.7

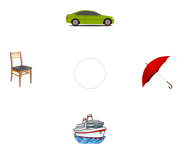
Item Nr.8

Item Nr.9

Item Nr.10

Item Nr.11

Item Nr.12



(m) Item Nr.13

(n) Item Nr.14

(o) Item Nr.15

(p) Item Nr.16

(q) Item Nr.17

(r) Item Nr.18

(s) Item Nr.19

(t) Item Nr.20

[1] All (royalty free) pieces of clip art used in any of our experiments have been retrieved from: http://www.clipartkid.com, http://clipart-library.com, http://www.clipartpanda.com, https://openclipart.org, https://pixabay.com.

Figure B.2 *Visual stimuli for* ***Experiment 4***. Tables in A indicate the matching linguistic stimuli presented simultaneously with the pictures.

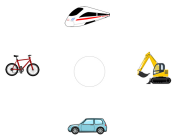
Item Nr.1/0Targets

Item Nr.1/1Target

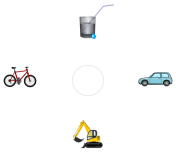
Item Nr.1/3Targets

Item Nr.1/4Targets

Item Nr.2/0Targets



Item Nr.2/1Target



Item Nr.2/3Targets

Item Nr.2/4Targets

Item Nr.3/0Targets

Item Nr.3/1Target

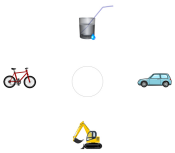
Item Nr.3/3Targets

Item Nr.3/4Targets

Item Nr.4/0Targets

Item Nr.4/1Target

Item Nr.4/3Targets



Item Nr.4/4Targets

Item Nr.5/0Targets

Item Nr.5/1Target

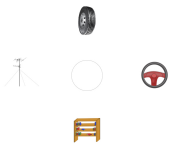
Item Nr.5/3Targets

Item Nr.5/4Targets

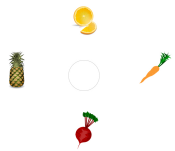
Item Nr.6/0Targets

Item Nr.6/1Target

Item Nr.6/3Targets



Item Nr.6/4Targets

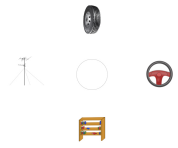


Item Nr.7/0Targets

Item Nr.7/1Target

Item Nr.7/3Targets

Item Nr.7/4Targets

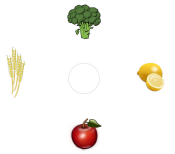


Item Nr.8/0Targets

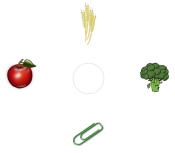
Item Nr.8/1Target

Item Nr.8/3Targets

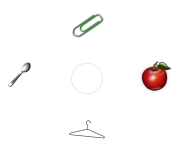
Item Nr.8/4Targets



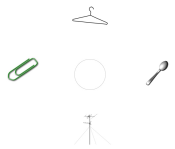
Item Nr.9/0Targets



Item Nr.9/1Target

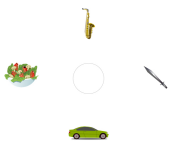


Item Nr.9/3Targets

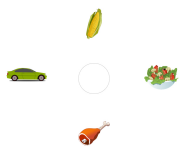


Item Nr.9/4Targets

Item Nr.10/0Targets

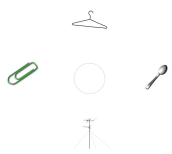


