

The Role of Epistemic Monitoring in Language Comprehension

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Philosophie (Dr. phil.)

eingereicht am Fachbereich Humanwissenschaften
der Universität Kassel

vorgelegt von
Dipl.-Psych. Maj-Britt Isberner
Dezember 2013

Erstgutachter: Prof. Dr. Tobias Richter

Zweitgutachterin: Prof. Dr. Barbara Kaup

Tag der Disputation: 26.02.2014

For Marc

Table of Contents

Summary	1
Chapter 1: Introduction and Aim	3
Chapter 2: Literature Review	39
Isberner, M.-B., & Richter, T. (in press). Comprehension and validation: Separable stages of information processing? A case for epistemic monitoring in language comprehension. In D. N. Rapp & J. Braasch (Eds.), <i>Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences</i> . Boston, MA: MIT Press.	
Chapter 3: Study 1	81
Isberner, M.-B., & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. <i>Acta Psychologica</i> , 142, 15-22.	
Chapter 4: Study 2	111
Isberner, M.-B., & Richter, T. (in press). Does validation during language comprehension depend on an evaluative mindset? <i>Discourse Processes</i> .	
Chapter 5: Study 3	137
Isberner, M.-B., Richter, T., & Kaakinen, J. (2013). Epistemic modality in sentence comprehension: Effects of epistemic adverbs on eye movements. <i>Manuscript submitted for publication</i> .	
Chapter 6: General Discussion	181
Erklärung zum Eigenanteil	211
Eidesstattliche Versicherung und Erklärung	213
Acknowledgements	215
Appendix	217

• SUMMARY •

Summary

Most modern theories of language comprehension agree that to understand a text, readers need to integrate text information with their knowledge about the world to construct a *situation model* of what the text is about (Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). However, the information readers are exposed to in their everyday lives can be more or less consistent with what they know or believe about the world. Thus, they may sometimes come to realize that what they are reading is false or implausible with regard to their world knowledge. It is yet unclear whether this realization always requires an intentional evaluation of the validity of the information by the reader, or whether it can also result incidentally as a byproduct of the comprehension process. A widespread assumption is that the evaluation of information for truth or plausibility (*epistemic validation*; Richter, Schroeder, & Wöhrmann, 2009) is a strategic, optional process subsequent to comprehension (e.g., Gilbert, 1991; Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993; Herbert & Kübler, 2011). However, a growing number of psycholinguistic studies directly or indirectly call this two-step model of comprehension and validation as nonoverlapping stages of information processing into question. In particular, recent evidence of Stroop-like stimulus response compatibility effects that emerge when positive and negative responses are required orthogonally to the task-irrelevant truth of a sentence (e.g., a positive response after reading a false sentence or a negative response after reading a true sentence) suggests that readers nonstrategically monitor the validity of information during comprehension, in the sense that they cannot ignore validity even when it is irrelevant to their reading goal (*epistemic Stroop effect*; Richter et al., 2009).

Based on these findings, the aim of this thesis was to further test the notion that language comprehension comprises a nonstrategic, routine, knowledge-based validation process (*epistemic monitoring*; Richter et al., 2009). For this purpose, three empirical studies were conducted to test predictions derived from this assumption.

The first study investigated the question of whether evidence for epistemic monitoring can also be found for information that is not clearly true or false, but merely more or less plausible with regard to readers' world knowledge. Using the epistemic Stroop paradigm introduced by Richter et al. (2009), Study 1 established a Stroop-like effect of the compatibility of task-irrelevant plausibility on the latencies of positive and negative responses in two unrelated tasks, suggesting that epistemic monitoring is also sensitive to more gradual differences in the consistency of

• SUMMARY •

information with world knowledge than the distinction between true and false. In addition, the epistemic Stroop effect was observed in two rather different experimental tasks, an orthographical judgment task and a nonlinguistic color judgment task, corroborating the notion that the interference of epistemic monitoring with positive and negative responses is not tied to a specific kind of task. Finally, Study 1 provided strong evidence that the observed epistemic Stroop effect is indeed attributable to plausibility rather than to differences in predictability between plausible and implausible information.

The goal of Study 2 was to further probe the assumption that epistemic monitoring is not dependent on an evaluative mindset of the reader. For this purpose, it combined the epistemic Stroop paradigm with an entirely nonevaluative task in which participants did not have to perform any kind of judgments, but were simply asked to respond to the probe words “true” and “false” after reading true versus false sentences. In contrast to results reported by other authors (Wiswede, Koranyi, Müller, Langner, & Rothermund, 2013), a Stroop-like compatibility effect of task-irrelevant truth on response latencies emerged in this nonevaluative task, suggesting that epistemic monitoring does not hinge on an evaluative mindset. This difference may be owing to the fact that in our study, an adequate depth of semantic processing of the stimuli was ensured by comprehension questions that required comprehension but not validation, suggesting that epistemic monitoring may be closely tied to the depth of processing.

Study 3 approached the question of the relation between comprehension and validation from a slightly different angle by investigating the online effects of plausibility and predictability on eye movements while reading short texts, as well as their potential modulation by epistemic markers signaling the certainty of the communicated information (e.g., *certainly* or *perhaps*). In line with the assumption of a fast and nonstrategic epistemic monitoring process, plausibility was found to affect indicators of early comprehension processes. In addition, both early and late effects of plausibility were modulated by the presence of epistemic markers, suggesting that the certainty of communicated information is taken into account by the monitoring process.

Taken together, the results speak against a conceptualization of comprehension and validation as nonoverlapping stages of information processing with validation being an optional and subsequent stage. Rather, an assessment of real-world truth or plausibility seems to be, at least to some extent, an obligatory and nonstrategic component of comprehension. Implications for current models of language comprehension and directions for future research on comprehension and validation are discussed.

Chapter 1

Introduction and Aim

Introduction

“In understanding, we match what we read with what we know about the world. We implicitly ask ourselves whether the text is true, plausible, or possible. We implicitly test the truth, plausibility, or possibility of the text. Understanding a text is explaining why the information conveyed by the text is true, plausible, or possible. Understanding actions and events implies that one can explain them: that one can specify reasons, motivations, and causal antecedents for the events and actions (Craik, 1943; Schank, 1986). These explanations depend on the knowledge of the reader (Graesser et al., 1994; Singer, Halldorson, Lear, & Andrusiak, 1992).”

- Noordman & Vonk (1998, p. 196)

Can you understand a sentence without judging whether it is true or at least plausible, given its context and your knowledge about the real world? This is the question that lies at the core of the present thesis. In current psycholinguistic research, the relation between *comprehension* – that is, constructing a mental representation of a communicated state of affairs – and *validation* – that is, assessing the real-world truth or plausibility of that state of affairs – is still controversial and rarely explicitly addressed in models of language comprehension (cp. Kendeou, 2013). Nonetheless, three different views regarding this relation can be identified in the current literature.

One rather influential position, which I will refer to as *comprehension without judgment*, is that any assessment of truth or plausibility based on general world knowledge is deferred until comprehension has terminated. In other words, this position assumes a two-step model of comprehension and validation in which validation operates on the *output* of the comprehension process in a separate, subsequent stage of information processing (e.g., Connell & Keane, 2006) that is usually assumed to be voluntary, offline, and downstream. Such a two-step model goes back to Descartes (1644/2003), who proposed that people initially represent ideas without judging their truth or falsity (holding them *in aequilibrio*) and only later subject them to a rational analysis. Theoretically, this two-step model is closely related to the notion that language comprehension occurs in some kind of encapsulated language module (Fodor, 1983) which relies purely on semantic knowledge (i.e., word meanings stored in some kind of mental lexicon), while general world knowledge (i.e., factual knowledge stored in some kind of mental encyclopedia) is only accessed after comprehension has terminated. Questions about the relation between comprehension and validation thus map onto the debate about whether such a structural separation between semantic and general world knowledge exists, and onto the more general question of what types of knowledge are routinely brought to bear on language comprehension (e.g., Chwilla &

Kolk, 2005; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Hagoort & van Berkum, 2007). In line with the notion that semantic knowledge is privileged over world knowledge in language comprehension, some studies investigating the time course of the influences of different kinds of knowledge on comprehension have reported relatively late effects of real-world plausibility compared to semantic anomaly (e.g., Joseph et al., 2008; Rayner, Warren, Juhasz, & Livens, 2004; Warren & McConnell, 2007).

A variant of this position, which I will call *comprehension entailing acceptance*, assumes that comprehended information is by default initially accepted as true and can only be effortfully “unbelieved” later. This view goes back to Spinoza (1677/1997) and has prominently been advocated by Gilbert and colleagues (Gilbert, 1991; Gilbert, Krull & Malone, 1990; Gilbert, Tafarodi & Malone, 1993). It is similar to the previous one in conceptualizing validation as a processing stage that is optional and subsequent to comprehension. However, the crucial difference is that it assumes that linguistic input is by default initially represented as true in the course of comprehension. In line with this view, readers have been found to exhibit an affirmation bias when judging the accuracy of information which they have previously learned under time-pressure or cognitive load. That is, they tend to mistakenly judge false information as true but not vice versa, suggesting that cognitive load or time pressure can prevent the effortful “unbelieving” of false information (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). Further evidence for this view comes from studies that demonstrate readers’ susceptibility to false information embedded in narratives or distorted questions, even in the face of sometimes blatant violations of their knowledge (e.g., Bottoms, Eslick, & Marsh, 2010; Marsh, Meade, & Roediger, 2003). Not only do readers sometimes fail to notice such violations of their prior knowledge, but they have also been found to subsequently use the false information to answer general knowledge questions, even when they exhibited accurate knowledge at an earlier or later point (e.g., Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013; Marsh et al., 2003) and expressed highest confidence in their knowledge of the correct facts (Fazio et al., 2013). These findings have usually been explained with reference to Gilbert’s (1991) proposal that readers need to invest cognitive effort to notice and reject false information, and that they often do not do so during normal reading.

Thus, a common assumption of these traditional two-step models is that readers need to actively question the accuracy of information to notice inconsistencies with their world knowledge. An alternative view, which I will term *comprehension entailing validation*, is that real-world truth or plausibility is assessed as a regular part of comprehension (Richter, 2003; Richter,

2011; Richter, Schroeder, & Wöhrmann, 2009; Schroeder, Richter, & Hoever, 2008; Singer, 2006; Singer, 2013). In line with this position, there is growing body of evidence that demonstrates early effects of real-world plausibility on on-line measures of reading such as reading times, event-related potentials (ERPs), and eye movements (e.g., Bicknell, Elman, Hare, McRae, & Kutas, 2010; Ferguson & Sanford, 2008; Hagoort et al., 2004; Hagoort & van Berkum, 2007; Matsuki et al., 2011; McRae & Matsuki, 2009; Murray & Rowan, 1998; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Traxler & Pickering, 1996; van Petten, Coulson, Rubin, Plante, & Parks, 1999). Moreover, a number of studies have reported that readers detect inconsistencies within a text (e.g., Albrecht & O'Brien, 1993; O'Brien & Albrecht, 1992; O'Brien, Rizzella, Albrecht, & Halleran, 1998) and validate causal relationships implied by a text against general world knowledge (e.g., Singer, 1993; Singer & Halldorson, 1996; Singer, Halldorson, Lear, & Andrusiak, 1992) without being instructed to do so, suggesting that the validation of information against prior knowledge and antecedent text (*epistemic validation*; Richter et al., 2009) is not a goal-dependent activity. In fact, many studies actually use the detection of implausibility as an indicator of semantic integration (e.g., Staub et al., 2007; van Berkum, Hagoort, & Brown, 1999), thereby implicitly assuming an analysis of plausibility to constitute an integral part of comprehension.

What these different theoretical approaches and seemingly contradictory empirical findings demonstrate is that the extent to which real-world truth or plausibility is routinely assessed in the course of comprehension is still unclear and merits further investigation. The goal of the present thesis is to empirically test three main predictions that follow from conceptualizing epistemic validation as an inherent part of language comprehension (*comprehension entailing validation*): (1), that epistemic validation is a nonstrategic process which occurs even when it is not explicitly encouraged or when it is discouraged by the reading task, (2) that it is not dependent on evaluative mindset, and (3), that it is evident early in information processing.

Before describing the aim and scope of the present thesis in more detail, I will first provide a brief overview of how current theories define and conceptualize language comprehension, with an emphasis on how world knowledge is assumed to be brought to bear on comprehension. Against this background, I will then present the theoretical framework I have adopted in my thesis and review empirical evidence for and against its core assumptions to motivate the rationale of the research presented in this thesis.

Current Theories of Language Comprehension

I will begin with an outline of how language comprehension is conceptualized according to modern psycholinguistic theories, which constitutes the basis for considering the role of epistemic validation in language comprehension.

How Do Readers Represent Textual Information?

Current theories of text comprehension assume that successful comprehension does not only entail the construction of a representation of the text itself (i.e., the specific wording and surface structure of the text) and of an abstract propositional representation of the text content (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986; van Dijk & Kintsch, 1983), but also the construction of a coherent mental representation of the state-of-affairs described by the text (e.g., Glenberg, Kruley, & Langston, 1994; Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). This mental representation of the communicated state-of-affairs is called a *mental model* (Johnson-Laird, 1983) or *situation model* (van Dijk & Kintsch, 1983; I will henceforth use the term situation model). The situation model is assumed to be organized in analogy to the structure of the situation, rather than the structure of the text, preserving relations on a number of dimensions. A famous example that illustrates this point is taken from a study by Bransford, Barclay, and Franks (1972): They found that people had trouble indicating which one of two sentences they had previously heard if the sentences described the same situation but with slightly different wording:

Three turtles rested on a floating log, and a fish swam beneath them vs.

Three turtles rested on a floating log, and a fish swam beneath it

In contrast, people had no difficulties differentiating between two sentences if the same change in wording resulted in different spatial arrangements:

Three turtles rested beside a floating log, and a fish swam beneath them vs.

Three turtles rested beside a floating log, and a fish swam beneath it

This result cannot be explained by differences on the surface level or on the propositional level, which are the same for both sentence pairs. Thus, this finding indicates that people

remembered the situation described by the sentence that they had heard (including its spatial properties), rather than the surface structure or the abstract propositional content of the sentence.

Although the spatial dimension of situation models has attracted the largest amount of research (Kaup, Kelter, & Habel, 1999), situation models are assumed to represent relations on a number of other dimensions such as time, causality, motivation, protagonists and objects (Zwaan & Radvansky, 1998). Studies have shown that discontinuities in these dimensions, such as time shifts within a narrative, lead to longer reading times, suggesting that readers monitor these dimensions during normal reading and update their situation model accordingly (e.g., Magliano, Zwaan, & Graesser, 1998; Radvansky & Copeland, 2010; Theriault, Rinck, & Zwaan, 2006; Zwaan, 1999; Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998). Moreover, inconsistencies within the dimensions, such as actions that violate characters' goals (e.g., O'Brien et al., 1998) or that are inconsistent with a character's spatial location (e.g., de Vega, 1995; O'Brien & Albrecht, 1992) also lead to longer reading times compared to a consistent condition. This is the case even when multiple text paragraphs intervene between the conflicting text passages and even when local coherence (i.e., coherence with up to the previous three sentences) is maintained, suggesting that situation models require both local and global coherence (Albrecht & O'Brien, 1993).

Unlike representations of the text surface and the propositional content of a text, situation models are assumed to be constructed by integrating text-derived information with general world knowledge (e.g., Kintsch, 1998). An important but often overlooked implication of this assumption is that the process of constructing a situation model must be sensitive to the goodness of fit between incoming information and prior knowledge: "Because situation models are constructed on the basis of information derived from the text as well as the comprehender's general knowledge (van Dijk & Kintsch, 1983), the fit of incoming information with the comprehender's prior knowledge affects the integration process" (Zwaan & Madden, 2004, p. 284). Hence, situation models seem to require not only local and global coherence, but also coherence with general world knowledge.

How Do Prior Knowledge and Antecedent Text Information Become Available During Reading?

A large body of research has been concerned with how global coherence of the discourse representation can be achieved, given that not all of the information contained in a text can be held

active in working memory. How, then, do readers notice a conflict between a current text passage (e.g., *Mary orders a cheeseburger*) and antecedent text passages (e.g., *Mary is a vegetarian*; Albrecht & O'Brien, 1993) that are no longer active in working memory? The fact that readers detect such global coherence breaks even when local coherence is maintained suggests that the reactivation of antecedent information is not the result of an active memory search; rather, the evidence suggests that antecedent text information is reactivated by a fast, memory-based, passive, continuous, automatic access process, without strategic effort on the part of the reader (Linderholm, Virtue, Tzeng, & van den Broek, 2004; Myers & O'Brien, 1998; van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011). However, it is important to note that many of the inconsistencies used in the studies on memory-based processing are not direct contradictions (e.g., *Mary is a vegetarian* and *Mary is not a vegetarian*), but only in relation to relevant world knowledge (cp. Cook & Guéraud, 2005; Cook & O'Brien, 2013; Kaup & Foss, 2005). For example, the inconsistency between Mary being a vegetarian and ordering a cheeseburger only becomes apparent when the knowledge that cheeseburgers usually contain meat and that vegetarians usually do not eat any meat whatsoever is accessed as well. Thus, the activation of information during reading does not seem to be restricted to prior text information, but it also appears to include general world knowledge.

In line with this notion, the *resonance model* assumes that the currently processed text serves as an activation cue by resonating with information stored in long-term memory (Albrecht & Myers, 1998; Cook, Halleran, & O'Brien, 1998; Gerrig & McKoon, 1998; Gerrig & O'Brien, 2005; Myers & O'Brien, 1998; O'Brien et al., 1998). Concepts resonate with the current input "as a function of their strength and their degree of match to the input" (Albrecht & Myers, 1995, p. 1460). Once resonance exceeds a certain threshold, the resonant information is autonomously incorporated into working memory (Myers & O'Brien, 1998). The resonance process is assumed to be "dumb" in the sense that it activates information via spreading activation in a non-goal-directed way, regardless of whether it is relevant or up-to-date (e.g., Cook et al., 1998; Cook & O'Brien, 2013; Kendeou, Smith, & O'Brien, 2013; O'Brien et al., 1998), and of whether it hinders or facilitates processing of the current input (Myers & O'Brien, 1998). It is also unrestricted, in the sense that it resonates with information from both the episodic representation of the text and from general world knowledge (Myers & O'Brien, 1998). The resonance process is assumed to be continuous, resulting in constantly changing fluctuations in the accessibility of information during reading (Myers & O'Brien, 1998).

A computational model of how the activation of information waxes and wanes during the processing of a text is the *Landscape Model* proposed by van den Broek and colleagues (e.g., Linderholm et al., 2004; van den Broek, Risdén, Fletcher, & Thurlow, 1996; van den Broek, Young, Tzeng, & Linderholm, 1998). Similar to the resonance model, it assumes fast, passive, memory-based activation of elements from prior text or background knowledge during reading. Rather than featural overlap, the central principle that is assumed to govern the association of information in memory is cohort activation – that is, how often concepts have been simultaneously activated. The memory representation of a text thus consists of a network of nodes, with the strength of connections between the nodes being a function of the frequency of co-activation.

An often overlooked implication of the routine detection of inconsistencies during reading is that the antecedent information and world knowledge that, in conjunction, cue the inconsistency are not only routinely activated during reading, but they are also used to evaluate the current linguistic input. In other words, the knowledge activated by memory-based processing in the service of situation model construction *affords* validation of the currently processed information (Singer, 2006). Based on this notion, a number of authors (e.g., Cook & O'Brien, 2013; Richter, 2011; Richter et al., 2009; Singer, 2006) have proposed that knowledge-based validation should be explicitly conceptualized as an inherent and obligatory component of language comprehension rather than as a separate stage of information processing. The theoretical framework adopted in this thesis is the *epistemic view of language comprehension* by Richter and colleagues (Richter, 2003; Richter et al., 2009; Schroeder et al., 2008), which I will present in the following.

The Epistemic View of Language Comprehension

According to the epistemic view of language comprehension, the epistemic validation of incoming text information rests on two types of processes: *epistemic monitoring* and *epistemic elaboration*.

Epistemic Monitoring

The core assumption of the epistemic view of language comprehension is that a routine epistemic monitoring process operates during situation model construction, which checks the consistency of incoming information with the current situation model as well as with prior knowledge and portions of text that are activated by memory-based processing (Richter, 2011; Richter et al., 2009; Schroeder et al., 2008; see Figure 1). This process is conceptualized as an

‘epistemic gatekeeper’, which ensures that only consistent information is incorporated into the situation model, thus protecting the situation model from contamination with inconsistent or false information. Once information has passed the epistemic gatekeeper, it becomes part of the background knowledge against which incoming information is validated, promoting the stability of the mental representation. In contrast, if an inconsistency is detected, the incoming information is routinely rejected by the monitoring process and hence not integrated into the situation model.

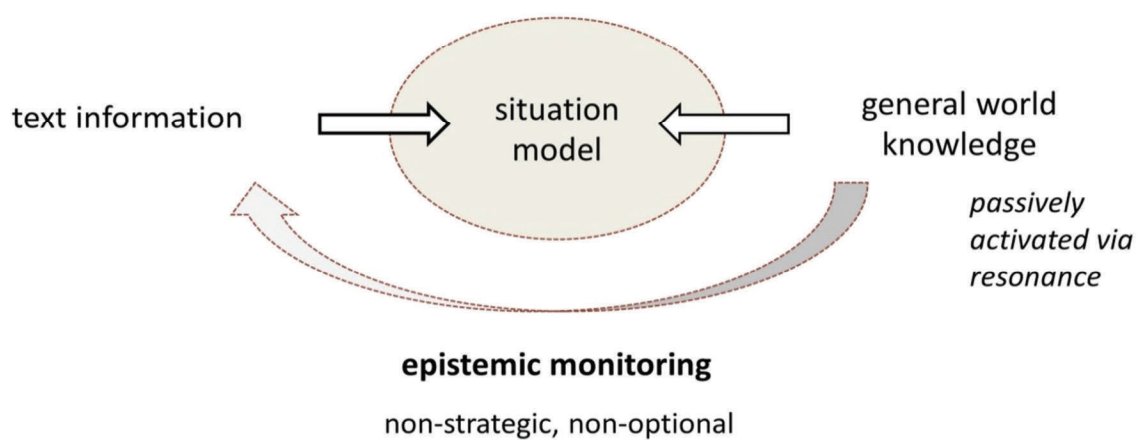


Figure 1. The epistemic view of language comprehension (Richter et al., 2009).

The fact that the activation of the knowledge used for validation is assumed to be passive implies that epistemic monitoring does not require an explicit validation goal, and that it is effortless, making low demands on the reader’s cognitive resources. However, it also implies that, under normal reading conditions, validation is only based on the knowledge which is passively activated by the resonance process – namely, on information that is easily accessible and strongly linked to the cueing information. In contrast, conflicts with less easily accessible knowledge may well go unnoticed. Likewise, the fact that information that has passed the consistency check is used for the validation of new information also has a downside: If false information does happen to pass the epistemic gatekeeper, it may subsequently hinder the acquisition of new correct information.

Epistemic Elaboration

The epistemic view of language comprehension assumes that epistemic monitoring operates by default during normal reading. However, to account for the fact that readers can also actively retrieve and use their knowledge to reason about and evaluate the accuracy of information, Richter and colleagues (Richter, 2011; Richter et al., 2009) propose a strategic validation process termed epistemic elaboration that can become operative in addition to epistemic monitoring depending on the reader's goals. In contrast to epistemic monitoring, this process is slow, strategic, and resource-demanding, and thus depends on sufficient (meta-)cognitive and motivational resources. In particular, epistemic elaboration may become operative when a conflict has been detected by epistemic monitoring. For example, a reader may deliberately attempt to resolve the detected inconsistency by elaborating hypothetical truth conditions – that is, conditions under which the ostensibly implausible information might be plausible – or draw inferences that reconcile the conflicting pieces of information. Depending on the result of these reasoning processes, readers may deliberately accept the new information and integrate it with their knowledge, or revise their knowledge to accommodate the new information, or they may consciously reject the information as false or implausible (Richter, 2011).

Superficial Versus Elaborative Epistemic Processing

Based on these two processes, Richter (2011) distinguishes two modes of processing conflicting information: *Superficial epistemic processing* and *elaborative epistemic processing*. In the default mode of assimilative epistemic processing, which readers adopt when their (meta-)cognitive and motivational resources are low, readers rely on epistemic monitoring only, meaning that information which is perceived as inconsistent with their knowledge is routinely rejected and thus, unlikely to be integrated into the situation model. Elaborative epistemic processing, in contrast, is marked by the joint operation of epistemic monitoring and epistemic elaboration. In this mode of processing, inconsistent information detected by epistemic monitoring is more likely to be actively elaborated on rather than simply rejected. This mode of processing is assumed to result in a richer, more balanced mental representation of the state-of-affairs described by a text.

The Framework of Epistemic Validation As an Alternative Two-Step Model

Thus, the model proposed by Richter and colleagues (Richter, 2003, 2011; Richter et al., 2009; Schroeder et al., 2008) can essentially be described as an alternative two-step model, with a first

stage of comprehension entailing fast and efficient epistemic monitoring based on knowledge that is passively activated via memory-based processing, and a second, optional, and goal-dependent stage of epistemic elaboration based on knowledge that is actively retrieved. It is important to note that this two-step model crucially differs from traditional two-step models of comprehension and validation in that it assumes the representation stage to already entail validation (*comprehension entailing validation*), unlike traditional two-step models which either assume that comprehension is completely non-evaluative (*comprehension without judgment*) or that it entails a positive evaluation (*comprehension entailing acceptance*), i.e., initial acceptance of the information, regardless of how implausible it may be (Gilbert, 1991). The theoretical differences between the epistemic view of language comprehension and traditional two-step models of comprehension and validation are displayed in Figure 2.

Theoretical position	Representation stage (obligatory)	Assessment stage (optional)
<i>Comprehension without judgment</i> (e.g., Descartes, 1644/2003)	Comprehension (non-evaluative)	Evaluation: Acceptance or rejection
<i>Comprehension entailing acceptance</i> (e.g., Gilbert, 1991; Spinoza, 1677/1997)	Comprehension (accepting)	Evaluation: „Unacceptance“ if necessary
<i>Comprehension entailing validation</i> (e.g., Richter, 2011; Richter et al., 2009)	Comprehension & Evaluation for consistency with world knowledge (epistemic monitoring)	Epistemic elaboration: Attempts at conflict resolution & deliberate acceptance or rejection

Figure 2. Three alternative two-step models of comprehension (representation) and validation (assessment).

Empirical Support for the Epistemic View of Language Comprehension

Evidence for efficiency and independence from reading goals: Epistemic monitoring under cognitive load. Richter et al. (2009) provided empirical evidence for their core assumption of a routine epistemic monitoring process using two kinds of paradigms. First of all, using the

paradigm introduced by Gilbert et al. (1990), they were able to show that the availability of easily accessible knowledge moderates the affirmation bias reported by Gilbert et al. In Gilbert et al.'s experiments, participants learned fictitious facts with associated truth values (e.g., *A twyrin is a doctor – FALSE*) and later had to verify those facts. Fictitious facts were used to rule out influences of prior knowledge (Gilbert et al., 1990, p. 603). In the verification task, participants tended to erroneously judge false facts that had been learned under cognitive load or time pressure as true but not vice versa. Gilbert et al. interpreted this as evidence that “unbelieving” requires effort and that if this effortful process is interrupted because cognitive resources are depleted, information remains represented as true.

Richter et al. (2009) proposed that this affirmation bias may be attributable to the fact that participants did not have easily accessible knowledge that might have allowed them to validate the facts efficiently. Therefore, Richter et al. used true and false assertions associated with either strong (e.g., *Perfume contains scents*) or weak (e.g., *Toothpaste contains sulfur*) real-world knowledge and had participants learn these facts with their associated truth values (Experiment 1). As in Gilbert et al.'s (1990) study, for some assertions, learning was interrupted by a secondary task which induced cognitive load. Richter et al. were able to replicate the affirmation bias for assertions associated with weak knowledge, but not for assertions associated with strong knowledge. This indicates that when people have strong pertinent knowledge that allows them to judge the validity of assertions fast and efficiently, cognitive load during the encoding phase does not lead to an affirmation bias. The results suggest that epistemic monitoring based on easily accessible knowledge indeed requires little cognitive effort and operates successfully even when cognitive resources are depleted. Moreover, Richter et al. (2009, Experiment 2) were able to show that no affirmation bias occurred even when participants were simply asked to memorize the true and false facts (without truth values) for recognition, which did not encourage any kind of assessment of the accuracy of the facts during the learning phase. These results suggest that the validity of information is routinely monitored during comprehension when easily accessible knowledge is available.

Evidence for the nonstrategic character of epistemic monitoring. Despite this initial evidence for the efficient and routine character of epistemic monitoring, it is still possible that epistemic monitoring can be strategically suppressed when it is irrelevant or impedimental to the task at hand. Therefore, Richter et al. (2009) directly tested the nonstrategic character of epistemic monitoring by using an adaptation of the Stroop paradigm, a standard tool for investigating

nonstrategic processes (MacLeod, 1991; Stroop, 1935). The logic underlying their *epistemic Stroop paradigm* was that if epistemic monitoring is nonstrategic, it should create interference if it is impedimental to an unrelated task. More precisely, Richter et al. assumed that merely comprehending a sentence that is true or false should result in a corresponding positive or negative response tendency. If, at the same time, an incompatible response (i.e., a positive response after reading a false sentence or a negative response after reading a true sentence) is required in an unrelated task, the nonstrategic epistemic monitoring process should interfere with the response and thus lead to longer response latencies and, potentially, more errors (the rationale of the epistemic Stroop paradigm is illustrated in Figure 3).

	Congruent	Incongruent
Stroop paradigm (Stroop, 1935)	blue green red yellow	blue green red yellow
Epistemic Stroop paradigm (Richter, Schroeder, & Wöhrmann, 2009)	True sentence/ Positive response False sentence/ Negative response	True sentence/ Negative response False sentence/ Positive response
Predictions	Response Latencies Error rates	< < Response Latencies Error rates

Figure 3. Rationale of the epistemic Stroop paradigm (Richter et al., 2009), in comparison to the original Stroop paradigm (Stroop, 1935).

To test their predictions, Richter et al. (2009) presented true (e.g., *Perfume contains scents*) and false (e.g., *Soft soap is edible*) assertions word by word on a computer screen and asked participants to judge the orthographical correctness of one target word per item. In experimental items, the target word was the last word of the sentence, which determined the validity of the assertion. In line with the predictions, response latencies and error rates were higher in incompatible than in compatible conditions, indicating that participants routinely assessed the truth

value of the assertions even though this was irrelevant and impedimental to their task of judging orthographical correctness, which indicates the nonstrategic character of epistemic monitoring.

Further Direct and Indirect Evidence for the Assumptions of the Epistemic View

Evidence for a Close Relation Between Validation and Situation Model Construction

In line with the idea that information is validated before being integrated into the situation model, and that already acquired knowledge is used for validating new information, a study by Schroeder et al. (2008) provided evidence for a close bi-directional relationship between the situation model for expository texts and the plausibility of text information. In their study, plausible information was more likely to be integrated into the situation model (*plausibility bias*), while information that was already part of the situation model was more likely to be judged as plausible. A similar plausibility bias has also been found for multiple text comprehension (Isberner et al., 2013; Maier & Richter, 2013a).

In a related vein, Ferretti, Singer, and Harwood (2013) used ERPs to investigate how readers process information that is true, false, or indeterminate with regard to antecedent text and found evidence for situation model updating only for indeterminate but plausible information, as indicated by a late positivity that was absent for true or false information.

Moreover, it has been shown that *outdated* information, which is information that has been explicitly discredited or corrected by subsequent information, continues to influence people's understanding, inferences and judgments. For example, in a study by Johnson and Seifert (1994), people were presented with fictitious news reports about a warehouse fire in which initially provided information about a potential cause of the fire (namely, that a closet contained volatile material) was explicitly corrected by subsequent information (namely, that the closet was actually empty). Nonetheless, participants still used the outdated information to later answer questions such as *For what reason might an insurance claim be refused?*, even when the correcting information was presented immediately after the outdated information. This continued influence of misinformation effect (Johnson & Seifert, 1994; Ross, Lepper, & Hubbard, 1975) can be interpreted as an indication that readers rely on already acquired knowledge to assess the validity or plausibility of new information, which hinders the acquisition of new correcting information (cp. Richter, 2011, and Schroeder et al., 2008).

Bridging Inferences Are Routinely Validated

In line with the assumptions of the epistemic view of language comprehension, studies by Singer (1993) and Singer et al. (1992) have reported evidence that bridging inferences which connect causally related sentences are routinely generated and validated during comprehension. In their studies, participants read sentences such as *Dorothy poured the bucket of water on the fire. The fire went out* (consistent with world knowledge) or *Dorothy poured the bucket of water on the fire. The fire grew hotter* (inconsistent with world knowledge). Afterwards, participants were faster to answer the question *Does water extinguish fire?* than when they had read sentences that did not imply a causal relationship (*Dorothy placed the bucket of water by the fire. The fire went out / grew hotter*). This supports the notion that readers routinely activate and use their world knowledge to validate causal relationships implied by a text.

Reading Times Are Sensitive To Truth and Negation

Singer (2006) also argued and provided evidence for a tacit and routine verification of text ideas during reading. He used a self-paced reading task in which participants read target sentences which were either consistent or inconsistent with antecedent text, and either contained negation or not (e.g., *Ken and his brother gobbled some oranges/apples. [...] The coach figured that it was/was not oranges that Ken ate.*). Even though the readers' task was to merely comprehend the stories, reading times on the target sentences were sensitive to truth and negation, which are factors that have been shown to influence overt sentence verification (e.g., Carpenter & Just, 1975). Singer thus proposed that fully comprehending a text entails its verification, or in other words, that verification emerges not from an evaluative processing goal but from "the fundamentals of the cognition of reading" (p. 589).

In addition, Singer (2006) investigated the influence of discourse-pragmatic factors on validation. He assumed that factive verbs (e.g., *know*) place greater demands on the comprehender to validate their complement than nonfactive verbs (e.g., *figured*) because only factive verbs entail the truth of their complement (e.g., *She knows that the world is round*). As a consequence, *knowing* something false should be perceived as pragmatically infelicitous, but only *figuring* something true should also appear "mildly incongruent" (p. 581). In line with this notion, verb factivity modulated the reading time pattern: Only figuring something true or determining something false incurred pragmatic costs that led to longer reading times in the pragmatically infelicitous conditions, which were superimposed on the interaction of truth and negation. These

results were corroborated by an ERP study, which found interactive effects of verb factivity and truth on both early and later potentials (Ferretti, Singer, & Patterson, 2008).

Comprehension and Verification Require Similar Amounts of Time for Certain Types of Sentences

Similar to the epistemic view of language comprehension, Wyer and Radvansky (1999) proposed a theory of the comprehension of social information which assumes that certain types of information (namely, statements that are well-known to be true or false) are spontaneously recognized as true or false in the course of comprehension, without any need for goal-directed evaluation. According to their theory, all types of declarative knowledge, not just semantic knowledge, can be accessed during comprehension before any goal-directed processing takes place (p. 92). They also assume comprehension to comprise a component termed the *Model Constructor*, which, amongst others, has the capacity to “evaluate the compatibility of new information with previously formed situation models” (p. 101). “To this extent, the recognition of a statement’s truth is a by-product of the same processes used in comprehension.” (p. 103). In line with this notion, Wyer and Radvansky (1999) reported evidence that for sentences describing states of affairs that are known to be clearly true or false by the reader (e.g., *Jane Fonda does aerobics / plays pro hockey*), validity judgments do not take significantly longer than comprehensibility judgments; only when validity was unclear (e.g., *Jane Fonda rides a motorcycle*) did readers take longer to judge validity than comprehensibility. Wyer and Radvansky (1999) proposed that this is because under these circumstances, readers need to engage in the more time-consuming process of comparing the features of the subject with previously formed situation models of the predicate to make validity judgments.

Immediate Effects of Plausibility on Comprehension

The epistemic view of language comprehension is in line with a large body of evidence that, although not directly investigating epistemic validation, has demonstrated rapid effects of plausibility and validity on comprehension in the absence of an evaluative processing goal. **Chapter 2** provides a detailed overview over this research. In these studies, plausibility is usually manipulated in the service of investigating other phenomena, such as syntactic analysis, situation model updating, memory-based processing, the processing of negation, the use and time course of availability of different kinds of knowledge during comprehension, or the integration of new information with its context. Interestingly, many of these studies use the detection of

implausibility as an indicator of how or how fast certain constructions are interpreted (for example, ambiguous or garden path sentences; e.g., Pickering & Traxler, 1998; Traxler & Pickering, 1996; Trueswell, Tanenhaus, & Garnsey, 1994), which implicitly contradicts the assumption of traditional two-step models that validation is merely an optional process subsequent to comprehension.

The kinds of knowledge violations that have been investigated in this literature comprise world knowledge violations, semantic anomalies, (direct or indirect) inconsistencies with antecedent text, logical inconsistencies, and implausibility. However, it must be noted that the terms have not been used systematically in the literature: As will be discussed in more detail in **Chapter 2**, some studies use semantic anomalies that by other definitions would be considered world knowledge violations or inconsistencies with the context, and the term implausibility has been used synonymously for all of these types of knowledge violations. Doubts about whether a clear distinction can be drawn between semantic and world knowledge violations have also been raised from a theoretical perspective (e.g., Jackendoff, 2002). For the purpose of the review, I will follow Warren's (2011) approach to map the findings onto the dimension of *plausibility* or *knowledge violation severity*. Anomalous, impossible or severely implausible information can be located on one end of this continuum, while the other end comprises highly plausible, typical or even predictable information. Although there may be qualitative differences between these types of information, both theoretical considerations (e.g., Jackendoff, 2002) and empirical results (e.g., Hagoort et al., 2004) support the assumption of more gradual differences in the way these types of information are processed by readers.

In particular indicators of language processing with a high temporal resolution, such as eye movements and ERPs, have revealed that plausible and implausible information elicits differential effects on comprehension at a very early stage – as early as word meaning is accessed (Hagoort et al., 2004) and sometimes even before the isolation point for identification of the implausible word (van Petten et al., 1999) or before the implausible word itself (Kennedy, Murray, & Boissiere, 2004; Murray, 1998; Murray & Rowan, 1998). To summarize the findings, invalid, implausible, or inconsistent information in comparison to valid, plausible, or consistent information has been found to elicit:

- slower reading times in self-paced reading (e.g., Albrecht & Myers, 1995; Albrecht & O'Brien, 1993; Kaup & Foss, 2005; Myers, O'Brien, Albrecht, & Mason, 1994; O'Brien & Albrecht, 1992; O'Brien et al., 1998; Singer, 2006)
- longer fixations on and/or immediately after the implausible word in eye-tracking, including in measures of early comprehension processes such as first fixation duration on the implausible word (e.g., Matsuki et al., 2011; Staub et al., 2007), or even before the implausible word (Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998), as well as more and longer regressions out of implausible regions (e.g., Ni, Fodor, Crain, & Shankweiler, 1998; Rayner et al., 2004; Speer & Clifton, 1998; Warren, McConnell, & Rayner, 2008)
- an enhanced N400 ERP (e.g., Chwilla & Kolk, 2005; Ferretti et al., 2013; Hagoort et al., 2004; Hald, Steenbeek-Planting, & Hagoort, 2007; Macizo & Herrera, 2010; Nieuwland, 2013; van Berkum et al., 1999; Wiswede, Koranyi, Müller, Langner, & Rothermund, 2013); moreover, an enhanced P600 has been reported under some circumstances (e.g., Kolk & Chwilla, 2007; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010; van Herten, Chwilla, & Kolk, 2006; van Herten, Kolk, & Chwilla, 2005; Vissers, Chwilla, & Kolk, 2007; Vissers, Kolk, van de Meerendonk, & Chwilla, 2008)
- slower positive responses in lexical decision or in an orthographical task (Richter et al., 2009; West & Stanovich, 1982)

Thus, all of these measures can be seen (albeit not exclusively) as indicators of validation processes. Two of these indicators will also be used in this thesis: response latencies and eye movements.

Evidence Challenging the Epistemic View of Language Comprehension

At the same time, in apparent contradiction to those previously discussed results, there is also abundant evidence that demonstrates readers' lack of sensitivity to text-internal inconsistencies and violations of their general world knowledge, as well as their susceptibility to false or contradictory information.

Failures to Notice False or Inconsistent Information

One finding that challenges the epistemic view of language comprehension is that people sometimes fail to notice that information is false or inconsistent with regard to prior text information or their world knowledge. Multiple lines of research have reported that such failures are quite common; these findings will be presented in the following.

Research on comprehension monitoring. Evidence that raises doubts about the existence of a routine epistemic monitoring process comes, amongst others, from studies on comprehension monitoring by Baker and colleagues (for an overview, see Baker, 1989). To investigate whether readers notice different kinds of problems during reading, Baker and colleagues introduced various errors into texts whose identification was assumed to reflect the application of different standards for monitoring comprehension. The embedded errors were nonsense words (probing for use of a lexical standard, which “involves checking that the meaning of an individual word is understood”; Baker, 1989, p.14), prior knowledge violations (probing for use of an external consistency standard, which “involves checking that the ideas in the text are true or plausible with respect to what one already knows”; Baker, 1989, p.14), and text-internal contradictions (probing for use of an internal consistency standard, which “involves checking that the ideas expressed in the text are logically consistent with one another”; Baker, 1989, p.14). Although comprehension monitoring is conceptualized as a metacognitive, strategic, and controlled activity, the kinds of errors that it was hypothesized to detect were thus highly similar to those that should be routinely detected by epistemic monitoring (Richter et al., 2009). Baker (1985) found that while nonsense words were detected relatively frequently, most of the prior knowledge violations (78%) and even more of the contradictions (88%) went unnoticed, even though readers were explicitly instructed to look for problems in the texts. Other studies on comprehension monitoring (Baker & Anderson, 1982; Glenberg, Wilkinson, & Epstein, 1982; Grabe, Antes, Thorson, & Kahn, 1987) reported similarly low detection rates, averaging around only 50% across studies (cp. Baker, 1989). These low detection rates for external and internal consistency violations appear difficult to reconcile with the assumption that readers routinely monitor the validity and consistency of what they comprehend, especially as these failures to detect inconsistencies usually do not keep people from rating their comprehension as high (*illusion of knowing*; Glenberg et al., 1982).

Further evidence for readers’ failure to detect prior knowledge violations. Another line of research which has demonstrated readers’ failure to detect knowledge violations is research on so-called “semantic illusions”. These are cases in which readers miss rather blatant violations

of their prior knowledge. A famous example is the Moses illusion, in which readers do not detect the false presupposition in the question “How many animals of each kind did Moses take on the Ark?”, although they know that it was Noah and not Moses who built the Ark (Erickson & Mattson, 1981; for an overview, see Park & Reder, 2004). Another well-known example is the airplane crash scenario by Barton and Sanford (1993), in which people do not notice the anomaly in the question where authorities should bury the *survivors* of a plane crash. Surprisingly, readers have been shown to fall victim to such semantic illusions even when they are explicitly warned against them (e.g. Marsh & Fazio, 2006).

This phenomenon has been explained in terms of shallow processing (Barton & Sanford, 1993): Because the fit of the anomalous word with the scenario is high, it does not receive an in-depth analysis, i.e., its meaning is not exhaustively retrieved. In line with this view, detection rates are much higher when the word *survivors* occurs in a bicycle crash scenario in which its fit is much poorer (Barton & Sanford, 1993). In this way, pragmatics can override local semantic analysis (Sanford & Garrod, 1998), meaning that people sometimes rely on a plausibility heuristic to construct a mental representation of the input (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; cf. Nieuwland & van Berkum, 2005).

Further evidence for readers’ failure to notice text-internal inconsistencies. Readers’ failures to notice text-internal inconsistencies has also been demonstrated in a study by Otero and Kintsch (1992). In this study, students read short expository paragraphs that contained direct contradictions between the second and last sentence of the paragraph (e.g., *Superconductivity is obtained by cooling materials [...] Superconductivity is obtained by increasing the temperature of materials*). Overall, 40.3% of these inconsistencies went unnoticed. However, there were large individual differences between students: most students were either detectors, who reported almost all of the contradictions, or nondetectors, who did not notice any of the contradictions; only few students noticed some contradictions while missing others. Moreover, this study allowed insight into the kind of representation that readers build from inconsistent text: In their summaries, nondetectors either mentioned only one or neither of the conflicting sentences, or they drew unwarranted inferences to resolve the conflict. Otero and Kintsch (1992) attributed failures to notice the inconsistencies to exaggerated weighting of one of the contradictory statements or of the reader’s prior beliefs, which resulted in the suppression of conflicting information.

In a related vein, Kaup and Foss (2005) investigated the processing of text-internal inconsistencies in narrative comprehension. Specifically, they examined whether readers are more

likely to detect direct inconsistencies (e.g., a sweater being described as *blue* in one part of the text and as *green* in another part) or relational inconsistencies (e.g., the sweater being first described as *too large* for a character and then later as *too small* for the same character). They hypothesized that readers would be more likely to notice direct inconsistencies because direct properties should be part of the mental representation of the object to which they pertain, whereas relational properties pertain to specific situations and should therefore only be activated within those situations. Contrary to the predictions, readers seemed to spontaneously detect more relational than direct inconsistencies as evidenced by an increase in reading times; however, when readers were specifically instructed to look for and report inconsistencies, detection rates for direct inconsistencies (on average 68%) resembled those for relational inconsistencies (on average 70%). Nonetheless, these results again confirm that readers miss a considerable number of inconsistencies even when they are explicitly instructed to monitor for them. Kaup and Foss (2005) hypothesized that this may be due to some properties fading from the mental representation over time.

Persuasion Through Narratives and Reliance on Misinformation

Failures to detect false information might not be so problematic if this information did not have any further impact on the reader. However, such failures have also been shown to be associated with negative memorial consequences (Bottoms et al., 2010). For example, readers have been shown to use undetected false presuppositions in distorted questions to later answer general knowledge questions (e.g., Bottoms et al., 2010). The same has been reported for misinformation embedded in narratives, even when readers demonstrate accurate knowledge of the correct facts before or after reading the misleading information (Fazio & Marsh, 2008; Marsh et al., 2003) with the highest degree of confidence in the accuracy of their knowledge (Fazio et al., 2013). Therefore, reliance on misinformation has been explained as *knowledge neglect*: Readers possess accurate knowledge but fail to activate it (Fazio et al., 2013). The mechanism that has been suggested to explain the subsequent use of misinformation is that the false information becomes temporarily more easily accessible than the correct information, and that readers mistake the ease of retrieval of the false information for truth (Kelley & Lindsay, 1993; see Bottoms et al., 2010; Fazio et al., 2013; Marsh et al., 2003). In line with this notion, the reliance on misinformation decreases over time as activation of the false information fades (Barber, Rajaram, & Marsh, 2008; Marsh et al., 2003). Nonetheless, the implications of the misinformation effect are worrisome as readers have

been shown to generate deductive inferences from misinformation and integrate them into their knowledge (Butler, Dennis, & Marsh, 2012).

In a similar vein, research on persuasion through narratives has demonstrated that if a character in a story expresses inaccurate beliefs (e.g., “mental illness is contagious”), readers later take longer to assess the real-world truth of the assertions in a verification task (Gerrig & Prentice, 1991) and their beliefs shift in the direction of the views expressed in the story as measured by agreement ratings (e.g., Appel & Richter, 2007; Prentice, Gerrig, & Bailis, 1997; Wheeler, Green, & Brock, 1999). Moreover, Appel and Richter (2007) presented evidence that these effects are not short-lived, but rather increase over time as memory for the source of the belief change declines, suggesting that the acquired beliefs are integrated into real-world knowledge.

These findings are often interpreted as evidence for Gilbert’s (1991) model that people are bound to initially believe everything they read, and that they often do not invest the cognitive effort required to “unbelieve” what they have read (e.g., Gerrig, 1993; Marsh et al., 2003). In other words, it is assumed that readers must “actively construct disbelief” (Prentice et al., 1997, p. 416), which they are not necessarily motivated to do.

Necessity of an Evaluative Mindset

In line with these concerns, a recent study by Wiswede et al. (2013) suggested that the epistemic Stroop effect reported by Richter et al. (2009) may be dependent on an evaluative mindset. Wiswede et al. found an epistemic Stroop effect for true and false sentences similar to those used by Richter et al. (e.g., *Saturn is a planet / continent*) only when one of two randomly intermixed tasks explicitly required the evaluation the sentences for truth, whereas no such effect was found when neither of the two tasks made such an evaluation necessary (a more detailed description of this study will be provided in **Chapters 2 and 3**). Based on these results, they suggested that the orthographic task used by Richter et al. (i.e., judging whether words were spelled correctly or incorrectly) may have induced an evaluative mindset that is absent under normal reading conditions. Again, Gilbert (1991)’s model was invoked to explain these findings.

Delayed Effects of Real-World Plausibility on Comprehension

Finally, some research on the integration of different kinds of knowledge during comprehension (e.g., Joseph et al., 2008; Rayner et al., 2004; Warren & McConnell, 2007) has suggested that real-world implausibility has delayed disruptive effects on reading compared to semantic anomaly, calling into question the idea that plausibility is immediately assessed during

comprehension. In line with these findings, the notion that such an assessment is subsequent to comprehension is still present in some more recent theoretical accounts and computational models of language processing (e.g., Bornkessel & Schlesewsky, 2006; Connell & Keane, 2006).

Summary

To summarize, the extant research shows that on the one hand, readers often detect inconsistencies with their knowledge or within a text without having been instructed to monitor for errors, suggesting that to some extent, monitoring for such inconsistencies is an integral part of normal reading activity. On the other hand, it shows that readers sometimes fail to detect such inconsistencies even when they are rather blatant, and even in spite of explicit warnings against their presence, suggesting that this monitoring is not always successful. Given these findings, it is apparent that readers monitor the real-world validity or plausibility of information to some extent during comprehension, but the precise nature and the boundary conditions of this monitoring process still seem to be unclear. Therefore, the goal of the present thesis is to further elucidate this process.

Aim and Scope of This Thesis

The present thesis addresses the question of whether epistemic monitoring can be considered an inherent component of language comprehension. **Chapter 2** provides an extensive review of studies that, directly or indirectly, allow conclusions regarding this central question. The three subsequent chapters present empirical work which was aimed at further exploring the scope of the framework of epistemic validation (Richter, 2011; Richter et al., 2009; Schroeder et al., 2008) by testing predictions derived from the three primary implications (1) that epistemic monitoring is obligatory and therefore nonstrategic, (2) that it does not require an evaluative mindset, and (3) that the effects of epistemic monitoring are evident early in language processing.

The studies presented in **Chapters 3 and 4** are explicitly concerned with testing the nonstrategic character of epistemic monitoring. For this purpose, variations of the epistemic Stroop paradigm introduced by Richter et al. (2009) were developed and employed. The rationale of the epistemic Stroop paradigm is that if validation is nonstrategic, it should interfere with performance on an unrelated task when it produces incompatible responses – that is, if the task requires a positive response after reading a false or implausible sentence, or vice versa. In **Chapter 3**, I will present two experiments that address the question of whether epistemic monitoring is restricted to

clearly true or false information or whether it extends to information that is more or less plausible information with regard to knowledge about real-world events (*event-based plausibility*; Matsuki et al., 2011). In addition, these experiments were designed to rule out the alternative explanation that the epistemic Stroop effect is attributable to predictability rather than truth or plausibility, as true or false (and plausible or implausible) stimuli usually differ in their predictability. Finally, this study provides a test of whether epistemic validation can interfere not only with responses in linguistic tasks (such as the orthographical task employed by Richter et al.) that rely on partly the same processes as validation (e.g., word identification), but also with positive and negative responses in a completely unrelated nonlinguistic task, namely the task of judging whether or not a word has changed color.

The study presented in **Chapter 4** addresses the question of whether epistemic monitoring is dependent on an evaluative mindset, as has recently been suggested by Wiswede et al. (2013). In addition, it investigates whether the epistemic Stroop effect reflects facilitation in compatible conditions, interference in incompatible conditions, or a combination of both, which allows conclusions regarding the processes underlying epistemic monitoring.

The research reported in **Chapter 5** is concerned with the time-course of the effects of plausibility and predictability on language comprehension. Plausibility and predictability are two variables that reflect different aspects of the fit of incoming information with prior knowledge. To examine their respective influences on online comprehension, this study uses eye-tracking technology as a method for investigating reading processes with a high temporal and spatial resolution. In addition, this study explores potential modulations of the influences of these two variables by the presence of epistemic markers which signal the certainty of the focal information. As has been demonstrated by Singer and colleagues, the degree to which verbs entail the truth of their complements (e.g., *know* vs. *believe*) influences validation processes (e.g., Singer, 2006). Similarly, the degree to which epistemic adverbs imply the certainty of communicated information (e.g., *certainly* vs. *perhaps*) may affect both prediction and validation processes during language comprehension.

References

- Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1459–1468.
- Albrecht, J. E., & Myers, J. L. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes*, 26(2-3), 87–107.
- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1061–1070.
- Appel, M., & Richter, T. (2007). Persuasive effects of fictional narratives increase over time. *Media Psychology*, 10(1), 113–134.
- Baker, L., & Anderson, R. I. (1982). Effects of inconsistent information on text processing: Evidence for comprehension monitoring. *Reading Research Quarterly*, 17(2), 281–294.
- Barber, S. J., Rajaram, S., & Marsh, E. J. (2008). Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience. *Memory*, 16(8), 934–946.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21(4), 477–487.
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63(4), 489–505.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821.
- Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, 18(6), 670–678.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3(2), 193–209.
- Butler, A. C., Dennis, N. A., & Marsh, E. J. (2012). Inferring facts from fiction: Reading correct and incorrect information affects memory for related information. *Memory*, 20(5), 487–498.

- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1), 45–73.
- Chwilla, D. J., & Kolk, H. H. J. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25(3), 589–606.
- Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science*, 30(1), 95–120.
- Cook, A. E., & Guéraud, S. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes*, 39(2&3), 265–278.
- Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Processes*, 26(2&3), 109–129.
- Cook, A. E., & O'Brien, E. J. (2013). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2013.855107
- Craik, K. (1943). *The nature of explanation*. Cambridge, England: Cambridge University Press.
- de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 373–385.
- Descartes, R. (2003). *Selections from the principles of philosophy* (translated by J. Veitch). Retrieved March 31, 2009, from <http://www.gutenberg.org/etext/4391> (Original work published 1644)
- Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 540–551.
- Fazio, L. K., Barber, S. J., Rajaram, S., Ornstein, P. A., & Marsh, E. J. (2013). Creating illusions of knowledge: Learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General*, 142(1), 1–5.
- Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, 15(1), 180–185.
- Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, 58(3), 609–626.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15.
- Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, 1(1-2), 71–83.

- Ferretti, T. R., Singer, M., & Harwood, J. (2013). Processes of discourse integration: Evidence from event-related brain potentials. *Discourse Processes*, 50(3), 165–186.
- Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition*, 108(3), 881–888.
- Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes*, 13(2), 175–190.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, Massachusetts and London, England: The MIT Press.
- Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven, CT: Yale University Press.
- Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26(2&3), 67–86.
- Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39(2&3), 225–242.
- Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, 2(5), 336–340.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46(2), 107–119.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59(4), 601–613.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65(2), 221–233.
- Glenberg, A. M., Kruley, P., & Langston, W. E. (1994). Analogical processes in comprehension: Simulation of a mental model. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 609–640). San Diego, CA, US: Academic Press.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10(6), 597–602.
- Grabe, M., Antes, J., Thorson, I., & Kahn, H. (1987). Eye fixation patterns during informed and uninformed comprehension monitoring. *Journal of Reading Behavior*, 19(2), 123–140.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395.

- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438–441.
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 801–811.
- Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, 1146, 210–218.
- Isberner, M.-B., Richter, T., Maier, J., Knuth-Herzig, K., Horz, H., & Schnotz, W. (2013). Comprehending conflicting science-related texts: graphs as plausibility cues. *Instructional Science*, 41(5), 849–872.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, New York: Oxford University Press.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1420–1436.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.
- Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 61(5), 708–723.
- Kaup, B., & Foss, D. J. (2005). Detecting and processing inconsistencies in narrative comprehension. In D. T. Rosen (Ed.), *Trends in experimental psychology research* (pp. 67–90). Hauppauge, NY: Nova Science Publishers.
- Kaup, B., Kelter, S., & Habel, C. (1999). Taking the functional aspects of mental models as a starting point for studying discourse comprehension. In G. Rickheit & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 93–112). Amsterdam: North Holland.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32(1), 1–24.

- Kendeou, P. (2013). Validation and comprehension: An integrated overview. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2013.855874
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 854–865.
- Kennedy, A., Murray, W., & Boissiere, C. (2004). Parafoveal pragmatics revisited. *European Journal of Cognitive Psychology*, 16(1-2), 128–153.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, 100(3), 257–261.
- Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*, 37(2), 165–186.
- Macizo, P., & Herrera, A. (2010). What is heavier, a kilo of feathers or a kilo of lead? What you know determines what you understand. *NeuroReport*, 21(14), 938–942.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203.
- Magliano, J. P., Zwaan, R. A., & Graesser, A. C. (1998). The role of situational continuity in narrative understanding. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 219–245). Hove, UK: Psychology Press.
- Maier, J., & Richter, T. (2013). How nonexperts understand conflicting information on social science issues: The role of perceived plausibility and reading goals. *Journal of Media Psychology: Theories, Methods, and Applications*, 25(1), 14–26.
- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, 34(5), 1140–1149.
- Marsh, E. J., Meade, M. L., & Roediger, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, 49(4), 519–536.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 913–934.

- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, 3(6), 1417–1429.
- Murray, W. S. (1998). Parafoveal pragmatics. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 181–199). Oxford, England: Elsevier.
- Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, 27(1), 1–22.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26(2-3), 131–157.
- Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 876–886.
- Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, 27(5), 515–539.
- Nieuwland, M. S. (2013). “If a lion could speak ...”: Online sensitivity to propositional truth-value of unrealistic counterfactual sentences. *Journal of Memory and Language*, 68(1), 54–67.
- Nieuwland, M. S., & van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3), 691–701.
- Noordman, L. G. M., & Vonk, W. (1998). Memory-based processing in understanding causal information. *Discourse Processes*, 26(2-3), 191–212.
- O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(4), 777–784.
- O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(5), 1200–1210.
- Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, 3(4), 229–235.
- Park, H., & Reder, L. M. (2004). Moses illusion: Implication for human cognition. In R. F. Pohl (Ed.), *Cognitive Illusions* (pp. 275–291). Hove, UK: Psychology Press.

- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 940–961.
- Prentice, D. A., Gerrig, R. J., & Bailis, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review*, 4(3), 416–420.
- Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 210–216.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290–1301.
- Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen*. Lengerich: Pabst.
- Richter, T. (2011). Cognitive flexibility and epistemic validation in learning from multiple texts. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility* (pp. 125–140). Dordrecht: Springer Netherlands.
- Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96(3), 538–558.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32(5), 880–892.
- Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26(2-3), 159–190.
- Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, 25(3), 279–294.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59(3), 237–255.

- Singer, M. (1993). Causal bridging inferences: Validating consistent and inconsistent sequences. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47(2), 340–359.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54(4), 574–591.
- Singer, M. (2013). Validation in reading comprehension. *Current Directions in Psychological Science*, 22(5), 361–366.
- Singer, M., & Halldorson, M. (1996). Constructing and Validating Motive Bridging Inferences. *Cognitive Psychology*, 30(1), 1–38.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, 31(4), 507–524.
- Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition*, 26(5), 965–978.
- Spinoza, B. (1997). *Ethics demonstrated in geometric order* (translated by R. H. M. Elwes). Retrieved April 4, 2009, from <http://www.mtsu.edu/~rbombard/RB/Spinoza/ethica-front.html> (Original work published 1667)
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6), 1162–1169.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.
- Therriault, D. J., Rinck, M., & Zwaan, R. A. (2006). Assessing the influence of dimensional focus during situation model construction. *Memory & Cognition*, 34(1), 78–89.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35(3), 454–475.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33(3), 285–318.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.

- van de Meerendonk, N., Kolk, H. H. J., Vissers, C. Th. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1), 67–82.
- van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. J. Schraw (Eds.), *Text relevance and learning from text*. (pp. 123–139). Greenwich, CT: Information Age Publishing.
- van den Broek, P., Ridsen, K., Fletcher, C. R., & Thurlow, R. (1996). A “landscape” view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B.K. Britton & A.C. Graesser (Eds.), *Models of understanding text* (pp. 165–187). Mahwah, NJ: Erlbaum.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1998). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Hove, UK: Psychology Press.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18(7), 1181–1197.
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241–255.
- van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 394–417.
- Vissers, C. Th. W. M., Chwilla, D. J., & Kolk, H. H. J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75(1), 8–18.
- Vissers, C. Th. W. M., Kolk, H. H. J., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture–sentence matching task. *Neuropsychologia*, 46(4), 967–982.

- Warren, T. (2011). The influence of plausibility and anomaly on eye movements in reading. In S. P. Livversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook on Eye Movements* (pp. 911–923). USA: Oxford University Press.
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, 14(4), 770–775.
- Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 1001–1010.
- West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 385–399.
- Wheeler, C., Green, M. C., & Brock, T. C. (1999). Fictional narratives change beliefs: Replications of Prentice, Gerrig, and Bailis (1997) with mixed corroboration. *Psychonomic Bulletin & Review*, 6(1), 136–141.
- Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, 8(6), 647–653.
- Wyer, R. S., & Radvansky, G. A. (1999). The comprehension and validation of social information. *Psychological Review*, 106(1), 89–118.
- Zwaan, R. A. (1999). Five dimensions of narrative comprehension: The event-indexing model. *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso*, 93–110.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6(5), 292–297.
- Zwaan, R. A., & Madden, C. J. (2004). Updating Situation Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 283–288.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 386–397.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.

• CHAPTER 1 • INTRODUCTION AND AIM •

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, 2(3), 199–220.

Chapter 2

Literature Review

Comprehension and validation: Separable stages of information processing? A case for epistemic monitoring in language comprehension

A version of this chapter will be published as:

Isberner, M.-B., & Richter, T. (in press). Comprehension and validation: Separable stages of information processing? A case for epistemic monitoring in language comprehension. In D. N. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Boston, MA: MIT Press.

Introduction

How and when do we realize that something we comprehend is inconsistent with our knowledge about the world? Is this realization a part of comprehension, or is it a voluntary decision process subsequent to comprehension? Is it strategic – that is, dependent on an evaluative processing goal – or nonstrategic – that is, relatively fast, effortless and difficult to suppress? Clearly, we cannot properly judge the truth or plausibility of something we do not comprehend. But can we comprehend something without also judging its truth or plausibility?

Evaluation of information is widely considered an offline, downstream, voluntary process that is subsequent to comprehension (e.g., Gilbert, Krull, & Malone, 1990; Gilbert, Tatarodi, & Malone, 1993; Herbert & Kübler, 2011; Sparks & Rapp, 2011; Wiswede, Koranyi, Müller, Langner, & Rothermund, 2013). Underlying this conception is a two-step model of comprehension and evaluation, in which comprehension is non-evaluative and any knowledge-based (epistemic) evaluation of information is strategic and delayed until after comprehension has finished.

Consistent with this idea, there has been an implicit division of epistemic labor in psychological research: Whereas cognitive psychology mainly focuses on investigating phenomena of comprehension, processes of information evaluation are primarily investigated in social psychology (e.g., in the framework of the Elaboration Likelihood Model; Petty & Cacioppo, 1986; Petty & Wegener, 1999). In contrast, evaluative processing is often used in psycholinguistic research to *measure* comprehension in the first place: In studies on the organization of semantic memory, the time it takes to verify a sentence is often taken as an indicator of how long it takes to comprehend it (e.g., Kintsch, 1980; Kounios, Osman, & Meyer, 1987). This approach is consistent with the idea that meaning is conveyed when the truth conditions of a sentence are understood (e.g., Davidson, 2001). Moreover, readers' ability to evaluate information fast and incrementally with regard to their world knowledge is often utilized (seemingly naturally) in psycholinguistic studies to investigate a variety of phenomena, such as the time course of availability of different kinds of information during comprehension (e.g., Fischler, Childers, Achariyapaopan, & Perry, 1985; Hagoort, Hald, Bastiaansen, & Petersson, 2004; O'Brien, Rizzella, Albrecht, & Halleran, 1998; Rayner, Warren, Juhasz, & Liversedge, 2004), memory-based processing in situation model updating (e.g., Albrecht & O'Brien, 1993; O'Brien et al., 1998), the processing of negation (e.g., Nieuwland & Kuperberg, 2008), the question of when a sentence's meaning is integrated with its context (Nieuwland & van Berkum, 2006; van Berkum, Zwitserlood, Hagoort, & Brown, 2003), or

syntactic analysis (e.g., Pickering & Traxler, 1998; Speer & Clifton, 1998; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Traxler & Pickering, 1996; van Gompel, Pickering, & Traxler, 2001), to name just a few topics. Many of the results reported in these studies call the two-step model into question, although this is often overlooked or taken for granted in interpretations of the findings. It seems thus that the relationship between comprehension and evaluation of information, which is rarely explicitly addressed in the literature, merits a closer examination.

In this chapter, we would like to provide a systematic overview of studies that allow conclusions regarding this issue. From our point of view, many of these studies support the assumption that comprehension comprises a routine, nonstrategic validation process that detects knowledge violations, which we will term *epistemic monitoring* (Richter, 2011; Richter, Schroeder, & Wöhrmann, 2009). Our literature review will span violations of factual world knowledge (e.g., *Soft soap is edible*), implausibility (e.g., *Frank has a broken leg. He calls the plumber*), inconsistencies with antecedent text (e.g., *Mary is a vegetarian. [...] She orders a cheeseburger*), and semantic anomalies (e.g., *Dutch trains are sour*). Moreover, we will briefly touch on the validation of self-referential statements (e.g., *My name is Ira*) and of statements that refer to a person's value system (e.g., *Euthanasia is acceptable/unacceptable*). We will also consider the role of negation, predictability, and typicality in both comprehension and validation.

We will start by proposing a framework for integrating the findings on these various topics by projecting them onto the common dimension of plausibility. We will then review evidence which demonstrates the sensitivity of early comprehension processes to plausibility, challenging the two-step model's assumption of a non-evaluative comprehension stage. In doing so, we will integrate research from studies spanning three decades that have used a variety of dependent variables (reading times, response latencies, eye movements, and event-related potentials or ERPs) to investigate a variety of phenomena (e.g., semantic integration, syntactic analysis, situation model construction, and prediction). Finally, we will try to reconcile the assumption of nonstrategic validation in language comprehension with findings that show people's failure to notice even blatant violations of their knowledge under certain circumstances by discussing the limitations of epistemic monitoring.

The Role of Knowledge-Based Validation in Language Processing

With knowledge-based (or epistemic) validation, we mean the evaluation of information with regard to its consistency with stored knowledge, which entails the detection of knowledge violations. These knowledge violations can take several forms: They may comprise information that is clearly false (based on semantic or world knowledge), information that is merely implausible, or information that is not false or implausible per se but inconsistent with antecedent text.

It is difficult to draw clear distinctions between these types of knowledge violations. First of all, whether something is perceived as clearly false or merely as implausible depends on the knowledge of the reader and on the certainty of that knowledge. For example, a reader can only recognize a particular sentence as false if he or she possesses the specific knowledge required to assess its truth, while plausibility judgments can be based on less specific or certain knowledge.

Moreover, the detection of inconsistencies with antecedent text sometimes not only requires the reactivation of previous text information but also of relevant semantic or world knowledge. Thus, this type of violation seems to be a subtype of false or implausible information in which the discourse context becomes part of the background against which incoming information is evaluated (in line with the notion that information is immediately related to the widest available context during language comprehension; e.g., Hagoort & van Berkum, 2007; Just & Carpenter, 1980; Nieuwland & van Berkum, 2006; van Berkum, Hagoort, & Brown, 1999; van Berkum et al., 2003). For example, in a classic study by O'Brien et al. (1998), participants first read that either *Mary is a vegetarian* or *Mary is a fast food lover*, and later in the text read that *she orders a cheeseburger*. Reading times were longer when the behavior was inconsistent with the previously described trait, which is generally taken as evidence that the trait is reactivated by a passive memory-based retrieval process. However, it is important to note that the inconsistency only becomes apparent when relevant world-knowledge is also activated – namely, that cheeseburgers usually contain meat and that vegetarians usually do not eat any meat whatsoever (see also Cook & Guéraud, 2005).

For the purpose of integrating the findings reviewed in this chapter, it seems to us that the best way of systematizing the various kinds of knowledge violations is to project them onto the dimension of plausibility. In the extant literature, plausibility has been defined as the goodness of fit with prior knowledge (Connell & Keane, 2004) or as the “relative potential truthfulness of

incoming information compared to our existing mental representations.” (Lombardi, 2012, p. 3). Some researchers use “plausibility” synonymously with “sensibility” (e.g., Speer & Clifton, 1998), while others have obtained plausibility judgments by asking participants to rate how “realistic” (e.g., van Gompel et al., 2001) or how “likely” (e.g., Matsuki et al., 2011; Warren & McConnell, 2007) a described situation is. According to Matsuki et al.,

... plausibility, in its most general form, can be defined as the acceptability or likelihood of a situation or a sentence describing it, as a whole. Plausibility usually is measured by asking participants to rate, on a Likert scale, “How likely it is [*sic*] that the described event occurs in the real world?” (p. 926)

Their definition points to the interesting phenomenon that plausibility of linguistic information is influenced not only by its content, but also by the pragmatic felicity of the utterance itself. Even a true sentence, if it is pragmatically infelicitous, will be perceived as implausible, such as in the case of implausible negatives for which it is difficult to imagine a context in which they are plausibly uttered (e.g., *A sparrow is not a vehicle*; Fischler, Bloom, Childers, Roucos, & Perry, 1983; Wason, 1965).

Effects of plausibility on language processing have been widely acknowledged and well-documented in the literature; however, it is still a point of contention whether these effects reflect downstream, offline, strategic processing or on-line, nonstrategic processing.

Two-Step Models of Comprehension and Validation

Why is the view that evaluative processing is delayed with regard to comprehension so widespread? One major reason for this is the popularity of two-step models of comprehension and validation, which differentiate between separate comprehension and evaluation stages (Connell & Keane, 2006; Gilbert, 1991; Herbert & Kübler, 2011; Wiswede et al., 2013). According to these models, the evaluation of information comes in only after the (non-evaluative) comprehension stage is completed. Moreover, evaluation is usually assumed to be an intentional and therefore optional decision process. This assumption is based on a theoretical distinction between linguistically relevant lexical knowledge (e.g., knowledge about selectional restrictions), which is assumed to be accessed for comprehension, and world knowledge, which is assumed to be accessed for evaluation (e.g., Chomsky, 1965; Rayner et al., 2004).

In line with the two-step model, effects of world knowledge on language processing are sometimes found to be delayed in comparison to effects of semantic knowledge (Rayner et al., 2004; Warren & McConnell, 2007). Moreover, evidence has been presented which suggests that people are bound to believe everything they read at first, and can only effortfully “unbelieve” it at a later point if they have the motivation and cognitive resources to do so (Gilbert et al., 1990, 1993). Consistent with this idea, there are findings suggesting that people sometimes fail to use their general world knowledge adequately even when they are explicitly instructed to do so (e.g., Rapp, 2008), and under some conditions, fall victim to even blatant inaccuracies (e.g., Marsh, Meade, & Roediger, 2003). Moreover, response latencies and ERPs associated with verifying or reading negated sentences have often been found to be more sensitive to semantic mismatches than to a sentence’s truth value, suggesting that word-level matches are privileged over message-level truth or plausibility, at least at an early stage of comprehension (Clark & Chase, 1972; Fischler et al., 1983; Wiswede et al., 2013). Finally, much evaluative processing is indeed offline and deliberate, generally when relevant knowledge for judging truth or plausibility is not (easily) available.

However, from a theoretical perspective, the conceptualization of a non-evaluative comprehension stage is problematic. First of all, the distinction between semantic and world knowledge has been called into question by a number of researchers (e.g., Hagoort et al., 2004; Jackendoff, 2002; Matsuki et al., 2011; this will be discussed in more detail at a later point).

Secondly, modern theories of language comprehension, like the situation model approach (Johnson-Laird, 1983; van Dijk & Kintsch, 1983), assume that world knowledge is already activated in the course of comprehension: Situation models (or mental models) are conceptualized as referential representations of the state of affairs described in the text, which are constructed by integrating text information with prior knowledge, including relevant world knowledge (e.g., Zwaan & Radvansky, 1998). In this way, situation models specify how linguistic expressions relate to the world, and thus represent extensional aspects of meaning, i.e. reference and truth (Johnson-Laird, Herrman, & Chaffin, 1984).

The construction of a situation model has been found to entail the monitoring of multiple dimensions such as time, space, or characters’ goals (e.g., Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998). Discontinuities (e.g., time shifts) or inconsistencies (e.g., conflicts between characters’ goals and their actions) in these dimensions have been shown to slow down processing, which is usually interpreted as

comprehension or integration difficulty. This is also the case when inconsistencies arise between the currently processed text information and antecedent text information that is no longer active in working memory, suggesting that situation models require global coherence, and that previous text information can be reactivated by a passive resonance process (memory-based processing; e.g., Albrecht & O'Brien, 1993; Gerrig & McKoon, 1998; Gerrig & O'Brien, 2005; McKoon, Gerrig, & Greene, 1996; McKoon & Ratcliff, 1998; O'Brien, 1995; O'Brien & Albrecht, 1992; O'Brien et al., 1998; Ratcliff, 1978).

An Alternative Two-Step Model: Epistemic Monitoring Followed by Epistemic Elaboration

Based on the assumption that comprehension entails the construction of a situation model, our central claim in this chapter is that comprehenders monitor incoming information routinely, nonstrategically, and on-line during comprehension with respect to its internal consistency and its consistency with their knowledge and beliefs about the world (Richter, 2003).

In line with Singer (2006), we assume that this monitoring process relies on the activation of knowledge through memory-based processing: "The passive retrieval associated with the memory-based analysis [...] affords the reader with the opportunity to evaluate each discourse constituent in the context of referent text and knowledge" (Singer, 2006, p. 585). However, it is important to note that not all knowledge that is potentially relevant for evaluation will be activated during situation model construction; rather, the activation will be a function of accessibility. It is reasonable to assume that the accessibility of knowledge will, among other things, be influenced by how recently it was previously activated, by its typicality given the reader's experience of the world (e.g., Matsuki et al., 2011), by how well connected it is with other stored knowledge (e.g., Kendeou, Smith, & O'Brien, 2013), by how the currently processed text information is phrased (resulting in more or less surface overlap with knowledge in long-term memory; e.g., Albrecht & O'Brien, 1993; O'Brien et al., 1998), by how focussed the information is in the text (e.g., Sanford, 2002), and by the depth of processing required by the task (e.g., Sanford & Sturt, 2002). Thus, violations of knowledge that is not activated for situation model construction may well go unnoticed.

Accordingly, we do not propose that comprehension entails a full analysis of the (potential) truth of information, but rather a quick and incomplete analysis based on the knowledge that is activated for situation model construction. If an inconsistency with the activated knowledge is detected, the incoming information is initially rejected to protect the situation model from

contamination with false information (Richter et al., 2009). However, if the reader's motivation and cognitive capacity are high, or if the reader pursues an explicit evaluation goal (e.g., a patient reading about his or her medical condition), the detection of implausibility may be followed by a deliberate attempt to resolve the conflict. We call this kind of offline deliberation *epistemic elaboration* (Richter, 2011). It may range from reasoning about the conflicting information (plausible reasoning; e.g., Collins & Michalski, 1989) to attempts to ascertain its validity with the help of external sources (e.g., looking up information in a text book or encyclopedia, or searching the Internet). In this way, epistemic elaboration allows for taking more factors into account than just the immediately accessible knowledge, and thereby making a more elaborate (and probably more justified) judgment about the plausibility of information.

Thus, we propose an alternative two-step model of an evaluative comprehension stage (comprising epistemic monitoring) followed by an optional stage of epistemic elaboration (depending on the reader's goals and motivations). In particular, we would like to argue against the still very prevalent idea that any effects of plausibility on language processing are downstream and reflect intentional evaluative processing. For example, Connell and Keane (2004) proposed a two-step model with a comprehension and a subsequent assessment stage:

For example, if someone is asked to assess the plausibility of the statement *The bottle rolled off the shelf and smashed on the floor*, he or she might make the inference that the bottle rolling off the shelf *caused* it to smash on the floor. Then he or she might consider this elaborated description to be highly plausible because past experience has suggested that falling fragile things often end up breaking when they hit floors. In short, the description has a certain conceptual coherence. (p. 186)

However, it seems more reasonable to us to assume that, rather than making the inference and then judging its plausibility, the plausibility of the inference (based on world knowledge) is the reason *why* it is made in the first place. In this way, plausibility already influences the comprehension stage of the assumed two-step model by affecting which inferences are drawn during text comprehension (e.g., Thorndyke, 1976). Moreover, to the extent that plausibility affords on-line prediction (DeLong, Urbach, & Kutas, 2005; van Berkum et al., 2005), it can exert effects even *prior* to comprehension of a particular linguistic input (for example, in the form of predictive inferences; e.g., Cook, Limber, & O'Brien, 2001).

In fact, the idea that the validity or plausibility of information is monitored during language comprehension is not entirely new. Some researchers have at least alluded to the idea that such a process may exist. For example, according to Fischler et al. (1983), “The negativity associated with anomalous sentences [...] suggests that a basic process in sentence comprehension is the monitoring of the consistency or validity of the propositions asserted by the sentence, with a negativity associated with the disruption of that process” (p. 401). Similarly, West & Stanovich (1982) further investigated the finding that words which are incongruent in their context (e.g., *The accountant balanced the people*) lead to slower yes-responses in lexical decision. They suggested that “responses in the lexical-decision task are affected by postlexical message-level processes that detect incongruity” (p. 385). Fischler and Bloom (1980) pointed out that the interference produced by a detection of incongruity “... would be of value to the reader as a signal that perception or comprehension has failed, and that some reanalysis is called for.” (p. 224). In a similar vein, Murray and Rowan (1998) argued that plausibility effects in eye movements reflect “early, mandatory, pragmatic processing” (p. 1), in contrast with the view that pragmatic plausibility is “extralinguistic, outside the language module and optional in [its] operation” (p. 3). Rather, they concluded that “plausibility effects are not restricted to low-level phrasal units and that they appear to arise as a necessary consequence of the process deriving basic sentence meaning” (p. 1).

Furthermore, Singer and colleagues (Singer, 1993, 2006; Singer, Halldorson, Lear, & Andrusiak, 1992) have provided evidence that readers routinely validate the causal relationships implied by a text (e.g., *Dorothy poured the bucket of water on the fire. The fire went out/grew hotter - Does water extinguish fire?*), and that in self-paced reading, reading times are sensitive to truth and negation, which are factors that have been shown to influence overt sentence verification (e.g., Carpenter & Just, 1975). Based on these findings, Singer (2006) proposed that fully comprehending text entails its verification, and that verification emerges not from an evaluative processing goal but from “the fundamentals of the cognition of reading” (p. 589).

Nonetheless, the two-step model of comprehension and validation is still very prevalent in psycholinguistic research. In the following, we will present a summary of what we perceive as the greatest challenges for this conceptualization.

Challenges for a Two-Step Model of Comprehension and Validation

Findings by Gilbert and Colleagues

Evidence for the view that validation is subsequent, optional, and cognitively effortful primarily comes from studies by Gilbert and colleagues, which show a so called “affirmation bias” in sentence verification: When people learn sentences and their associated truth values under speeded conditions or when they are distracted by a secondary task, they have a tendency to later judge false sentences as true, but not vice versa (Gilbert et al., 1990, 1993). This finding is explained by the two-step-model of comprehension and validation: First, the linguistic message is understood, that is, its meaning is computed based purely on semantic knowledge. Then, in the second step, the computed meaning can be compared with real world knowledge to assess its truth value and – if necessary – “unbelieve” the message. This second step, however, is assumed to be cognitively demanding and therefore only carried out when the reader has the motivation and the cognitive resources to do so. Thus, under speeded conditions, or when readers are put under additional cognitive load, this “unbelieving” is disrupted, leading to a higher amount of falsely accepted sentences. Gilbert et al. (1990, 1993) found support for these assumptions in several studies, using fictitious facts with arbitrarily assigned truth values, for example fictitious Hopi word definitions such as *A monishna is a star*, which were presented to the participants as being either true or false.

However, multiple arguments against the generality of Gilbert et al.’s (1990, 1993) findings have been raised. For example, Richter et al. (2009) pointed out the fact that Gilbert et al. eliminated effects of prior knowledge by using fictitious facts which limits the applicability of their findings. In contrast, when people do have prior knowledge, this may enable them to reject false statements even when they are distracted by a secondary task. To test this assumption, Richter et al. replicated Gilbert et al.’s study with facts for which participants had either weak (e.g., *Toothpaste contains sulphur*) or strong knowledge (e.g., *Perfume contains scents*) and showed that cognitive load produces an affirmation bias only for facts associated with weak knowledge. Sperber et al. (2010) also raised the question of relevance:

Even if the participants could muster some interest for statements about the meaning of Hopi words (and there is nothing in either the experimental situation or the participants’ background knowledge which makes it likely that they would), the information that one of

these statements (e.g. “A Monishna is a star”) is false would still be utterly irrelevant to them. From the knowledge that such a statement is false, nothing follows. With other statements, things may be different. If you had prior reasons for thinking that a certain statement was true, or if it described a normal state of affairs, it is easy to see how you might find it relevant to be told that it is false. (p. 8)

In line with this objection, an experiment by Hasson, Simmons, and Todorov (2005) showed that there is no affirmation bias when the false version of a statement is informative. From this, the authors concluded that it does in fact appear to be possible to suspend belief in comprehended statements, in contrast to what Gilbert et al. (1990, 1993) claimed. In conclusion, it seems that the two-step model proposed by Gilbert and colleagues only applies to the processing of certain kinds of information, but certainly not to all.

Semantic versus World Knowledge Violations – Is Semantic Knowledge Temporally Privileged?

Another main reason for the widespread assumption of a two-step model of comprehension and validation is that some research has suggested that the integration of world knowledge in language comprehension is delayed relative to semantic knowledge. However, as we will show in the following, it is debatable whether the two types of violations can actually be clearly distinguished on a theoretical basis, and whether their operationalization in psycholinguistic studies has been appropriate.

From a theoretical point of view, the distinction between semantic knowledge and world knowledge is based on a distinction between meaning and truth, or in other words, on the distinction between the construction of meaning and the verification of that meaning (Hagoort et al., 2004). Meaning construction is assumed to rely purely on semantic knowledge, whereas verification also draws on (extralinguistic) world knowledge. Semantic violations are assumed to be violations of purely semantic knowledge: They arise when a word or phrase violates the selectional restrictions placed by the context. For example, as Hagoort et al. (2004) pointed out, the sentence *The favorite palace of the present queen of England is divorced* violates the selectional restrictions of the predicate *is-divorced* because being divorced requires an animate object as its argument. Therefore, this sentence is semantically malformed. This malformation, according to different definitions, causes the sentence to be incapable of having a sense (“senseless

sentences”; Kutas & Hillyard, 1980, p. 203), of being true (“Selectional restriction violations lead to impossibility”; Warren & McConnell, 2007, p. 770), or of having a truth value because their truth conditions are unclear (Asher, 2011). In contrast, world knowledge violations are perfectly sensible and could therefore theoretically be true, but it is the (current) state of affairs in the real world that renders them false (e.g., *The present queen of England is divorced*; Hagoort et al., 2004).

However, the problem is that not all stimuli clearly fall into one category or the other. For example, Kutas and Hillyard (1980) used “strongly anomalous” sentence completions (e.g., *You can’t make a silk purse out of a cow’s chair* or *He took a sip from the transmitter*) and “moderately anomalous” completions (e.g., *You can’t make a silk purse out of a cow’s skin* or *He took a sip from the waterfall*). It is important to note, however, that the latter examples seem to violate world knowledge rather than selectional restrictions. Similarly, van Berkum et al. (1999) referred to words that were implausible in their context (e.g., *Jane told her brother that he was exceptionally slow*, when the brother had previously been described as having been very quick) as “discourse-dependent semantic anomalies” (p. 657), although these words did not render the context sentences senseless. Likewise, the so-called Moses illusion (i.e., the often overlooked error in the question “How many animals of each kind did Moses take on the ark?”; Erickson & Mattson, 1981) has often been described as a semantic anomaly, when it is clearly world knowledge that renders it false (it should be Noah instead of Moses). As these examples show, the terminology in the literature has at the very least been blurry. Matsuki et al. (2011) point towards a similar problem by remarking that some of the selectional restriction violations that have been used in psycholinguistic studies (e.g., Warren & McConnell, 2007) involve violations of highly verb-specific restrictions such as *inflatable*, *catchable*, *cookable*, *mixable*, or *edible*, rather than of abstract verb-general features such as *animacy* or *humanness*. However, knowing which objects fulfil these verb-specific requirements seems to qualify as world or event knowledge rather than as specifically semantic knowledge.

An additional problem is that different definitions of semantic violations exist in linguistics and psycholinguistics: A sentence such as *He spread the warm bread with socks* (Kutas & Hillyard, 1980), which has been considered a semantic violation in psycholinguistics, would be categorized as a typical world knowledge violation in linguistics because “it follows all the semantic rules of English, and the reason it sounds odd is simply that socks do not have the right chemical make-up to function as a spread” (Pylkkänen, Oliveri, & Smart, 2009, p. 1314).

Moreover, it is always possible to construct a context – at least a fictional or metaphorical one – in which an anomaly can appear perfectly sensible. For example, it is not unusual for waiters to refer to their customers by the names of the dishes they ordered, so the utterance “The ham sandwich is getting impatient” would not be perceived as anomalous in this context (Asher, 2011, p. 237). This notion is supported by results showing that animacy violations which are consistent with the discourse context (e.g., a peanut described as being in love in a cartoon-like discourse) elicit a smaller N400 (an ERP component that indexes semantic integration) than information which is consistent with world knowledge, but does not fit well with the context (e.g., the same peanut described as being salted; Nieuwland & van Berkum, 2006; see also Filik & Leuthold, 2008). Consequently, the context can obviously override local animacy requirements.

Moreover, Hagoort et al. (2004) found that the time-course and amplitude of the N400 are highly similar for the processing of world knowledge violations (e.g., *Dutch trains are white*) and semantic violations (e.g., *Dutch trains are sour*) as compared to correct controls (e.g., *Dutch trains are yellow*). Thus, whether or not a clear distinction is feasible, the kinds of knowledge usually separated into semantic and world knowledge in psycholinguistics appear to be integrated simultaneously during language comprehension. This speaks against a two-step model of comprehension and validation in which first meaning and then truth value is computed.