Item Nr.10/1Target

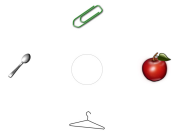


Item Nr.10/3Targets

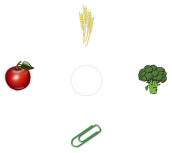
Item Nr.10/4Targets



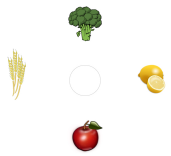
Item Nr.11/0Targets



Item Nr.11/1Target

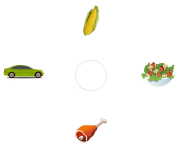


Item Nr.11/3Targets

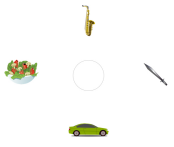


Item Nr.11/4Targets

Item Nr.12/0Targets

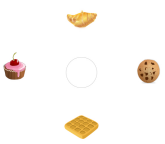


Item Nr.12/1Target



Item Nr.12/3Targets

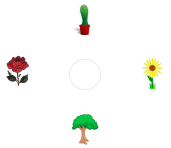
Item Nr.12/4Targets



Item Nr.13/0Targets

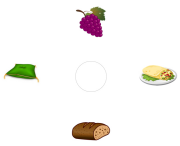
Item Nr.13/1Target

Item Nr.13/3Targets



Item Nr.13/4Targets

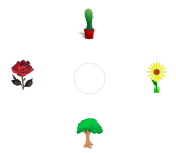
Item Nr.14/0Targets



Item Nr.14/1Target

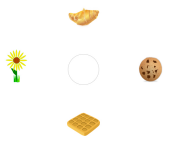
Item Nr.14/3Targets

Item Nr.14/4Targets

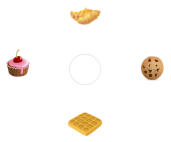


Item Nr.15/0Targets

Item Nr.15/1Target



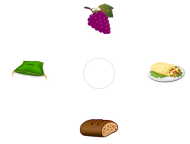
Item Nr.15/3Targets



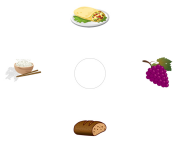
Item Nr.15/4Targets

Item Nr.16/0Targets

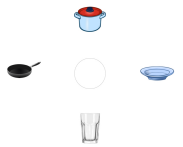
Item Nr.16/1Target



Item Nr.16/3Targets



Item Nr.16/4Targets



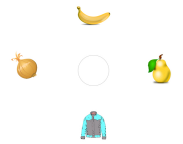
Item Nr.17/0Targets

Item Nr.17/1Target

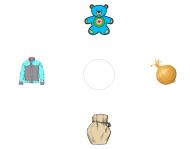
Item Nr.17/3Targets

Item Nr.17/4Targets

Item Nr.18/0Targets



Item Nr.18/1Target

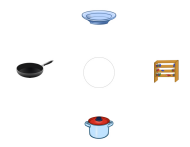


Item Nr.18/3Targets

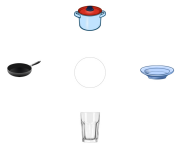
Item Nr.18/4Targets

Item Nr.19/0Targets

Item Nr.19/1Target

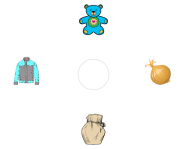


Item Nr.19/3Targets



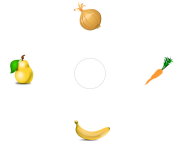
Item Nr.19/4Targets

Item Nr.20/0Targets



Item Nr.20/1Target

Item Nr.20/3Targets



Item Nr.20/4Targets

Figure B.3 *Visual stimuli for* ***Experiment 5*** *(EEG).* Tables in A indicate the matching linguistic stimuli presented simultaneously with the pictures.