However, not all studies support this conclusion. Using eye-tracking technology, Warren and McConnell (2007) and Rayner et al. (2004) found that semantic violations have immediate disruptive effects on reading, whereas those of world knowledge violations are delayed. Matsuki et al. (2011) tried to reconcile the contradictory results by proposing that the typicality of the items used may be key. They hypothesized that in order to find early effects of plausibility, it is crucial that the plausible items are typical of the reader’s experience, and that the plausible items used by Warren and McConnell (2007) and Rayner et al. (2004) may not have been typical enough given participants’ real-world event-knowledge (e.g., items about dusting miniatures or catching a goose with a trap). To ensure typicality, Matsuki et al. (2011) based their stimuli on production norms in addition to plausibility ratings (e.g., by asking “what do you cut with a knife?”). With these stimuli, they found immediate effects of plausibility (or typicality) in both eye-tracking and self-paced reading. These results suggest that the goodness of fit with readers’ knowledge and experience (and consequently, the accessibility of knowledge) may be the key variable, rather than the type of knowledge necessary for validation.

However, semantic and world knowledge may gradually differ on the dimensions of commonality, stability, and accessibility: knowledge which is at the core of word meanings tends to be shared by many language users, to be more stable, and may therefore be on average more easily accessible. Moreover, the verification of world knowledge violations may on average require more specialized knowledge. To verify the sentence *The favourite palace of the present queen of England is divorced*, one merely needs to draw on the (rather general) knowledge that a palace is a building and buildings cannot be divorced. However, to verify the sentence *The present queen of England is divorced*, one needs to possess and invoke more specific knowledge about the present queen of England and her biography.

Nonetheless, it is unclear at this point whether a clear distinction between semantic and world knowledge violations is possible and useful. It is not our goal in this chapter to advance this debate, but we do find it worthy of noting here because we think that the heterogeneous use and blurry operationalizations of terms obscure the relevance of some existent studies for validation in language comprehension. For the present purpose, it seems both fruitful and parsimonious to view the different types of violations on a continuous dimension of plausibility, rather than trying to separate them into different categories. This may seem like an oversimplification at first glance, but plausibility ratings obtained in various studies (e.g., Rayner et al., 2004; Warren & McConnell, 2007) support the idea that both semantic and world knowledge violations can be adequately arranged on a continuum of plausibility, with semantic violations usually receiving lower plausibility ratings. It is also in line with the terminology chosen by Warren (2011), who uses implausibility and anomaly virtually interchangeably and locates different kinds of knowledge violations on a dimension of “plausibility violation severity” (p. 914). Moreover, as we hope to show in this chapter, the results of these seemingly different types of knowledge violations can be integrated into a surprisingly coherent picture, supporting the idea that there is a common underlying dimension (e.g., Fischler, Bloom, Childers, Arroyo, & Perry, 1984).

Immediate Effects of Plausibility on Comprehension

A highly consistent finding in the literature is that implausible information is processed more slowly as compared to plausible information. Specifically, it leads to longer reading times in self-paced reading (e.g., Albrecht & O’Brien, 1993; O’Brien et al., 1998; Singer, 2006) and longer fixations in eye-tracking (e.g., Cook & Myers, 2004; Murray & Rowan, 1998; Rayner et al., 2004; Staub et al., 2007; Warren & McConnell, 2007), as well as to more regressions and re-reading

(e.g., Braze, Shankweiler, Ni, & Palumbo, 2002). Processing difficulties due to implausibility are immediate, suggesting an incremental assessment of the plausibility of linguistic input (Staub et al., 2007; Traxler & Pickering, 1996), which is unaffected by the distance and structural relationship between the words cueing the implausibility (Patson & Warren, 2010). Moreover, the reading time penalty caused by an implausibility correlates with offline plausibility ratings (Staub et al., 2007), suggesting that the plausibility judgment – which can be computed in an intentional, offline decision process – is in fact in some form immediately available during on-line comprehension (see also Isberner & Richter, 2013).

In line with the findings demonstrating implausibility-related comprehension disruptions, Black, Freeman, and Johnson-Laird (1986) have shown that the more implausible a text is, the more it is judged as surprising and as difficult to comprehend, and the less it is remembered. Implausible information has also been shown to be less likely to be integrated into the situation model (Maier & Richter, 2013; Schroeder, Richter, & Hoever, 2008). Other work from our own lab (Isberner & Richter, 2013, in press; Richter et al., 2009) as well as results obtained in lexical decision tasks (e.g., West & Stanovich, 1982) also suggest that implausibility makes any kind of subsequent positive responses more difficult, even if the task that requires them is completely unrelated to plausibility. Moreover, plausibility guides the knowledge-based inferences that are generated during comprehension, such as thematic or predictive inferences (e.g., Federmeier & Kutas, 1999; Hannon & Daneman, 1998; Long, Oppy, & Seely, 1994; Peracchi & O'Brien, 2004; Thorndyke, 1976).

Furthermore, plausibility has been found to influence syntactic analysis and reanalysis. For example, it can bias analysis in syntactically ambiguous sentences (van Gompel et al., 2001). In the so-called “garden path” phenomena, it affects the commitment to the initial analysis and therefore, recovery from an implausible misanalysis is faster than from a plausible one (Pickering & Traxler, 1998). Plausibility has also been present in everyday reasoning contexts (plausible reasoning; e.g., Collins & Michalski, 1989) and has been shown to be used as a shortcut in recognition when the retrieval of an exact memory match is difficult or costly (Lemaire & Fayol, 1995; Reder, 1982; Reder, Wible, & Martin, 1986).

Another finding that has been associated with implausibility is an elevated N400 ERP (Ferretti, Singer, & Patterson, 2008; Fischler et al., 1984, 1985; Hagoort et al., 2004; Nieuwland & Kuperberg, 2008; van Berkum et al., 1999, 2003). This ERP component was elevated for all types of knowledge violations that we discussed in the previous sections, supporting our decision to

subsume them on a single dimension of plausibility. In addition, violations of self-referential knowledge (e.g., *My name is Ira*; Fischler et al., 1984) and statements contradicting (moral) beliefs (e.g., *I think euthanasia is an acceptable/unacceptable course of action*; van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009) have also demonstrated an elevated N400, suggesting that these types of knowledge are also accessed during comprehension and used for validation. In line with this idea, van Berkum et al. (2009) concluded from their data that “strong disagreement rapidly influences the ongoing analysis of meaning, indicating that even very early processes in language comprehension are sensitive to a person’s value system” (p. 1092). As Lombardi (2012) has pointed out, “Plausibility judgments do not rely on absolute definitions of and distinctions between knowledge and belief” (p. 4), suggesting that there may be little difference between the processing of information that is inconsistent with one’s knowledge and the processing of information that is inconsistent with one’s beliefs.

We propose that these findings should not merely be seen as showing the processing costs of implausibility, but rather as evidence for a purposeful on-line and nonstrategic validation process (epistemic monitoring) which protects the system from false information and thereby – in general – promotes accurate and stable mental representations (Schroeder et al., 2008).

Does the N400 Index Validation or Detection of Semantic Mismatches? Evidence from Studies on Negation, Self-Referential Statements, and Knowledge Acquired in the Lab

A finding which has often been interpreted as evidence for a two-step model of comprehension and validation is that in tasks crossing truth value and negation, the N400 has been found to be more sensitive to semantic matches versus mismatches than to truth value (e.g., Fischler et al., 1983; Wiswede et al., 2013). However, the problem with many of the negated sentences that have been used in these studies is that they violate pragmatic rules. According to Wason (1965), negation is typically used to deny an assumption that the recipient is assumed to hold – for example, to point out exceptions to a rule or deny plausible misconceptions (e.g., “A whale is not a fish”). In contrast, denying something that makes no sense in the first place violates the conversational maxim to be informative (Grice, 1975; Sperber & Wilson, 1995). Against this background, a sentence such as *A sparrow is not a vehicle* (Fischler et al., 1983), which is representative of the kind of stimuli used in these studies, seems pragmatically implausible. Fischler et al. (1983) also refer to these types of negated sentences as implausible negatives to point out that the negative itself is implausible, even if the overall sentence is true according to

formal logic. Thus, in this particular case, the plausibility of the utterance per se runs counter to the formal truth value of its content.

Unlike most other studies on the processing of negation, Nieuwland and Kuperberg (2008) used negated sentences that were pragmatically licensed (e.g., *With proper equipment, scuba diving is not very dangerous*). Under these conditions, the N400 was modulated by truth value rather than by semantic matches or mismatches, which speaks against a two-step model of comprehension and validation, in which “nonpropositional semantic processes precede the decision processes that compute sentence truth value” (Nieuwland & Kuperberg, 2008, p. 1213). Rather, under these conditions, people appear to be able to validate the negated sentences without any delay.

Fischler and colleagues used two other methods to avoid confounds between semantic mismatch and truth value. In one of their studies (Fischler et al., 1984), they investigated the verification of statements that referred to facts about the participants themselves, such as *My name is Ira*. Although these stimuli contained neither semantic mismatches nor semantic anomalies, false statements were marked by a larger N400 than true statements. In another study, Fischler et al. (1985) investigated the processing of knowledge acquired in the lab, whose truth and falsity was determined arbitrarily. An elevated N400 distinguished false from true sentences, regardless of whether the participants were explicitly asked to verify the sentences, or to “continue to look at the statement and read the words, but not make any decision or response about its truth value or meaning” (p. 87), or to respond incorrectly (i.e., respond with “false” to a true statement, and vice versa). Based on these results, the authors concluded that comprehension is automatic: “It is concluded that attending to a presented word results in an automatic analysis of its meaning in the context of a preceding verbal input, and that ERPs can indicate the nature of the output of that analysis” (p. 83). It is important to note, however, that this conclusion is actually based on the automaticity of the true/false discrimination. Thus, Fischler et al. directly equated analysis of meaning with the determination of truth value, consistent with Singer’s (2006) proposal that fully comprehending a sentence entails its verification.

Prediction: Readers Use Their World Knowledge to Predict Upcoming Words Before They Appear

Knowledge-based prediction in language comprehension represents another phenomenon that speaks against two-step models of comprehension and validation. For example, many studies

suggest that people make predictive inferences based on their world knowledge (e.g., Cook et al., 2001; Peracchi & O'Brien, 2004). These experiments usually have participants read texts which strongly imply a particular consequence (e.g., a person falling from the 15th story of a building) and then test for the activation of the predictive inference by having readers name or perform lexical decision on a word which is assumed to represent the inference (e.g., "dead"). Faster naming or lexical decisions compared to a baseline condition are taken as evidence that the inference was indeed activated. However, few studies are conclusive regarding the question of whether this processing advantage may simply reflect easier post-lexical integration rather than prediction. An exception is a study by van Berkum et al. (2005) on on-line prediction of specific upcoming words. Using an ingenious paradigm, they were able to show effects of prediction before the predicted word actually appeared by making use of the fact that adjectives in Dutch are gender-inflected and positioned before the noun. If readers or listeners predict a particular word in a certain context, they should struggle when they encounter an adjective whose gender does not match that of the expected word. Van Berkum et al. found evidence for this disruption both in ERPs and reading times, and both for reading and listening. These results clearly show that world knowledge not only exerts an influence after comprehension has terminated, but that readers use it on-line during comprehension to predict what will be said (or written) next. Thus the fit between their world knowledge and unfolding discourse immediately affects subsequent processing.

Evidence for Nonstrategic Validation

One may raise the concern that many of the above studies may have triggered evaluative processing by repeatedly exposing participants to knowledge violations. However, this concern can be rejected on the basis of the results of studies reporting early effects of local implausibility despite using globally plausible stimuli (e.g., Pickering & Traxler, 1998; Matsuki et al., 2011; Staub et al., 2007). This suggests a nonstrategic, fast, word-by-word plausibility assessment that is not triggered by repeated exposure to unusual stimuli. Moreover, in a study by Fischler and Bloom (1979, Experiment 2), effects of implausibility on lexical decision latencies disappeared when all of the stimuli were implausible, suggesting that a large proportion of implausible stimuli in an experiment reduces validation rather than encouraging it, at least when participants are not explicitly instructed to validate. Thus, it is unlikely that epistemic monitoring is merely triggered by implausible or anomalous stimuli.

Evidence from our own lab: The epistemic Stroop effect. The aforementioned studies suggest that evaluative processing during comprehension is the default mode of processing (i.e., whenever readers are instructed to read for comprehension), but they still leave open the possibility that readers can “switch off” validation when it is strategically useful to do so. However, there is evidence from our own lab that speaks against this possibility.

To test whether it is possible for readers to ignore validity or plausibility when it is irrelevant to a task, we used a Stroop-like paradigm (Stroop, 1935) in which participants were required to respond positively or negatively after reading true (plausible) or false (implausible) sentences. Crucially, the actual experimental task did not require or encourage validation: It involved an orthographic task in which participants were asked to indicate whether a particular word was spelled correctly or not (Richter et al., 2009; Isberner & Richter, 2013), a nonlinguistic color judgment task in which participants judged whether or not a word had changed color (Isberner & Richter, 2013), or a simple probe task in which participants were required to respond to the words “true” and “false” (as introduced by Wiswede et al., 2013; Isberner & Richter, in press). Participants were presented with the sentences word-by-word using rapid serial visual presentation (RSVP) and were prompted to perform the assigned task at varying points during sentence presentation. In experimental items, which varied in validity or plausibility, the prompt always appeared immediately after the end of the sentence. We predicted that if validation is a nonstrategic process, it should be more difficult to respond positively after reading invalid/implausible information because the validation process should bias the system towards a negative response (Figure 1). In line with this prediction, a consistent finding in all of the tasks was that task-irrelevant plausibility affected response latencies, resulting in a significant interaction of plausibility and required response (Figure 2 shows representative data from two of our experiments). This was the case regardless of whether the stimuli used were true or false sentences (referring to factual knowledge), plausible or implausible scenarios (referring to event-based knowledge), and whether plausibility was manipulated by the intra-sentential (e.g., *Soft soap is edible*) or extra-sentential context (e.g., *Frank has a broken pipe/leg. He calls the plumber.*). In particular, positive responses were always much slower after reading implausible stimuli, which suggests that implausible information elicits a negative response tendency which interferes with any kind of positive response.

In fact, similar interference has been found in lexical decisions for “yes” responses to words that are incongruous in their context (e.g., Fischler & Bloom, 1979; West & Stanovich,

1982), the difference being that these incongruous completions, unlike ours, mostly comprised more severe plausibility violations which some might classify as semantic violations (e.g., *We stayed until the pants*). Moreover, in lexical decision, validation processes are confounded with access to word meaning (West & Stanovich, 1982). Therefore, tasks that are non-semantic or even non-linguistic, such as the tasks we have used so far in our Stroop-like paradigm, appear to be better suited for directly investigating nonstrategic validation. Nonetheless, the consistency of our findings with these previous results (despite using very different stimuli and tasks) is noteworthy. Moreover, this effect was present even when participants were instructed to treat the context reading and lexical decision as different tasks (Fischler & Bloom, 1979, Experiment 5), suggesting that epistemic monitoring is difficult to suppress. However, the effect disappeared when the participants only saw incongruous (or implausible) sentences (Fischler & Bloom, 1979, Experiment 2), suggesting that epistemic monitoring can be reduced by strong manipulations of the task context.

Are the effects of validation due to prediction? Another question we have attempted to answer in our experiments is whether the effects of plausibility may be due to prediction rather than validation. As previously discussed, readers can exploit their world knowledge and situation model to predict upcoming words during on-line comprehension (DeLong et al., 2005; van Berkum et al., 2005), and words that are implausible in their context are often surprising, i.e., not predictable (Black et al., 1986). Therefore, the disruptive effects of implausibility may be simply due to the fact that the implausible completion is unexpected.

Plausibility and predictability are empirically very difficult to disentangle (see also Matsuki et al., 2011), as they are both related to the goodness of fit with what one knows about the world. However, implausible completions are always unpredictable, while unpredictable completions are not necessarily implausible. Taking advantage of this fact, we tried to disentangle effects of plausibility and predictability in our own experiments by varying the predictability of the target words in the plausible condition (Isberner & Richter, 2013). To do so, we obtained cloze norms for sentence pairs (e.g., *Frank has a broken pipe. He calls the...*) which were strongly biased towards a particular completion (*plumber*) but at the same time permitted other plausible completions (e.g., *tradesman*). Based on these norms, we selected completions that had either a high ($M = 75\%$) or a low ($M = 6\%$) cloze value in the plausible context, but were equally plausible when paired with the plausible context sentence (e.g., *Frank has a broken pipe. He calls the plumber/tradesman*) and equally implausible when paired with the implausible context sentence (e.g., *Frank has a broken*

leg. He calls the plumber/tradesman). If the epistemic Stroop effect reported in the previous section was due to disconfirmations of specific lexical predictions, then it should emerge only for highly predictable completions, resulting in a three-way interaction with predictability. Alternatively, a two-way interaction of predictability and required response analogous to the predicted interaction of plausibility and required response might emerge. Contrary to these notions, the epistemic Stroop effect was not modulated by predictability, suggesting that it hinges on message-level implausibility rather than on the disconfirmation of a specific lexical prediction.

These results are consistent with results presented by West and Stanovich (1982), who found that “yes” responses in a lexical decision task are delayed when the word is presented in an incongruous context. Given that there is no such delay in naming tasks, they interpreted that it is not a delay (or inhibition) of lexical access but rather a bias towards a “no” response because incongruity is detected on the message-level. They were also able to confirm that it is congruity at the message level rather than “a mismatch between the stimulus word and lexical-level expectations” (p. 385) which produces this effect, because there was no such inhibition for unexpected but congruous words. Thus, these convergent findings suggest differential effects of plausibility and predictability.

Open questions – Future directions. In future studies, it would be desirable to establish more clearly to what extent interference and facilitation contribute to the observed Stroop-like effect, as the present results are not fully conclusive regarding this question. For this purpose, it would be useful to introduce an adequate neutral condition into the design. It would also be interesting to investigate the processing of self-referential or belief-related statements with our paradigm. There are ERP studies suggesting validation also occurs rapidly in these types of sentences (Fischler et al., 1984; van Berkum et al., 2009), and we would assume that our behavioral measures should be sensitive to these types of violations as well.

In any case, our paradigm has proven to be a useful tool for investigating non-strategic validation in language comprehension. As such, it can be used to investigate how validation is affected by context and task demands (e.g., by the text genre or the required depth of comprehension). It could also shed light on (gradual or categorical) differences in the processing of knowledge violations, for example, in the time course or strength of the interference effects of semantic and world knowledge violations.

Does Validation Always Work?

The studies presented in the previous sections may foster the impression that validation is, so to speak, “infallible.” However, this conclusion would be inconsistent with a large number of studies where validation appears to be absent. The most extreme examples are so called semantic illusions, in which a blatant violation of world knowledge (which should be easily accessible) goes unnoticed, such as *The authorities were trying to decide where to bury the survivors* (Barton & Sanford, 1993). But there are also less extreme examples of people’s susceptibility to false information. For example, Marsh and colleagues have reported evidence suggesting that readers sometimes rely on previously read information which they should know to be false to later answer general knowledge questions (e.g., Bottoms, Eslick, & Marsh, 2010; Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013; Fazio & Marsh, 2008; Marsh & Fazio, 2006; Marsh et al., 2003). Similarly, findings by Rapp (2008) suggest that prior knowledge does not protect readers from being affected by false information. Warning readers about this in advance or explicitly encouraging them to use their prior knowledge does not seem to eliminate these effects (e.g., Marsh & Fazio, 2006; Rapp, 2008).

More evidence that seemingly contradicts the notion of non-strategic validation comes from a study by Wiswede et al. (2013). They used a Stroop-like paradigm similar to ours in which participants were asked to respond to a “true” or “false” probe presented after a true or false sentence (orthogonal to its truth value). Two groups of participants performed the same probe task, but for each group, it was randomly intermixed with a second task: The experimental group performed a truth evaluation task (“Is the sentence true or false?”), whereas the control group performed a sentence comparison task (“Is this the sentence that you’ve just seen?”). Until the prompt appeared (1500 ms after the final word), the participants did not know which task they would have to perform. Thus, for the experimental group, evaluation of the sentences was strategically useful on half of the trials, whereas for the control group, it was never useful. For the experimental group, Wiswede et al. found a compatibility effect in the probe task similar to our Stroop-like effect, with responses being faster when the required response matched the actual truth value of the sentence. For the control group, however, there was no such effect. Wiswede et al. thus concluded that validation is only conditionally automatic, i.e., depending on an evaluative mindset of the reader, which they assume to be induced by an evaluative task such as the orthographic task we used in our studies.

We think, however, that alternative interpretations of the results are conceivable: Unlike the study by Nieuwland and Kuperberg (2008) discussed above, Wiswede et al. (2013) used pragmatically unlicensed negatives (Wason, 1965). Therefore, the plausibility of the negative sentences themselves ran counter to their formally derived truth value. Second, it is possible that the sentence comparison task not only reduced evaluative but also semantic processing in general. Notice that it would be possible to perform this task in a foreign language on a purely perceptual level. This interpretation is supported by an attenuation of the influence of semantic mismatches on the amplitude of the N400 in the control group in Wiswede et al.'s (2012) study. This suggests that epistemic monitoring is closely tied to the depth of comprehension. We will discuss this issue in more detail in the following section.

Limits of Epistemic Monitoring - Reconciling the Contradictory Findings

We hope we have been able to show in this book chapter that routine validation can protect the mental system from false/implausible information to some extent (epistemic vigilance, Sperber et al., 2010), but also that this protection is far from perfect. The biggest challenge at the present time seems to be the integration of findings showing people's resistance to correcting information when they have acquired false knowledge on the one hand (continued influence of misinformation; Ross, Lepper, & Hubbard, 1975; Johnson & Seifert, 1994) – which is in line with the notion that previously acquired knowledge is used for validating new information – and people's apparent susceptibility to false information on the other hand, even when they have relatively strong and certain knowledge which should prevent this. However, in light of the abundance of evidence for validation in language comprehension, we would agree with Singer (2006) that the conditions under which validation fails “tend to specify the factors that regulate text verification rather than diagnose readers' systematic failure to scrutinize text” (p. 588). He identified factors on the part of the text, the reader, and the task, some of which we will reiterate here.

1. Validation can be based on false/subjective beliefs and, hence, contribute to the persistence of such beliefs. We suggested that prior knowledge and beliefs are used for validating incoming information in order to protect the mental system from contamination with false information. However, this conversely implies that in the case of false beliefs, validation can actually hinder the acquisition of correct knowledge (e.g., Maier & Richter, 2013) because prior beliefs immediately influence the analysis of meaning (van Berkum et al., 2009). This may also

explain people's resistance to revising their beliefs or previously acquired knowledge in the face of correcting information (Ross et al., 1975; Johnson & Seifert, 1994).

2. *Validation is moderated by available knowledge/beliefs.* Naturally, only knowledge that is available and activated during comprehension can be used for validation. Knowledge can be activated either through memory-based processes (independent of reading goals, O'Brien & Myers, 1999) or through strategic memory retrieval (dependent on reading goals). For the detection of inconsistencies in a text, the co-activation of the conflicting information is crucial (van den Broek & Kendeou, 2008). As a consequence, inconsistencies will not be detected if conflicting information is not co-activated (e.g., Otero & Kintsch, 1992). Consequently, a reader characteristic that is likely to play a role for validation is working memory capacity, as it limits the extent to which information can be co-activated (e.g., Hannon & Daneman, 2001; Singer, 2006). Daneman and colleagues also showed that reading skill (Daneman, Lennertz, & Hannon, 2007) and the ability of readers to access prior knowledge from long-term memory (Hannon and Daneman, 2001) moderate the detection of semantic anomalies. But characteristics of the text also play a role: False information is particularly likely to be missed when it is semantically strongly related to its context (Hannon & Daneman, 2001). Moreover, to the extent that readers cannot memorize all of the details mentioned in a text (Kintsch & van Dijk, 1978), inconsistencies with antecedent text may go unnoticed when the text is complex (Glenberg, Sanocki, Epstein, & Morris, 1987; Otero & Kintsch, 1992) or when it does not provide sufficient retrieval cues (e.g., lack of surface overlap with antecedent text; Albrecht & Myers, 1995). Similarly, implausibility may be overlooked in syntactically complex sentences, such as *No head injury is too trivial to be ignored* (Wason & Reich, 1979), whose exact meaning is so difficult to compute that readers may rely on a shallow analysis. In fact, it seems that in shallow semantic processing, pragmatics in the form of situation-specific plausibility can actually *override* local semantic or syntactic processing (Sanford, 2002; Sanford & Sturt, 2002), which is clearly at odds with a two-step model of comprehension and validation.

3. *Routine validation processes may be conditionally automatic* (Wiswede et al., 2013). Validation is assumed to be routine in the sense of conditionally automatic processes (Bargh & Chartrand, 1999). As such, it is nonstrategic (involuntary, i.e., it does not require specific processing goals). Nevertheless, it still depends on certain conditions: It may be modulated by mindsets (e.g., it may be influenced by text genre or by perceived credibility of the text source), but further research is necessary to elucidate the modulating conditions. For example, readers

seem to be particularly susceptible to misinformation and persuasion when they read narratives (e.g., Appel & Richter, 2007; Gerrig & Prentice, 1991; Green & Brock, 2000), which suggests that epistemic monitoring of incoming information might be suppressed to some extent in this text genre.

4. *Readers can fall victim to false information when it is sufficiently plausible.* It must be noted that studies demonstrating readers' susceptibility to misinformation usually probe knowledge that is on average not held with high certainty among the participants, and they always use lures that are relatively plausible (e.g., Marsh et al., 2003). Since plausibility judgments correlate with effects of plausibility on on-line comprehension (Staub et al., 2007), it seems possible that the more plausible false information is, the less likely it is to be reliably rejected by epistemic monitoring. In line with this idea, the plausibility of false information has been shown to affect the probability with which readers accept it as (potentially) correct (Hinze, Slaten, Horton, Jenkins, & Rapp, submitted). Moreover, epistemic monitoring is based on easily accessible knowledge, not on a complete analysis of all principally available knowledge. Unfortunately, this makes it prone to error by changes in the accessibility of inaccurate knowledge. It has been shown that episodic memory traces of false information can interfere with the retrieval of correct information from long-term memory when the false information is more readily available, even if the correct information is held with a relatively strong degree of certainty (e.g., Marsh et al., 2003). As Fazio et al. (2013) point out, it seems that the false information does not overwrite the correct knowledge, but that the two representations co-exist in memory and that temporarily higher accessibility of the false information (for example, because it was more recently encountered) explains why readers may inaccurately rely on this information. In line with this idea, readers' reliance on the false information decreases as its activation fades over time (e.g., Barber, Rajaram, & Marsh, 2008; Marsh et al., 2003).

5. *Validation is based on a quick and incomplete analysis.* Naturally, we are not suggesting that all kinds of evaluative processes occur as part of comprehension, or that epistemic monitoring allows a complete analysis of the (potential) truth of information. Rather, epistemic monitoring only detects (and initially rejects) inconsistencies with easily accessible knowledge. Whether a reader then elaborates on a detected inconsistency (epistemic elaboration) depends on his or her goals and strategies (Maier & Richter, 2013; Richter, 2011).

6. *Readers perform validation to the extent that they understand a linguistic message.* It is important to note that we do not propose validation to be a routine component of listening or

reading per se, but of comprehension. Language is often processed in a shallow manner (Barton & Sanford, 1993), which can lead to “good-enough” representations but sometimes also to false representations (Ferreira, Bailey, & Ferraro, 2002). Shallow processing of language can lead to failures to notice even blatant errors, as in the case of the previously mentioned semantic illusions. For example, in the case of the Moses illusion (Erickson & Mattson, 1981), the wrong kind of knowledge seems to be activated and used for interpretation as well as for validation, causing both comprehension and validation to fail and resulting in an inaccurate representation of what is being asked. Moreover, linguistic focus is likely to have an influence on the success of epistemic monitoring: False or implausible information that is not in linguistic focus, especially when it is marked as presupposed or “given” (Haviland & Clark, 1974) as in the Moses illusion, is more likely to be missed by the monitoring process. Consistent with this idea, detection rates are much higher when the implausible information is focalized (e.g., “It was Moses who put two of each kind of animal on the ark”; Bredart & Modolo, 1988). Linguistic focus has been suggested to affect the specificity of the mental representation constructed during comprehension (Sanford & Garrod, 2005). Thus, failures to notice such errors may be due to the construction of an underspecified mental representation (Bohan & Sanford, 2008; Sanford, 2002; Sanford & Graesser, 2006; Sanford, Leuthold, Bohan, & Sanford, 2011).

Because of this close relationship between comprehension and validation, one may ask whether epistemic monitoring may not be more accurately termed comprehension monitoring. However, it is important then to keep in mind that comprehension itself already seems to comprise evaluative processing with regard to world knowledge, as we have hopefully been able to show in this chapter. Moreover, it is the detection of implausibility that prompts people to reanalyze a syntactically or semantically ambiguous sentence when they have initially chosen the wrong interpretation (e.g., when reading the ambiguous sentence “He gave her cat food”) – *not* because of a failure to extract any meaning at all, but because of a failure to extract a meaning that is plausible with regard to world knowledge and the current situation model. In this way, validation seems to serve as a means for preserving the usefulness of language for its primary purpose: successful communication about states of affairs in the real world.

Summary

We propose that the disruption of comprehension by implausible information, which has been shown in many studies, does not just reflect processing costs of implausibility, but rather a highly purposeful validation process that protects the mental system from false information. This validation process (termed *epistemic monitoring*) appears to be incremental, immediate, context-sensitive, and nonstrategic. However, the protection it provides is far from perfect, given that epistemic monitoring relies on easily accessible knowledge and seems to hinge on a minimum depth of processing.

Acknowledgements

The research reported in this chapter was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, grants RI 1100/4-2 and RI 1100/5-2).

References

- Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1459–1468.
- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1061–1070.
- Appel, M., & Richter, T. (2007). Persuasive effects of fictional narratives increase over time. *Media Psychology*, 10, 113-134.
- Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge: Cambridge University Press.
- Barber, S. J., Rajaram, S., & Marsh, E. J. (2008). Fact Learning: How information accuracy, delay and repeated testing change retention and retrieval experience. *Memory*, 16, 934-946.
- Bargh, J.A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462-479.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21, 477-487.
- Black, A., Freeman, P., & Johnson-Laird, P.N. (1986). Plausibility and the comprehension of text. *British Journal of Psychology*, 77, 51-62.
- Bohan, J., & Sanford, A. (2008). Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *The Quarterly Journal of Experimental Psychology*, 61, 232-239.
- Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses Illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, 18, 670-678.
- Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, 31, 25-44.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67, 135-144.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82, 45–73.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3, 472–517.
- Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science*, 13, 1-49.
- Connell, L., & Keane, M. T. (2004). What plausibly affects plausibility? Concept coherence and distributional word coherence as factors influencing plausibility judgments. *Memory & Cognition*, 32, 185-197.
- Connell, L., & Keane M. T. (2006). A model of plausibility. *Cognitive Science*, 30, 95-120.
- Cook, A. E., & Guéraud, S. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes*, 39, 265-278.
- Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language*, 44, 220-234.
- Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, 50, 268-288.
- Daneman, M., Lennertz, T., & Hannon, B. (2007). Shallow semantic processing of text: Evidence from eye movements. *Language and Cognitive Processes*, 22, 83-105.
- Davidson, D. (2001). *Inquiries into truth and meaning* (2nd ed.). Oxford, England: Oxford University Press.
- DeLong K. A., Urbach T. P., & Kutas M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.
- Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540-552.
- Fazio, L. K., Barber, S. J., Rajaram, S., Ornstein, P. A., & Marsh, E. J. (2013). Creating illusions of knowledge: Learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General*, 142, 1-5.
- Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, 15, 180-185.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469-495.
- Ferreira, F., Bailey, K., & Ferraro, V. (2002). Good enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.

- Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the timecourse of verifying text ideas. *Cognition*, 108, 881-888.
- Filik R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, 45, 554-558.
- Fischler, I., & 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, 1-20.
- Fischler, I., & Bloom, P. A. (1980). Rapid processing of the meaning of sentences. *Memory & Cognition*, 8, 216-225.
- Fischler, I., Bloom, P.A., Childers, D.G., Arroyo, A.A., & Perry, N.W. (1984). Brain potentials during sentence verification: Late negativity and long-term memory strength. *Neuropsychologia*, 22, 559–568.
- Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E., & Perry, N.W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20, 400–409.
- Fischler, I., Childers, D.G., Achariyapaopan, T., & Perry, N.W. (1985). Brain potentials during sentence verification - automatic aspects of comprehension. *Biological Psychology*, 21, 83–105.
- Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26, 67–86.
- Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39, 225–242.
- Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, 2, 336–340.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46, 107–119.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59, 601–613.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65, 221-233.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116, 119–136.
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79, 701–721.

- Grice, P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Seminar.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.
- Hagoort, P. & van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society Series B: Biological Sciences*, 362, 801-811.
- Hannon, B., & Daneman, M. (1998). Facilitating knowledge-based inferences in less-skilled readers. *Contemporary Educational Psychology*, 23, 149-172.
- Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-differences perspective. *Memory & Cognition*, 29, 449-461.
- Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: on the possibility of suspending belief. *Psychological Science*, 16, 566–571.
- Haviland, S. E., & Clark, H. H. (1974). What's new ? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512–521.
- Herbert, C., & Kübler, A. (2011). Dogs cannot bark: Event-related brain responses to true and false negated statements as indicators of higher-order conscious processing. *PloS one*, 6, e25574.
- Hinze, S.R., Slaten, D.G., Horton, W.S., Jenkins, R., & Rapp, D.N. (submitted). Pilgrims Sailing the Titanic: Plausibility effects on memory for facts and errors.
- Isberner, M.-B., & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, 142, 15-22.
- Isberner, M.-B., & Richter, T. (in press). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes*.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press: Oxford, England.
- Johnson, H.M., & Seifert, C.M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1420-1436.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inferences, and consciousness*. Cambridge: Harvard University Press.

- Johnson-Laird, P. N., Herrmann, D. J., & Chaffin, R. (1984). Only connections: A critique of semantic networks. *Psychological Bulletin*, 96, 292-315.
- Just, M. A., Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating During Reading Comprehension: Why Causality Matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 854-865.
- Kintsch, W. (1980). Semantic memory: A tutorial. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 595-620). Hillsdale, NJ: Erlbaum.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kounios, J., Osman, A. M., & Meyer, D. E. (1987). Structure and process in semantic memory: New evidence based on speed-accuracy decomposition. *Journal of Experimental Psychology: General*, 116, 3-25.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Lemaire, P., & Fayol, M. (1995). When plausibility judgments supersede fact retrieval: The example of the odd-even effect on product verification. *Memory & Cognition*, 23, 34-48.
- Lombardi, D. A. (2012). *Students' conceptions about climate change: Using critical evaluation to influence plausibility reappraisals and knowledge reconstruction*. Unpublished dissertation, University of Nevada, Las Vegas.
- Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1456-1470.
- Maier, J. & Richter, T. (2013). Text-belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, 31, 151-175.
- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, 34, 1140-1149.
- Marsh, E. J., Meade, M. L., & Roediger III, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, 49, 519-536.

- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 913-934.
- McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 919-932.
- McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, 49, 25-42.
- Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, 27, 1-22.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19, 1213-1218.
- Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098-1111.
- O'Brien, E. J. (1995). Automatic components of discourse comprehension. In R. E Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 159-176). Hillsdale, NJ: Erlbaum.
- O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 777-784.
- O'Brien, E.J., & Myers, J.L. (1999). Text comprehension: A view from the bottom up. In S.R. Goldman, A.C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 35-53). Mahwah, NJ: Erlbaum.
- O'Brien, E., Rizzella, M. L., Albrecht, J. E. & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1200-1210.
- Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, 3, 229-235.
- Patson, N.D. & Warren, T. (2010). Eye movements when reading implausible sentences: Investigating potential structural influences on semantic integration. *Quarterly Journal of Experimental Psychology*, 63, 1516-1532.

- Peracchi, K. A., & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory & Cognition*, 32, 1044-1052.
- Petty, R.E., & Cacioppo, J.T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.
- Petty, R.E., & Wegener, D.T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 41-72). New York: Guilford.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 940-961.
- Pylkkänen, L., Oliveri, B., & Smart, A. (2009). Semantics vs. World Knowledge in Prefrontal Cortex. *Language and Cognitive Processes*, 24, 1313-1334.
- Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition*, 36, 688-701.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1290–1301.
- Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, 89, 250-280.
- Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 72-81.
- Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen*. Lengerich: Pabst.
- Richter, T. (2011). Cognitive flexibility and epistemic validation in learning from multiple texts. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility* (pp. 125-140). Berlin: Springer.
- Richter, T., Schroeder, S., Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96, 538-558.

- Ross, L., Lepper, M.R. & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32, 880-892.
- Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, 17, 188-206.
- Sanford, A. J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes*, 39, 205–224.
- Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes*, 42, 99-108.
- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23, 514-523.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in cognitive sciences*, 6, 382-386.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59, 237-259.
- Singer, M. (1993). Causal bridging inferences: Validating consistent and inconsistent sequences. *Canadian Journal of Experimental Psychology*, 47, 340-359.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574-591.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences. *Journal of Memory and Language*, 31, 507–524.
- Sparks, J.R., & Rapp, D.N. (2011). Readers' reliance on source credibility in the service of inference generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 230-247.
- Speer, S. R., & Clifton, C., Jr. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition*, 26, 965-978.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25, 359-393.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford, England: Blackwell.

- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1162-1169.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Thorndyke, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 15, 437-446.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory & Language*, 35, 454-475.
- van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443-467.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M., (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657-671.
- van Berkum, J. J. A., Holleman, B., Nieuwland, M. S., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, 20, 1092 -1099.
- van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17, 701-718.
- van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, 22, 335-351.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45, 225-258.
- Warren, T. (2011). The influence of plausibility and anomaly on eye movements in reading. In S. P. Livsedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook on Eye Movements* (pp. 911-923). USA, Oxford University Press.

- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eyemovements in reading. *Psychonomic Bulletin & Review*, 14, 770–775.
- Wason, P. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7–11.
- Wason, P. & Reich, S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, 31, 591-597.
- West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 385-399.
- Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, 8(6), 647–653.
- Zwaan, R. A., Langston, M. C. & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, 292-297.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Figure Captions

Figure 1. Rationale of our epistemic Stroop paradigm.

Figure 2. Mean correct response latency as a function of required response (positive, negative) a) when participants were required to judge the orthographical correctness of the last word of valid vs. invalid sentences (adapted from “You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information”, by T. Richter, S. Schroeder, and B. Wöhrmann, 2009, *Journal of Personality and Social Psychology*, 96, pp. 538-558. Copyright 2009 by the American Psychological Association. Adapted with permission) and b) when participants were required to indicate whether the last word of plausible vs. implausible sentences had changed color (reprinted from *Acta Psychologica*, 142, M.-B. Isberner and T. Richter, “Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension”, pp. 15-22. Copyright 2013, with permission from Elsevier). Error bars correspond to ± 1 standard error of the mean.

Figure 1

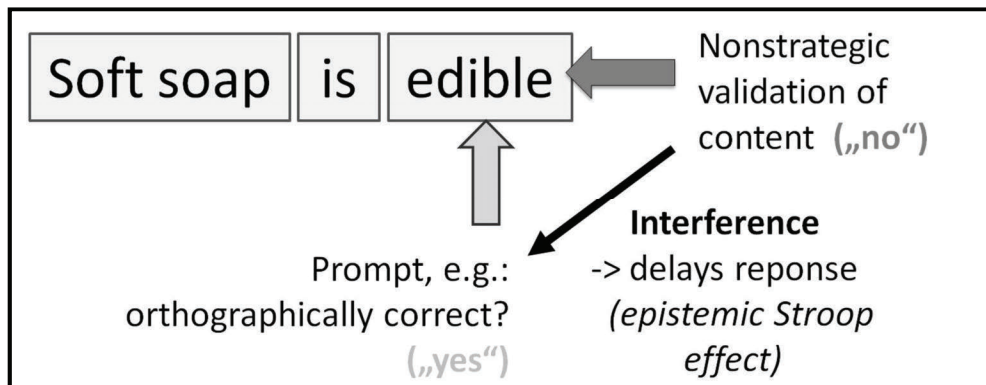
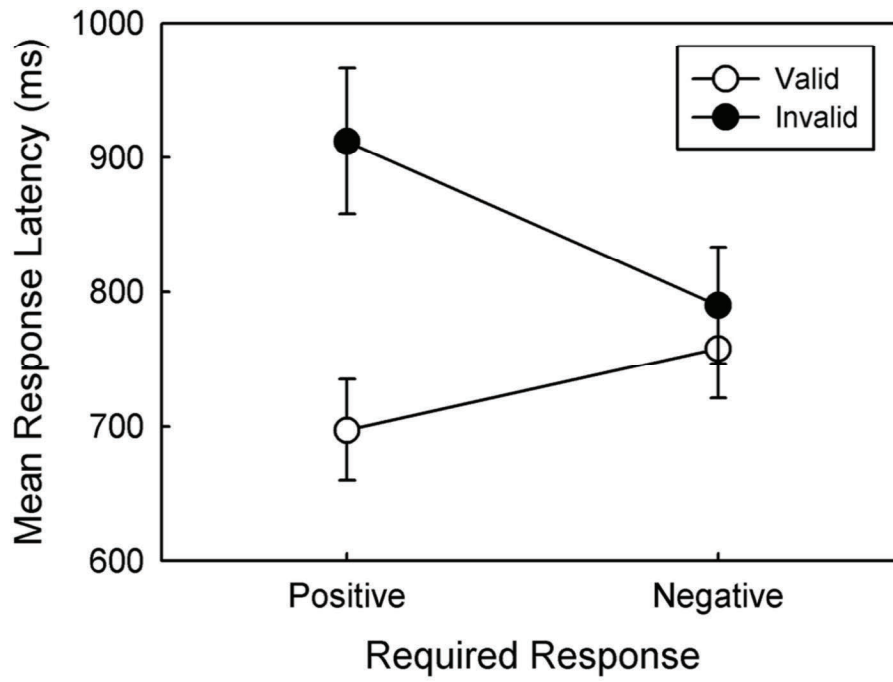
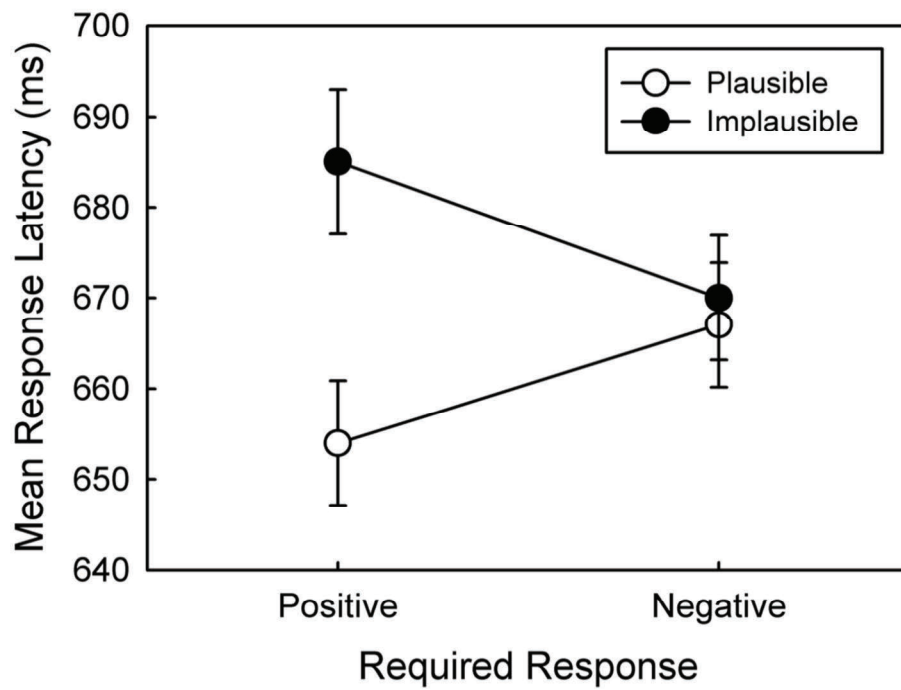


Figure 2

a)



b)



Chapter 3

Study 1

Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension

A version of this chapter is published as:

Isberner, M.-B., & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, 142, 15-22.

Abstract

We present evidence for a nonstrategic monitoring of event-based plausibility during language comprehension by showing that readers cannot ignore the implausibility of information even if it is detrimental to the task at hand. In two experiments using a Stroop-like paradigm, participants were required to provide positive and negative responses independent of plausibility in an orthographical task (Experiment 1) or a nonlinguistic color judgment task (Experiment 2) to target words that were either plausible or implausible in their context. We expected a nonstrategic assessment of plausibility to interfere with positive responses to implausible words. ANOVAs and linear mixed models analyses of the response latencies revealed a significant interaction of plausibility and required response that supported this prediction in both experiments, despite the use of two very different tasks. Moreover, it could be shown that the effect was not driven by the differential predictability of plausible and implausible words. These results suggest that plausibility monitoring is an inherent component of information processing.

Keywords: language comprehension – plausibility monitoring – event knowledge –predictability – verification – validation – context

Introduction

Whether world or event knowledge is immediately accessed during language comprehension is still a point of contention. While some studies report immediate effects of such knowledge on various measures of reading comprehension including reading times, eye tracking measures, and event related potentials (ERPs) (e.g., Hagoort et al., 2004; McRae et al., 1998; Matsuki et al., 2011; Rapp, 2008; van Berkum et al., 2005), other studies come to the conclusion that its influence in language comprehension is delayed in comparison to semantic knowledge. For example, Rayner et al. (2004) and Warren and McConnell (2007) found early effects of semantic violations on eye movements, but not of implausibility, suggesting that semantic knowledge is temporally privileged in language comprehension, whereas the access of world knowledge is slightly delayed.

A recent study by Matsuki et al. (2011) attempted to reconcile these seemingly contradictory findings by scrutinizing the typicality of the events described in the stimuli used in different studies. Their hypothesis was that typicality might be the key to explain the differences in the obtained results: In order to obtain early plausibility effects in reading times and eye tracking measures, they proposed that it is crucial that the plausible stimuli describe situations which are typical of people's world experience. The authors ensured typicality of their own stimuli by using production norms in addition to rating norms. Since the focus of their study was on instrument-action combinations, they asked their participants to "List the things or people that have the following actions done to them with the specified instruments" (p. 916). Based on the responses, they created minimal pairs of stimuli which reflected typical and atypical (but not anomalous) real world events, such as *Donna used the hose/shampoo to wash her filthy car/hair* (typical) and *Donna used the shampoo/hose to wash her filthy car/hair* (atypical). With these stimuli, in contrast to Rayner et al. (2004) and Warren and McConnell (2007), they found rapid effects of event-based plausibility (or typicality) in both self-paced reading and eye tracking, suggesting that there is in fact no delay in the access of event knowledge when this knowledge is typical of the readers' experience.

Similarly, Staub et al. (2007) report immediate effects of plausibility in an ingenious study that used sentences which were always globally plausible, but contained noun-noun compounds (e.g., *cafeteria manager*) whose modifier was either plausible or implausible in its context if it was initially analysed as a head noun (e.g., *The new principal visited / talked to the cafeteria manager*).

Plausibility had very rapid effects on eye movements, much faster than effects usually found in ERP studies, with implausibility resulting in an increase in reading time on the initially implausible word. Moreover, the size of the reading time penalty correlated with offline ratings of the implausibility of the word in the context leading up to it. It is important to note that these effects obtained although readers were merely asked to read for comprehension and although all sentences were globally plausible. Staub et al. (2007) interpreted their findings as evidence that the rapid effects of plausibility were not due to strategic factors.

These results are fascinating for two reasons: First, because they suggest that plausibility is monitored in the absence of an explicit evaluative processing goal. Second, this monitoring seems to follow the same principles as intentional plausibility ratings, suggesting that the plausibility assessment that can be computed in an intentional decision process is in fact – in some form – immediately available as a word is comprehended in its context. These findings are in stark contrast with two-step models of sentence verification which assume that any kind of evaluation is delayed with regard to comprehension, or in other words, withheld until the comprehension process has terminated (e.g., Gilbert et al., 1993). Rather, they suggest that language comprehension comprises a routine, online plausibility monitoring process that operates nonstrategically and fast on a word-by-word basis as the linguistic input unfolds.

Interestingly, Staub et al. (2007) do not draw a distinction between implausible and semantically anomalous sentences, which is in line with Matsuki et al.'s (2011) conclusion that this kind of distinction may in fact be arbitrary (see also Jackendoff, 2002, and Hagoort et al., 2004). However, if one inspects their stimuli, at least some of the local implausibilities are due to animacy violations (including the aforementioned example), which are generally considered semantic violations. It thus remains an open question whether the same nonstrategic process underlies the plausibility effects that were obtained by Matsuki et al. (2011).

In the present study, our goal is to investigate this question by testing whether event-based plausibility is routinely monitored during language comprehension. We do this by testing the nonstrategic nature of the proposed routine plausibility monitoring process with a Stroop-like paradigm (Stroop, 1935) in which an assessment of plausibility is irrelevant or even detrimental to task performance. Specifically, we test the potential interference of plausibility monitoring with an unrelated task that requires responses which are orthogonal to plausibility. Beyond the question of the time course of plausibility effects, we thereby attempt to elucidate what actually happens in the

reader when he or she encounters implausible information (besides taking longer to process it than plausible information).

What kind of interference is to be expected from a routine, nonstrategic monitoring of event-based plausibility? We assume that, if readers indeed routinely assess plausibility, they will react to information that is inconsistent with their event knowledge with a negative response tendency. This negative response tendency, in turn, should make it more difficult to provide positive responses of any kind, even if the responses are completely unrelated to plausibility. To test this hypothesis, we make use of the so-called epistemic Stroop paradigm, an adaptation of the Stroop paradigm introduced by Richter et al. (2009) for testing the interference of factual knowledge with an unrelated judgment task. In their study, participants were asked to judge the orthographical correctness of words embedded in assertions that were presented word by word on a computer screen and were either valid or invalid with regard to common factual knowledge. In experimental items, the word that had to be judged was the last word of the assertion, and it was spelled either correctly or incorrectly (e.g., *Perfume contains scents / sents* or *Soft soap is edible / eddible*; the original sentences were in German: *Parfüm enthält Duftstoffe/duftstoffe* and *Schmierseife ist essbar/essbahr*). Although the validity of the assertions was irrelevant to the orthographical task, responses were delayed when the word to be judged was presented at the end of an invalid assertion but required a positive (“correct”) response. This resulted in a significant interaction of validity and orthographical correctness.

If our assumption of a nonstrategic plausibility monitoring process holds, we should find a similar effect for stimuli which tap into readers’ event knowledge. Specifically, we expect to find slower latencies for positive (i.e., affirmative) responses in the unrelated task when a word (for example, the word *plumber*) is implausible in its context (*Frank has a broken leg. He calls the plumber.*) compared to when it is plausible (*Frank has a broken pipe. He calls the plumber.*). We therefore expect a significant interaction of plausibility and required response which conforms to this pattern.

In Experiment 1, we test this hypothesis with the same orthographical judgment task used by Richter et al. (2009). However, if it is true that the interference of plausibility monitoring hinges on the positive/negative character of the response rather than on other task characteristics, it should obtain in any kind of task that requires positive and negative responses and is independent of plausibility. In order to test this hypothesis, we go one step further in Experiment 2 and investigate the interference of plausibility monitoring with a completely different, nonlinguistic

task which is even more obviously independent of plausibility than the orthographical task: The task of judging whether or not a word that is plausible or implausible in its context has changed color.

As discussed by Matsuki et al. (2011), a variable that is often confounded with plausibility is predictability. Although, as the authors point out, these two dimensions are practically extremely difficult to disentangle, we nonetheless attempt to do this by varying the predictability of the target word in the plausible context while keeping plausibility constant. According to Matsuki et al. (2011), “One way to differentiate the two would be to contrast implausible items with plausible ones for which cloze values of all targets is zero” (p. 926). However, since it is, as the authors state, “virtually impossible” (p. 926) to construct plausible targets with a cloze value of zero (particularly in minimal pairs of stimuli that differ only regarding the target word), since even atypical or implausible targets usually have cloze values higher than that, our goal was to approximate zero as much as possible without creating unnatural stimuli.

Moreover, in order to keep the plausible and implausible conditions strictly parallel, we designed our stimulus material in such a way that the same target sentences could be used in both conditions. This was achieved by varying the plausibility of each target sentence by means of a context sentence, which rendered the same target sentence either plausible or implausible. In this way, and in extension of the aforementioned studies, our experiments also allowed testing whether the extrasentential linguistic context routinely becomes part of the background against which incoming information is monitored for plausibility.

Experiment 1

The primary goal of Experiment 1 was to investigate whether event-based plausibility is nonstrategically monitored by testing its interference with an orthographical task unrelated to plausibility, using a Stroop-like paradigm introduced by Richter et al. (2009). We assume that if this is the case, information that is implausible with regard to a comprehender’s event knowledge should elicit a negative response tendency. The negative response tendency, in turn, should interfere with positive responses in the unrelated task. Thus, we expect participants to take longer to indicate that a word is spelled correctly when it is implausible in its context than when it is plausible.

Method

Participants. Participants were 70 psychology undergraduates at the University of Cologne (52 women and 18 men). All participants were native speakers of German. Their average age was 24.2 years ($SD = 4.8$).

Stimulus material. Stimuli were pairs of context and target sentences describing situations that were either plausible or implausible with regard to common event knowledge. For each of the experimental items, four different versions were constructed. First, there were two versions of each context sentence. One version rendered the last word of the target sentence plausible and the other one rendered it implausible (e.g., *Frank has a broken pipe / leg. He calls the plumber.*). Second, there were two versions of each target sentence. One version ended with a word that was assumed to have a high predictability in the plausible context, and the other ended with a word that was equally plausible but had a low predictability in the plausible context (e.g., *Frank has a broken pipe. He calls the plumber / tradesman.*). Of each of the four versions of each item, an orthographically incorrect version was constructed by inserting, exchanging, or removing one letter or changing the case of the last word of the target sentence, while maintaining the phonology of the correct word (such as *shampoo* instead of *shampoo*; the actual stimuli were in German, e.g., *Shampoo* instead of *Shampoo*). In addition to the experimental items, 160 filler items were constructed. These were also pairs of context and target sentences, of which 80 described plausible situations and 80 described implausible situations. Of the plausible as well as the implausible filler items, half contained a word with a spelling mistake. This word served as the target word for the orthographical task. The procedure for inserting the spelling mistakes was based on the same principles as in the experimental items. The position of the misspelled word within each filler item was selected randomly, excluding the first word of the context sentence and the last word of the target sentence. Following the same principle, one word was selected as the target word in each of the 80 remaining filler items but maintained in its orthographically correct form.

Norming study. A norming study was conducted to select experimental items with both an effective plausibility and an effective predictability manipulation out of a pool of 97 items. The participants of the norming study (14 psychology undergraduates not identical to the experimental sample) completed a questionnaire with two tasks. First, there was a cloze test to assess the predictability of the final word in each item. Participants were asked to read each item and spontaneously fill in the last word of the target sentence, which had been substituted by a blank. Second, they were asked to rate the plausibility of each of the four (orthographically correct)

sentence pairs that resulted from pairing both versions of the context sentence with both versions of the target sentence ($4 \times 97 = 388$ sentence pairs). Participants were asked to indicate for each sentence pair whether they found it plausible (“yes”) or implausible (“no”). The sentence pairs were presented in the same order to all participants but mixed randomly within the questionnaire. Based on these data, 64 out of the 97 items were selected in which both the plausibility manipulation as well as the predictability manipulation proved to be effective. These were items in which the mean agreement with the assumed plausibility was high for all versions of the item and in which the cloze values were high only for the predictable word in the plausible condition and low in all other conditions. The norms for the selected items are displayed in Table 1.

Procedure. All items were presented word by word on a computer screen using Rapid Serial Visual Presentation (RSVP) with a fixed rate of 600 ms per word. Each word was presented in bold black letters in the font type Arial (approximate height 1 cm) in a white 13 x 6 cm square placed in the middle of the screen against a silver background. The viewing distance was approximately 60 cm. Each trial was preceded by a fixation cross presented for 250 ms and followed by a blank screen presented for 500 ms. At one word per trial (the target word), the presentation stopped and participants were prompted by the question *Spelling?*, which appeared above the target word 300 ms after the onset of the target word, to indicate whether or not the word was spelled correctly. The prompt and the target word remained on the screen until the participant provided a response. Participants were instructed to provide their responses as fast and as accurately as possible by pressing ‘k’ for correct spelling and ‘d’ for incorrect spelling, and to keep their fingers on the two response keys throughout the whole experiment. As a reminder for which of the two keys to press for which response, the prompt was accompanied by a label *correct* in green font inside a white box with a green frame on the right hand side, and a label *incorrect* in red font inside a white box with a red frame on the left hand side. On half of the trials, the target word was spelled correctly, requiring a “correct”-response, and on the other half of the trials, the presented word was spelled incorrectly, requiring an “incorrect”-response. In experimental trials, the target word was always the final word of the item. In filler trials, the target word was at a randomly selected position within the item (see 2.1.2 *Stimulus Material*). The purpose of the filler items was to ensure that participants would not be able to guess at which word of the item they would be asked to provide a response. To encourage correct responses, participants received a feedback on the accuracy of each of their responses, which was presented for 600 ms after each response. The trial either ended with a blank screen (experimental items) or continued with the

next word of the item (filler items). The first six items presented to each participant were practice items that were not included in the analysis.

Design. The design was a 2(*plausibility*: plausible vs. implausible) X 2(*predictability*: predictable vs. unpredictable) X 2(*required response*: positive vs. negative) within-subjects design. Dependent variables were the response latency and the accuracy of the responses. Assignments of experimental items to experimental conditions were counterbalanced across participants by eight item lists. Each participant saw eight experimental items in each of the eight experimental conditions. Experimental and filler items were presented in random order.

Results and Discussion

Type-I error probability was set at .05 for all hypothesis tests. Under the assumption of a medium effect size ($f = .25$ according to Cohen, 1988) and medium correlations ($\rho = .5$) between the levels of the independent variables in the population, the design and sample size of Experiment 1 yielded a power ($1-\beta$) of .98 for detecting the focal interaction of plausibility and required response in the ANOVA based on subjects as the units of analysis (power computed with the software G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007). We conducted ANOVAS for repeated measurements with both participants (F_1 , by-subjects) and items (F_2 , by-items) as the source of random variance. The reported means and standard errors are based on subjects as the units of analysis. Standard errors of the mean were computed for within-subjects designs (Morey, 2008).

In addition to the ANOVA analyses, we conducted a linear mixed models (LMM) analysis for the response latencies and a generalized linear mixed models (GLMM) analysis with logit link for the error rates with subjects and items included as random factors, i.e. the means of subjects as well as items were allowed to vary randomly. This type of analysis accounts for the fact that both subjects and items represent samples of larger populations. Unlike the F_1 - and F_2 -ANOVA, the LMM/GLMM analysis with crossed random effects for subjects and items does not decrease power but allows for an adequate and stringent test of the hypothesized effects of the independent variables in one single model (for further discussion, see Baayen et al., 2008). We included all three independent variables as contrast-coded predictors with fixed effects in the model (plausibility: 1 = plausible, -1 = implausible; predictability: 1 = predictable, -1 = unpredictable; required response: 1 = positive, -1 = negative). In addition, the presentation position of each item was included in the model as centered predictor (fixed effect) to control for position effects. The

LMM/GLMM analysis was conducted with the *lmer* command of the *lme4* package for R (Bates et al., 2011). For the sake of conciseness, only significance tests associated with the fixed effects (main and interaction effects) of the independent variables are reported as these are directly relevant for our hypotheses (data files and R-scripts for both experiments are available from the authors upon request). Please note that no degrees of freedom are reported for the t-values of the LMM analysis because it is still unclear how these should be derived. However, given the large number of observations in the present experiments (items times participants), it is safe to assume that the distribution of t-values approximates the standard normal distribution (z-distribution; see Bayen et al., 2008, Note 1). Thus, the standard normal distribution was assumed for significance tests of fixed effects in the LMM analysis.

Response latencies. Response latencies were calculated for correct responses (93% of the responses in experimental trials). Response latencies deviating more than three standard deviations from either the subject or item mean (1.8% of all correct latencies) were treated as outliers and removed from the data set. Figure 1 shows the mean correct response latencies as a function of plausibility and required response; Table 2 displays the means and standard deviations associated with the by-subjects analysis. We found significant main effects for all of the three independent variables. Plausible target words ($M = 962$ ms, $SE = 6$ ms) elicited faster responses than implausible target words ($M = 1035$ ms, $SE = 6$ ms), $F_1(1, 69) = 48.30, p < .001, \eta_p^2 = .41, F_2(1, 63) = 24.41, p < .001, \eta_p^2 = .28$ (LMM analysis: $t = -6.20, p < .05$). Predictable words ($M = 947$, $SE = 8$) elicited faster responses than non-predictable words ($M = 1050$ ms, $SE = 8$ ms), $F_1(1, 69) = 53.58, p < .001, \eta_p^2 = .44, F_2(1, 63) = 21.26, p < .001, \eta_p^2 = .25$ (LMM analysis: $t = -8.88, p < .05$). Furthermore, negative responses to incorrectly spelled words ($M = 951$ ms, $SE = 11$ ms) were faster than positive responses to correctly spelled words ($M = 1046$ ms, $SE = 11$ ms), $F_1(1, 69) = 22.50, p < .001, \eta_p^2 = .25, F_2(1, 63) = 12.80, p < .01, \eta_p^2 = .17$, (LMM analysis: $t = 7.75, p < .05$).

However, the main effects of plausibility and required response were qualified by a significant interaction of the two variables, $F_1(1, 69) = 6.77, p < .05, \eta_p^2 = .09, F_2(1, 63) = 5.69, p < .05, \eta_p^2 = .08$ (LMM analysis: $t = -2.97, p < .05$). Planned contrasts revealed that the pattern underlying the interaction conformed to the hypothesized Stroop-like effect. Positive responses in the orthographical task were significantly slower for implausible ($M = 1103$ ms, $SE = 18$ ms) compared to plausible words ($M = 990$ ms, $SE = 14$ ms), $F_1(1, 69) = 31.62, p < .001, \eta_p^2 = .31, F_2(1, 63) = 30.61, p < .001, \eta_p^2 = .33$. Negative responses were also slower for implausible ($M = 968$ ms, $SE = 16$ ms) compared to plausible words ($M = 934$ ms, $SE = 14$ ms), but with $F_1(1, 69) =$

4.13, $p < .05$, $\eta_p^2 = .06$, $F_2(1, 63) = 3.51$, $p = .07$, $\eta_p^2 = .05$, this difference was much smaller than for positive responses and non-significant in the by-items analysis. Furthermore, there was no three-way interaction with predictability, $F_1(1, 69) = 2.35$, $p = .13$, $F_2(1, 63) < 1$, $p = .47$ (LMM analysis: $t = 1.82$, $p > .05$).

Error rates. The error rates were low overall ($M = .07$, $SD = .11$). There was a significant main effect of required response, $F_1(1, 69) = 32.92$, $p < .001$, $\eta_p^2 = .32$, $F_2(1, 63) = 19.14$, $p < .001$, $\eta_p^2 = .23$ (ANOVAs performed on arc-sine transformed proportions; GLMM analysis: $z = 7.42$, $p < .001$). More errors were made in the judgment of orthographically incorrect words, that is, when the required response was negative ($M = .097$, $SE = .006$) compared to orthographically correct words, that is, when the required response was positive ($M = .043$, $SE = .006$). Furthermore, there was a significant main effect of predictability in the by-subjects ANOVA, $F_1(1, 69) = 6.93$, $p = .01$, $\eta_p^2 = .09$, $F_2(1, 63) = 3.29$, $p = .08$, $\eta_p^2 = .05$ (GLMM analysis: $z = 4.03$, $p < .001$). More errors were made in the judgment of non-predictable words ($M = .082$, $SE = .004$) compared to predictable words ($M = .058$, $SE = .004$). In contrast to the results for the response latencies, there was no interaction effect of plausibility and required response, $F_1(1, 69) < 1$, $p = .88$, $F_2(1, 63) < 1$, $p = .35$ (GLMM analysis: $z = 0.28$, $p = .78$). Thus, there was no indication of a speed-accuracy trade-off in our data.

Discussion. The delay of positive responses to words that are implausible in their context supports the hypothesis that event-based plausibility is routinely monitored during language comprehension and results in the detection and rejection of implausible information. However, the fact that both positive and negative responses were faster when the target word was plausible compared to when it was implausible prevents a fully conclusive interpretation of this effect. This pattern indicates that the orthographical task was easier for plausible than for implausible words, which may be due to the fact that words are generally easier to recognize when they are congruent with a context than when they are incongruent (e.g., Stanovich & West, 1981, 1983). Thus, plausible words may have been easier to recognize and check for orthographical correctness. However, this makes the orthographical task somewhat suboptimal for investigating effects of nonstrategic plausibility monitoring because it might attenuate the expected difference between the effects of plausibility on positive and negative responses. We ran Experiment 2 to clarify this issue.

It may also seem unusual that there was no advantage for affirmative responses in our task, which is often found in other types of tasks, such as lexical decision. In fact, negative responses in

our task were significantly faster than positive responses. This main effect of required response in the orthographical task was also found by Richter et al. (2009, Experiment 3). It may be attributable to the fact that the misspelled words were phonologically and orthographically very similar to the original words so that they remained easy to recognize, while the orthographical errors were blatant enough to be easy to spot for native speakers with a regular school education. This interpretation is supported by the high accuracy rate despite the speeded response conditions (93%). It is also important to note that our task was quite different from lexical decision with regard to both the stimuli and the instructions. Most importantly, there were no nonwords in our task, unless one would like to define the misspelled words as nonwords. Even so, the instruction for the orthographical task would have led participants to perceive them as real but misspelled words rather than as meaningless nonwords (such as those that are usually used in lexical decision). Thus, the processing induced by our task instruction and stimuli should have been rather different from the processing required by a lexical decision task.

Importantly, the main effect of required response does not limit the interpretation of the results because the critical comparisons were those between the two plausibility conditions for the same response type, rather than between positive and negative responses. However, because we were not interested in effects other than those produced by the assumed plausibility monitoring process, a task in which there is no general advantage for one response or the other would be preferable, and we tried to achieve this in Experiment 2.

Experiment 2

Experiment 2 was designed to eliminate potential problems of Experiment 1 by using a different kind of task. The fact that in Experiment 1, negative responses were also slower when the word was implausible in its context indicates that plausibility might have been confounded with task difficulty. Therefore, we will use a different task in Experiment 2 whose difficulty should be unaffected by plausibility. Moreover, for the purpose of testing the generalizability of the interference effect, it is advantageous to use a task which strongly differs from the orthographical task. Therefore, we chose the nonlinguistic task of judging whether or not the target word changes color.

Finally, despite the fact that the orthographical task did not require any semantic (let alone plausibility) judgment, the presentation rate of one word per 600 ms used in Experiment 1 might have provided participants with sufficient time to engage in some kind of strategic evaluation of

the message prior to seeing the target word. For this reason, Experiment 2 used a presentation rate of one word per 300 ms which roughly corresponds to the average fixation duration during reading (Rayner, 1998). Thus, the presentation rate in Experiment 2 was sufficiently short to minimize any strategic processing during sentence reading besides the focal color judgment task.

Method

Participants. Participants were 67 undergraduates (native speakers of German) at the University of Kassel. The average age of the participants (44 women and 23 men) was 24.2 years ($SD = 5.7$).

Stimulus material. The orthographically correct versions of the experimental and filler items of Experiment 1 were used. The target words were the same as in Experiment 1 (i.e., the final word in experimental items and a randomly selected word in filler items) but they now either changed color or remained black when the response prompt appeared.

Procedure. The procedure was identical to the procedure used in Experiment 1, except for the following differences: First, the presentation time for each word in the RSVP and for the feedback was reduced to 300 ms. Second, 300 ms after the target word appeared, instead of the orthographical judgment participants were now prompted to indicate whether or not the word had changed color as the prompt appeared (50% of the trials required a yes response; in the other half of the trials, the word remained black). In the color change trials, colors were chosen randomly from a list of 9 colors which had been approved for readability on a white background.

Design. Design and dependent variables were the same as in Experiment 1.

Results and Discussion

Type-I error probability was set at .05 for all hypothesis tests. The design and sample size of Experiment 2 yielded a power of .98 for detecting the focal interaction of plausibility and required response (with $f = .25$ and $\rho = .5$) in a by-subjects ANOVA. Due to a programming error, the presentation of one of the 64 experimental items was faulty in one of the eight conditions. For this reason, this item was discarded from all further analyses. As in Experiment 1, ANOVAS were conducted for repeated measurements with both participants (F_1 , by-subjects) and items (F_2 , by-items) as the source of random variance. The reported means and standard errors were computed with subjects as the units of observation. Standard errors of the mean were computed for within-subjects designs (Morey, 2008). In addition, the fixed effects from an LMM/GLMM analysis with crossed random effects of subjects and items (Baayen et al., 2008) are reported.

Response latencies. Response latencies were calculated for correct responses (96.8% of the responses in experimental trials). Latencies deviating more than three standard deviations from either the subject or item mean (2.1% of all correct latencies) were removed from the data set. Figure 2 shows the mean correct response latencies as a function of plausibility and required response; Table 3 displays the means and standard deviations based on subjects as the units of observation. As in Experiment 1, there was a main effect of plausibility which was significant in the by-subjects and the LMM analysis. Plausible target words ($M = 661$ ms, $SE = 4$ ms) were responded to faster than implausible target words ($M = 677$ ms, $SE = 4$ ms), $F_1(1, 66) = 6.09$, $p < .05$, $\eta_p^2 = .08$, $F_2(1, 62) = 3.59$, $p = .06$, $\eta_p^2 = .06$ (LMM analysis: $t = -2.25$, $p < .05$).

Moreover, the analysis revealed an interaction of plausibility and required response which was significant in the by-subjects analysis, $F_1(1, 66) = 5.18$, $p < .05$, $\eta_p^2 = .07$, but missed significance using items as a random source of variance, $F_2(1, 62) = 2.66$, $p = .11$, $\eta_p^2 = .04$. Most importantly, however, the interaction of plausibility and required response was significant in the LMM analysis which includes subjects as well as items as sources of random variance ($t = -2.27$, $p < .05$). In order to interpret the interaction, we conducted planned contrasts which revealed that the pattern underlying the interaction was similar to the pattern found in Experiment 1. As before, positive responses were slower for implausible ($M = 685$ ms, $SE = 8$ ms) compared to plausible words ($M = 654$ ms, $SE = 7$ ms), $F_1(1, 66) = 9.45$, $p < .01$, $\eta_p^2 = .13$, $F_2(1, 62) = 5.32$, $p < .05$, $\eta_p^2 = .08$. Crucially, and in contrast to Experiment 1, the latencies of negative responses to plausible ($M = 667$ ms, $SE = 7$ ms) and implausible target words ($M = 670$ ms, $SE = 7$ ms) did not differ significantly from each other, $F_1(1, 66) < 1$, $p = .71$, $F_2(1, 62) < 1$, $p = .90$. Moreover, there was again no three-way interaction with predictability, $F_1(1, 66) < 1$, $p = .56$, $F_2(1, 62) < 1$, $p = .56$ (LMM analysis: $t = -0.51$, $p > .05$).

Error rates. Again, the error rates were low overall ($M = .03$, $SD = .07$) and showed no indication of a speed-accuracy trade-off: The interaction of plausibility and required response was not significant $F_1(1, 66) < 1$, $p = .98$, $F_2(1, 62) < 1$, $p = .87$ (ANOVAs performed on arc-sine transformed proportions; GLMM analysis: $z = 0.18$, $p = .86$). All other effects were also non-significant, with all p -values exceeding .10, except for the interaction of plausibility and predictability, $F_1(1, 66) = 9.18$, $p < .01$, $\eta_p^2 = .12$, $F_2(1, 62) = 10.10$, $p < .01$, $\eta_p^2 = .14$ (GLMM analysis: $z = 2.78$, $p < .01$). This interaction was due to more errors being made in response to non-predictable words in the plausible condition ($M = .047$, $SE = .007$) compared to non-predictable words in the implausible condition ($M = .025$, $SE = .005$), $F_1(1, 66) = 8.08$, $p < .01$, $\eta_p^2 = .11$, $F_2(1,$

62) = 6.70, $p < .05$, $\eta_p^2 = .10$, as well as compared to predictable words in the plausible condition ($M = .021$, $SE = .005$), $F_1(1, 66) = 9.49$, $p < .01$, $\eta_p^2 = .13$, $F_2(1, 62) = 6.67$, $p < .05$, $\eta_p^2 = .10$. As we had no hypotheses concerning this interaction, and it does not affect the interpretation of the response latency data, we simply point it out here without further interpretation.

Discussion. These results are an important extension of Experiment 1. First, the similarity of the patterns in the two experiments, despite the fact that the tasks were entirely different (i.e., a linguistic orthographical task vs. a nonlinguistic color judgment task), is striking. This confirms our assumption that the only task dimension which produces the pattern is the requirement of positive and negative responses independent of plausibility. Second, and most importantly, the pattern that emerged in Experiment 2 clearly indicates that the effect hinges on a delay of positive responses to implausible words, since the negative responses were unaffected by plausibility. Third, the effect occurred despite the fact that the presentation rate in Experiment 2 was much shorter than in Experiment 1, reducing the likelihood of strategic processing even further. Finally, there was no main effect of required response as in Experiment 1, which suggests that this effect was due to the specific demands of the orthographical task.

General Discussion

We assumed that the influence of event-based plausibility in comprehension, as found by Matsuki et al. (2011), reflects a routine plausibility monitoring process that is nonstrategic and inherent in language comprehension. In order to give this tacit process a “voice”, we tested the interference of its assumed negative outcome for implausible information with incongruent positive responses in an unrelated judgment task using a Stroop-like paradigm adapted from Richter et al. (2009). In Experiment 1, the task we used was an orthographical judgment task as in the original Richter et al. (2009) study. In Experiment 2, we used a nonlinguistic color judgment task and increased the presentation rate in order to rule out potential alternative explanations and test the generalizability of the results.

In line with our predictions, responses were delayed in both tasks when the task required a positive response to a target word that was implausible in its context, compared to when it was plausible, resulting in an interaction of plausibility and required response. However, in the orthographical task, negative responses were also slower for implausible compared to plausible words, suggesting a higher overall task difficulty for implausible words. This may have been due to the fact that words are generally easier to recognize when they are plausible in their context

(e.g., Stanovich & West, 1981, 1983), which makes the orthographical task somewhat suboptimal for investigating the effects of nonstrategic plausibility monitoring. For this reason, and to test the generalizability of our results, we chose a nonlinguistic color judgment task in Experiment 2. In spite of the entirely different nature of the task, the global pattern of results was strikingly similar. Although the interaction of plausibility and required response fell short of significance in the by-items analysis, it was significant in the by-subjects analysis as well as in a Linear Mixed Models analysis which takes both subjects and items into account as sources of random variation. Moreover, the critical planned contrasts produced the same results in the F1 and F2 analyses, with positive responses being slower for implausible compared to plausible words and – in contrast to Experiment 1 – negative responses being unaffected by plausibility. This confirms that we indeed managed to find a task whose difficulty does not vary with plausibility and thus in principle allows the interference effect to emerge even more clearly. In addition, this task also eliminated the response time advantage for negative responses, which seemed to be specific to the orthographical task. Overall, the two experiments provide strong evidence for routine, nonstrategic plausibility monitoring during language comprehension.

These findings are in line with both the Matsuki et al. (2011) findings that event-based plausibility is immediately accessed in language comprehension, as well as with the Staub et al. (2007) findings that plausibility effects on language comprehension are nonstrategic. In addition, our results bridge both findings by suggesting that, despite differences between the stimuli, the same nonstrategic process may be underlying the rapid plausibility effects obtained in both studies. Beyond questions of the time course of access to different kinds of knowledge, our results suggest that event knowledge and the assessment of plausibility based on this knowledge are routine and obligatory in language comprehension. An interesting extension of the Matsuki et al. (2011) and the Staub et al. (2007) results is that while in those studies, plausibility of the target word hinged on the intrasentential context, in our study it was manipulated by the extrasentential context (i.e., the preceding context sentence). The fact that the effect obtained nonetheless is in line with other findings that people immediately relate linguistic input to the widest available context (e.g., Hagoort & van Berkum, 2007; Just & Carpenter, 1980; Nieuwland & van Berkum, 2006; van Berkum et al., 1999; van Berkum et al., 2003).

Furthermore, we attempted to rule out the alternative explanation that the effect might be driven by the predictability rather than the plausibility of the target word by using target words that were similar in plausibility but with highly different cloze values. Naturally, the non-predictable

target words were still more predictable in the plausible than in the implausible condition; however, if the Stroop-like effect was driven by predictability (i.e., by a negative response tendency elicited by unexpected words) it would be expected to be much stronger for the predictable words. Alternatively, an interaction of predictability and required response analogous to the predicted interaction of plausibility and required response should emerge if predictability was indeed the crucial variable here. Contrary to this idea, neither of the experiments showed a modulation of the effect by predictability in terms of a three-way interaction or an interaction of predictability and required response. Hence, it seems unlikely that the effect obtained in our study is due to predictability differences between plausible and implausible items.

Despite the fact that the overall interaction of plausibility and required response and the corresponding planned comparisons are in line with our predictions, it must be noted that the interaction effect was slightly smaller in Experiment 2 than in Experiment 1 (which is evident in the by-items analysis). This pattern may point towards a disadvantage of the nonlinguistic color judgment task: it did not require comprehension of the stimuli and may thus have reduced semantic processing. It is important to note that while we argue that plausibility assessment is nonstrategic, we do not argue that it can occur without an adequate level of comprehension. Despite the proposed nonstrategic nature of plausibility monitoring, it is still reasonable to assume that more shallow semantic processing will reduce validation processes and hence their interference with other tasks. A way to avoid this problem and ensure deeper semantic processing while still using a nonlinguistic task would be to include questions which require comprehension but not plausibility assessment of the sentences. This would also open up the possibility of directly exploring the relationship between depth of semantic processing and nonstrategic plausibility assessment, which our results suggest to be a promising endeavor for future experiments.

One further issue worth noting is the asymmetry of the effects obtained for positive and negative responses. In our hypotheses, we predicted the interference of a negative response tendency evoked by implausible information with positive responses. We did not expect a converse interference of plausible information with negative responses because we assumed the monitoring process to respond negatively to implausible information (in terms of an error detection process) rather than positively to plausible information. Nonetheless, one might have expected facilitation for negative responses after implausible information, which is clearly not present in either of the experiments. A possible interpretation of this result is that it might point towards a special status of implausible information: It could be that readers react to implausible

information with reduced acceptance rather than with outright rejection because they cannot be certain whether the sentence – although implausible – is actually false. For example, it is implausible but not impossible that in the example event *Frank has a broken leg. He calls the plumber*, Frank did (for unknown but conceivable reasons) call the plumber after breaking his leg. Plausibility comes into play only when there is uncertainty (e.g., Friedman & Halpern, 2001) and this uncertainty may prevent a clear rejection of implausible information. Thus, it may be more difficult to affirm implausible information (compared to plausible information) but not necessarily easier to reject it.¹ If this is the case, then one might find a different pattern for stimuli that describe events which are impossible rather than merely implausible (a terminology which Warren & McConnell, 2007, use to discriminate between violations of semantic vs. world knowledge), in which sentences describing impossible events evoke a clear negative response tendency which also leads to facilitation for negative responses. For this purpose, it would be useful to include an adequate neutral condition in future experiments to determine precisely the extent to which interference and facilitation contribute to the observed pattern.

It is important to note here that the present study was not aimed at contributing to the debate on whether there is a distinction between semantic and world or event knowledge, but rather focused on the specific question of whether event knowledge is used nonstrategically to assess plausibility during on-line comprehension. However, as outlined above, our paradigm offers a novel tool that might be useful to elucidate processing differences between different types of knowledge violations in future research.

Conclusion

In conclusion, our results suggest that plausibility monitoring is a routine, nonstrategic process that is invariably interwoven with language comprehension. As such, our findings are in line with Singer's (2006) proposal that the verification of linguistic messages is not dependent on an evaluative processing goal but "rather emerges from the fundamentals of the cognition of reading" (p. 589). In this way, our study elucidates an aspect of plausibility effects that has so far received relatively little attention, namely the extent to which these effects are nonstrategic and may reflect more than simple "processing costs" of implausible information: Rather, they point towards a highly purposeful monitoring process that promotes the accuracy and stability of the

¹ We would like to thank an anonymous reviewer for suggesting this possibility.

mental representations which are constructed during language comprehension (Schroeder et al., 2008).

Role of the funding source

The research reported in this article was supported by the German Research Association (Deutsche Forschungsgemeinschaft, DFG, grants RI 1100/4-1 and RI 1100/4-2).

Acknowledgements

We would like to thank Marc Grosjean, Julia Knoepke, and Sascha Schroeder for helpful discussions, David Nitz for technical assistance, and Claudia Beck, Ricarda Moll, Sarah Rom, Miriam Schwarzenenthal, and Monika Wieland for their help in collecting data.

References

- Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Bates, D., Maechler, M., Bolker, & B. (2011). *Linear mixed-effects models using S4 classes* [Software]. R Foundation for Statistical Computing, Vienna, Austria.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum, Hillsdale, NJ.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Friedman, N. & Halpern, J. Y. (2001). Plausibility measures and default reasoning. *Journal of the ACM*, 48, 648-685.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65, 221-233.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.
- Hagoort, P., & van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society. Series B: Biological Sciences*, 362, 801-811.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, Oxford, England.
- Just, M. A., & Carpenter, P. A., (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 913-934.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283-312.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4, 61-64.

- Nieuwland, M. S. & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098-1111.
- Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition*, 36, 688-701.
- Rayner, K. (1998). Eye-movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1290-1301.
- Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96, 538-558.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59, 237-259.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574-591.
- Stanovich K. E., & West R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 658-672.
- Stanovich K. E., & West R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, 112, 1-36.
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1162-1169.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- van Berkum, J. J. A., Brown, C. M., Zwitterlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443-467.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657-671.

- van Berkum, J. J. A., Zwitterlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17, 701-718.
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eyemovements in reading. *Psychonomic Bulletin & Review*, 14, 770-775.

Table 1

Norms for Plausibility (Mean Proportion of “Plausible” Judgments in %) and Predictability (Mean Cloze Value in %) of the Selected Items

Condition	Plausibility <i>M (SD)</i>	Predictability <i>M (SD)</i>
<i>Plausible</i>		
Predictable	97.85 (4.60)	75.22 (20.16)
Non-predictable	96.02 (5.66)	5.91 (9.40)
<i>Implausible</i>		
Predictable	4.10 (6.73)	1.45 (4.06)
Non-predictable	4.39 (6.00)	0.11 (0.89)

Table 2

Results (Means and Standard Deviations by Experimental Condition) of Experiment 1

Condition	Plausible		Implausible	
	RT	Error Rate	RT	Error Rate
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
<i>Predictable</i>				
Positive Response	953 (280)	.029 (.057)	1034 (333)	.025 (.055)
Negative Response	873 (261)	.095 (.117)	927 (331)	.086 (.101)
<i>Non-predictable</i>				
Positive Response	1026 (293)	.048 (.086)	1171 (371)	.070 (.106)
Negative Response	995 (317)	.096 (.131)	1009 (310)	.113 (.140)

Note. Means and standard deviations are based on participants as units of observation.

Table 3

Results (Means and Standard Deviations by Experimental Condition) of Experiment 2

Condition	Plausible		Implausible	
	RT	Error Rate	RT	Error Rate
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
<i>Predictable</i>				
Positive Response	647 (143)	.019 (.054)	694 (198)	.033 (.067)
Negative Response	655 (151)	.024 (.059)	666 (152)	.032 (.083)
<i>Non-predictable</i>				
Positive Response	662 (149)	.059 (.095)	675 (179)	.029 (.067)
Negative Response	678 (173)	.036 (.072)	673 (153)	.021 (.058)

Note. Means and standard deviations are based on participants as units of observation.

Figure Captions

Figure 1. Mean correct response latency as a function of plausibility (plausible, implausible) and orthographical correctness (correct, incorrect) in the orthographical judgment task of Experiment 1. Error bars correspond to ± 1 standard error of the mean computed for within-subjects designs (Morey, 2008).

Figure 2. Mean correct response latency as a function of plausibility (plausible, implausible) and required response (positive, negative) in the color judgment task of Experiment 2. Error bars correspond to ± 1 standard error of the mean computed for within-subjects designs (Morey, 2008).

Figure 1

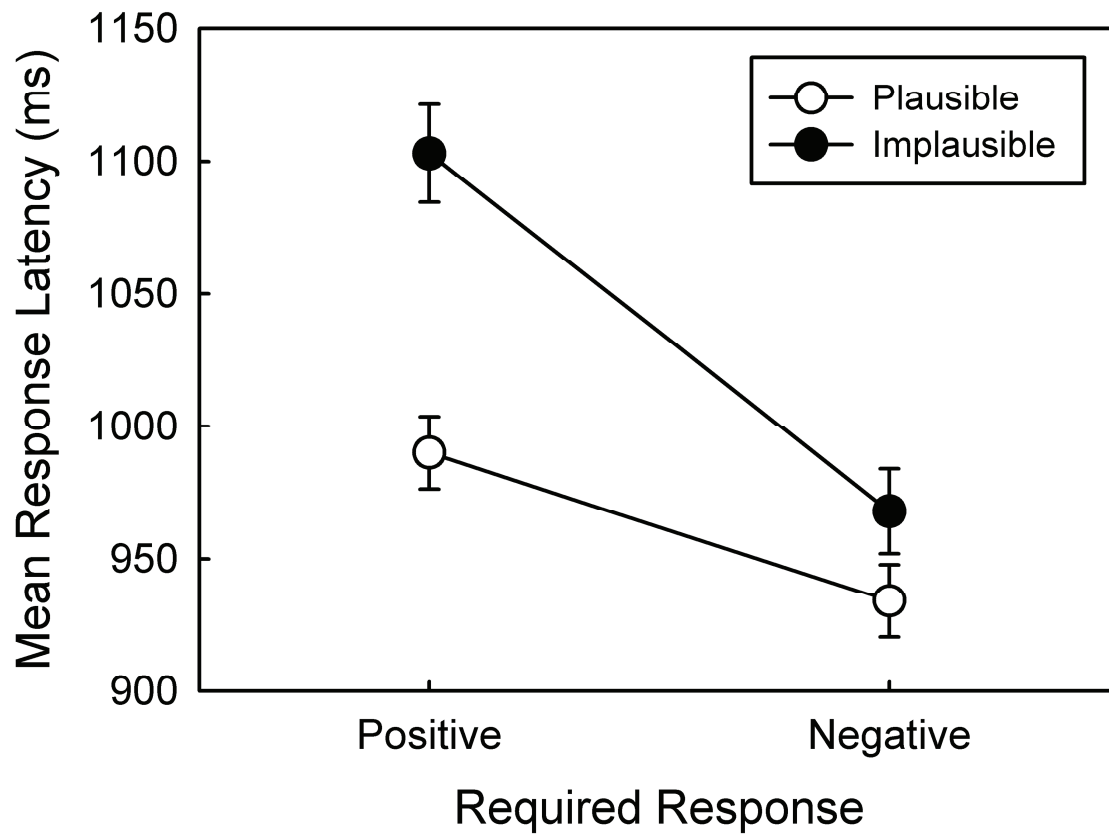
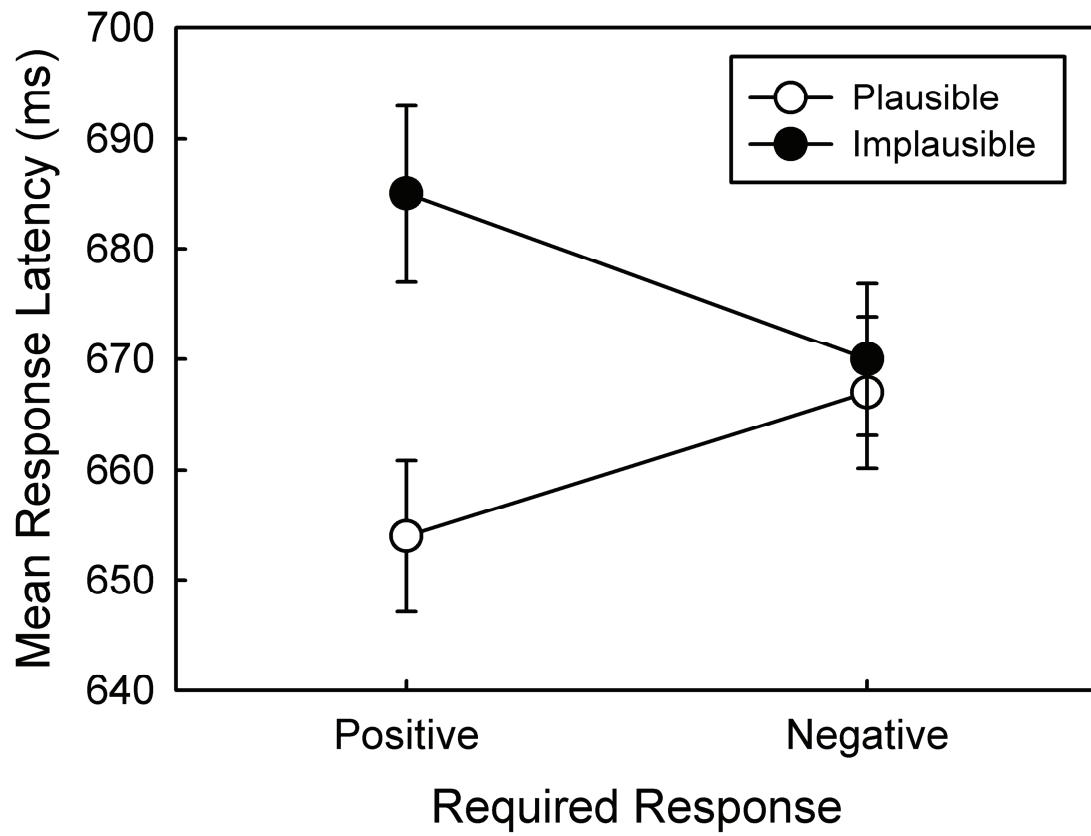


Figure 2



Chapter 4

Study 2

Does validation during language comprehension depend on an evaluative mindset?