Item Nr.1/0Targets

Item Nr.1/1Target

Item Nr.1/3Targets

Item Nr.1/4Targets

Item Nr.2/0Targets

Item Nr.2/1Target

Item Nr.2/3Targets

Item Nr.2/4Targets

Item Nr.3/0Targets

Item Nr.3/1Target

Item Nr.3/3Targets

Item Nr.3/4Targets

Item Nr.4/0Targets

Item Nr.4/1Target

Item Nr.4/3Targets

Item Nr.4/4Targets

Item Nr.5/0Targets

Item Nr.5/1Target

Item Nr.5/3Targets

Item Nr.5/4Targets

Item Nr.6/0Targets

Item Nr.6/1Target

Item Nr.6/3Targets

Item Nr.6/4Targets

Item Nr.7/0Targets

Item Nr.7/1Target

Item Nr.7/3Targets

Item Nr.7/4Targets

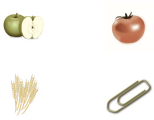
Item Nr.8/0Targets

Item Nr.8/1Target

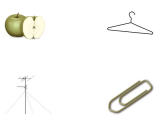
Item Nr.8/3Targets

Item Nr.8/4Targets

Item Nr.9/0Targets



Item Nr.9/1Target



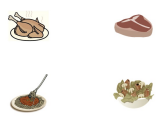
Item Nr.9/3Targets

Item Nr.9/4Targets

Item Nr.10/0Targets

Item Nr.10/1Target

Item Nr.10/3Targets

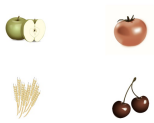


Item Nr.10/4Targets

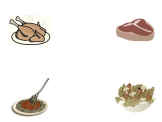
Item Nr.11/0Targets

Item Nr.11/1Target

Item Nr.11/3Targets



Item Nr.11/4Targets



Item Nr.12/0Targets

Item Nr.12/1Target

Item Nr.12/3Targets

Item Nr.12/4Targets

Item Nr.13/0Targets

Item Nr.13/1Target

Item Nr.13/3Targets

Item Nr.13/4Targets

Item Nr.14/0Targets

Item Nr.14/1Target

Item Nr.14/3Targets

Item Nr.14/4Targets

Item Nr.15/0Targets

Item Nr.15/1Target

Item Nr.15/3Targets

Item Nr.15/4Targets

Item Nr.16/0Targets

Item Nr.16/1Target

Item Nr.16/3Targets

Item Nr.16/4Targets

Item Nr.17/0Targets

Item Nr.17/1Target

Item Nr.17/3Targets

Item Nr.17/4Targets

Item Nr.18/0Targets

Item Nr.18/1Target

Item Nr.18/3Targets

Item Nr.18/4Targets

Item Nr.19/0Targets

Item Nr.19/1Target

Item Nr.19/3Targets

Item Nr.19/4Targets

Item Nr.20/0Targets

Item Nr.20/1Target

Item Nr.20/3Targets

Item Nr.20/4Targets

Figure B.4 *Visual stimuli for* ***Experiment 6***. Tables in A indicate the matching linguistic stimuli presented simultaneously with the pictures.

Item Nr.1/1Target

Item Nr.1/2Target

Item Nr.1/4Targets

Item Nr.1/7Targets

Item Nr.2/1Target

Item Nr.2/2Target

Item Nr.2/4Targets

Item Nr.2/7Targets

Item Nr.3/1Target

Item Nr.3/2Target

Item Nr.3/4Targets

Item Nr.3/7Targets

Item Nr.4/1Target

Item Nr.4/2Target

Item Nr.4/4Targets

Item Nr.4/7Targets

Item Nr.5/1Target

Item Nr.5/2Target

Item Nr.5/4Targets

Item Nr.5/7Targets

Item Nr.6/1Target

Item Nr.6/2Target

Item Nr.6/4Targets

Item Nr.6/7Targets

Item Nr.7/1Target

Item Nr.7/2Target

Item Nr.7/4Targets

Item Nr.7/7Targets

Item Nr.8/1Target

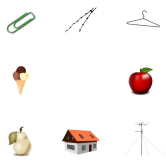
Item Nr.8/2Target

Item Nr.8/4Targets

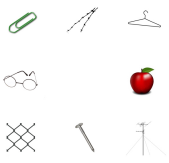
Item Nr.8/7Targets

Item Nr.9/1Target

Item Nr.9/2Target



Item Nr.9/4Targets



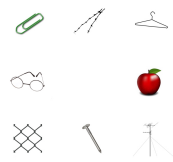
Item Nr.9/7Targets

Item Nr.10/1Target

Item Nr.10/2Target

Item Nr.10/4Targets

Item Nr.10/7Targets



Item Nr.11/1Target

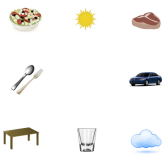
Item Nr.11/2Target

Item Nr.11/4Targets

Item Nr.11/7Targets

Item Nr.12/1Target

Item Nr.12/2Target



Item Nr.12/4Targets

Item Nr.12/7Targets

Item Nr.13/1Target

Item Nr.13/2Target

Item Nr.13/4Targets

Item Nr.13/7Targets

Item Nr.14/1Target

Item Nr.14/2Target

Item Nr.14/4Targets

Item Nr.14/7Targets

Item Nr.15/1Target

Item Nr.15/2Target

Item Nr.15/4Targets

Item Nr.15/7Targets

Item Nr.16/1Target

Item Nr.16/2Target

Item Nr.16/4Targets

Item Nr.16/7Targets

Item Nr.17/1Target

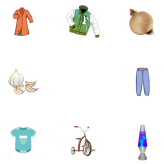
Item Nr.17/2Target

Item Nr.17/4Targets

Item Nr.17/7Targets

Item Nr.18/1Target

Item Nr.18/2Target



Item Nr.18/4Targets

Item Nr.18/7Targets

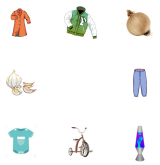
Item Nr.19/1Target

Item Nr.19/2Target

Item Nr.19/4Targets

Item Nr.19/7Targets

Item Nr.20/1Target



Item Nr.20/2Target

Item Nr.20/4Targets

Item Nr.20/7Targets

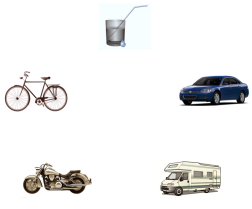
Figure B.5 *Visual stimuli for* ***Experiment 7 (A B).*** 7 B uses the exact same stimuli with a fixation cross in the centre. Tables in A indicate the matching linguistic stimuli presented simultaneously with the pictures.