A version of this chapter will be published as:

Isberner, M.-B., & Richter, T. (in press). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes*.

Abstract

Whether information is routinely and nonstrategically evaluated for truth during comprehension is still a point of contention. Previous studies supporting the assumption of nonstrategic validation have used a Stroop-like paradigm in which participants provided yes/no judgments in tasks unrelated to the truth or plausibility of the experimental sentences. Other studies using a nonevaluative task failed to support this assumption. This leaves open the possibility that validation is conditional on an evaluative mindset of the reader. In the present study, we investigated this question directly by using a nonevaluative probe task. Participants responded to the probe words "true" or "false" with two different keys after reading true or false sentences for comprehension. Results provide evidence for routine validation even when it is not encouraged by the task but they also suggest that semantic processing is critical for validation to occur. These results can be taken as evidence for a close connection between validation and comprehension rather than validation being a goal-dependent process.

Introduction

Whether information is routinely evaluated for validity (or truth) during language comprehension is still a point of contention. A widely accepted view is that validation – that is, computing truth values or plausibility based on relevant world knowledge – is a strategic, optional process which is subsequent to comprehension (Gilbert, 1991; Gilbert, Krull, & Malone, 1990; Gilbert, Tatarodi, & Malone, 1993; Herbert & Kübler, 2011). Based on this idea, two-step models of comprehension and validation either assume that comprehension proceeds without any evaluative component (e.g., Connell & Keane, 2006) or that the linguistic input is by default initially accepted as true and can only effortfully be “unbelieved” at a later point (e.g., Gilbert et al., 1990, 1993). However, many psycholinguistic studies implicitly or explicitly call into question the conceptualization of comprehension and validation as nonoverlapping stages of information processing. In fact, validation is often utilized in psycholinguistic studies to *measure* comprehension in the first place: Sentence verification has been a popular tool to assess the time it takes to understand a sentence, thus allowing conclusions regarding the organization of semantic memory (e.g., Kintsch, 1980; Kounios, Osman, & Meyer, 1987). Similarly, readers’ ability to detect inconsistencies with their knowledge or with prior text information is frequently utilized to study the kinds of knowledge or portions of prior discourse that are accessed during comprehension, and the time courses of their activation (Fischler, Childers, Achariyapaopan, & Perry, 1985; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Nieuwland & van Berkum, 2006; O’Brien, Rizzella, Albrecht, & Halleran, 1998; Rayner, Warren, Juhasz, & Liversedge, 2004; van Berkum, Zwitserlood, Hagoort, & Brown, 2003), as well as how (quickly) particular syntactic structures are interpreted (e.g., Nieuwland & Kuperberg, 2008; Pickering & Traxler, 1998; Speer & Clifton, 1998; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Traxler & Pickering, 1996; van Gompel, Pickering, & Traxler, 2001).

A prominent example is the study of memory-based processes: These have been demonstrated by introducing an inconsistency with prior information into a text (e.g., *Mary is a vegetarian. [...] She orders a cheeseburger*), which generally results in longer reading times on the inconsistent sentence (e.g., Albrecht & O’Brien, 1993; O’Brien et al., 1998). This finding is taken as evidence for the reactivation of prior text information by memory-based processes, but it also shows how routinely readers detect inconsistencies during comprehension. In a similar vein, studies on how and when linguistic input is related to the wider discourse have used

inconsistencies, which generally elicit an elevated N400 event-related potential (ERP) component, to show the immediate integration of new information into its context (Nieuwland & van Berkum, 2006; van Berkum, Zwitserlood, Hagoort, & Brown, 2003). Following a similar logic, Hagoort et al. (2004) provided evidence for simultaneous integration of semantic knowledge and world knowledge in language comprehension by showing that semantic violations (*Dutch trains are sour*) and world knowledge violations (*Dutch trains are white*) elicit an N400 of similar time course and magnitude compared to correct control information (*Dutch trains are yellow*). This not only calls the traditional distinction between semantic knowledge (relevant for comprehension) and world knowledge (relevant for validation) into question (see also Jackendoff, 2002), but also clearly speaks “against a nonoverlapping two-step interpretation procedure in which first the meaning of a sentence is determined, and only then is its meaning verified in relation to our knowledge of the world” (Hagoort et al., 2004, p. 440).

Consistent with this notion, Singer (2006) proposed that “Memory-based processes afford the verification of the current text constituent.” (p. 587). Essentially, this implies that the information which is passively activated for the comprehension of incoming information concurrently allows its validation. Singer and colleagues have shown that this is true for so-called bridging inferences, which causally link sentences (e.g., Singer, 1993; Singer, Halldorson, Lear, & Andrusiak, 1992): When participants read sentences such as *Dorothy poured the bucket of water on the fire. The fire went out* or *Dorothy poured the bucket of water on the fire. The fire grew hotter*, which imply a causal relationship that is either consistent or inconsistent with general world knowledge, they are subsequently faster to answer the question *Does water extinguish fire?* than after reading sentences which did not imply a causal relationship (i.e., *Dorothy placed the bucket of water by the fire. The fire went out / grew hotter*). This suggests that causal bridging inferences are not only routinely generated during comprehension but at the same time validated against relevant world knowledge.

All of these results suggest a close connection or even an overlap between comprehension and validation, which is rarely explicitly addressed (but see Fischler & Bloom, 1980; Fischler et al, 1983; Isberner & Richter, 2013; Murray & Rowan, 1998; Richter, Schroeder, & Wöhrmann, 2009; Singer, 2006; West & Stanovich, 1982, for exceptions). Moreover, they suggest that validation is a routine component of comprehension under normal reading conditions – that is, without the explicit instruction to validate incoming information in relation to world knowledge or prior discourse. In line with this idea, Richter et al. (2009) and Isberner and Richter (2013) found

evidence for nonstrategic validation in terms of Stroop-like stimulus response compatibility effects (Stroop, 1935). In these experiments, participants read sentences presented word-by-word and were prompted to respond with a positive or negative response (independent of validity) at varying points during sentence presentation. In experimental items, which varied in validity (e.g., *Perfume contains scents/Soft soap is edible*) or plausibility (e.g., *Frank has a broken pipe/leg. He calls the plumber*), the prompt always appeared immediately after the end of the sentence. With this paradigm, positive and negative responses in an orthographical judgment task (*Is the word spelled correctly?* Isberner & Richter, 2013; Richter et al., 2009) and a color judgment task (*Did the word change color?* Isberner & Richter, 2013) were shown to be slower when they were incongruent with the truth value or plausibility of the sentence read prior to responding (i.e., a positive/‘correct’/‘yes’ response after a false/implausible sentence or a negative/‘incorrect’/‘no’ response after a true/plausible sentence) than when they were congruent. This suggests that readers cannot ignore validity or plausibility even when it is irrelevant to the experimental task (irrelevant stimulus-response compatibility). Based on these findings, Richter et al. suggested that comprehension comprises an epistemic monitoring process, which detects inconsistencies with easily accessible prior knowledge and thus protects the mental representation (situation model; Zwaan & Radvansky, 1998) from contamination with false or implausible information (Schroeder, Richter, & Hoever, 2008).

However, the generality of these findings has recently been called into question by a study by Wiswede, Koranyi, Müller, Langner, and Rothermund (2013) which seems to show that the Stroop-like compatibility effect is conditional on an evaluative mindset. Although validity was irrelevant in the orthographical task used by Richter et al. (2009), Wiswede et al. noted that the correct/wrong orthography decision may have induced an evaluative mindset which may have encouraged evaluation of the stimuli for validity. Thus, Wiswede et al. attempted to show that effects of automatic validation hinge on an evaluative mindset of the reader.

For this purpose, they asked participants to read obviously true and false sentences (e.g., *Africa is a continent* or *Saturn is not a planet*) presented word-by-word using Rapid Serial Visual Presentation. After the presentation of each sentence, which was followed by a blank screen of 1500 ms, a prompt was presented signaling which of two randomly intermixed tasks participants had to perform on the current trial. One of the two tasks was a simple probe word identification task in which participants had to respond to a probe which either read “TRUE” or “FALSE” with the associated response key. Importantly, the probe was independent of the actual truth value of

the sentence; that is, it matched the truth value on a random half of the trials (e.g., *Africa is a continent* – TRUE) and on the other half it did not (e.g., *Africa is a continent* – FALSE). Participants only had to respond to the probe, regardless of whether it matched the truth value of the previous sentence or not.

To induce an evaluative or nonevaluative mindset, Wiswede et al. (2013) intermixed this probe task randomly with a second task, which differed between two groups of participants. In the evaluative mindset group, the second task was a truth evaluation task, in which the participants were prompted to decide about the truth value of the sentence (e.g., *Africa is a continent* – True or false?). Thus, evaluating the sentences regarding their truth value was encouraged in this group because it was useful on one of the two tasks (i.e., on a random half of the trials). In the nonevaluative mindset group, the second task was a sentence comparison task. Participants were shown a sentence which was either the same as the one they had read before (e.g., *Africa is a continent* – *Africa is a continent*) or a slightly different one (e.g. *Africa is a continent* – *Africa is a planet*) and had to indicate whether the second sentence was the same as the first (“Is this the sentence that you’ve just seen?”). Thus, in this task, evaluating the truth value of the sentences was not beneficial for completing either of the two tasks. Wiswede et al. assumed that as participants did not know which of the two tasks they would have to perform on each trial until the response prompt appeared, the demands of the second task (or the mindset induced by the second task) would affect sentence processing in the probe task as well, which should result in performance differences in this task between the two groups. Specifically, they expected that only the evaluative mindset group would exhibit interference if the probe (TRUE or FALSE) did not match the actual truth value of a sentence, whereas the nonevaluative mindset group would not spontaneously evaluate the sentences (given that it was not required by either of the two tasks), and thus not show any interference.

In line with this prediction, Wiswede et al. (2013) found a compatibility effect in terms of a Truth X Probe interaction as well as ERP evidence for validation in the evaluative mindset group, but not in the nonevaluative mindset group. Thus, they concluded that validation is conditional on an evaluative mindset of the reader.

In the present study, we call this conclusion into question. We argue that the two different tasks Wiswede et al. (2013) intermixed with the probe task differed not only in whether they encouraged evaluation, but more generally regarding the depth of semantic processing that was required. Consider, for example, that the sentence comparison task could also be performed in a

foreign language at a purely perceptual level (even though it must be noted that the first sentence was presented with Rapid Serial Visual Presentation while the second sentence was presented all at once). This idea is supported by the fact that effects of semantic mismatches on the amplitude of the N400, which is associated with semantic processing (e.g., Kutas & Federmeier, 2011), were significantly reduced in the nonevaluative mindset group ($\eta_p^2 = .84$ in the evaluative mindset group vs. $\eta_p^2 = .36$ in the nonevaluative mindset group). Naturally, if the depth of semantic processing is reduced, the effectiveness of validation will be impaired as well. Therefore, in order to investigate the conditionality of validation, it would be more appropriate to use a task which, while still requiring an adequate depth of processing of the stimuli, does not explicitly encourage validation. Our main goal in the present study was to find such a task, and to test whether it produces compatibility effects as a function of validity and required response, which are consistent with routine validation. This finding would support the idea that merely understanding a sentence by default entails its validation, provided that a reader has easily accessible knowledge which allows assessing its validity.

To test this assumption, we used the probe task by Wiswede et al. (2013) but combined it with comprehension questions which did not require validation of the sentences – namely, whether or not a particular sentence involved an animate object. In this way, our task ensured comprehension of the sentences without encouraging validation.

Another open question is whether the reported compatibility effects reflect facilitation for compatible conditions, interference for incompatible conditions, or both. Thus, a second goal of our study was to address this question. As has been noted in the Stroop literature, in order to study interference and facilitation effects, it is necessary to use an adequate neutral condition (e.g., MacLeod, 1991). For the purpose of the present study, our idea was to use as a neutral condition stimuli for which participants have no (or little) knowledge which would allow them to assess the validity of the sentences (e.g., *Toothpaste contains sulfur*). Thus, these sentences should not create interference with or facilitation of positive and negative responses because participants do not possess easily accessible knowledge required for assessing validity.

If our assumptions hold, we should thus find a compatibility effect in the response latencies but only for items for which participants have high knowledge. Thus, we expect a three-way interaction of knowledge, validity, and required response, driven by a two-way interaction of the latter two variables (compatibility effect) emerging only in the high knowledge condition.

Moreover, compatibility effects may show not only in the response latencies, but also in the error rates (e.g., Richter et al., 2009).

Method

Participants

Participants were 42 students of various subjects at the University of Kassel. The data of 9 nonnative speakers of German were excluded from the analysis. The average age of the 33 remaining participants (21 female) was 23.3 years ($SD = 2.9$; range 19-33 years). Participants provided informed consent at the beginning of the experiment and were reimbursed with 6 € after its completion.

Stimulus Material

The stimuli were valid (true) and invalid (false) sentences of the structure “[a] [*concept noun*] [is/has/causes/contains] [a] [*concept noun/adjective*].” - for example *Perfume contains scents* (the actual stimuli were in German; e.g., *Parfüm enthält Duftstoffe*). The materials were taken from the study by Richter et al. (2009, Experiment 4), having already been normed for validity and knowledge. The 12 participants in that norming study were asked to indicate for 288 items in total (144 true and 144 false) whether the sentences were true or false and how certain they were in their judgment on a 6-point scale ranging from 1 (*very uncertain*) to 6 (*very certain*). This norming study allowed for grouping the items according to knowledge (high knowledge: high agreement between participants and high average judgment certainty vs. low knowledge: low agreement between participants and low average judgment certainty).

Experimental items. From this pool of items, 96 experimental items were drawn. Half of these were associated with high knowledge, i.e., the true items had consistently been judged as true (mean agreement: 100%) with high judgment certainty ($M = 5.73$, $SD = 0.18$), e.g., *Perfume contains scents*, and the false items had consistently been judged as false (mean agreement: 98%) with high judgment certainty ($M = 5.78$, $SD = 0.10$), e.g., *Soft soap is edible*. These items were identical to the 48 experimental items used in Experiment 4 by Richter et al. (2009).

In addition, we used 24 true and 24 false items for which participants in the norming study had exhibited low knowledge (e.g., *Krypton is a noble gas* or *Toothpaste contains sulfur*); that is, the truth value of these items had been judged inconsistently in the norming study (mean agreement: 56%) and with on average low judgment certainty ($M = 2.83$, $SD = 0.95$). These were

part of the filler items in Experiment 4 by Richter et al. (2009) but were used as experimental items in the present study.

Filler items. For the filler trials, we used 56 additional items from the Richter et al. (2009) material, of which 32 were associated with low knowledge and 24 were associated with high knowledge. Of each of these, half were true and half were false.

Procedure

Participants were tested in a computer lab in groups of up to 5 people. They were asked to rest the index fingers of their left and right hand on the two response keys throughout the experiment, and to respond as fast and as accurately as possible. All stimuli were presented in an individually randomized order. Every 40 trials, participants were allowed to take a short break. The first 8 trials were practice trials, after which participants had the opportunity to ask questions before starting the actual experiment.

On all trials, the stimuli were presented word-by-word on a computer screen using Rapid Serial Visual Presentation with a fixed rate of 300 ms per word; all words were presented in black font (Arial, approximate height 1 cm) against a white background. Every trial was followed by a blank screen presented for 1000 ms. Figure 1 displays the trial structure of experimental and filler trials.

Experimental trials. Our Stroop-like task combined the procedure employed by Richter et al. (2009) with the probe identification task used by Wiswede et al. (2013). In the experimental 96 trials, the probe "***Richtig**" ("True") or "***Falsch**" ("False") appeared after the third and final word of the stimulus sentence, prompting the participants to respond with the corresponding key ("k" for the "True" probe, "d" for the "False" probe). Half of the trials were presented with the "True" probe, the other half with the "False" probe. The probe was presented orthogonally to the validity of the sentence; that is, it was independent of validity and thus either matched or mismatched the actual truth value of the sentence. Importantly, participants were only required to identify the probe word and press the corresponding key, regardless of whether the probe matched or mismatched the validity of the sentence. The probe remained on the screen until the participant provided a response.

Filler trials. In the 56 filler trials, as in the study by Richter et al. (2009), the probe appeared after the first word (28 trials) or the second word (28 trials) of the sentence, in order to make the appearance of the probe less predictable and the goal of the study (investigating effects

of the match or mismatch of the probe with the actual truth value of the sentence) less transparent. Half of the filler trials were presented with the “True” probe, whereas the other half were presented with the “False” probe.

Comprehension questions. In each of the 28 filler trials on which the prompt appeared after the first word, a comprehension question was presented immediately after the sentence which – crucially – required comprehension but not validation of the sentence (see Figure 1). Thus, comprehension was ensured without inducing an evaluative mindset. Specifically, participants were asked to indicate whether the sentence had referred to an animate object; five of the comprehension questions required a yes-response². Participants were informed before the experiment that they would be asked comprehension questions and were instructed to process the content of the sentences to be able to answer these questions. To make sure that participants understood the importance of reading for comprehension before starting the actual experiment, 2 of the 8 practice trials comprised comprehension questions.

Design

The design was a 2 (knowledge: high vs. low) X 2 (validity: valid vs. invalid) X 2 (required response: positive vs. negative) within-subjects design. The assignment of the probes “True” and “False” to the stimuli was counterbalanced via two item lists. Response latencies and accuracy in the probe task were recorded as dependent variables.

Results

Type-I error probability was set at .05 for all hypothesis tests. ANOVAs were conducted for repeated measurements with both participants (F_1 , by-subjects) and items (F_2 , by-items) as the source of random variance. The reported means and standard errors were computed with subjects as the units of observation.

Comprehension Questions

On average, participants answered 85.9 % ($SD = 7.8$ %) of the comprehension questions correctly.

² The imbalance between items requiring yes -and no- responses was due to the fact that we used the already normed material from Richter et al. (2009), which had not been constructed to be balanced regarding animacy of the objects. Given that participants received the questions in a random order, however, we think that it is unlikely that this induced a response strategy in the reader.

Response Latencies

Response latencies were included for correct responses (94.9% of the responses in experimental trials). Latencies deviating more than three standard deviations from either the subject or item mean (2.1% of all correct latencies) were treated as outliers and removed from the data set.

The full data for the response latencies are displayed in Figure 2. In the overall 2 X 2 X 2 ANOVA, we found main effects of all three independent variables: Items for which participants had high knowledge ($M = 700$, $SE = 43$) were responded to faster than items for which participants had low knowledge ($M = 720$, $SE = 46$), $F_1(1, 32) = 5.92$, $p < .05$, $\eta_p^2 = .16$, $F_2(1, 92) = 7.05$, $p < .01$, $\eta_p^2 = .07$; valid items ($M = 693$, $SE = 41$) were responded to faster than invalid items ($M = 727$, $SE = 48$), $F_1(1, 32) = 12.94$, $p < .01$, $\eta_p^2 = .29$, $F_2(1, 92) = 8.47$, $p < .01$, $\eta_p^2 = .08$; and positive responses ($M = 695$, $SE = 43$) were given faster than negative responses ($M = 725$, $SE = 46$), $F_1(1, 32) = 5.44$, $p < .05$, $\eta_p^2 = .15$, $F_2(1, 92) = 5.29$, $p < .05$, $\eta_p^2 = .05$. In addition, there was a two-way interaction between validity and required response in the by-subjects analysis, $F_1(1, 32) = 4.69$, $p < .05$, $\eta_p^2 = .13$, $F_2(1, 92) = 2.12$, $p = .15$. However, this interaction was qualified by a three-way interaction between all three independent variables, $F_1(1, 32) = 4.26$, $p < .05$, $\eta_p^2 = .12$, $F_2(1, 92) = 6.07$, $p < .05$, $\eta_p^2 = .06$. We followed up on this finding by running separate analyses for the high and low knowledge conditions.

High versus low knowledge. Whereas in the low knowledge condition, there were no significant effects, all $p > .10$, in the high knowledge condition there were significant main effects of validity, $F_1(1, 32) = 12.31$, $p < .01$, $\eta_p^2 = .28$, $F_2(1, 46) = 16.27$, $p < .001$, $\eta_p^2 = .26$, with valid items ($M = 675$, $SE = 39$) being responded to faster than invalid items ($M = 726$, $SE = 48$), and required response, $F_1(1, 32) = 9.52$, $p < .01$, $\eta_p^2 = .23$, $F_2(1, 46) = 4.42$, $p < .05$, $\eta_p^2 = .09$, with positive responses ($M = 685$, $SE = 42$) given faster than negative responses ($M = 716$, $SE = 45$). More importantly, and as predicted, there was an interaction between validity and required response, $F_1(1, 32) = 4.79$, $p < .05$, $\eta_p^2 = .13$, $F_2(1, 46) = 7.97$, $p < .01$, $\eta_p^2 = .15$. Planned comparisons revealed that this interaction was due to positive responses after valid sentences ($M = 625$, $SE = 30$) being significantly faster compared to negative responses after valid sentences ($M = 724$, $SE = 52$), $F_1(1, 32) = 8.71$, $p < .01$, $\eta_p^2 = .21$, $F_2(1, 46) = 12.13$, $p < .01$, $\eta_p^2 = .21$, as well as compared to positive responses after invalid sentences ($M = 744$, $SE = 58$), $F_1(1, 32) = 9.08$, $p < .01$, $\eta_p^2 = .22$, $F_2(1, 46) = 18.96$, $p < .001$, $\eta_p^2 = .29$.

This pattern is in line with the predicted Stroop-like effect, with responses in the congruent condition *valid sentence/positive response* being faster than in the two incongruent conditions *valid sentence/negative response* and *invalid sentence/positive response*. However, it remains unclear whether this pattern reflects facilitation in the congruent condition, interference in incongruent conditions, or both. Given that there were no significant effects in the low knowledge condition – as indeed there should not be, if our participants did not have knowledge concerning the validity of the sentences – it appears that this condition may be a suitable neutral condition to test for interference and facilitation effects. Therefore, we reran the analyses comparing responses in the high knowledge conditions to responses in the respective low knowledge (control) conditions. For this purpose, we ran separate ANOVAs for valid and invalid sentences.

Facilitation versus interference. For valid sentences, there was a significant interaction between knowledge and required response, $F_1(1, 32) = 5.87, p < .05, \eta_p^2 = .16, F_2(1, 46) = 5.78, p < .05, \eta_p^2 = .11$. This interaction was driven by a large facilitation effect for positive responses in the high knowledge condition ($M = 625, SE = 30$) compared to the low knowledge control condition ($M = 706, SE = 48$), $F_1(1, 32) = 11.83, p < .01, \eta_p^2 = .27, F_2(1, 46) = 14.82, p < .001, \eta_p^2 = .24$, with a small numerical trend of interference for negative responses in the high knowledge condition ($M = 724, SE = 52$) as compared to the low knowledge control condition ($M = 716, SE = 42$) being nonsignificant, $F_1(1, 32) = 0.22, p = .64, F_2(1, 46) = 0.59, p = .45$. In addition, there were main effects of knowledge, $F_1(1, 32) = 15.78, p < .001, \eta_p^2 = .33, F_2(1, 46) = 10.18, p < .01, \eta_p^2 = .18$, with responses being faster in the high knowledge ($M = 675, SE = 39$) than in the low knowledge ($M = 711, SE = 44$) condition, and of required response, $F_1(1, 32) = 6.14, p < .05, \eta_p^2 = .16, F_2(1, 46) = 8.56, p < .01, \eta_p^2 = .16$, with positive responses ($M = 665, SE = 38$) being faster than negative responses ($M = 720, SE = 47$).

For invalid sentences, the numerical pattern indicates both interference for positive responses (high knowledge: $M = 744, SE = 58$, low knowledge: $M = 707, SE = 45$) and facilitation for negative responses (high knowledge: $M = 709, SE = 41$, low knowledge: $M = 750, SE = 55$); however, the interaction between knowledge and required response fell short of significance, $F_1(1, 32) = 2.05, p = .16, F_2(1, 46) = 1.44, p = .24$.

Error Rates

The full data for the error rates are displayed in Figure 3. The error rates were low overall ($M = .051, SD = .108$); ANOVAs were performed on arc-sine transformed proportions. The only

significant effect that was reliable in both the F_1 and F_2 analyses was an interaction of knowledge and validity, $F_1(1, 32) = 5.18, p < .05, \eta_p^2 = .14, F_2(1, 92) = 7.54, p < .01, \eta_p^2 = .08$. This interaction was due to a significant difference between the error rates for valid and invalid sentences in the high knowledge condition: There was a higher error rate in the valid ($M = .062, SE = .014$) than in the invalid condition ($M = .042, SE = .017$), $F_1(1, 32) = 6.02, p < .05, \eta_p^2 = .16, F_2(1, 92) = 6.42, p < .05, \eta_p^2 = .07$.

In addition, there was a two-way interaction between validity and required response in the by-items analysis, $F_1(1, 32) = 1.55, p = .22, F_2(1, 92) = 4.99, p < .05, \eta_p^2 = .05$, which was qualified by a three-way interaction of all variables, $F_1(1, 32) = 3.62, p = .07, \eta_p^2 = .10, F_2(1, 92) = 11.32, p < .01, \eta_p^2 = .11$.

High versus low knowledge. We followed up on this result by running separate by-items analyses for the high and low knowledge conditions. Similar to the results for the response latencies, there was a significant interaction of validity and required response only in the high knowledge condition, $F_2(1, 46) = 17.29, p < .001, \eta_p^2 = .27$, but not in the low knowledge condition, $F_2(1, 46) < 1$. The interaction in the high knowledge condition was driven by the fact that after invalid sentences, more errors were made when the required response was positive ($M = .063, SE = .030$) than when it was negative ($M = .020, SE = .008$), $F_2(1, 46) = 5.14, p < .05, \eta_p^2 = .10$, and after valid sentences, more errors were made when the required response was negative ($M = .091, SE = .027$) than when it was positive ($M = .033, SE = .008$), $F_2(1, 46) = 13.05, p < .01, \eta_p^2 = .22$.

Facilitation versus interference. To address the question of facilitation vs. interference, as for the response latencies, we ran separate by-items ANOVAs for valid and invalid sentences.

For invalid sentences, the interaction of knowledge and required response was significant, $F_2(1, 46) = 6.13, p < .05, \eta_p^2 = .12$. Participants made significantly less errors when a negative response was required and knowledge was high ($M = .020, SE = .008$) compared to the low knowledge control condition ($M = .068, SE = .020$), $F_2(1, 46) = 10.42, p < .01, \eta_p^2 = .19$. The trend for a higher error rate for positive responses when knowledge was high ($M = .063, SE = .030$) compared to the low knowledge control condition ($M = .043, SE = .015$) was nonsignificant, $F_2(1, 46) < 1$.

For valid sentences, the interaction between knowledge and required response was also significant, $F_2(1, 46) = 5.19, p < .05, \eta_p^2 = .10$, and it was again the negative responses that were affected by a compatibility effect: Participants made significantly more errors when the required

response was negative and knowledge was high ($M = .091$, $SE = .027$) compared to the low knowledge control condition ($M = .048$, $SE = .015$), $F_2(1, 46) = 10.80$, $p < .01$, $\eta_p^2 = .19$.

Discussion

The fact that a compatibility effect in terms of a validity x probe interaction obtained in a nonevaluative task in our study suggests that validation can also occur without an evaluative mindset, that is, without a task that explicitly encourages evaluation. Thus, in contrast to the conclusions by Wiswede et al. (2013), our study suggests that an evaluative task is not a prerequisite for validation and that merely understanding a sentence which can be judged as obviously true or false based on easily accessible knowledge is sufficient to produce a compatibility effect. It was already suggested by Singer (2006) that memory-based processes afford the verification of incoming information. The present study produced evidence that this verification, which operates on the information activated by the passive retrieval processes during comprehension, is nonstrategic, meaning that it operates without the reader's intention. As such, validation itself appears to be a passive process and by default a routine component of comprehension, in the sense that comprehension cannot occur independently from validation as both processes rely on the same knowledge activated by memory-based processes.

However, our results do not rule out the possibility that validation may be conditional in other ways. Quite to the contrary, it appears that a condition which must be fulfilled for validation to occur (or to be successful) is a certain depth of processing, i.e., a minimum level of comprehension. Shallow processing results in the activation of less information via memory-based processing which, in turn, results in less information on which validation can operate. When comprehension is impaired, for example because the experimental task requires only a relatively shallow level of processing (as seems to have been the case in the control group in the study by Wiswede et al., 2013), then validation appears to be impaired as well. Thus, in contrast to what Wiswede et al. proposed, the conditionality of validation does not seem to refer to an evaluative goal or mindset, but rather to a certain level of processing, which points towards a close relationship between comprehension and validation.

One might object that the presentation of the "True" and "False" probes or the use of obviously true and false sentences may have been sufficient to induce an evaluative mindset by making the validity dimension salient to the reader. However, the fact that the same probe task and

similar stimuli did not induce compatibility effects in the nonevaluative mindset group of the study by Wiswede et al. (2013) speaks against this notion.

In extension of previous studies, the present study also sheds light on the question of whether the Stroop-like validity/response compatibility effects reported by Richter et al. (2009), Wiswede et al. (2013), and Isberner and Richter (2013) are attributable to facilitation for congruent conditions, interference for incongruent conditions, or a combination of both. While the present study only produced clear evidence for facilitation of positive responses after valid sentences in the response latencies, the overall pattern of results for the response latencies and error rates suggests both facilitation of congruent and interference with incongruent responses. However, as there were no other interference or facilitation effects that were reliable in both the by-subjects and the by-items analyses in the present experiment, further research on this issue seems desirable.

One major difference between the experiment by Wiswede et al. (2013) and our study should be noted. To avoid interference with the EEG recordings, Wiswede et al. presented the probe 1800 ms after the onset of the final word, whereas it was presented only 300 ms after the onset of the final word in our experiment. This makes the behavioral data somewhat difficult to compare, as it is possible that compatibility effects produced by routine validation change over such a long time course. Nonetheless, the fact that a compatibility effect obtained in Wiswede et al.'s evaluative mindset condition despite the relatively long post-sentence delay suggests that the effects of validation in language comprehension are stable over quite some time. Alternatively, it is possible that the only reason why the "true" and "false" evaluations were kept active over such a long time is because they were relevant for one of the two tasks in the "evaluative mindset group," namely the truth evaluation task. If this assumption is correct, then compatibility effects should disappear at some point between 300 ms and 1800 ms after the sentence if the evaluation is irrelevant for the task (as in our experiment). An interesting direction for future research thus seems to be systematically investigating the time course of the effects by varying the stimulus onset asynchrony (SOA) of the probe. It is also possible that the pattern of interference and facilitation found in the present study merely represents a snapshot and would vary over different SOAs, which would allow insight into the time course of the positive and negative evaluations that arise from epistemic monitoring.

Overall, our results provide strong support for the idea that language comprehension entails a routine, nonstrategic validation process (epistemic monitoring; Richter et al., 2009), because readers do not seem to be able to ignore validity when they have easily accessible knowledge,

even when assessing validity is irrelevant or even detrimental to the experimental task. This speaks against a conceptualization of comprehension and validation as nonoverlapping stages of information processing (e.g., Connell & Keane, 2006; Gilbert, 1991; Gilbert et al., 1990, 1993; Herbert & Kübler, 2011; Wiswede et al., 2013). However, it may seem at odds with studies showing readers' susceptibility to false information (e.g., Bottoms, Eslick, & Marsh, 2010; Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013; Fazio & Marsh, 2008; Marsh & Fazio, 2006; Marsh et al., 2003; Rapp, 2008), as well as examples of readers' failures to sometimes notice even blatant inconsistencies with their knowledge (e.g., Barton & Sanford, 1993; Erickson & Mattson, 1981). Our study may point towards a potential way of reconciling these seemingly contradictory findings: As validation seems to hinge on a minimum depth of comprehension, it is possible that such failures of validation are due to the construction of an underspecified mental representation (Bohan & Sanford, 2008; Sanford, 2002; Sanford & Graesser, 2006; Sanford, Leuthold, Bohan, & Sanford, 2011). In line with this idea, it has been shown that factors which influence the specification of a mental representation, such as linguistic focus (e.g., Sanford & Garrod, 2005), also influence the extent to which false information is detected (Bredart & Modolo, 1988).

Moreover, despite the fact that an evaluative processing goal is not necessary for validation to occur, our results are still compatible with the idea that validation is conditional in other ways. For example, it may be affected by the text genre (narrative vs. expository texts) or by the perceived credibility of a text source (encoding under distrust; e.g., Schul, Mayo, & Burnstein, 2004). For example, people seem to be particularly susceptible to false information and persuasion when reading or viewing narratives (e.g., Appel & Richter, 2007; Gerrig & Prentice, 1991; Green & Brock, 2000; Umanath, Butler, & Marsh, 2012), which suggests that epistemic monitoring might be suppressed to some extent in narrative (as opposed to argumentative) texts. Future research should focus more explicitly on the conditions under which validation succeeds or fails, with the goal of reconciling evidence for readers' apparent susceptibility to false information with the abundant evidence for routine validation in language comprehension.

Acknowledgments

This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, grant RI 1100/4-2). We would like to thank Benjamin Piest for his help in data collection.

References

- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1061–1070.
- Appel, M., & Richter, T. (2007). Persuasive effects of fictional narratives increase over time. *Media Psychology*, *10*, 113-134.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477-487.
- Bohan, J., & Sanford, A. (2008). Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *The Quarterly Journal of Experimental Psychology*, *61*, 232-239.
- Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses Illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, *18*, 670-678.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*, 135-144.
- Connell, L., & Keane M. T. (2006). A model of plausibility. *Cognitive Science*, *30*, 95-120.
- Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*, 540-552.
- Fazio, L. K., Barber, S. J., Rajaram, S., Ornstein, P. A., & Marsh, E. J. (2013). Creating illusions of knowledge: Learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General*, *142*, 1-5.
- Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, *15*, 180-185.
- Fischler, I., & Bloom, P. A. (1980). Rapid processing of the meaning of sentences. *Memory & Cognition*, *8*, 216-225.
- Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E., & Perry, N.W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*, 400–409.
- Fischler, I., Childers, D.G., Achariyapaopan, T., & Perry, N.W. (1985). Brain potentials during sentence verification - automatic aspects of comprehension. *Biological Psychology*, *21*, 83–105.

- Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, 2, 336–340.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46, 107–119.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59, 601–613.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65, 221–233.
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79, 701–721.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Herbert, C., & Kübler, A. (2011). Dogs cannot bark: Event-related brain responses to true and false negated statements as indicators of higher-order conscious processing. *PloS one*, 6, e25574.
- Isberner, M.-B., & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, 142, 15–22.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press: Oxford, England.
- Kintsch, W. (1980). Semantic memory: A tutorial. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 595–620). Hillsdale, NJ: Erlbaum.
- Kounios, J., Osman, A. M., & Meyer, D. E. (1987). Structure and process in semantic memory: New evidence based on speed–accuracy decomposition. *Journal of Experimental Psychology: General*, 116(1), 3–25.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621–647.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, 109, 163–203.
- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, 34, 1140–1149.

- Marsh, E. J., Meade, M. L., & Roediger III, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, 49, 519-536.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4, 61-64.
- Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, 27, 1-22.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19, 1213-1218.
- Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098-1111.
- O'Brien, E., Rizzella, M. L., Albrecht, J. E. & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1200-1210.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 940-961.
- Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition*, 36, 688-701.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1290-1301.
- Richter, T., Schroeder, S., Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96, 538-558.
- Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, 17, 188-206.
- Sanford, A. J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes*, 39, 205-224.
- Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes*, 42, 99-108.

- Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23, 514-523.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59, 237-259.
- Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, 86, 668-679.
- Singer, M. (1993). Causal bridging inferences: Validating consistent and inconsistent sequences. *Canadian Journal of Experimental Psychology*, 47, 340-359.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574-591.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences. *Journal of Memory and Language*, 31, 507-524.
- Speer, S. R., & Clifton, C., Jr. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition*, 26, 965-978.
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1162-1169.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory & Language*, 35, 454-475.
- Umanath, S., Butler, A. C., & Marsh, E. J. (2012). Positive and negative effects of monitoring popular films for historical inaccuracies. *Applied Cognitive Psychology*, 26, 556-567.
- van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17, 701-718.
- van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45, 225-258.

- West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 385-399.
- Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, 8, 647-653.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Figure Captions

Figure 1. Trial structure of the experiment. (A) Experimental trials. (B) Filler trials with probe after the second word. (C) Filler trials with probe after the first word.

Figure 2. Mean correct response latency as a function of validity (invalid, valid) and required response (positive, negative) for a) low knowledge and b) high knowledge. Error bars correspond to ± 1 standard error of the mean computed for within-subjects designs (Morey, 2008).

Figure 3. Mean error rates as a function of validity (invalid, valid) and required response (positive, negative) for a) low knowledge and b) high knowledge. Error bars correspond to ± 1 standard error of the mean computed for within-subjects designs (Morey, 2008).

Figure 1

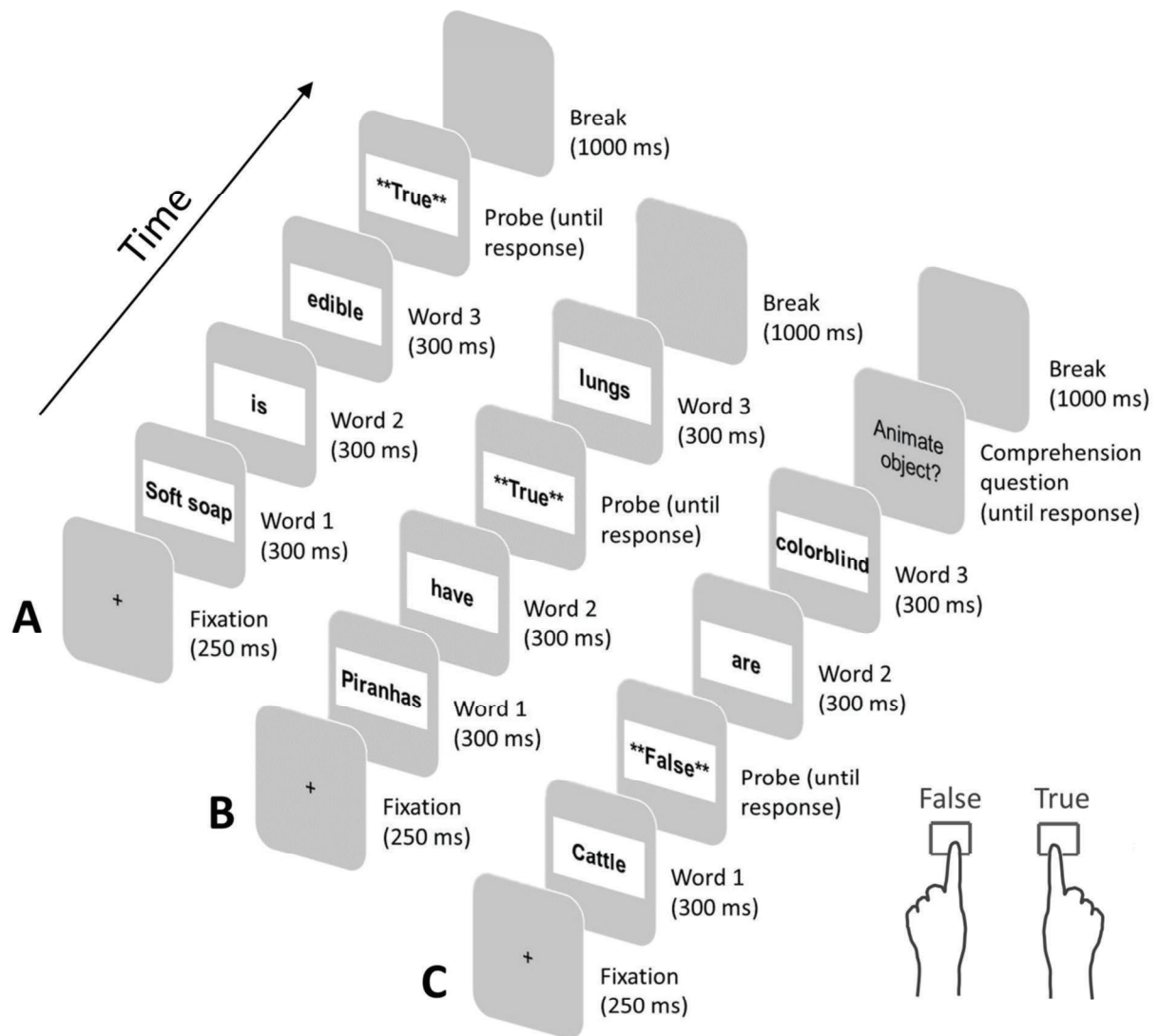
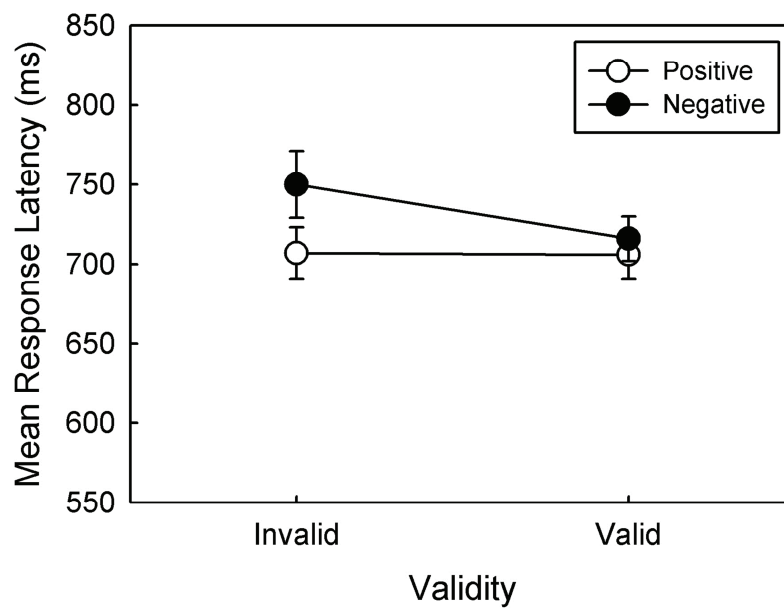


Figure 2

a)



b)

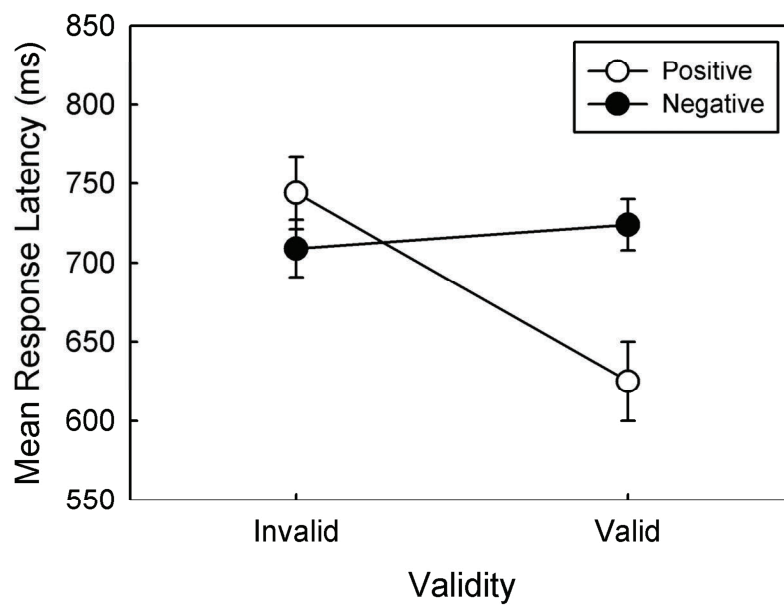
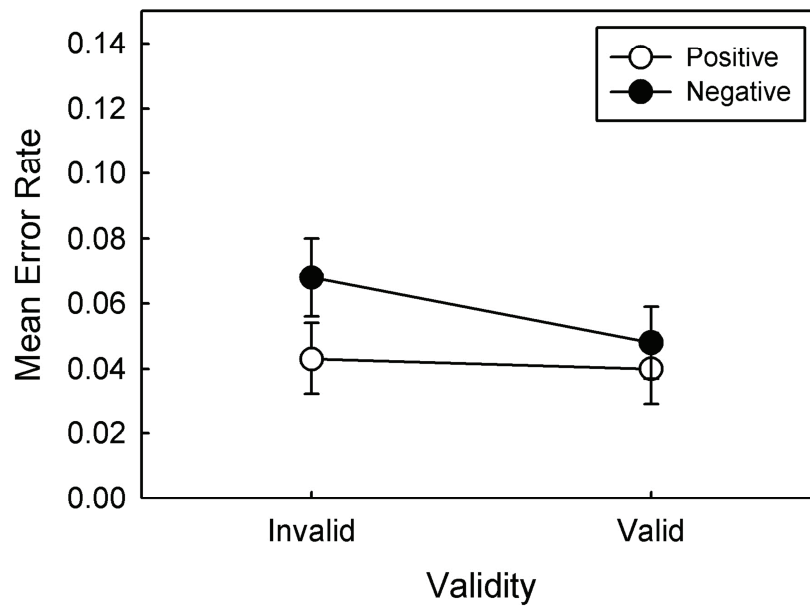
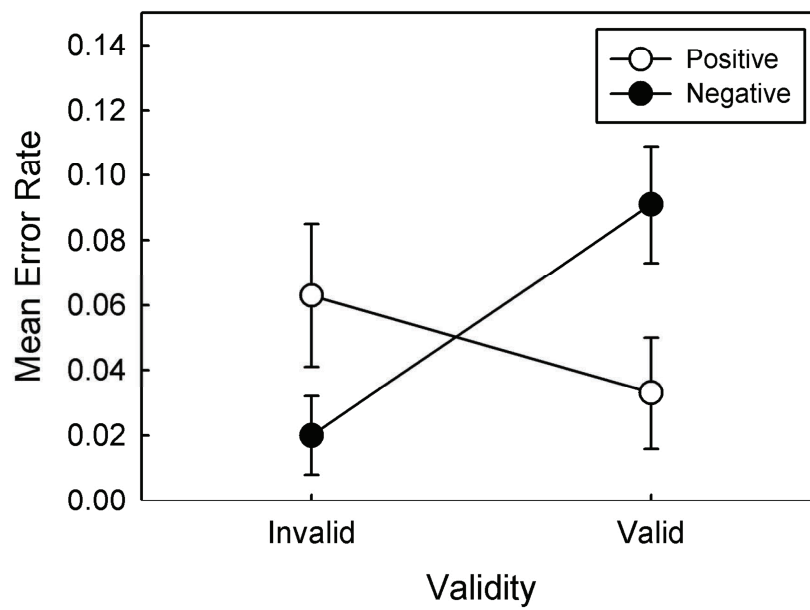


Figure 3

a)



b)



Chapter 5

Study 3

Epistemic modality in sentence comprehension: Effects of epistemic adverbs on eye movements

A version of this chapter is submitted as:

Isberner, M.-B., Richter, T., & Kaakinen, J. (2013). Epistemic modality in sentence comprehension: Effects of epistemic adverbs on eye movements. *Manuscript submitted for publication*.*

*The first and the second author contributed equally to this article.

Abstract

Epistemic adverbs are markers of epistemic modality that signal the certainty of communicated information. The present study used eye movements to investigate the effects of epistemic markers on on-line comprehension. Participants read mini-stories whose second sentence was always the target sentence. Target sentences included target words that were predictable versus unpredictable (Experiment 1) or plausible versus implausible (Experiment 2) within the context. In addition, the presence of epistemic markers in the target sentence was manipulated. Experiment 1 explored whether certainty (e.g., *certainly*) and uncertainty (e.g., *perhaps*) markers (compared to no markers) modulate effects of predictability on sentence processing. Experiment 2 explored whether uncertainty markers (compared to neutral adverbs, e.g., *presently*, or no markers) modulate effects of plausibility. Predictability and plausibility affected indicators of both early and late comprehension processes; however, only plausibility effects were modulated by epistemic markers. The results suggest that epistemic adverbs serve an alerting function and influence validation processes during comprehension.

Keywords: Epistemic modality – evidentiality – eye movements – language comprehension – predictability – plausibility – validation.

Introduction

Speakers can tell others what they know, think, or believe by asserting propositions about some state of affairs in the world. However, knowledge is fallible as a matter of principle and speakers can be more or less confident about whether what they say is actually true. Accordingly, all languages of the world provide speakers with a rich array of grammatical or lexical devices to express their degree of commitment and, thus, to qualify the propositional content of their utterances with an epistemic judgment. In linguistic semantics, this capacity of language is called *epistemic modality* (e.g., Lyons, 1977; Nuyts, 2001; Palmer, 2001). Speakers of English can use modal verbs (*It may/must have rained earlier today*), mental verbs (*I know/think/doubt that it has rained earlier today*), or adverbials (*Certainly/Probably/Perhaps it has rained earlier today*) to express the degree to which they are committed to what they say. Together, these epistemic modal markers (or epistemic markers) form an epistemic scale ranging from the speaker expressing certainty that the asserted proposition is in fact true over an agnostic stance to certainty that the asserted proposition is in fact false (Nuyts, 2001, p. 22). The epistemic scale seems to provide a universal semantic structure of epistemic modality, regardless of the fact that its grammatical or lexical realization varies greatly across languages (Palmer, 2001). Epistemic markers are also used quite frequently. For example, Chafe (1986) found in a corpus analysis of US-American conversations and academic writing that roughly 4% of the words in conversations and academic writing were epistemic markers (i.e. modal verbs, mental verbs, or adverbials).

Despite their cross-linguistic relevance and pervasiveness in actual language use, epistemic markers have attracted very little theoretical and empirical research in the psychology of language (with few exceptions, Britt & Larson, 2003; Ferretti, Singer, & Patterson, 2008; Singer, 2006, 2009). Against this background, the present research explored two basic theoretical possibilities of how epistemic markers in the form of modal adverbials might affect on-line processing of declarative sentences. The first possibility is that epistemic markers may guide comprehenders in using their world knowledge to predict and integrate upcoming parts of discourse. As a consequence, knowledge-based predictability effects (e.g., van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) might be modulated by the presence of epistemic markers. The second possibility is that epistemic markers may alert comprehenders to scrutinize the validity of the communicated information, hence modulating knowledge-based plausibility effects (e.g.,

Singer, 2006). These two theoretical possibilities were tested in an eye-tracking experiment with adverbial epistemic markers; their specific predictions will be discussed in turn.

Epistemic Markers and Knowledge-Based Predictability Effects

One way how epistemic markers might affect sentence comprehension is by modulating knowledge-based effects of predictability, i.e. constraints provided by parts of the text preceding the sentence. A large body of research suggests that comprehenders make *predictive inferences* about what will happen next in an unfolding narrative if the semantic context is sufficiently constraining (e.g., Calvo, 2001; Calvo & Castillo, 1996; Campion, 2004; Fincher-Kiefer, 1993). Other studies concerned with prediction have shown that readers do not only make predictions at a conceptual level, but that they can also use textual cues to predict what words they will encounter next as a sentence unfolds (DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Pickering & Garrod, 2007). An experiment by van Berkum et al. (2005) is particularly informative in this respect. In the experiment, participants read two-sentence mini-stories. When the first sentence strongly suggested that a particular noun would appear at a particular position in the second sentence, participants took longer to read a preceding adjective that did not agree with the gender of the predictable noun. Given that these effects occurred before the critical noun was processed, they cannot be explained by a facilitation of integrative processes that occur only after word recognition. Thus, the results by van Berkum et al. (2005) may be regarded as evidence for top-down prediction processes which are based on comprehenders' world knowledge and are triggered by the linguistic context. An eye-tracking study by Ehrlich and Rayner (1981) suggests a similar conclusion. Here, words that were predictable from the context not only received shorter fixations but they were also skipped more often than words that could not be predicted. The skip rate is an indicator of early word recognition processes that occur even before the word is fixated. Thus, the higher skip rate for predictable words suggests that comprehenders indeed engaged in knowledge-based predictions. A number of eye-tracking studies have provided additional evidence for facilitating effects of predictability (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Frisson, Rayner, & Pickering, 2005; Rayner & Well, 1996). However, it must be noted that unlike the study by van Berkum et al. (2005), these studies focused on fixation measures on the predictable words, which might reflect prediction as well as integration processes in varying degrees. Nevertheless, given that predictability effects are reliably obtained on early eye-tracking measures (first fixation

duration and gaze duration), the eye-tracking studies may be taken as tentative support for the assumption that predictability facilitates prediction and not only integration.

Thus, predictability greatly benefits language comprehension. However, when prediction fails, it can increase processing costs and disrupt the flow of comprehension. Similarly, when inadequate knowledge is activated and used for integrating text information, comprehension is likely to fail as illustrated by the many instances of the Moses illusion (*How many animals did Moses put on the ark?*, Erickson & Mattson, 1981). This is why epistemic markers might serve an important role in on-line processing. If prediction and integration are to assist comprehension, it would be dysfunctional for comprehenders to use their knowledge blindly. Rather, they should be sensitive to textual signals which indicate whether it makes sense to rely on activated knowledge to predict or integrate textual information. From this perspective, epistemic markers may be regarded as signals that modulate the effects of predictability on prediction and integration of information. We will coin this the *predictability modulation hypothesis*.

According to the predictability modulation hypothesis, the likelihood that comprehenders rely on the constraints provided by the linguistic context depends on the degree of certainty signaled by an epistemic marker. By using epistemic markers that signal a high degree of certainty, speakers make it explicit that what they are saying is backed up by commonly available knowledge (common ground, Clark & Brennan, 1991). In this way, they advise comprehenders that it is safe to rely on this knowledge for interpreting incoming linguistic information but also for making predictions about upcoming parts of discourse. Consider, for example, the following pair of sentences:

(1a) Simon is overweight.

(1b) Naturally, his doctor recommends a diet to make him feel better.

Here, the adverbial epistemic marker *naturally* signals that the doctor is likely to recommend something that is commonly known to reduce weight. Consequently, the use of this marker prompts comprehenders to readily use their world knowledge to predict that the word *diet* will appear at the object position of the sentence and, later on, to integrate the information provided in Sentence 1b in the mental representation of the text content.

Likewise, speakers can use epistemic markers that express a lower degree of certainty to forewarn comprehenders that upcoming propositions are not firmly grounded in commonly available knowledge. Consider, for example, Sentences 1a and 1b with the epistemic marker

naturally replaced by *perhaps*. In that case, the predictability modulation hypothesis implies that encountering such a marker causes comprehenders to abstain from using the knowledge activated by the discourse context for prediction as well as integration of text information.

Epistemic Modality and Validation of Communicated Information

Epistemic markers might also serve as cues guiding processes involved in the validation of information rather than (or in addition to) modulating predictability effects. The concept of validation refers to the ability of comprehenders to check whether communicated information is consistent with world knowledge which is made available by fast and efficient memory-based processes (Richter, Schroeder, & Wöhrmann, 2009; Singer, 2006). Several lines of research suggest that the knowledge-based validation of incoming information is indeed a routine and early component of language comprehension. For example, ERP studies have demonstrated that comprehenders determine the plausibility of three-word scenarios (e.g., *director – bribe – dismissal*) as quickly as they react to semantic relations, as indicated by a congruent pattern of N400 effects (Chwilla & Kolk, 2005). Similarly, Hagoort, Hald, Bastiaansen, and Peterson (2004) found an N400 effect for false (but meaningful) sentences (e.g., *Dutch trains are white*) that was highly similar in time-course and amplitude to the N400 effect for semantic anomalies (e.g., *Dutch trains are sour*). These experiments not only show that pertinent world knowledge is activated early in comprehension but also that comprehenders are quite efficient in judging the validity or plausibility of communicated information based on this knowledge. In addition, there is evidence that they do so automatically even if validation interferes with the focal task. In experiments by Richter et al. (2009), participants judged whether words presented one by one on a computer screen were spelled correctly. Sequences of words formed simple assertions which were either true, i.e. consistent with common world knowledge (e.g., *Cognac contains alcohol*), or false, i.e. inconsistent with common world knowledge (e.g., *Computers have emotions*). Response times for judging the spelling of the last word of each assertion were increased when the task required an affirmative response (i.e., the word was spelled correctly) but the assertion was false. Isberner and Richter (2013) found a similar result for words that rendered stimuli merely implausible rather than false (e.g., *Frank has a broken leg. He calls the plumber*) compared to words that were plausible in their context (e.g., *Frank has a broken pipe. He calls the plumber*). In a similar vein, Isberner and Richter (in press) found that participants were slower to respond to the probe word “TRUE” when it was presented after a false rather than a true sentence, even though validity was

irrelevant to the task of correctly identifying the probe word. Thus, these studies show a Stroop-like interference effect demonstrating that comprehenders routinely monitor the validity of information.

An ERP study by Ferretti, Singer, & Patterson (2008) suggests a similar conclusion for the validation of sentences embedded in larger texts. Participants read stories that contained a target sentence (e.g., *The coach determined/figured that it was oranges that Ken ate*) that was either consistent or inconsistent with the situation established by a sentence provided earlier (e.g., *On this day, it was very hot and Ken and his brother gobbled some oranges/apples*). Inconsistent sentences produced a strong N400 effect following the noun where the inconsistency became apparent. Moreover, when the target sentence contained a factive verb (e.g., *determined*) as opposed to a non-factive verb (e.g., *figured*), the N400 effect produced by the inconsistent nouns lasted longer. Thus, comprehenders validate information they encounter in a discourse context and they seem to do so immediately when a truth value can be computed rather than deferring their judgment until the end of the sentence. In addition, Ferretti et al. (2008) have shown that validation processes are influenced by a specific type of epistemic markers, i.e. factive vs. non-factive mental verbs (for similar results with a reading time paradigm, see Singer, 2006). They also showed that certain combinations of factivity and truth incur pragmatic processing costs, namely the combination of a factive verb with something false (e.g., *Jill knew that the world is flat*) and of a non-factive verb with something true (e.g., *Jill believed that the world is round*).

The epistemic markers investigated in the present study, i.e. adverbial markers signaling the certainty of information, might affect validation processes as well. According to the *alerting hypothesis*, the presence of epistemic adverbs will alert comprehenders to direct their attention to the validity of the propositions asserted in the sentence. Given that unmarked sentences are the default case, both certainty and uncertainty markers may be assumed to exert this effect. Considering the sample Sentence pair 1, the alerting hypothesis would predict that regardless of whether the epistemic markers *naturally* or *perhaps* are used, the presence of an epistemic marker per se will alert comprehenders to scrutinize whether it is really plausible that the doctor recommends a diet. Moreover, in line with the results by Singer (2006) and Ferretti et al. (2008), we might expect that the higher processing costs typically associated with implausible sentences (see also Rayner, Warren, Juhasz, & Liversedge, 2004) would be reduced when these sentences are accompanied by an epistemic marker that signals uncertainty: Consider the following sentence pair:

(2a) Sebastian is a butcher.

(2b) *Perhaps* he sells flowers in his shop.

Here, the adverbial epistemic marker *perhaps* signals that Sebastian selling flowers is possible but not very likely. In other words, the assertion is moved closer to the midpoint of the epistemic scale. As a consequence, the inconsistency with the readers' common knowledge about butchers should be reduced compared to the Sentence pair 2 without the qualification by an epistemic marker.

Rationale of the Present Experiments

We conducted two eye-tracking experiments with short texts such as the one provided in Sentence pairs 1 and 2 to test the predictions of the two accounts. The predictability modulation hypothesis implies that the use of context-triggered knowledge for the prediction and integration of textual information is less likely if an uncertainty marker is present in the sentence. As a consequence, predictability effects which are typically found for eye-tracking measures that indicate early comprehension processes such as first fixation duration and gaze duration (Ehrlich & Rayner, 1981), which arguably reflect prediction as well as integration processes, should be reduced in size. The alerting hypothesis predicts that epistemic markers increase attention to words in a sentence that are crucial for its validity or plausibility, and result in longer eye fixation times on these words. Due to the fact that validation seems to rest on efficient memory-based processes (e.g., Chwilla & Kolk, 2005), alerting effects might occur in eye-tracking measures of early comprehension processes. However, this does not preclude the possibility that alerting effects might also be detectable in later processes such as sentence integration.

We ran two experiments to test these two possible ways in which epistemic markers might influence language processing. The experimental texts in both experiments each comprised (1) a context sentence that introduced a person and (2) a target sentence that ascribed a property to that person or described an action performed by the person which was designated by a single target word. In Experiment 1, the *predictability* of that property or action was manipulated by means of the context sentence (in order to test the predictability modulation hypothesis) whereas in Experiment 2, the *plausibility* of that property or action was manipulated by the context sentence (in order to test the alerting hypothesis). Moreover, the presence of epistemic markers in the target sentence was varied. The primary region of interest was the target word denoting the property or action. However, if evidential markers affect later integrative stages of processing, the effects

might also be observable on sentence-final words (the so-called sentence wrap-up effect, Rayner, Kambe, & Duffy, 2000). For this reason, we included fixation measures for sentence-final words in our analyses as well.

Depending on their position in a sentence, epistemic markers can direct linguistic focus (Rooth, 1992) either to the entire sentence (e.g., *Perhaps she gets flowers*) or only to a specific part of the sentence (e.g., *She gets perhaps flowers*). Linguistic focus is known to influence the depth of semantic processing (Sturt, Sanford, Stewart, & Dawydiak, 2004; Ward & Sturt, 2007) and has been shown to affect the detection of knowledge violations such as semantic illusions (Bredart & Modolo, 1988), suggesting that it may also modulate the hypothesized effects of epistemic markers. Therefore, in Experiment 1, the focus of epistemic markers was included as a control variable; in Experiment 2, it was varied systematically and a condition with non-epistemic adverbial markers (e.g., *presently*) was introduced to disentangle the effects of linguistic focus and the presence of an epistemic marker.

Experiment 1

Method

Participants. Thirty-six undergraduate students of the University of Turku (31 women, 5 men) participated in the experiment for credit in an introductory psychology course. All participants were native speakers of Finnish.

Apparatus. A desk-mount EyeLink 1000 eyetracker (SR Research Ltd., Canada) with a chin-forehead rest was used to collect the eye movements. Eye position was sampled at a rate of 1000 Hz with an average spatial accuracy better than 0.5°. The experimental stimuli were presented on a 21" CRT monitor using a resolution of 1024 x 768 and 150 Hz refresh rate. The viewing distance was approximately 86 cm.

Text material. Experimental materials were 84 short Finnish texts consisting of two sentences (target sentence and preceding context sentence) as in Sentence pair 1. The target sentences always ascribed a property to a person introduced in the context sentence or described an action performed by the person. The focal action or property was designated by one particular word (the target word) that was the third, fourth, or fifth word of the target sentence. The target word was never the final word of the sentence; on average, it was followed by another 3.39 words ($SD = 1.66$; in characters: $M = 22.99$, $SD = 10.97$). There were three versions of each target sentence. One version did not contain an epistemic marker. In the second version, an adverbial

epistemic marker signaling high certainty of the communicated information was placed in the first part of the sentence that preceded the target word (certainty marker: *tietenkin* [certainly], *tietysti* [naturally], or *totta kai* [of course]). In the third version, an adverbial epistemic marker signaling a lower certainty of the communicated information was included (uncertainty marker: *ehkä* [perhaps], *kenties* [maybe], *kaiketi* [probably]). In the versions with an epistemic marker, 1.32 words ($SD = 1.30$; in characters: $M = 7.42$, $SD = 7.35$) intervened on average between the marker and the target word.

For each of the target sentences, two alternative context sentences preceding the target sentence were constructed. One of the two context sentences provided strong situational constraints allowing readers to use their world knowledge to predict the target word in the target sentence (average length in characters: $M = 30.70$, $SD = 10.35$; in words: $M = 4.52$, $SD = 1.46$). The alternative context sentence provided only weak situational constraints allowing no such prediction (average length in characters: $M = 30.18$, $SD = 9.95$; in words: $M = 4.49$, $SD = 1.38$). Here is a sample experimental text for illustrating the manipulation of predictability and epistemic markers (the target word is printed in bold face):

2a) Ossi on vastuuntuntoinen pyöräilijä.

Ossi is a responsible cyclist. (High-constraining context)

2a*) Ossi on luonteeltaan varovainen.

Ossi is a careful person. (Low-constraining context)

2b) (*Tietenkin/Kenties*) hän käyttää aina **kypärää** pyöräillessään.

(*Naturally/Perhaps*) He always wears a **helmet** when he is riding his bike. (No marker, certainty marker, uncertainty marker).

In addition to signaling certainty or uncertainty, the epistemic markers also differed in their focus. They could either refer to the target word or the noun phrase which the target word was part of (49%), or they could refer to the entire target sentence (51%). None of our hypotheses referred to the focus of the evidential marker but effects of this variable on eye movements were explored in additional analyses.

Norming of materials. The six epistemic markers used in the experiment were selected on the basis of a norming study with 15 university students who rated for 25 different adverbial markers how certain a speaker using the expression would feel about the information that he or she is communicating. The three certainty markers were consistently rated to signal a high certainty of

the communicated information ($M = 84$, $SD = 16$ on a scale from 0-100). In contrast, the three uncertainty markers received consistently low certainty ratings ($M = 28$, $SD = 15$), $t(14) = 15.49$, $p < .001$, $d = 3.61$.

The predictability of the target words in the two versions of the experimental texts (without the epistemic markers) was tested in a web-based norming study with a total of 278 student participants and 240 different texts. The study was based on a cloze completion task. Participants received a subset of a total of 8 lists that contained 30 texts presented in random order. They could complete up to 4 lists. The lists were assigned to participants in a way that each participant never received the high-constraining and the low-constraining version of one text. On average, 35 ratings were collected for each text version. For the experiment, we selected the 84 texts with the greatest predictability differences between the high-constraining and the low-constraining versions. For the selected texts, the mean predictability of the target word was .66 ($SD = .13$) in the predictable context and .09 ($SD = .10$) in the unpredictable context, respectively. The predictability difference was significant, $t(83) = 34.60$, $p < .001$, $d = 4.91$.

Procedure. The experimental session started with calibrating the eye-tracker with a 9-point calibration grid that covered the entire computer screen. The 84 experimental texts were presented one-by-one in black letters (font Courier New 20; each letter subtended approximately 0.53° of vertical visual angle), double-spaced on light gray background in random order on the computer screen. Four practice trials preceded the actual experimental trials. Prior to each trial, a fixation point was presented on the lower left corner of the area where the text appeared to check the calibration. When the calibration was no longer accurate (i.e., when the fixation did not fall on or very close to the fixation point), the calibration procedure was repeated as often as necessary to achieve an average angular error of less than 0.5° . Participants moved on to the next text by pressing a key on a response box. Participants were instructed to read the texts for comprehension. To make sure that participants actually followed this instruction, a comprehension question that required participants to respond with yes or no by pressing one of two buttons on the response box was presented on average after every five experimental texts.

Design. The experimental design was a 3 (*epistemic marker*: certainty marker, uncertainty marker, no marker) X 2 (*predictability*: high-constraining context vs. low-constraining context) within-subjects design. Assignment of target sentences to experimental conditions was counterbalanced across participants by means of six different items lists. For exploratory purposes,

the focus of the epistemic marker (entire sentence vs. noun/noun phrase) was included as additional independent variable.

Results

Comprehension questions. Participants answered on average 97.2 % ($SD = 1.7$ %) of the comprehension questions correctly, with no participant scoring less than 85 %.

Eye-movement measures. In order to get a comprehensive picture of how epistemic markers affect on-line sentence comprehension, we included five different fixation measures in our analyses. These measures referred to fixations on or originating from the target word, i.e. the noun in the object position of the second sentence where knowledge-based predictability effects should become apparent. The target word was also a crucial location for computing the plausibility of the target sentence.

First fixation duration (the duration of the first fixation on a word) and *gaze duration* (the sum of all fixations made on a word during first-pass reading) are usually regarded as indicative of early processes in reading comprehension. These measures reflect the amount of attention that a word receives when it is first encountered during on-line sentence comprehension. In contrast, total fixation duration, probability of regressions to the target word and regression path duration can be interpreted as indicative of rather late comprehension processes. *Total fixation duration* is the sum of all fixations that a participant made on the word, including second-pass reading and regressions back to it. Similar to this measure, the *probability of regressions to the target word* after the eyes have moved to later parts of the sentence reflect the attention that the target word receives in integrative processing that occurs late in sentence comprehension. *Regression path duration* includes gaze duration, the time spent on earlier parts of the sentence after regressive eye-movements following first-pass reading, and the duration of refixations to the target word before the eyes move on to words beyond the target word. Accordingly, this measure captures late cognitive processes that integrate the target word with earlier parts of the sentence.

In order to examine the possibility that evidential markers influence the later integrative processing of sentences (as observed as an increased sentence wrap-up effect), we also analyzed gaze duration, total fixation time, and regression path duration on sentence-final words.

Analyses. To test our hypotheses, we conducted a linear mixed models (LMM) analysis for all fixation durations and a generalized linear mixed models (GLMM) analysis with logit link for proportions (e.g., the probability of regressions). These analyses were conducted using the lmer

command of the lme4 package for R (Bates et al., 2011). Subjects and items were included as random factors to account for the fact that both subjects and items represent samples of larger populations. Unlike F_1 - and F_2 -ANOVAs, this type of analysis allows for an adequate and stringent test of the hypothesized effects in a single model (for further discussion, see Baayen et al., 2008). As it is still unclear how the degrees of freedom for the t -values of the LMM analyses should be derived, no degrees of freedom will be reported for these analyses. However, the large number of observations in the present experiment (items times participants) makes it safe to assume that the distribution of t -values approximates the standard normal distribution (z -distribution; see Baayen et al., 2008, Note 1). Therefore, the standard normal distribution was assumed for significance tests of fixed effects in the LMM analysis. A type-I-error probability α of .05 was chosen for all hypothesis tests.

Predictability modulation hypothesis. For testing the hypothesis that epistemic markers modulate knowledge-based prediction (predictability modulation hypothesis), the independent variable manipulating the presence of an epistemic marker was decomposed into two orthogonal contrasts. One contrast, coding the *presence of an epistemic marker*, compared the mean of the conditions with a certainty or uncertainty marker (coded with 1) to the conditions without a marker (coded with -2). The other contrast, coding the *type of epistemic marker*, compared the condition with a certainty marker (1) to the condition with an uncertainty marker (-1). The second independent variable (predictability) was contrast-coded as well (1 = high-constraining context vs. -1 = low-constraining context). In addition, the two variables were allowed to interact in the model. The hypothesis that the presence of uncertainty markers prompts comprehenders to refrain from knowledge-based prediction implies that the interaction of the variable that codes the type of epistemic marker (certainty vs. uncertainty marker) with predictability should be significant, with a weaker predictability effect when the target sentence contains an uncertainty marker. Given that predictability affects early stages of word processing, modulating effects of epistemic markers should become apparent in eye-fixation measures that indicate early rather than late comprehension processes.

Alerting hypothesis. If epistemic markers serve a general alerting function, there should be a significant main effect of the variable coding the presence of an epistemic marker (marker vs. no marker), with longer fixation times on the target word in sentences with an epistemic marker. In principle, alerting effects of epistemic markers could occur with fixation measures that indicate

early as well as those that indicate late processes (including integrative processing on sentence-final words).

Linguistic focus. For exploring whether effects of epistemic markers depend on their focus (noun phrase or sentence), the focus of epistemic markers was included as another contrast-coded predictor (1 = focus on target phrase, -1 = focus on target sentence). In addition, the interactions of linguistic focus with predictability and with the type of epistemic marker, as well as with the interaction term of the latter two variables, were included in the model.

Follow-up analyses with dummy-coded predictors. To follow up on significant interactions, we tested for simple main effects by decomposing the contrast-coded predictors into dummy-coded variables so that the effects of the remaining variables were estimated on each level of the experimental factors (see Aiken & West, 1991).

Exclusion of trials with skipped markers. As we were interested in effects of the presence of an epistemic marker, we excluded all trials in which the marker (if a marker was present) was skipped in first-pass reading (6.21 % of all trials) from the analysis.

Fixation measures on target words.

First fixation duration. Target words predictable from the context sentence received shorter first fixations compared to target words that were not predictable, $t = -1.99$, $p < .05$ (Figure 1a). Neither of the interactions of predictability with the two predictors coding the presence and type of epistemic markers reached significance (both $t \leq |1.62|$). Thus, there was a predictability effect but it was not moderated by the presence or type of epistemic markers. However, in line with the alerting hypothesis, first fixations on the target word were generally longer when an epistemic marker was present compared to when no epistemic marker was present, $t = 2.23$, $p < .05$. First fixation durations did not differ between sentences with certainty and uncertainty markers, $t = -0.24$, $p > .05$. Linguistic focus neither exerted a main effect, $t = -0.29$, $p > .05$, nor was it involved in any significant interactions (all $|t| \leq |1.81|$).

In sum, the first fixation durations provided partial support for the alerting hypothesis. The presence of epistemic markers generally prolonged first fixations on the target word, an indicator of comprehension processes occurring relatively early on the target word. In addition, there was a facilitating effect of predictability on first fixation durations but this effect seemed to be independent of the presence or type of epistemic markers.

Gaze duration. Similar to first fixations, gaze durations on predictable target words were shorter compared to unpredictable target words, $t = -2.26$, $p < .05$ (Figure 1b). Again, the

predictability effect was not modulated by the presence or type of epistemic markers, all $|t| \leq |1.46|$, which is inconsistent with the predictability modulation hypothesis. Like for first fixation durations, the presence of epistemic markers tended to generally increase gaze durations, but this time the effect fell short of significance, $t = 1.53$, $p > .05$. Again, there were no differences between certainty and uncertainty markers, $t = 0.91$, $p > .05$, and no main effect or interactions of linguistic focus, all $|t| \leq |1.36|$.

Probability of skips of the target word. As discussed earlier, predictability may also affect the probability of skips of the target word. However, the analyses revealed that there were no skips of target words in Experiment 1, which may be due to the fact that the target words in the Finnish stimuli were rather long ($M = 7.98$ characters, $SD = 2.40$ characters) and therefore less likely to be skipped.

Total fixation duration. Total fixation durations (Figure 1c) on the target words showed a predictability effect as well: Fixations on target words that were predictable from the context sentence were shorter overall than fixations on target words that were not predictable, $t = -2.50$, $p < .05$. Again, there was no indication of a modulation of this predictability effect by the presence or type of epistemic markers, as none of the interactions reached statistical significance, all $|t| \leq |1.76|$. Similar to the results for first fixation duration and gaze duration, target words in sentences with epistemic markers tended to receive overall longer fixations compared to target words in sentences with no epistemic markers, but again, this effect fell short of significance, $t = 1.87$, $p > .05$. The type of epistemic marker did not significantly influence total fixation durations, $t = 1.25$, $p > .05$. There was no main effect of linguistic focus and no interactions involving this variable, all $|t| \leq |1.66|$.

In sum, the data on total fixation duration, an indicator of rather late comprehension processes, were similar to those for the two indicators of early comprehension processes: Despite the fact that there was an overall facilitating effect of predictability, no evidence was found for the predictability modulation hypothesis. Rather, the results tend to support the alerting hypothesis, with target words in sentences containing epistemic markers generally receiving longer fixations, although this effect only reached significance for first fixation durations.

Regression path duration. There was no main effect of predictability on regression path duration (Figure 1d), $t = -0.84$, $p > .05$, and none of the interactions involving predictability reached significance, all $|t| \leq |1.82|$. Durations of regression paths originating from the target word tended to be longer in sentences with epistemic markers compared to sentences that did not contain an epistemic marker, albeit not significantly so, $t = 1.89$, $p > .05$. Again, no differences were found

between certainty and uncertainty markers, $t = 0.50$, $p > .05$, and no significant effects involving linguistic focus, all $|t| \leq |-1.82|$.

Probability of regressions to the target word. The probability of regressions to the target word (Figure 1e) was lower when the target word was predictable from the context sentence compared to when it was not predictable, $z = -2.14$, $p < .05$. The predictability effect was not modulated by the presence of an epistemic marker (for all interactions: $|z| \leq |-1.18|$, $p > .24$), nor did the presence of an epistemic marker exert a significant main effect, $z = 0.74$, $p = .46$. There was a numerical tendency for uncertainty markers inducing a lower probability of regressions than certainty markers, but this difference was nonsignificant, $z = 1.66$, $p = .10$. Again, there were no significant effects involving linguistic focus, all $|z| \leq |-1.19|$, $p > .24$.

Fixation measures on sentence-final words. In order to explore whether effects of predictability and the presence and type of epistemic markers extend to integrative processes associated with sentence-wrap up, we ran three additional mixed models with gaze duration, total fixation duration, and regression path duration of the sentence-final words as dependent variables. However, these analyses did not yield any significant main effects or interactions of any of the predictors, all $|t| \leq |1.51|$.

Discussion

We found predictability effects on all early and late fixation measures of the target word, except for regression path duration. The predictability effects were unaffected by the presence or type of epistemic markers. In contrast, the presence of epistemic markers affected the duration of first fixations independent of predictability. In comparison to the condition without epistemic markers, certainty as well as uncertainty markers caused longer first fixations on the target word. Both types of epistemic markers also caused numerically longer gaze durations, total fixation durations, and regression path durations, although these effects fell short of significance.

Importantly, we found no evidence for differential effects of certainty and uncertainty markers, suggesting that prediction and epistemic markers exert independent effects. Moreover, we found no evidence for a modulation of the effects of epistemic markers by whether they direct linguistic focus to the target phrase or to the whole sentence.

In summary, these results speak against a modulation of predictability effects by epistemic markers but they suggest that epistemic markers generally increase allocation of processing resources to the target word, which is in line with the alerting hypothesis. However,

the interpretability of the results in terms of the alerting hypothesis is limited for two reasons: First, it could be that it is the presence of an adverbial marker in general – rather than the presence of an epistemic marker – that causes the increase in fixation durations on the target word. Second, it is unclear what exactly this increase reflects. The assumption of the alerting hypothesis is that it reflects a modulation of validation processes in terms of heightened epistemic vigilance (Sperber et al., 2010), with more resources allocated to checking the validity or plausibility of the linguistic input. However, given that this experiment did not manipulate plausibility (which cannot be varied orthogonally with predictability) it does not allow any sound conclusions regarding the modulation of validation processes. Thus, to clarify these issues, we ran a follow-up experiment with two modifications: First, to disentangle the effects of epistemic markers and linguistic focus, the condition with certainty markers was replaced by a new baseline condition with non-epistemic adverbial markers. Second, to investigate whether it is indeed validation that is affected by the presence of epistemic markers, plausibility was systematically varied to explore its interactions with the presence of epistemic (uncertainty) markers.

Experiment 2

Method

Participants. Fifty-five undergraduate students of the University of Kassel participated in the experiment for course credit. Data were excluded from the analysis for eight participants who were not native speakers of German, seven participants who failed to produce accurate measurements and for two participants due to technical problems. All of the remaining 38 participants (30 women, 8 men) were native speakers of German, with an average age of 23.9 ($SD = 6.3$) years.

Apparatus. The technical specifications of the eye tracker (EyeLink 1000) were identical to the one used in Experiment 1. The experimental stimuli were presented on a 22" TFT monitor using a resolution of 1024 x 768 and 75 Hz refresh rate. The viewing distance was approximately 86 cm.

Text material. Experimental materials were 120 short German texts which consisted of three sentences each (context sentence, target sentence, final sentence). As in Experiment 1, the target sentences always ascribed a property to a person introduced in the context sentence or described an action performed by that person. The focal action or property was designated by one particular word (the target word) which was never the final word of the sentence; on average, the

target word was followed by another 3.64 words ($SD = 1.69$; in characters: $M = 18.91$, $SD = 8.28$). Target words had an average length of 6.85 characters ($SD = 2.41$) and an average frequency of 330.86 ($SD = 490.78$) according to the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995; values are based on the 114 of the 120 target words for which frequency data were available). There were five versions of each target sentence. One version did not contain any adverbial marker. Two versions contained an adverbial epistemic marker signaling uncertainty of the communicated information (*eventuell* [potentially], *möglicherweise* [possibly], *vielleicht* [maybe], *womöglich* [perhaps]). In one of these two versions, the epistemic marker was placed at the beginning of the sentence to direct linguistic focus to the whole sentence, whereas in the other version, the epistemic marker was placed right before the phrase containing the target word, thus directing focus to the phrase. Finally, there were two versions containing temporal adverbials that were neutral regarding the epistemic status of the communicated information (e.g., *gegenwärtig* [currently], *mittlerweile* [these days], *momentan* [presently], *gestern* [yesterday]); again, in one of these two versions the marker was placed at the beginning of the sentence and in the other it was placed right before the phrase containing the target. In versions with the (epistemic or neutral) marker placed at the beginning of the sentence, on average 3.52 words ($SD = 1.26$; in characters: $M = 14.34$, $SD = 5.99$) intervened between the marker and the target word; in versions with the marker placed right before the target phrase, this distance was on average 0.99 words ($SD = 0.82$; in characters: $M = 4.16$, $SD = 3.78$).

As in Experiment 1, for each of the target sentences, two alternative context sentences preceding the target sentence were constructed. One of the two context sentences rendered the target sentence implausible (average length in characters: $M = 28.36$, $SD = 9.19$; in words: $M = 5.06$, $SD = 1.85$), whereas the other rendered it plausible (average length in characters: $M = 29.13$, $SD = 7.87$; in words: $M = 5.11$, $SD = 1.54$), without being highly constraining in order to keep the degree of constraint comparable between plausible and implausible targets.

To avoid inducing unnatural reading strategies by repeatedly exposing participants to implausible material, we added a third and final sentence which was supposed to cancel out the implausibility of the target by providing an explanation for the trait or action described in the target sentence, but also made sense in combination with the plausible context.

Below is a sample experimental text for illustrating the manipulations of plausibility, presence of an uncertainty marker, and linguistic focus (the target word is printed in bold face, the adverbial marker in italics):

1a) Carmen hat eine Menge Begabungen.

Carmen has many talents. (Plausible context)

1b) Carmen hat eine krächzende Stimme.

Carmen has a scratchy voice. (Implausible context)

2a) Sie arbeitet als **Sängerin** in einer Band.

She works as a **singer** in a band. (No marker)

2b) *Womöglich/Mittlerweile* arbeitet sie als **Sängerin** in einer Band.

Perhaps/These days she works as a **singer** in a band. (Uncertainty/neutral marker, focus on sentence)

2c) Sie arbeitet *womöglich/mittlerweile* als **Sängerin** in einer Band.

She works *perhaps/these days* as a **singer** in a band. (Uncertainty/neutral marker, focus on phrase)

3) Ihre ungewöhnliche Stimme ist sehr gefragt.

Her unusual voice is very sought-after. (Final sentence cancelling out implausibility)

Norming of materials. The plausibility of the target sentences (without the adverbial markers) combined with the two versions of the context sentence was tested in a web-based norming study with 22 student participants. Each participant saw each target sentence paired with either the plausible or the implausible context (counterbalanced via item lists) and was asked to rate the plausibility of the content of the sentence pair on a 7-point-sliding-scale. The extremes of the scale were marked as “very implausible” (1) and “very plausible” (7), with the midpoint (4) labelled as “neither plausible nor implausible”. All 120 items were presented in random order. Combinations of target sentences with the plausible context received considerably higher plausibility ratings ($M = 5.78$, $SD = 0.59$) than combinations with the implausible context ($M = 3.07$, $SD = 0.73$), $t(21) = 22.66$, $p < .001$, $d = 4.08$.

Procedure. The experimental session started with calibrating the eye-tracker with a 9-point calibration grid that covered the entire computer screen. The 120 experimental texts were presented one-by-one in black letters (font Courier New in size 20; each letter subtended approximately 0.53° of vertical visual angle) with 2.5 line spacing on a light gray background in random order on the computer screen. Each of the three sentences of each text began in a new line; this was done to avoid the target word being the first or last word in a line. Four practice trials

preceded the actual experimental trials. Prior to each trial, a fixation point was presented on the upper left corner of the area where the text appeared to check the calibration. As in Experiment 1, recalibration was performed whenever the calibration appeared to be no longer accurate and as often as necessary to obtain an average angular error of less than 0.5°. Calibrations were accepted at an angular error of less than 0.5 degrees. After reading a text, participants moved on to the next text by pressing the spacebar on the keyboard. They were instructed to read the texts for comprehension. To make sure that participants actually followed this instruction, a comprehension question that required them to respond with yes or no on the keyboard was presented after a third of the experimental texts.