Item Nr.1/1Target

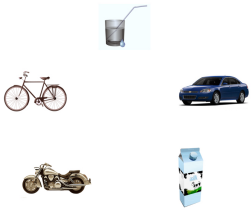
Item Nr.1/2Target

Item Nr.1/5Targets

Item Nr.2/1Target



Item Nr.2/2Target



Item Nr.2/5Targets

Item Nr.3/1Target

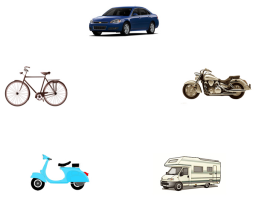
Item Nr.3/2Target

Item Nr.3/5Targets

Item Nr.4/1Target

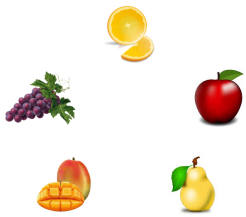
Item Nr.4/2Target

Item Nr.4/5Targets



Item Nr.5/1Target

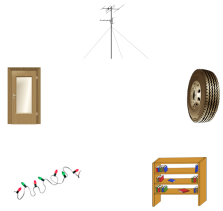
Item Nr.5/2Target



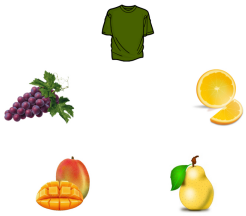
Item Nr.5/5Targets

Item Nr.6/1Target

Item Nr.6/2Target



Item Nr.6/5Targets



Item Nr.7/1Target

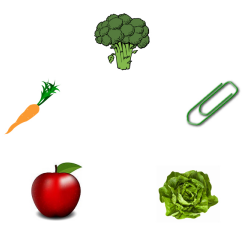
Item Nr.7/2Target

Item Nr.7/5Targets

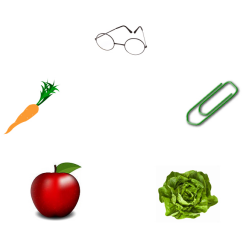
Item Nr.8/1Target

Item Nr.8/2Target

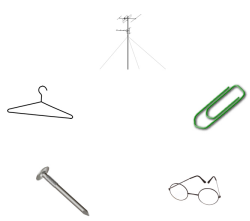
Item Nr.8/5Targets



Item Nr.9/1Target



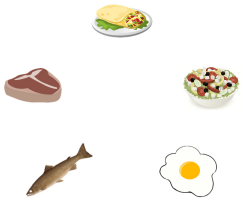
Item Nr.9/2Target



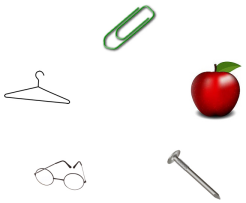
Item Nr.9/5Targets

Item Nr.10/1Target

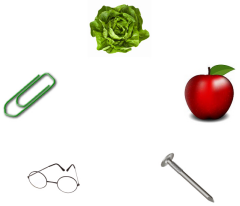
Item Nr.10/2Target



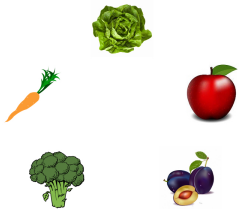
Item Nr.10/5Targets



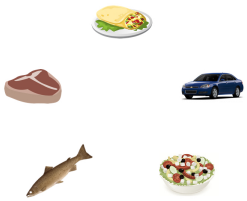
Item Nr.11/1Target



Item Nr.11/2Target



Item Nr.11/5Targets



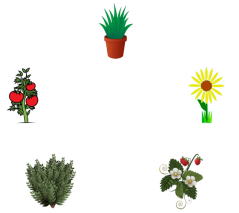
Item Nr.12/1Target

Item Nr.12/2Target

Item Nr.12/5Targets

Item Nr.13/1Target

Item Nr.13/2Target



Item Nr.13/5Targets

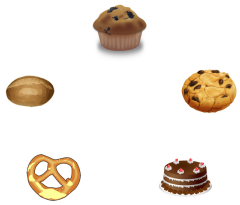
Item Nr.14/1Target

Item Nr.14/2Target

Item Nr.14/5Targets

Item Nr.15/1Target

Item Nr.15/2Target



Item Nr.15/5Targets

Item Nr.16/1Target

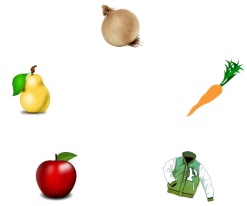
Item Nr.16/2Target

Item Nr.16/5Targets

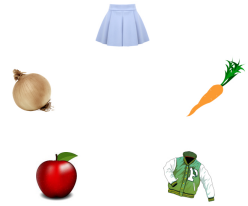
Item Nr.17/1Target

Item Nr.17/2Target

Item Nr.17/5Targets



Item Nr.18/1Target



Item Nr.18/2Target

Item Nr.18/5Targets

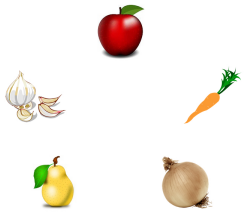
Item Nr.19/1Target

Item Nr.19/2Target

Item Nr.19/5Targets

Item Nr.20/1Target

Item Nr.20/2Target



Item Nr.20/5Targets