Design. The experimental design was a 2 (*plausibility*: plausible vs. implausible) X 3 (*presence of an uncertainty marker*: uncertainty marker vs. neutral marker vs. no marker) within-subjects design, with a third factor *linguistic focus* nested within the factor *presence of an uncertainty marker*: In sentences containing an uncertainty or a neutral marker, the *linguistic focus* (focus on target sentence vs. focus on target phrase) was varied by placing the marker either at the beginning of the sentence or right before the target phrase. This nested design resulted in ten experimental conditions. Assignment of target sentences to experimental conditions was counterbalanced across participants by means of ten different items lists.

Results

Comprehension questions. Participants answered on average 94.4 % ($SD = 4.5$ %) of the comprehension questions correctly, with no participant scoring less than 85 %.

Eye-movement measures. We included six different fixation measures in our analyses, three of which arguably reflect early processes (*first fixation duration*, *gaze duration* and *probability of skips*) and three of which are usually considered indicators of later integrative processes (*total fixation duration*, *probability of regressions* and *regression path duration*). All of these measures referred to fixations on or originating from the target word, i.e., the word in the target sentence which determined its plausibility or implausibility in combination with the context, and thus where effects of validation should become apparent. According to our assumption of nonstrategic validation, we expected disruptive effects of implausible targets on both early and late processes; however, we expected these effects to become reduced or even reversed when an epistemic marker was present that signaled uncertainty of the target information.

As further indicators of late integrative processing, we also analyzed gaze duration, total fixation duration, regression path duration, and probability of regressions to sentence-final words, as well as the probability of regressions to and total fixation duration on the context sentence, which determined the plausibility or implausibility of the target sentence.

As we were interested in effects of the presence of an epistemic marker, we excluded all trials in which the adverbial marker (epistemic or neutral) was skipped in first-pass reading (11.38 % of all trials) from the analyses.

Hypothesis tests and coding of predictors. As in Experiment 1, we conducted an LMM analysis for all fixation durations and a GLMM analysis with logit link for proportions, with subjects and items included as random factors. The standard normal distribution was assumed for significance tests of fixed effects in the LMM analysis.

Effects of plausibility. For testing the assumption of plausibility effects on early and late fixation measures, we entered plausibility into the model as a contrast-coded predictor (1 = plausible, -1 = implausible).

Modulation of plausibility effects by uncertainty markers. For testing the hypothesis that uncertainty markers modulate plausibility effects, the independent variable manipulating the presence of an uncertainty marker was decomposed into two orthogonal contrasts. First, the condition with an uncertainty marker (2) was contrasted with the mean of the conditions with a neutral marker (-1) or no marker (-1). Second, the condition with a neutral marker (1) was contrasted with the condition with no marker (-1). The hypothesis that uncertainty markers modulate effects of plausibility implies a significant interaction of plausibility with the first contrast (coding *presence of an uncertainty marker*). In principle, modulating effects of uncertainty markers could occur in fixation measures that indicate early as well as those that indicate late processes, including integrative processing on sentence-final words and reprocessing of the context sentence. The second contrast (*presence of a neutral marker*) was set to examine whether the effects of markers are specific to the evidential markers (in this experiment, uncertainty markers).

Linguistic focus. For investigating whether effects of uncertainty markers depend on whether they direct focus to the whole sentence or to the target phrase, the focus of the adverbial marker was included an additional contrast-coded predictor (1 = focus on target phrase, -1 = focus on target sentence) and allowed to interact with plausibility and the predictor coding *presence of an uncertainty marker*.

Follow-up analyses with dummy-coded predictors. As in Experiment 1, we followed up on significant interactions by decomposing the contrast-coded predictors into dummy-coded variables to estimate the effects of the remaining variables on each level of the experimental factors (Aiken & West, 1991).

Experimental half as additional predictor. In the initial analysis, we did not find any effects of plausibility on early fixation measures, which is surprising given that such effects have been reported in the literature (e.g., Murray & Rowan, 1998; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007). However, visual inspection of the data revealed that some data patterns actually became reversed from the first to the second half of the experiment. An explanation for this may be found in the nature of our stimulus material: To avoid inducing unnatural reading strategies by presenting participants repeatedly with implausible stimuli, we always included a final sentence which provided a reason for the action described in the target sentence, thus retroactively reducing its implausibility in the implausible condition (see the example provided in the section *Text Material*). However, from our post-experiment questionnaires, it is obvious that participants noticed this particularity of the material and probably adjusted their expectations and reading strategies accordingly. Thus, it seems likely that they no longer perceived implausible targets as quite that implausible because they realized that an explanation would always be provided for each implausible trait or behavior in the final sentence. For this reason, we reran our analyses including experimental half as an additional contrast-coded predictor (-1 = first half, 1 = second half). In line with our observation that the data patterns changed significantly from the first to the second half, we found several interactions with experimental half, which justifies our approach to include it in the model. In the following, we will therefore only report the results of the analyses including experimental half as an additional predictor.

Analyses of fixations on or originating from the target word.

First fixation duration. In first fixation durations on target words, we found a three-way interaction of plausibility, presence of an uncertainty marker, and experimental half, $t = -2.25$, $p < .05$. We followed up on this result by running separate analyses for the first (Figure 2a) and second half (Figure 2b) of the experiment. These analyses revealed that the interaction of plausibility and presence of an uncertainty marker, which indicates early effects of both variables on reading comprehension, was only significant in the first half of the experiment, $t = 1.96$, $p < .05$, but not in the second half, $t = -1.23$, $p > .05$. The interaction in the first half relied on the fact that in sentences with no marker or a neutral marker, implausible target words received longer first fixations than

plausible target words, $t = -2.11$, $p < .05$, whereas this was not the case in sentences containing an uncertainty marker, $t = 1.03$, $p > .05$. For these sentences, there was actually a trend in the opposite direction, with plausible targets receiving longer first fixations than implausible targets. The data pattern indicates that, in line with the alerting hypothesis, uncertainty markers increased early processing, but this was only the case for plausible targets.

In addition, we found a main effect of focus, $t = 3.02$, $p < .05$, with longer first fixations when the focus was on the target phrase rather than on the whole sentence, and of experimental half, $t = 2.07$, $p < .05$, with longer first fixations in the second half than in the first half of the experiment.

Gaze duration. In gaze duration, there was again a significant three-way interaction of plausibility, presence of an uncertainty marker, and experimental half, $t = -2.01$, $p < .05$. We followed up on this result by running separate analyses for the first (Figure 2c) and second half (Figure 2d) of the experiment. This time, however, the interaction of plausibility and presence of an uncertainty marker was neither significant in the first ($t = 1.39$, $p > .05$) nor in the second half ($t = -1.48$, $p > .05$), although the pattern was similar to that found for first fixation duration. However, there was a significant effect of the presence of an uncertainty marker in the first half, $t = 2.27$, $p < .05$, with uncertainty markers inducing longer gaze durations than neutral or absent markers. In addition, there was a main effect of the variable coding the difference between conditions with a neutral vs. no marker across both halves, $t = 2.21$, $p < .05$, which was due to longer gaze durations on the target in target sentences with a neutral marker as compared to target sentences without a marker.

Probability of skips of the target word. For probability of skips (Figure 2e), we found an interaction of plausibility and the presence of an uncertainty marker, $z = -2.05$, $p < .05$. This was due to the fact that there were significantly less skips of plausible targets when an uncertainty marker was present than when there was a neutral or absent marker, $z = -2.27$, $p < .05$. For implausible targets, there was a small trend in the opposite direction, but it was not significant, $z = 0.62$, $p > .05$.

Total fixation duration. In total fixation duration, we found a significant interaction of plausibility with experimental half, $t = 2.33$, $p < .05$. Follow-up analyses for each half revealed that there was a main effect of plausibility in the first half (Figure 2f), $t = -2.68$, $p < .05$, with longer total fixations on implausible targets, but this effect was no longer present in the second half (Figure 2g), $t = 0.64$, $p > .05$.

Moreover, the variable coding the difference between conditions with a neutral versus no marker also interacted with experimental half, $t = 2.09$, $p < .05$. This interaction was due to the fact that these conditions only differed significantly in the first half, $t = -2.61$, $p < .05$, with longer total fixation durations on the target in sentences with no marker than in sentences with a neutral marker. There was also a main effect of focus, $t = 2.21$, $p < .05$, which was due to focus on the target phrase increasing total fixation duration on the target. Finally, there was a strong negative effect of experimental half, $t = -5.87$, $p < .05$, with overall shorter total fixation durations on the target in the second half of the experiment.

Regression path duration. In regression path duration (Figure 2h), we found a main effect of plausibility, $t = -2.24$, $p < .05$, with longer regression path durations for implausible as compared to plausible targets. However, this effect was modulated by the presence of an uncertainty marker, $t = 3.09$, $p < .05$. Separate analyses for conditions with and without uncertainty markers revealed that the plausibility effect was only present in conditions without an uncertainty marker, $t = -2.90$, $p < .05$, but not in conditions with an uncertainty marker, $t = 1.23$, $p > .05$. Thus, the uncertainty marker appears to have reduced integrative processing of implausible information.

Moreover, there was a main effect of focus, $t = 1.98$, $p < .05$, which was due to longer regression path durations when focus was directed to the phrase rather than to the sentence. There was also a main effect of the variable coding the difference between conditions with neutral versus no markers, $t = -3.55$, $p < .05$, with shorter regression path durations for sentences with neutral markers. However, this effect was modulated by experimental half, $t = 3.09$, $p < .05$, due to the fact that it was only significant in the first half of the experiment, $t = -4.74$, $p < .05$, but not in the second half, $t = -0.31$, $p > .05$.

Finally, there was again a main effect of experimental half, $t = -3.67$, $p < .05$, with shorter overall regression path durations in the second half of the experiment.

Probability of regressions to the target word. For probability of regressions to the target word (Figure 2i), there was a main effect of the presence of an uncertainty marker, $z = -4.28$, $p < .001$. Target words in sentences containing an uncertainty marker had a lower probability of receiving regressions than sentences containing no marker or a neutral marker. This finding suggests that information that is marked as uncertain receives less attention in later integrative processing. The probability of regressions to the target word also differed between sentences with a neutral marker versus no marker, $z = -2.30$, $p < .05$, with sentences containing a neutral marker having a lower probability of receiving regressions than sentences containing no marker. Finally,

there was a main effect of experimental half, $z = -4.85$, $p < .001$, with an overall lower probability of regressions to the target word in the second than in the first half.

Analyses of fixations on or originating from sentence-final words. Because plausibility and uncertainty markers may also affect sentence wrap-up, we analyzed fixations on or originating from sentence-final words as well.

Gaze duration. In gaze duration on final words, there was only a main effect of experimental half, $t = -3.08$, $p < .05$, with longer durations in the first than in the second half.

Total fixation duration. In total fixation duration on final words, similarly, we only found a main effect of experimental half, $t = -6.10$, $p < .05$, with longer durations in the first than in the second half.

Regression path duration. In regression path duration on final words, we found a main effect of plausibility, $t = -2.09$, $p < .05$. Regression path durations were longer for implausible than for plausible target sentences. Moreover, there was again a main effect of experimental half, $t = -4.25$, $p < .05$, with longer regression path durations in the first than in the second half.

Probability of regressions to sentence-final words. As for the probability of regressions to the sentence-final words, plausibility interacted with the presence of an uncertainty marker, $z = -2.01$, $p < .05$. This interaction was due to the fact that for final words of implausible target sentences, there were more regressions to the final word when there was an uncertainty marker rather than a neutral or no marker, whereas for plausible target sentences, there was a higher probability of regressions to the final word when there was a neutral or no marker than when there was an uncertainty marker. However, none of the simple main effects reached significance, all $|z| \leq |1.88|$, $p > .05$.

As in the other measures, there was again a main effect of experimental half, $z = -3.49$, $p < .001$, with a lower probability of regressions to sentence-final words in the second half.

Analyses of fixations on the context sentence. Because implausibility may result in more rereading of previous text passages, we also analyzed fixations on the context sentence. However, we only analyzed measures which could reveal effects of the plausibility of the target sentence, i.e., total fixation duration and probability of regressions to the context sentence.

Total fixation duration. Plausibility exerted a main effect on total fixation duration on the context sentence, $t = -2.34$, $p < .05$. Total fixation durations were longer for stories containing an implausible target sentence as compared to stories containing a plausible target sentence. We also found a three-way interaction of plausibility, presence of an uncertainty marker, and linguistic

focus, $t = 2.11$, $p < .05$. We followed up on this result by running separate analyses for target sentences containing an uncertainty marker and target sentences containing a neutral or no marker. These analyses revealed that there were no effects for conditions with an uncertainty marker, while in conditions containing no marker or a neutral marker, there was both a significant main effect of plausibility, $t = -2.15$, $p < .05$, with longer total fixation durations on context sentences rendering the target sentence implausible, and a significant interaction of plausibility and focus, $t = -2.32$, $p < .05$. This interaction was due to the plausibility effect being significant only when focus was on the target phrase, $t = -2.52$, $p < .05$, rather than on the whole sentence, $t = 0.89$, $p > .05$.

Finally, we again found a main effect of experimental half, $t = -9.40$, $p < .05$, with longer total fixation durations in the first than in the second half.

Probability of regressions to the context sentence. There was a main effect of plausibility on the probability of regressions to the context sentence, $z = -2.52$, $p < .05$, with a higher probability of regressions when the target sentence was implausible than when it was plausible. Moreover, there was again a main effect of experimental half, $z = -7.85$, $p < .001$, with a lower overall probability of regressions in the second half.

Discussion

We found effects of plausibility and the presence of an uncertainty marker (including interactive effects) on both early and late fixation measures on target words. Especially the early effects, however, appeared to be sensitive to reading strategies developed over the course of the experiment, as they disappeared from the first to the second half of the experiment. Moreover, the effects of uncertainty markers on early processing were evident primarily for plausible targets whereas the effects of uncertainty markers on late processing were found primarily for implausible targets: They enhanced early processing of plausible targets (increasing first fixation durations and reducing skips) and reduced late processing of implausible targets (namely, reducing regression path duration). Effects of plausibility and uncertainty markers were also evident in fixation measures on sentence final words and in measures indicating reprocessing of the context sentence, suggesting that these manipulations also affected sentence wrap-up and integration across sentences.

Similar to Experiment 1, Experiment 2 showed little influence of linguistic focus on the effects of the presence of epistemic markers. Rather, linguistic focus seemed to increase processing of the target word independent of whether the marker was epistemic or not. The only

interaction effect of focus we found was with plausibility on total fixation durations on the context sentence: When there was a neutral or no marker, implausibility led to longer total fixation durations on the context sentence than plausibility but only when the focus was on the target phrase, suggesting that focus can also affect relatively late cross-sentence integrative processing.

Consistent with Experiment 1, these results provide evidence for the alerting hypothesis, which assumes that epistemic markers alert readers to process target words more thoroughly. However, they also extend the previous findings by showing that this seems to apply to plausible targets only. It is possible that the increased processing of plausible targets reflects pragmatic costs of something plausible being described as uncertain, comparable to the pragmatic costs of somebody only believing or figuring something true, as reported by Singer (2006) and Ferretti et al. (2008). Alternatively, it may indicate reduced reliance on prior knowledge when processing information that is marked as uncertain, thus slowing down the otherwise fast integration of plausible information. For implausible targets, in contrast, there seemed to be no such cost or even a slight facilitation in early measures, as well a clear reduction of later integrative processing, as indicated by shorter regression path durations. In addition, uncertainty markers reduced the overall probability of regressions to the target, suggesting that information marked as uncertain is either easier to integrate or it is given less weight in integrative processing.

General Discussion

Two eye-tracking experiments tested the assumptions that epistemic markers such as *perhaps* or *certainly* modulate effects of predictability, plausibility, or both. The aim of Experiment 1 was to investigate the role of adverbial markers of epistemic modality in predictability effects during sentence comprehension. To this end, we monitored participants' eye movements while they were reading declarative sentences preceded by a context sentence that either constrained the meaning of a target word or not. The sentences contained an epistemic adverb signaling a high degree of certainty of the speaker (such as *naturally*), an epistemic adverb signaling a low degree of certainty (such as *perhaps*), or no epistemic marker at all. Clear effects of predictability were found on almost all fixation measures, from early measures which arguably reflect prediction as well as integration processes to late measures which likely reflect integration processes only. By contrast, there was no evidence for the hypothesis that adverbial epistemic markers influence the use of knowledge for the prediction of upcoming words. However, both kinds of markers increased first fixation durations on the target word, which is in line with the

notion that epistemic markers serve an alerting function. Based on these results, Experiment 2 was designed to directly test the hypothesis that epistemic (uncertainty) markers alert comprehenders to scrutinize the plausibility of subsequently presented information more thoroughly. This time, the condition with certainty markers was replaced by a control condition including adverbial markers that were neutral with regard to the epistemic status of the target information. Moreover, plausibility of the target information was systematically manipulated by the context sentence. Analyses yielded both main and interactive effects of plausibility and the presence of uncertainty markers on indicators of both early and late comprehension processes.

The fact that we found early effects of plausibility supports the assumption of routine epistemic validation (Richter et al., 2009; Singer, 2006). This finding coheres well with the growing body of research that accounts for the routine and early character of validation in language comprehension (e.g., Chwilla & Kolk, 2005; Haagort et al., 2004; Isberner & Richter, 2013, in press; Richter et al., 2009; for an overview, see Singer, in press). At first sight, it seems to be at variance with earlier eye-tracking experiments that found effects of semantic anomaly on indicators of early comprehension processes, but not of implausibility (e.g., Rayner, Warren, Juhasz, & Liversedge, 2004). However, in contrast to the implausible sentences used by Rayner et al. which described untypical situations, the implausible sentences used in this study were rendered implausible by a context sentence but all described situations and events typical of readers' experience. A high typicality seems to be crucial for early plausibility effects to occur (Matsuki, Chow, Hare, Elman, Scheepers, & McRae, 2011), so it might very well account for the differences between our findings and those by Rayner et al. In addition, the situational context established by the context sentence may have provided a richer background for validation, thereby allowing faster detection of the implausibility of a target word in a given situation.

The modulation of some of the plausibility effects by the presence of uncertainty markers suggests that the validation process immediately takes into account the epistemic status of information. This indicates that effects of adverbial epistemic markers on comprehension are not confined to late and potentially strategic comprehension processes. Rather, they seem to include early comprehension processes. The pattern of results suggests that uncertainty markers impede the usually very quick integration of plausible information when it is presented as uncertain, whereas they tend to facilitate the processing of implausible information, especially in later integrative processing. The latter finding may reflect an efficient reading strategy in the face of

uncertainty: If implausible information is marked as uncertain, readers may invest less effort in trying to resolve the apparent inconsistency.

However, our results also show that readers can adapt their reading strategies when they notice certain patterns in presented information. This indicates that plausibility monitoring quickly adapts to the context: when reading material in which local implausibility is repeatedly explained later in the text, especially early effects of implausibility are reduced or even reversed over time. From a methodological point of view, it seems desirable to avoid inducing unnatural reading strategies by repeatedly exposing readers to implausible material. However, it must also be kept in mind that retroactively resolving these inconsistencies seems to induce strategies on the part of the reader as well, at least when readers notice this pattern in the material (which is difficult to avoid; but see Staub et al., 2007).

The results of both experiments suggest that the effects of epistemic markers on prediction and validation are not modulated by whether they direct linguistic focus to the target phrase or to the entire sentence. This result may seem surprising as focus has previously been shown to affect validation in terms of the detection of knowledge violations (Bredart & Modolo, 1988). However, there are two possible explanations for this result: First, Bredart and Modolo (1988) manipulated linguistic focus by using cleft sentences (e.g., *It was Moses who took two animals of each kind on the ark*). In comparison, focus induced by epistemic markers may be too subtle to affect validation processes. Second, the knowledge violations used in their study were so-called semantic illusions which, under normal circumstances, easily go unnoticed. In contrast, our plausibility manipulation was not designed to be subtle. Therefore, the detection of implausibility in our study was not dependent on whether the implausible information was in focus or not.

Despite the clear pattern of results, the experiments reported here suffer from certain limitations. In Experiment 1, we used a cloze procedure to construct the predictability manipulation, which inevitably leads to a confound of predictability and semantic relatedness because in real-world instances of language use predictability is correlated with semantic relatedness (that is why co-occurrences of words can be used as a basis for deriving semantic relationships, Landauer & Dumais, 1997). As a result, we cannot be sure whether and to what extent the robust predictability effects and the absence of any moderating effects of epistemic markers were in fact due to semantic priming. This limitation of our study also applies to the majority of eye-tracking studies on predictability but it certainly raises questions that should be addressed in further research. Second, the focal fixation measures in Experiment 1 were collected

on the target word so that it is difficult to clearly differentiate between prediction and integration processes. Future studies should manipulate predictability in a way that predictive processes can become apparent even before the target word (e.g., by using gender-marked adjectives before the target noun, see van Berkum et al., 2005). Third, although we did not find any evidence that certainty and uncertainty markers differentially influence prediction, it is still possible that they differentially influence validation. Our study does not allow conclusions regarding this issue as only uncertainty markers were included in Experiment 2 in order to keep the number of conditions manageable. However, in future experiments, it would be desirable to further explore the relationship between epistemic markers and validation by crossing the degree of certainty expressed by the marker with the plausibility of the target. The results reported here certainly (not just perhaps) encourage this endeavor.

Author Note

The first and the second author contributed equally to this article. The reported research was supported by a travel grant from the German Academic Exchange Service (Deutscher Akademischer Austauschdienst) and the Academy of Finland and by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, RI 1100/4-2). Johanna Kaakinen was supported by a grant from the Academy of Finland. We would like to thank Julia Knoepke for helpful discussions, Irina Fix, Susanne Hinz, Bianca Kramer, Ricarda Moll, and Sascha Müller for their help in preparing the text materials used in this study, and Theresa Hahn for collecting data.

References

- Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, 24, 477–492.
- Baayen, R.H., Davidson, D.J., Bates, D.M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.
- Bates, D., Maechler, M., Bolker, B. (2011). *Linear mixed-effects models using S4 classes* [Software]. R Foundation for Statistical Computing, Vienna, Austria.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67, 135-144.
- Britt, M.A., & Larson, A.A. (2003). Constructing representations of arguments. *Journal of Memory & Language*, 48, 798-810.
- Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, 9, 365–381.
- Calvo, M. G. & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes*, 22, 57-78.
- Campion, N. (2004). Predictive inferences are represented as hypothetical facts. *Journal of Memory and Language*, 50, 149-164.
- Chafe, W. (1986). Evidentiality in English conversation and academic writing. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 261-272). Norwood, NJ: Ablex.
- Chwilla, D.J., & Kolk, H.H.J. (2005). Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25, 589-606.
- Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during

- language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117-112.
- Ehrlich, S.F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 641-655.
- Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540-552.
- Federmeier, K.D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491-505.
- Ferretti, T.R., Singer, M., Patterson, C. (2008). Electrophysiological evidence for the time course of verifying text ideas. *Cognition*, 108, 881-888.
- Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes*, 16, 99-124.
- Frisson, S., Rayner, K., & Pickering, M.J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 862-877.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.
- Isberner, M.-B. & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, 142, 15-22.
- Isberner, M.-B. & Richter, T. (in press). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes*.
- Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lyons, J. (1977). *Semantics*, 2 vols. Cambridge: Cambridge University Press.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 913-934.
- Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, 27, 1-22.

- Nuyts, J. (2001). *Epistemic modality, language, and conceptualization*. Amsterdam: John Benjamins.
- Palmer, F.R. (2001). *Mood and modality*. New York: Cambridge University Press.
- Pickering, M.J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105-110.
- Rayner, K., & Well, A.D. (1996). Effects of contextual constraint on eye-movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504-509.
- Rayner, K., Kambe, G., & Duffy, S.A. (2000). The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology*, 53A, 1061-1080.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1290-1301.
- Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96, 538-558.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, 75-116.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59, 237-259.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574-591.
- Singer, M. (2009). Tacit verification of determinate and indeterminate text ideas. *Canadian Journal of Experimental Psychology*, 63, 185-192.
- Singer, M. (in press). Validation in reading comprehension. *Current Directions in Psychological Science*.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25, 359-393.
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1162-1169.
- Sturt, P., Sanford, A.J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, 11, 882-888.

- van Berkum, J.J.A, Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443-467.
- van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, 22, 335-351.
- Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye-movement study. *Memory & Cognition*, 35, 73-86.

Figure Captions

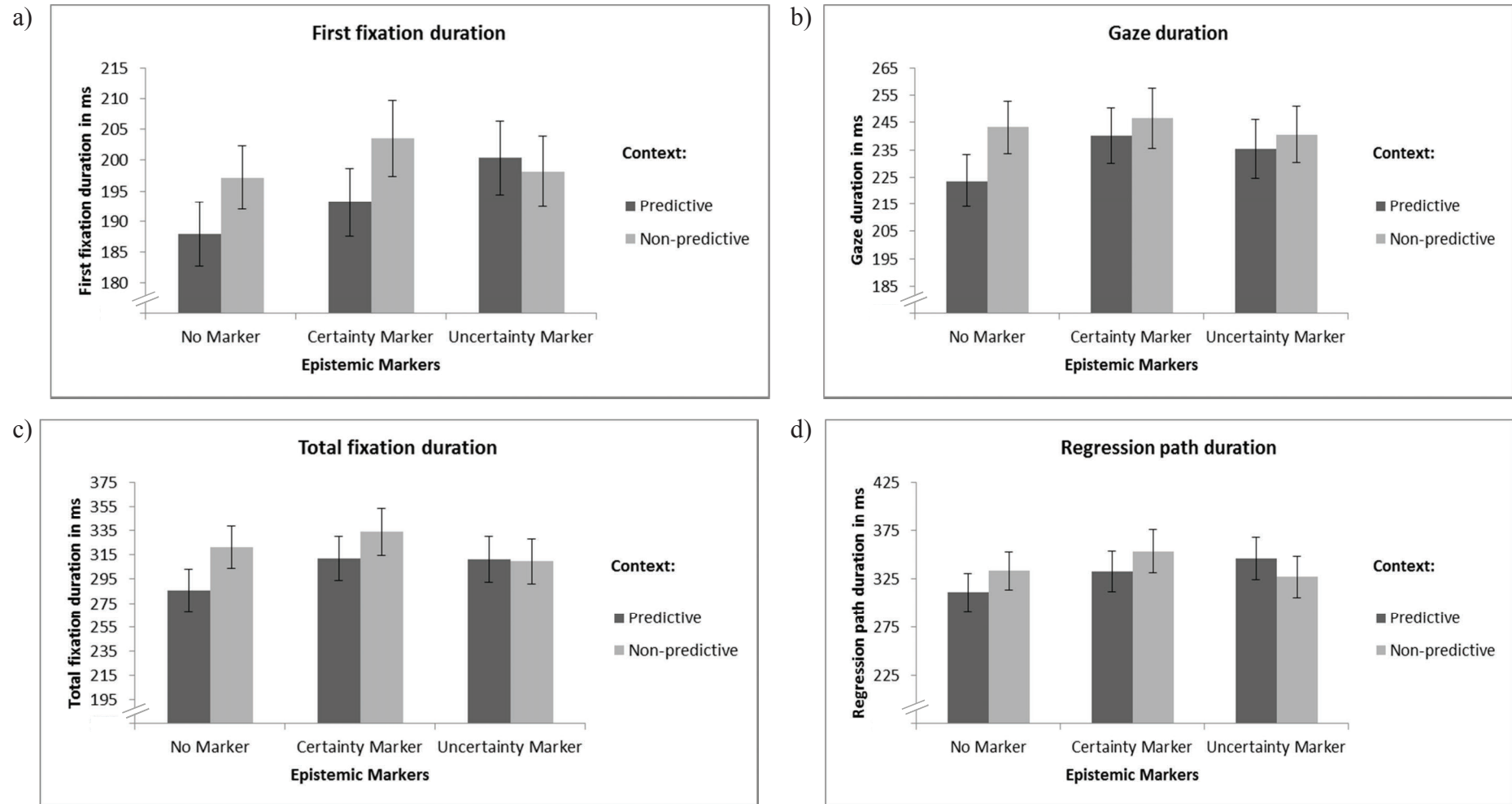
Figure 1. Effects of predictability and epistemic markers on fixation measures for the target word in Experiment 1 (model-based estimates with standard error of the mean): (a) First fixation duration, (b) gaze duration, (c) total fixation duration, (d) regression path duration, and (e) probability of regressions to the target word.

Figure 2. Effects of plausibility and epistemic markers on fixation measures for the target word in Experiment 2 (model-based estimates with standard error of the mean): (a) First fixation duration in the first half of the experiment, (b) first fixation duration in the second half of the experiment, (c) gaze duration in the first half of the experiment, d) gaze duration in the second half of the experiment, (e) total fixation duration in the first half of the experiment, (f) total fixation duration in the second half of the experiment, (g) regression path duration, (h) probability of skips, and (i) probability of regressions to the target word.

Figure 3. Effects of plausibility and epistemic markers on fixation measures for the final word of the target sentence in Experiment 2 (model-based estimates with standard error of the mean): (a) Regression path duration and (b) probability of regressions to the target word.

Figure 4. Effects of plausibility and epistemic markers on fixation measures for the context sentence in Experiment 2 (model-based estimates with standard error of the mean): (a) Total fixation duration, and (b) probability of regressions to the target word.

Figure 1



e)

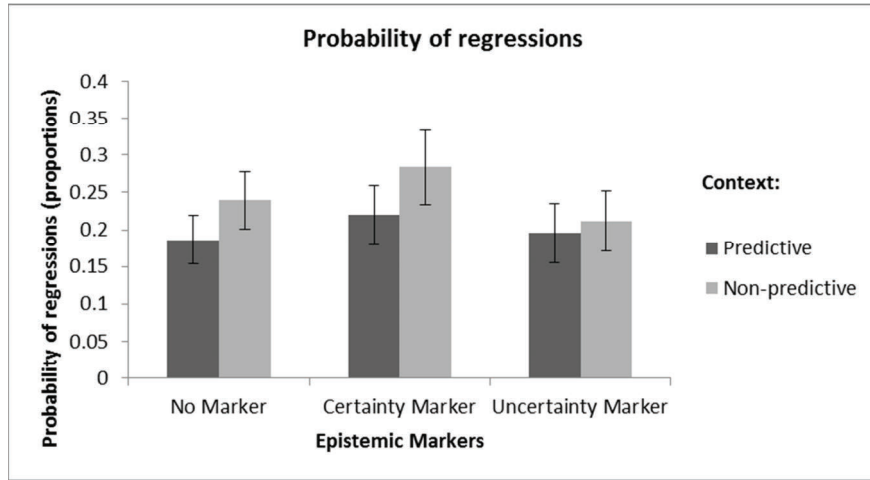
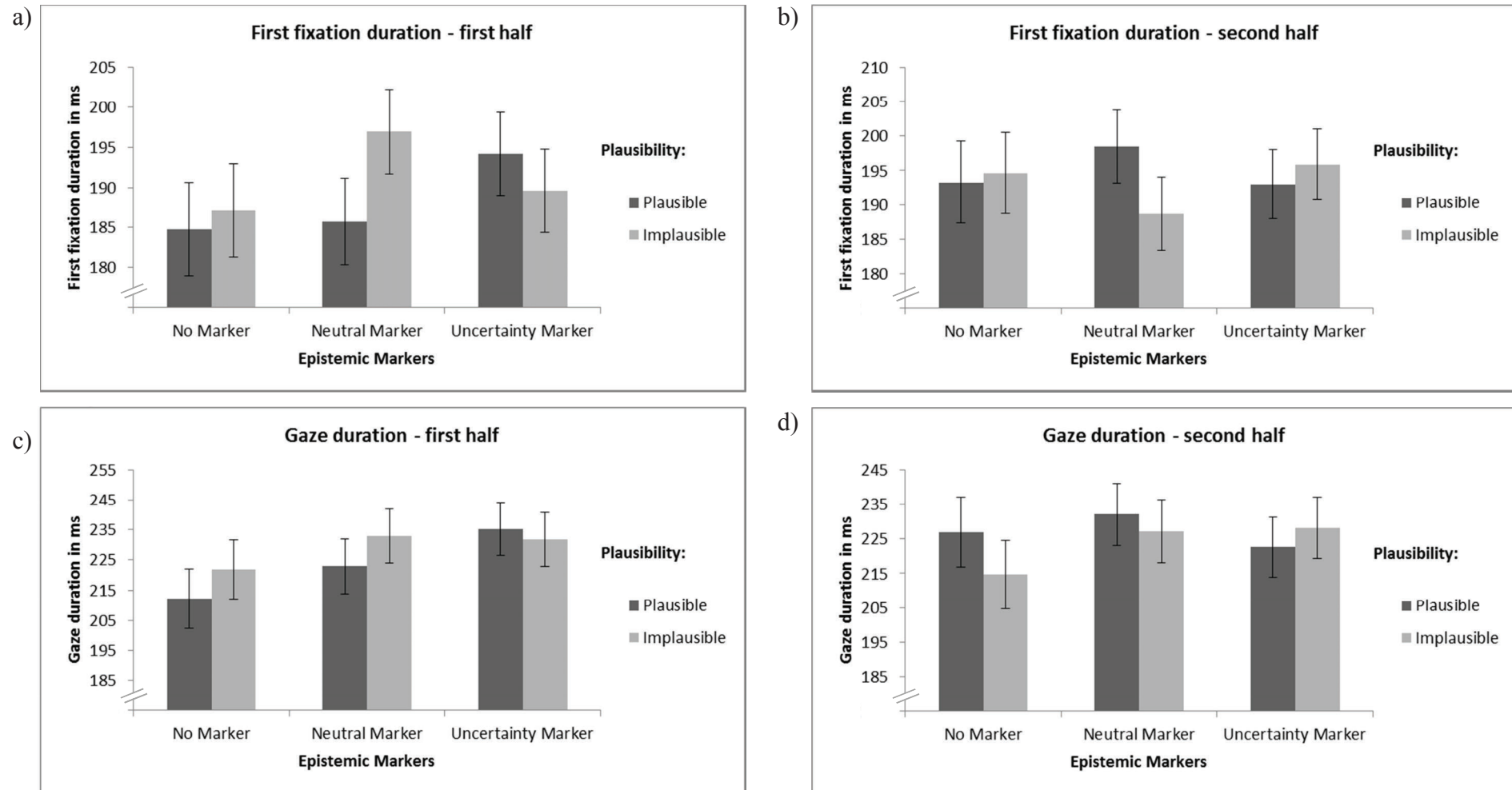
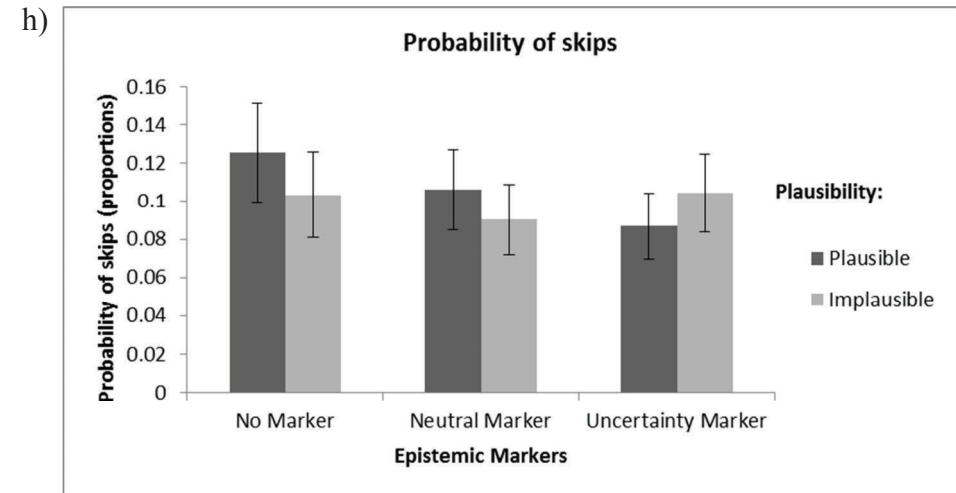
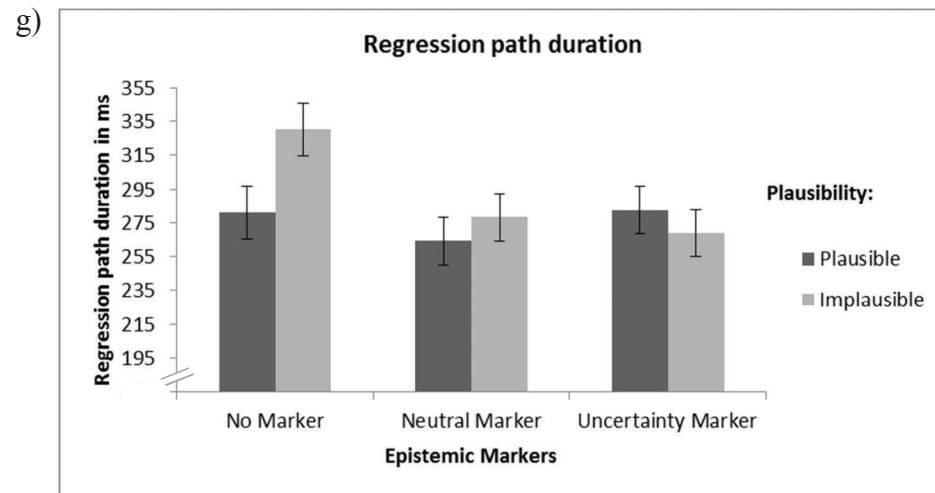
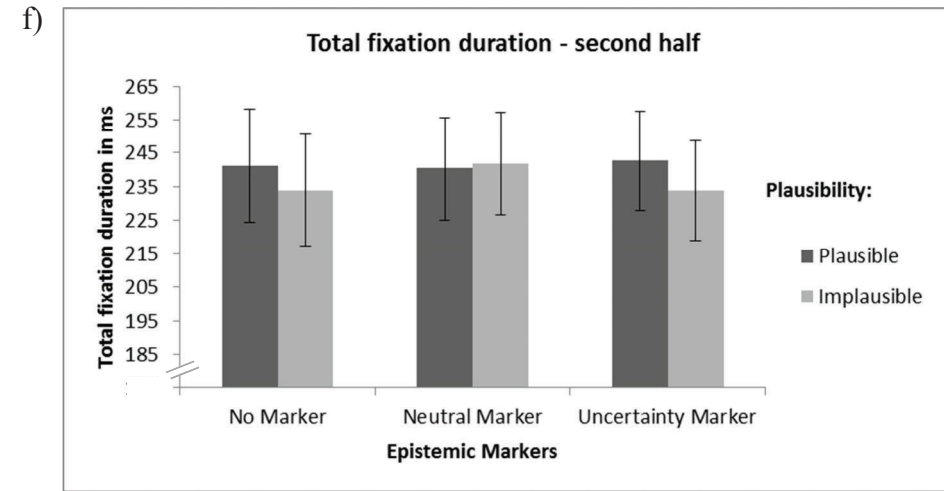
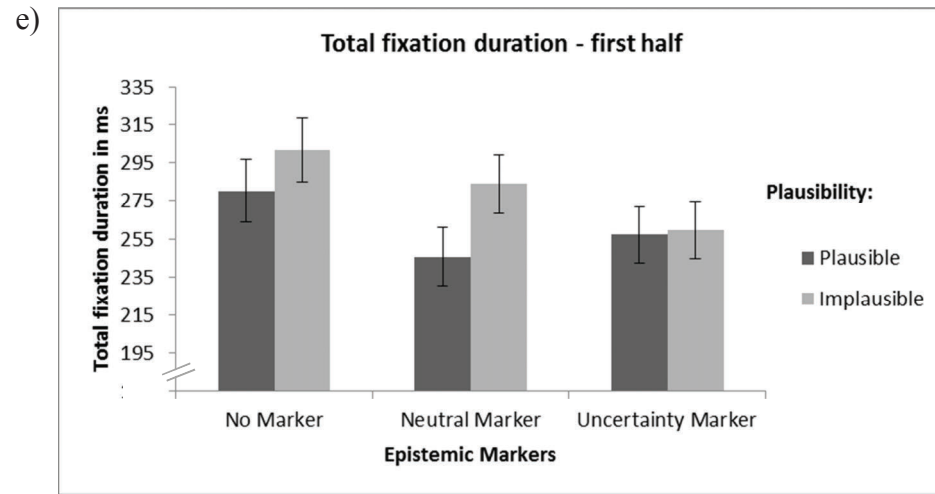


Figure 2





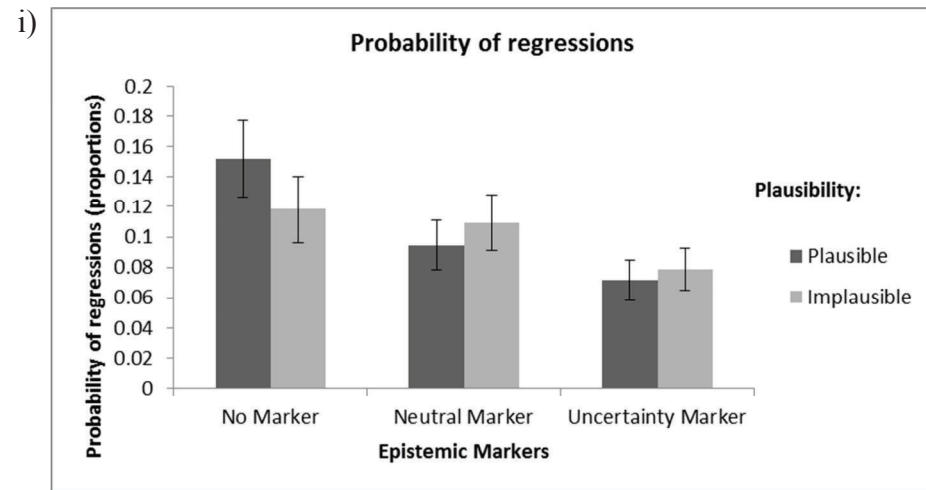


Figure 3

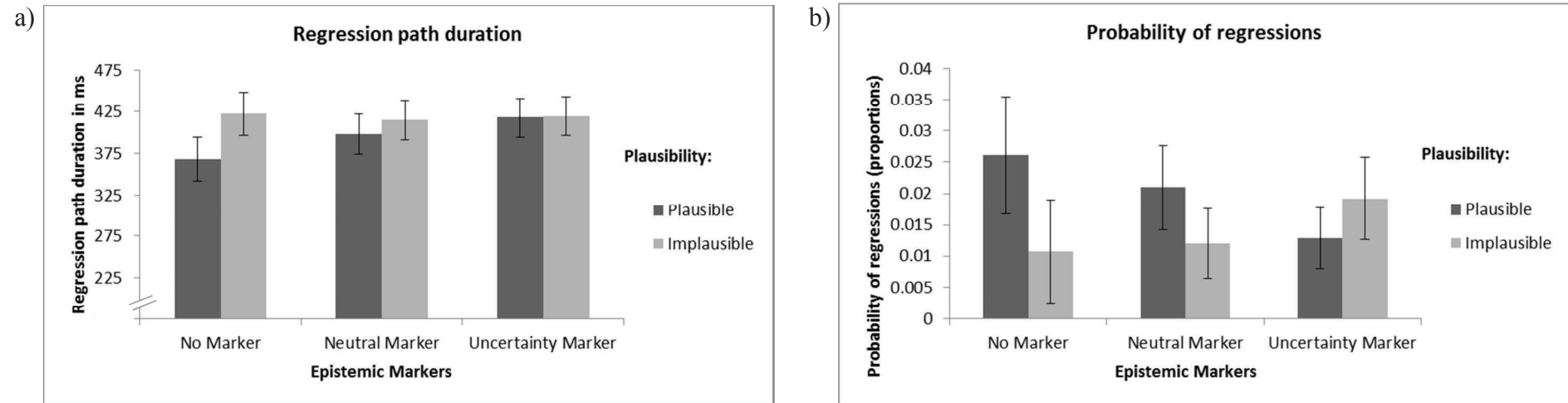
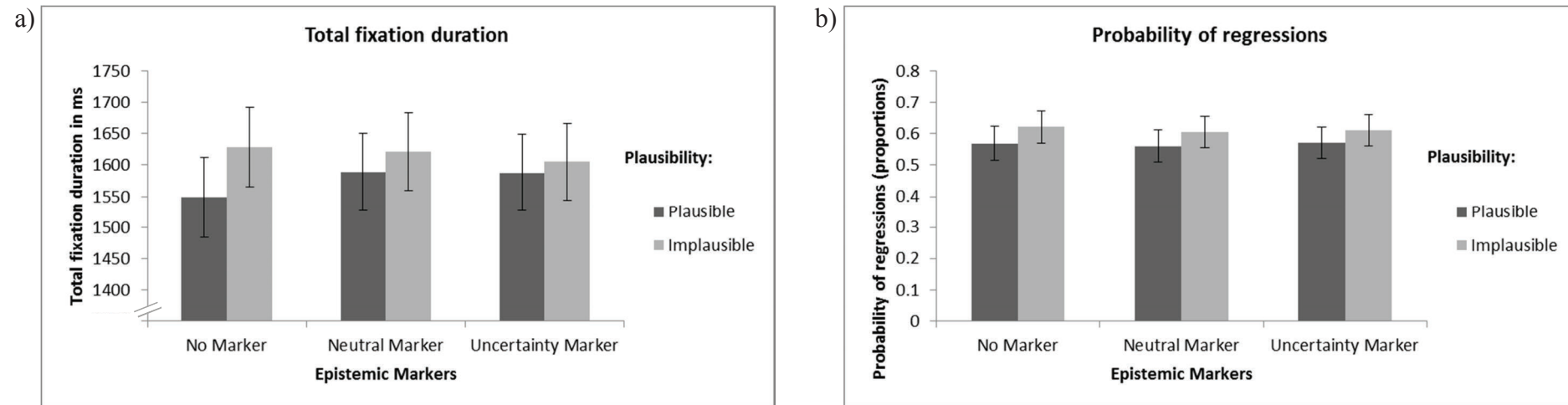


Figure 4



Chapter 6

General Discussion

Summary of Results

The aim of this thesis was to further elucidate the role of epistemic monitoring in language comprehension. Epistemic monitoring refers to the notion that readers routinely monitor the consistency of what they read with their world knowledge, with the current situation model, and with the text context (e.g., Richter, Schroeder, & Wöhrmann, 2009), which poses a challenge to the traditional view that such a consistency-check is optional and subsequent to, rather than a part of, comprehension (e.g., Connell & Keane, 2006; Gilbert, 1991). As reviewed in **Chapters 1** and **2**, there is a substantial body of evidence contradicting this traditional view. The goal of the present thesis was to further explore whether epistemic monitoring can be conceived of as an inherent component of language comprehension by testing the central implications of this notion (1) that epistemic monitoring nonstrategically accompanies language comprehension, (2) that it is not dependent on an evaluative mindset, and (3) that its effects on measures of comprehension emerge rapidly. In doing so, this thesis tested the generalizability of the results reported by Richter et al. (2009) across different types of stimuli and tasks. In addition, the effects of predictability, task-demands, and uncertainty markers on validation processes were investigated.

The experiments presented in **Chapters 3** and **4** directly tested the assumption of nonstrategic epistemic monitoring employing a Stroop-like paradigm. These experiments built on research presented by Richter et al. (2009) which demonstrated a Stroop-like compatibility effect for orthographical judgments on single words that rendered a sentence valid (true) or invalid (false) with regard to easily accessible world knowledge. Based on these findings, the experiments reported in **Chapter 3** explored the scope of epistemic monitoring by testing whether such a compatibility effect can also be found with different stimuli – namely, sentences which were not clearly valid or invalid, but merely plausible or implausible with regard to a context sentence – and with a different task – namely, a nonlinguistic color judgment task. Moreover, this study attempted to elucidate whether the compatibility effect may be due to the differential predictability of words in valid/plausible and invalid/implausible sentences. An interaction of plausibility and required response was found in both the orthographical (Experiment 1) and the color judgment task (Experiment 2), indicating that epistemic judgments are made along a continuum of plausibility. However, in the orthographical task, the interaction was ordinal, whereas in the color task it was semi-disordinal. Both experiments clearly showed that positive responses were much faster for words rendering the target sentence plausible (compatible condition) than for words rendering the target sentence implausible (incompatible condition). For negative responses, the results

were less clear: In the orthographical task, they were also faster for plausible words (incompatible condition) than for implausible words (compatible condition), while in the color task, there was no significant difference. Thus, in both tasks, the compatibility effect emerged only for positive responses, which allows several interpretations that will be addressed in more detail later.

The experiment presented in **Chapter 4** further probed the question of whether epistemic validation depends on an evaluative mindset. Wiswede, Koranyi, Müller, Langner, and Rothermund (2013) had raised the concern that any evaluative task (such as the orthographical task used by Richter et al., 2009) might induce an evaluative mindset and that the Stroop-like compatibility effect may be dependent on such a mindset. Therefore, they used a non-evaluative probe task (i.e., simply responding to the probe words TRUE and FALSE) and found a compatibility effect only when evaluation for truth was explicitly encouraged by a secondary task. Using this probe task introduced by Wiswede et al. (2013), we were able to show in our experiment that the compatibility effect is not dependent on an evaluative mindset, but that it requires an adequate depth of semantic processing. By adding simple questions that encouraged comprehension but not validation to the probe task, we found a Stroop-like compatibility effect in the response latencies (there was also a compatibility effect in the error rates, but it was only reliable in the by-items analysis). In addition, by introducing control items whose validity was unknown to participants, this experiment also provided insight into whether the Stroop-like effect is due to interference in incompatible conditions, facilitation in compatible conditions, or both. Although the numerical tendencies indicated both facilitation in compatible conditions and interference in incompatible conditions, the only effect that was reliable both by-items and by-subjects was the facilitation of positive responses after valid sentences.

The study reported in **Chapter 5** employed a different methodological approach to investigate validation processes. It used eye-tracking technology to examine the time course of the influences of plausibility and predictability on comprehension processes, as both variables reflect different aspects of the fit of information with prior knowledge. In addition, it investigated the potential modulation of these influences by the presence of adverbs that mark the epistemic status of the communicated information (e.g., *certainly* or *perhaps*). Although predictability and plausibility both affected indicators of early and late comprehension processes, only the effects of plausibility were modulated by the presence of epistemic markers: Uncertainty markers, as compared to neutral or no markers, increased early processing of plausible information (increasing first fixation durations and reducing skips),

but decreased late processing of implausible information (reducing regression path duration). However, particularly the early effects appeared to be sensitive to strategies developed over the course of the experiment, as they were only present in the first half of the experiment.

Theoretical Implications for the Epistemic View of Language Comprehension

The findings of early (Study 3) and nonstrategic (Studies 1 and 2) effects of validity or plausibility on eye movements (Study 3) and response latencies (Studies 1 and 2) are in line with the predictions derived from the epistemic view of language comprehension, which assumes that epistemic monitoring is a routine component of comprehension

Study 1 showed that validation is not limited to information which can be judged as clearly valid or invalid, but that it extends to information which is merely more or less plausible. This suggests that the output of the validation process is not dichotomous (true or false), but that evaluation occurs along a continuum ranging from positive to negative, and that epistemic monitoring is not restricted to situations in which the veracity of information can be assessed with certainty. The similarity of the results for valid versus invalid and plausible versus implausible stimuli indicates that the epistemic monitoring process is highly similar for both types of stimuli. In addition, Study 1 showed that the epistemic Stroop effect does not result from the differential predictability of valid/plausible and invalid/implausible information. More precisely, the interference with positive and negative responses does not seem to be attributable to the confirmation or disconfirmation of expectations concerning specific upcoming words. Finally, the epistemic Stroop effect was replicated with a task that was entirely different from the orthographical task, namely a nonlinguistic color judgment task, which does not confound lexical access and validation. This indicates that the response tendencies evoked by epistemic monitoring are of a general nature and do not only interfere with performance in other linguistic tasks that rely on partly the same processes (such as word recognition in the orthographical judgment task), but also with performance in completely unrelated and even nonlinguistic tasks.

Study 2 again replicated the epistemic Stroop effect with a different task that did not require any kind of evaluation on the part of the reader: A simple probe task that only required recognizing which of two probe words (TRUE or FALSE) was presented after a sentence and pressing the corresponding response key. As argued by Wiswede et al. (2013), this non-evaluative task should be unlikely to induce an evaluative mindset. To nonetheless ensure an adequate depth of comprehension, we presented questions that required comprehension but not validation after some of the trials. With this task, the epistemic Stroop

effect was obtained, suggesting that it does not hinge on an evaluative mindset of the reader, but rather on an adequate depth of processing. In addition, this experiment provided insight into the underlying pattern of facilitation and interference and suggested that the epistemic Stroop effect relies mainly on facilitation of positive responses after reading valid information. This finding has several potential implications, which will be discussed in more detail further below.

Using eye-tracking technology, Study 3 revealed plausibility effects on indicators of early comprehension processes when experimental half was taken into account as an additional factor. The effect of plausibility on first fixation duration coheres well with other research demonstrating very rapid effects of plausibility (e.g., Kennedy, Murray, & Boissiere, 2004; Murray, 1998; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007). Moreover, fewer regressions to information marked as uncertain suggested that uncertainty markers either facilitated processing or reduced the depth of processing of that information. In conjunction with evidence from a recent ERP study by Ferretti, Singer, and Harwood (2013), which suggests that information associated with nonfactive verbs does not precipitate situation model updating, this finding may also indicate that information marked as uncertain is not integrated into the mental representation. In terms of the framework of epistemic validation, the results of Study 3 might also suggest that uncertainty results in less epistemic elaboration, i.e., in reduced attempts to resolve the detected implausibility. However, these are speculations that need to be tested directly in future experiments by relating process measures of comprehension to reading outcomes.

Unexpected Findings

Overall, the findings are in line with the predictions derived from the epistemic view of language comprehension. However, a deviation from the predictions is noteworthy: Contrary to the notion that the primary function of epistemic monitoring is the rejection of implausible or invalid information, the pattern of results obtained in Studies 1 and 2 provides substantially more evidence for a positive response tendency in response to plausible or valid information, although there was also some indication of a negative response tendency in response to invalid or implausible information. In a similar vein, Study 3 revealed uncertainty markers to affect first fixations on plausible rather than on implausible target words, suggesting that they primarily affect the early processing of plausible (not of implausible) information.

There are several possible interpretations of these patterns: One simple explanation is that the interference of the negative response tendency may manifest itself more strongly in the error rates than in the response latencies. Indeed, the pattern of results obtained in Study 2 as well as by Richter et al. (2009) in their Experiments 3 and 4 supports this notion. If this interpretation is correct, it may be useful to compute an integrated measure that takes into account both the accuracy and latency of each participant in each condition in future experiments.

Second, it may be that the primary mechanism of epistemic monitoring is the matching of new input to prior knowledge, which results in a positive response tendency when new input is sufficiently consistent with prior knowledge. In fact, the more knowledge which is passively retrieved via resonance is consistent with the currently processed information, the more positive the evaluation of the message by epistemic monitoring may be, which could explain why plausible misinformation is much more likely to be missed by readers (e.g., Hinze, Slaten, Horton, Jenkins, & Rapp, 2013). It is even possible that the more knowledge a reader has that is consistent with the currently processed information, the less likely it is that inconsistent knowledge will be activated by passive resonance because a larger amount of consistent information should be more likely to draw activation away from the inconsistent information (cp. Kendeou, Smith, & O'Brien, 2013). As a consequence of this primary or initial focus on consistency, invalid or implausible input may not be rejected as strongly or as immediately as assumed. In light of findings by Pickering and Traxler (Pickering & Traxler, 2000; Traxler, Morris, & Seely, 2002; Traxler & Pickering, 1996) suggesting that readers commit more strongly to a plausible than to an implausible analysis (which is why a plausible analysis is initially processed faster but also leads to more disruption when it turns out to be wrong), this positive response tendency might also reflect the commitment to an analysis.

Another possibility is that the negative response tendency is either faster or slower than the positive response tendency and therefore emerges more strongly at an earlier or later point in time. Given that we always presented the response prompt after 300 ms, the results are but a snapshot of processes that presumably unfold continuously over a longer period of time. Thus, the respective response tendencies may wax and wane as a function of the time elapsed between the onset of the final word (at which point validity or plausibility can be determined) and the onset of the response prompt. The most plausible explanation seems that a negative judgment (i.e., the rejection of information as false or implausible) takes longer than recognizing information as true because it requires more mental operations (e.g., Carpenter & Just, 1975; Clark & Chase, 1972). This explanation is also in line with the theory

proposed by Wyer and Radvansky (1999), who proposed that comprehension and validation of true statements for which situation models already exist merely require identifying the corresponding representation in memory, whereas false statements require a comparison of the subject's features with situation models of the predicate. However, this does not necessarily imply that validation is slower for false than for true sentences; rather, both comprehension and validation of false sentences appear to be slower, as indicated by the finding that both comprehensibility and validity judgments take longer for false than for true sentences (Wyer & Radvansky, 1999).

There is, of course, also the possibility that the positive response tendency elicited by valid or plausible information is due to a contrast effect created by the concurrent presentation of invalid or implausible information. This alternative explanation is difficult to rule out because it is almost impossible to avoid exposing people to invalid or implausible information when one wants to study the effects of plausibility on comprehension, and while a large number of trials per condition is generally desirable, it may at the same time increase the negative side effects of implausible information such as inducing unnatural reading strategies. Our attempt in Study 3 to solve this problem by retroactively resolving the inconsistencies obviously had undesired effects on readers' expectancies, as indicated by the results. Thus, a more subtle way to induce and subsequently resolve local implausibility (but still strong enough to yield effects) seems desirable. This could be achieved with the method employed by Staub et al. (2007): They used compounds whose modifier was temporarily implausible if it was initially analyzed as a head noun (e.g., *She heard the mountain [implausible] lion [plausible]*). The epistemic Stroop paradigm could be used to probe for interference with positive and negative responses after both the modifier (*mountain*) and the head noun (*lion*), respectively. However, this approach also has some potential drawbacks: First, participants might notice the local implausibility and its later resolution nonetheless due to the rather strong demand characteristics of the task, resulting in similar problems as in our Study 3. Second, this approach is only feasible in languages in which compounds are separated into different words (like English, but unlike German). Third, it has to be noted that most of Staub et al. (2007)'s local implausibilities would qualify as semantic anomalies rather than as general world knowledge violations (if one wants to draw the distinction).

It is important to keep in mind that the apparent dominance of a positive response tendency after reading valid or plausible information does not constitute evidence for Gilbert (1991)'s position that people indiscriminately accept everything they read, as there clearly is an effect of validity or plausibility on this response tendency; invalid or implausible

information, in contrast, did not facilitate, but rather tended to interfere with this positive response tendency. Thus, readers do seem to discriminate already at an early stage between information that is valid or plausible versus invalid or implausible. However, the evidence for a *negative* response tendency in response to invalid or implausible information was weaker than expected. As discussed earlier, the most likely explanation is that this negative response tendency emerges more slowly, potentially because more cognitive operations are required for establishing a mismatch between new information and prior world knowledge than for identifying information as true or plausible (e.g., Carpenter & Just, 1975; Clark & Chase, 1972; Wyer & Radvansky, 1999). Alternatively, the initial detection of a knowledge violation may result in “withholding acceptance” rather than in outright rejection, as readers may first try to reanalyze the input in an attempt to make sense of it before they ultimately reject it.

Are Comprehension and Validation the Same?

The results presented in this thesis strongly support the notion that comprehension and validation are closely related, in the sense that validation often results as a “byproduct” of normal comprehension. However, it is important to note that this does not imply that validation is the same as comprehension. The distinction becomes clear when one considers what a failure of comprehension and a failure of validation means, respectively: Comprehension can be said to have failed when a situation model is constructed that is not in line with the text (as in the plane crash example by Barton & Sanford, 1993, in which readers clearly do not construct a situation model of actual *survivors* being buried), while validation can be said to have failed when a situation model is constructed that is fully in line with the text, but in contradiction with prior knowledge or antecedent text, and this contradiction is not detected. Comprehension is successful in this case (i.e., it produces an accurate situation model, at least for the current constituent), but validation is not (i.e., it produces an inaccurate evaluation of the validity or plausibility of the situation model). Whether validation fails or not depends on whether the relevant contradictory information is activated or not, but in either case validation is not (necessarily) an intentional activity; rather, it results from the passive activation of knowledge during comprehension.

Integrating the Findings With Other Research

The findings reported in this thesis provide strong evidence for a fast-acting and passive monitoring process that evaluates incoming information for its fit with prior knowledge, and thus cohere well with other research reporting rapid and nonstrategic effects

of real-world plausibility on language processing (see **Chapter 2**). They can also be integrated with a number of theoretical frameworks that imply a passive process similar to epistemic monitoring which evaluates information against some knowledge-based criterion, such as the resonance model (“Presumably, there is a second process that continually evaluates the contents of working memory. This process may register failure in certain circumstances, for example, [...] when propositions in working memory contradict each other”; Myers & O'Brien, 1998, p. 133), the landscape model (“The information activated through this autonomous process [...] is evaluated with respect to a reader’s standards of coherence”; van den Broek, Rapp, & Kendeou, 2005, p. 304; “[...] the Landscape model explicitly allows for the possibility that reactivation and hence inconsistency detection may take place as a side effect of other processes, such as an effort to comprehend the current sentence.”; van den Broek, Young, Tzeng, & Linderholm, 1998, p. 87), the situation model account (“A procedure which, if all the entities referred to in the assertion are represented in the current model, verifies whether the asserted properties or relations hold in the model”; Johnson-Laird, 1983, p. 249), the scenario-mapping and focus account (“[...] Barton and Sanford (1993) claimed that, when a word is encountered, the earliest thing to happen is that its relevance to the domain of discourse is established by a fast passive process”; Sanford & Garrod, 1998, p. 179) and the RI-Val view of comprehension (“We argue that activation, integration, and validation are passive parallel asynchronous processes, with activation preceding integration, and integration preceding validation”; Cook & O'Brien, 2013, para. 5). Similar to the epistemic view of language comprehension, the resonance model, the landscape model, and the scenario-mapping and focus account all assume that a negative outcome of this passive evaluative process is often followed by a more strategic processing of the information (e.g., coherence-based retrieval of information according to the landscape model or a deeper analysis of the input according to the scenario-mapping and focus account).

More generally, the results presented in this thesis are in line with accounts of language comprehension which assume that real world knowledge is immediately brought to bear on comprehension (e.g., Hagoort & van Berkum, 2007; McRae & Matsuki, 2009; Sanford & Garrod, 1998). In contrast, the results are in contradiction with two-step models of comprehension and validation which assume validation to be an optional processing stage separate from comprehension (e.g., Gilbert, 1991; Gilbert, Krull & Malone, 1990; Gilbert, Tatarodi & Malone, 1993).

Findings by Gilbert and Colleagues

Nonetheless, the findings can be reconciled with those reported by Gilbert and colleagues (Gilbert, 1991; Gilbert et al., 1990, 1993) if one takes into account that the availability of easily accessible relevant knowledge has been established as a moderator of the affirmation bias (Richter et al., 2009). The stimuli used in the present thesis strongly differed from those used by Gilbert et al. (1990, 1993) in that they were not fictitious facts, but assertions associated with strong and easily accessible background knowledge which has been shown to permit routine and effortless validation (Richter et al., 2009). Thus, the scope of the two theories appears to be complementary: Unlike the epistemic view of language comprehension, Gilbert (1991)'s theory seems to apply primarily to information which is not associated with easily accessible background knowledge (for a similar argument, see Richter, 2003; Richter et al., 2009).

However, Gilbert (1991) seems to have been right in proposing that readers have an overall affirmation bias, in the sense that they generally appear to expect information to be true and informative (in line with the cooperative principles defined by Grice, 1975) and therefore “go to great lengths to make sense of text” (Baker, 1989, p. 13). Moreover, there seems to be an affirmation bias with regard to (ostensibly) new information such that it is by default presumed true if no knowledge that is in conflict with it is retrieved in the course of comprehension (Jackendoff, 2002, p. 327; Johnson-Laird, 1983, p. 249). Thus, it may be true that readers tend to be somewhat “gullible”; nonetheless, they clearly do not indiscriminately believe everything they read. On the contrary, it is all the more noteworthy that readers *do* detect many inconsistencies without necessarily expecting them. Thus, the detection of false or implausible information reported in this thesis and in the reviewed studies does not seem to result from a generally skeptical or critical evaluative stance towards what is being read, but rather from the processes underlying comprehension itself: When readers attempt to construct a coherent mental representation of what they are reading by integrating new information with their prior knowledge, they cannot help but experience difficulties when they encounter new information which is inconsistent with knowledge that is recruited for comprehension.

Findings on the Misinformation Effect

Similarly, the findings in this thesis can be reconciled with research on misinformation and persuasion through narratives if one takes into account two factors that influence the detection of knowledge violations: The plausibility of the false information (e.g., Hinze et al.,

2013), and the ease of accessibility of the knowledge required to detect the false information (e.g., Kamas, Reder, & Ayers, 1996).

First of all, the false lures used in this research are usually designed to be highly plausible, i.e., to overlap quite strongly with prior knowledge (e.g., Marsh, 2004; cp. Hinze et al., 2013), which makes them likely to be missed by epistemic monitoring. In line with this notion, a recent study by Hinze et al. (2013) suggests that plausibility is a key moderator of reliance on misinformation from narratives. Think-aloud protocols revealed that implausible lures increased overall skepticism towards information from a text (including correct information), suggesting that implausible information generally increases vigilance (“evaluative spillover”, p. 3). Moreover, relating participants’ think-aloud responses to their reliance on misinformation provided insight into the mechanisms underlying the misinformation effect. Importantly, this study showed that some use of misinformation on the later knowledge test was actually due to the fact that participants actively accepted the false information during reading, which probably concerns mainly items for which participants did not have accurate knowledge before reading. Participants expressed acceptance for 17.3% of the implausible and 31.5% of the plausible lures on average, suggesting that plausible misinformation is more likely to be actively accepted by readers than implausible information. Moreover, when readers responded to target information with content-irrelevant responses, which indicated that they did not engage in explicit evaluation of that information, reliance on implausible misinformation was much lower (3.31%) than on plausible misinformation (29.03%), “suggesting a general effect of plausibility even without explicit critical evaluation” (p. 12). Thus, contrary to what Gilbert et al. (1990, 1993) contend, plausibility seems to be tacitly assessed even in the absence of explicit evaluation. Based on these findings, Hinze et al. (2013) concluded that reliance on misinformation does not necessarily speak against epistemic monitoring during comprehension, but rather that it to some extent reflects (active or passive) positive evaluations of the false information. However, they also pointed out an alternative interpretation of the plausibility effect on reliance on misinformation, which is that plausible misinformation is more easily retrieved at test because it aligns better with existing representations. Future research should attempt to explicitly test these two alternative explanations against each other.

In addition, the detection of the lures usually used in studies on the misinformation effect requires knowledge that may not be easily accessible for or even possessed by all readers (for example, the knowledge of which is the largest ocean, or the largest planet; e.g., Marsh, 2004). Whereas the stimuli used by Richter et al. (2009) and in this thesis were

normed to be recognized as true versus false or plausible versus implausible by virtually all participants of a sample from the same population as the experimental sample (with mean agreements of 95-100%), studies on misinformation usually use high and low knowledge items which are categorized based on norms from 1980 (Nelson & Narens, 1980), with high knowledge sometimes referring to items with an average of correct answers of only about 70% in the norming sample (e.g., Marsh, 2004). Even when analyses are only performed on items for which participants have demonstrated accurate knowledge (e.g., Bottoms, Eslick, & Marsh, 2010), the accessibility of that knowledge is presumably still considerably lower than for the items used by Richter et al. (2009) and in this thesis. This has important implications because the effortless retrieval of pertinent knowledge is assumed to be a precondition for the routine detection of a knowledge violation by the epistemic view of language comprehension, as well as by the landscape model (Linderholm, Virtue, Tzeng, & van den Broek, 2004) and the resonance model (Myers & O'Brien, 1998). Although the proportion of people who can answer a particular knowledge question correctly may be to some extent related to the accessibility of that knowledge in the average reader, response latencies for correct responses knowledge questions in addition to probabilities may be a much more suitable indicator of accessibility. In fact, Nelson and Narens (1980) pointed out that probability of correct responses and latency of correct responses are only moderately correlated ($r = -.30$; p. 367).

However, although it seems plausible, it must be noted that a failure to detect an inconsistency (as measured, for example, by reading times) does not necessarily mean that the knowledge cueing the inconsistency was not activated; alternatively, it may have been activated but not used appropriately for validation. Studies using probe statements to test for the activation of stored knowledge (e.g., Kendeou et al., 2013) have shown that inconsistent knowledge can also be activated *without* disrupting comprehension. Thus, a distinction should be made between the activation of relevant knowledge and the detection of an inconsistency with that knowledge, and verification times for probe statements clearly seem to be the most adequate measure of knowledge activation.

Moreover, the assumption of the epistemic view of language comprehension that false information is rejected from the situation model does not preclude the possibility that this information may nonetheless leave a memory trace that co-exists with the correct knowledge in memory and may be used on general knowledge tests because it is temporarily more easily accessible (Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013). This may be the case especially when the correct knowledge is not easily accessible in the first place.

Finally, besides the plausibility of the lures and a lack of (easily accessible) relevant knowledge, a number of other explanations have been put forward for readers' failures to detect false or inconsistent information, such as the construction of an underspecified mental representation due to shallow processing (Sanford & Garrod, 1998), only partial matching between the current input and knowledge stored in memory (Reder & Kusbit, 1991), working memory limitations (Hannon & Daneman, 2001), shallow standards of coherence (Linderholm et al., 2004), exaggerated weighting of prior knowledge (Otero & Kintsch, 1992), or the fading of relevant information from the mental representation (Kaup & Foss, 2005). It is likely that all of these explanations account for the reported cases of failures of epistemic monitoring in varying degrees depending on characteristics of the reader, the text, and the reading situation, and can thus contribute to a deeper understanding of how epistemic monitoring functions.

Limitations and Directions for Future Research

In the following, I will discuss a number of limitations of the studies presented in this thesis, which at the same time highlight important directions for future research.

Mechanisms Underlying Epistemic Monitoring

Although the present thesis further substantiated the claim that epistemic monitoring is an integral component of language comprehension, the underlying mechanisms of how knowledge is activated and used for validation were not directly investigated. A primary goal for future research should thus be to refine the epistemic view of language comprehension so that it allows for specific predictions as to when false or inconsistent information will be detected or missed by epistemic monitoring, taking into account the reported failures of epistemic monitoring. Extant models of knowledge activation during language comprehension (such as the landscape model; e.g., van den Broek et al., 1998), as well as the factors that have been identified to moderate the detection of false or inconsistent information, should inform these specifications. For example, studies on semantic illusions have elucidated both text and reader characteristics that influence the detection of false information. Factors that have been shown to influence detection rates are the semantic feature overlap between the correct and the incorrect term (Erickson & Mattson, 1981; Hannon & Daneman, 2001; van Oostendorp & de Mul, 1990), the semantic cohesion of the context (Hannon & Daneman, 2001; Reder & Kusbit, 1991; van Oostendorp & Kok, 1990), the syntactic structure of the sentence (Bredart & Modolo, 1988), processing load (Bohan, 2008), and task instructions (Bohan, 2008). Detection rates are also higher when errors are more common (Bottoms et al., 2010).

Another factor that has been reported to affect the detection of knowledge violations is whether the problematic information is marked as given (e.g., *The liver, which is an organ found only in humans, is often damaged by heavy drinking*) or new (*The liver, which is often damaged by heavy drinking, is an organ found only in humans*; Baker & Wagner, 1987). Interestingly, this factor appears to interact with the type of inconsistency, as detection rates for inconsistencies with prior knowledge (such as the one in the example) have been reported to be higher when the inconsistent information is marked as new rather than given, while the opposite pattern emerges for text-internal inconsistencies (e.g., Glenberg, Wilkinson, & Epstein, 1982; Zabucky, Moore, & Schultz, 1987; cp. Baker, 1989). These findings emphasize the relationship between depth of processing, which has been shown to be affected by linguistic focus (e.g., Sanford & Sturt, 2002; Sturt, Sanford, Stewart, & Dawydiak, 2004), and epistemic monitoring. In addition to characteristics of the message, reader characteristics that have been found to influence the detection of false or inconsistent information are working memory capacity and the ability to access and reason about prior knowledge (Hannon & Daneman, 2001).

Extant research has also provided insight into the mechanisms that may account for failures to detect false information. Research by Reder and colleagues, for instance, has suggested that these failures may be due to only partial matching of the new input to prior knowledge (Reder & Kusbit, 1991). It has also shown that, surprisingly, detection is not increased by simply improving the accessibility of the correct information in memory (Reder & Cleeremans, 1990), but by increasing the accessibility of features that differentiate the correct term and the false lure (e.g., by asking prior to the presentation of the anomalous Moses question, *What sea did Moses part?*; Kamas et al., 1996; Park & Reder, 2004). These findings can serve as valuable clues regarding the processes underlying epistemic monitoring. Similarly, the principles suggested by Wyer and Radvansky (1999) that specify conditions under which sentences should be spontaneously recognized as true or false during comprehension, as well as the mechanism underlying this recognition, merit further empirical investigation and could be integrated into the epistemic view of language comprehension.

Relating Epistemic Monitoring To Reading Outcomes and Reader Characteristics

One major limitation that must be acknowledged is that the research presented in this thesis only tested part of the process model of epistemic validation (Richter, 2011), namely the assumption that comprehension comprises a routine epistemic monitoring process that checks the consistency of new information with the current situation model, the antecedent

text, and general world knowledge. It did not investigate how detected inconsistencies are processed further. Thus, it focused on the question of whether readers detect knowledge violations rather than on how they resolve them. Although the eye-tracking study allows some conclusions regarding the later integrative processing of implausible information, these conclusions are tentative as no outcome of the reading process was measured that would have allowed to validate them. Valuable insights into how readers build mental representations from conflicting information have been provided by other studies (e.g., Braasch, Rouet, Vibert, & Britt, 2012; Hakala & O'Brien, 1995; Johnson-Laird, Girotto, & Legrenzi, 2004; Maier, 2013; Maier & Richter, 2013a; Maier & Richter, 2013b; Walsh & Johnson-Laird, 2009).

An interesting avenue for future research would be to directly relate measures of epistemic monitoring to reading outcomes. For example, it would be useful to employ a non-reactive method for measuring online comprehension, such as eye-tracking, to investigate the processing of misinformation in narratives and relate indicators of misinformation detection (e.g., longer fixations on misinformation) to whether readers subsequently rely on the false information in knowledge tests. Compared to think-aloud protocols, which were related to the reliance on misinformation by Hinze et al. (2013), eye movements would enable a more fine-grained analysis of online comprehension and validation processes. This approach would also allow investigating the effects of different reading goals, such as a receptive versus an epistemic reading goal (e.g., Maier & Richter, 2013a), on epistemic monitoring.

Another important next step would be an in-depth investigation of interindividual differences in validation, both in online measures and reading outcomes of processing false or implausible information. A recent study by Singer and Doering (2013) has provided initial evidence that online validation, as measured by reading times on inconsistent information, is modulated by readers' ability to access world knowledge in the service of comprehension. In a related vein, a study by Otero and Kintsch (1992), which relied on offline reports to measure inconsistency detection, reported large interindividual differences in detection rates between readers, with most readers either detecting almost all or none of the embedded inconsistencies. Thus, the average rates often reported for the detection of and reliance on misinformation may obscure potentially large interindividual differences in readers' ability to successfully validate text information.

A number of authors (e.g., Cook & O'Brien, 2013; Kendeou, 2013; Wittwer & Ihme, 2013) have suggested that individual differences in validation are moderated by readers' standards of coherence (van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011; van

den Broek, Risdén, & Husebye-Hartmann, 1995), which determine the degree of comprehension that a reader attempts to achieve (Linderholm et al., 2004). Standards of coherence are assumed to be the explicit or implicit standards readers use to evaluate their comprehension; if readers detect a violation of these standards, they are assumed to initiate strategic processing to try and resolve the coherence break (Linderholm et al., 2004; van den Broek et al., 2011). These standards of coherence, in turn, are assumed to depend on characteristics of the reader, the text and the reading situation (van den Broek et al., 2011).

In line with the notion of individual differences in validation, visual inspection of individual data obtained using the epistemic Stroop paradigm has shown that not all participants exhibit an epistemic Stroop effect. However, it is unclear at this point whether this can be taken as evidence for poor validation, as the absence of a Stroop-like pattern could either indicate that knowledge is not being efficiently retrieved and used for validation, or it could indicate the reader's ability to flexibly adapt his or her standards of coherence to the requirements of the task. In order to determine whether a reader's pattern of responses in the epistemic Stroop paradigm could be used as a diagnostic tool for individual validation processes, it would be important to relate it to online measures of validation (e.g., inconsistency detection as measured by reading times or eye movements) and to reading outcomes (e.g., the situation model or reliance on misinformation), as well as to other reader characteristics such as working memory, need for cognition (Cacioppo & Petty, 1982), and epistemological beliefs (Schommer, 1990), that may be relevant for evaluative processing (e.g., Bromme, Kienhues, & Porsch, 2010; Kardash & Scholes, 1996; Richter & Schmid, 2010; Strømsø, Bråten, & Samuelstuen, 2008).

Time Course of Epistemic Monitoring

The research presented in this thesis suggests that another potentially fruitful avenue might be to further explore the time course of epistemic monitoring. The results reported by Wiswede et al. (2013) suggest that the latent response tendencies remain active over quite some time (1800 ms), although it is important to bear in mind that this could also have resulted from participants in the evaluative mindset group actively holding the truth value in mind in case they were prompted to perform the truth evaluation task. Nonetheless, it can be assumed that the strength of the positive and negative response tendencies elicited by epistemic monitoring waxes and wanes as the semantic integration of information proceeds. The results of Studies 1 and 2 suggest that at an interval of 300 ms between the onset of the target word and the onset of the response prompt, a positive response tendency for plausible

or valid information is strongest, as indicated by the facilitation of positive responses after valid or plausible information. However, there was also some indication of a negative response tendency facilitating negative and interfering with positive responses; but this was clearly weaker than the positive response tendency as the effects fell short of significance. It appears that this matter still merits some further investigation to determine whether the pattern of facilitation and interference changes across different intervals, thus allowing conclusions regarding the development of the latent response tendencies (and, as a consequence, regarding the underlying processes of epistemic monitoring).

Relationship Between Depth of Processing and Validation

Based on the results presented in Study 2, it also seems desirable to further explore the relation between depth of comprehension and epistemic monitoring by varying the depth of processing required by the experimental task. The results of Study 2 suggests that using comprehension questions requiring different depths of processing is an appropriate method for this purpose as it avoids the problems of task-switching which were evident in the study by Wiswede et al. (2013). Thus, directly comparing the effects of different comprehension questions which encourage more or less deep processing on the epistemic Stroop effect would allow further conclusions regarding the relationship between comprehension and validation.

Potential Interactions of Certainty and Plausibility

The research presented in Study 3 should be followed-up by an investigation of potential interactions of the certainty of information (as signaled by epistemic markers) and its plausibility. As uncertainty and certainty markers were found to modulate the effects of predictability on comprehension indicators in a similar way in Experiment 1 of this study, only uncertainty markers were included in Experiment 2. However, it is conceivable that the two types of epistemic markers differentially affect validation processes, comparable to the modulating effects of factive versus nonfactive verbs reported in other studies (Ferretti et al., 2013; Ferretti, Singer, & Patterson, 2008; Singer, 2006). An eye-tracking study crossing the type of epistemic marker with plausibility would be useful to clarify this issue. In addition, relating the eye movements to outcome measures of comprehension would allow for a direct test of whether information marked as uncertain is less likely to be integrated into the situation model (Ferretti et al., 2013).

Potential Differences Between Validity and Plausibility

In this thesis, validity and plausibility were heuristically subsumed on a common dimension of goodness of fit with prior knowledge and assumed to have similar effects on comprehension and validation. However, some qualitative differences between these two concepts must be noted. First of all, truth is dichotomous and needs to be established whereas plausibility is continuous and needs to be computed. Second, plausibility and truth can be orthogonal, in the sense that information can be plausible but false. Conversely, false information can vary in its plausibility depending on the goodness of its fit with prior knowledge (e.g., Hinze et al., 2013). Moreover, only plausible information requires situation model updating as true information should already be represented in memory (Ferretti et al., 2013). As information that is judged as true should normally be redundant with the knowledge base, it may require retrieval of an existing situation model rather than construction of a new one (Wyer & Radvansky, 1999). Therefore, investigating potential differences in the processing of validity and plausibility may be a fruitful endeavor. It is important to keep in mind, however, that whether information is considered (im)plausible or (in)valid depends on the knowledge available to the reader: information that can be identified as true or false by knowledgeable readers may be merely plausible or implausible to less knowledgeable readers (e.g., sentences such as *Canberra is the capital of Australia*). To complicate matters further, readers have been shown to sometimes rely on a plausibility strategy rather than direct retrieval to judge the truth of statements (Reder, 1982), making it even more difficult to disentangle the two concepts.

Effects of Text Genre and Presentation Format on Epistemic Monitoring

This thesis focused on epistemic monitoring in short pieces of text ranging from one to three sentences, rather than on validation in larger text contexts. As a consequence, it did not consider influences of text source or genre on validation processes.

An interesting avenue for future research would thus be to investigate the impact of text genre on epistemic monitoring. Richter et al. (2009) proposed that because fictional stories mentally transport readers into a narrative world, they induce an epistemic mindset of suspension of disbelief which makes readers more susceptible to persuasion. In line with this notion, Green and Brock (2000) reported that readers who were more strongly transported into a story detected less misinformation, regardless of whether the story was presented as fact or fiction. They suggested that transportation hampers critical evaluation of story information as it results in “a loss of access to real-world information” (Green & Brock, 2000,

p. 703). On the other hand, recent evidence suggests that people rely less on misinformation from science-fiction texts than realistic fiction, indicating that the realism of a fictional story plays a role in the misinformation effect (Rapp, Hinze, Slaten, & Horton, 2013).

Van den Broek et al. (2011) suggested that text genre may influence the standards of coherence readers apply during comprehension: As expository texts are usually associated with study purposes and narrative texts with leisure reading, expository texts might generally evoke the application of stricter standards compared to narrative texts. Moreover, they suggest that narrative texts should induce more emphasis on standards of referential and causal coherence, while expository texts are assumed to emphasize logical coherence and integration with background knowledge. As a consequence, information that violates readers' prior knowledge should be more likely to be detected in expository texts than in narratives. Thus, the emerging question is whether text genre affects the operation of epistemic monitoring: Is epistemic monitoring reduced in narratives and is this reduction mediated by the level of transportation into the narrative (Appel & Richter, 2010; Richter et al., 2009)? This question is in the focus of a new series of experiments in our laboratory.

Besides text genre, another factor that has recently been shown to affect the processing of conflicting information is presentation format: In a study by Stadtler, Scharrer, Brummernhenrich, and Bromme (2013), readers exhibited better memory for conflicting information and were more likely to convey both sides of a conflict in an essay when the conflicting information was presented across multiple documents rather than within a single document. This indicates that epistemic monitoring may function differently depending on the presentation format of the information, which would be worthwhile to investigate with online measures of epistemic monitoring in future research.

Potential Interactions of Plausibility and Source Credibility

In a related vein, it may also be a fruitful endeavor to explore potential interactions between the credibility of a source and the plausibility of the text content: High (low) credibility of a source might reduce (enhance) epistemic monitoring; at the same time, the perceived plausibility of the text content might influence judgments of the overall credibility of the source. The study by Hinze et al. (2013) showed that the presence of implausible misinformation in a narrative elicited overall more skeptical responses in think-aloud protocols (even for correct information), suggesting at least some degree of “evaluative spillover” to other information presented within the same document. Similarly, Braasch et al. (2012) found that detection of within-text conflicts increases attention to source information.

Conversely, Lombardi, Seyranian, and Sinatra (2013) found a positive effect of source credibility on the perceived plausibility of scientific statements about climate change. Given these findings, it appears plausible that the detection of false, implausible, or inconsistent information in a text and the attention to source information such as credibility may reciprocally influence each other in a cyclical process, as speculated by Stadtler et al. (2013).

Epistemic Monitoring for Consistency With Pictures

Most studies that compare the processing of valid or plausible with the processing of invalid or implausible information suffer from the limitation that these types of information often differ in factors other than validity or plausibility. For instance, the semantic associations between words in valid or plausible assertions are usually stronger than between words in valid or implausible assertions, which may result in confounds of validity or plausibility with the degree of semantic association. The presence of this confound in the reported studies cannot be ruled out, although at least for the stimuli used in Study 2, Richter et al. (2009) found no significant differences in semantic association between valid and invalid items based on latent semantic analysis (Landauer & Dumais, 1997) cosines.

An alternative way to deal with this problem is to use assertions which are not valid or invalid per se, but only with reference to a specific situation, such as a simultaneously presented picture depicting a situation that renders the assertion true or false (comparable to the paradigms employed by Carpenter & Just, 1975, or Clark & Chase, 1972). A recent study in our lab has replicated the epistemic Stroop effect using this method; moreover, it has provided initial evidence that the epistemic Stroop effect extends to spoken language (Piest, 2013).

Epistemic Monitoring of Consistency With Personal Beliefs and Attitudes

It seems likely that readers do not only validate new information against their knowledge, but also against their personal attitudes and beliefs. In line with this idea, a study by van Berkum, Holleman, Nieuwland, Otten, and Murre (2009) showed that personal beliefs (such as *Euthanasia is acceptable / unacceptable*) affect even very early stages of meaning analysis, in the sense that information which contradicts a person's beliefs elicits an enhanced N400 ERP. This raises the question of whether the task-irrelevant consistency of information with a person's attitudes and beliefs would also elicit a compatibility effect in the epistemic Stroop paradigm. If readers cannot ignore the consistency of information with their personal attitudes and beliefs, they may find it more difficult to respond positively after reading information that contradicts their beliefs and to respond negatively after reading information

that is in line with their beliefs. In fact, the epistemic Stroop effect for personal attitudes and beliefs may be even stronger than for general declarative knowledge, because attitudes and beliefs also have an emotional component which might affect the strength of the respective response tendencies (Rosenberg & Hovland, 1960). In addition, it would be particularly informative to relate individual differences in the presence or size of the compatibility effect to reading outcomes for texts which concern the respective attitudes and beliefs. A series of experiments in our lab is currently investigating this issue.

Final Conclusions

The research presented in this thesis constitutes further evidence against the claim that “You Can’t Not Believe Everything You Read” (Gilbert et al., 1993, p. 221); in contrast, it provides additional support for the epistemic view of language comprehension (Richter, 2003; Richter, 2011; Richter et al., 2009; Schroeder, Richter, & Hoever, 2008), which assumes epistemic monitoring to be a regular component of comprehension (*comprehension entailing validation*). In addition, this thesis has further specified the scope of the epistemic view of language comprehension by providing evidence for early and nonstrategic effects of plausibility across different stimuli (valid/invalid vs. plausible/implausible information), different tasks (orthographical judgments, color change judgments, probe identification, and normal reading), and different process indicators (response latencies and eye movements). Finally, the reported results also allow tentative conclusions regarding the mechanisms underlying these effects and highlight a number of potentially fruitful avenues for future research.

The main conclusion of this thesis is that readers do not seem to need an evaluative processing goal to assess the consistency of linguistic input with their general world knowledge. In fact, a review of the literature has suggested that readers often, by default, do *not* pursue such an evaluative processing goal. Rather, the results presented here (in line with other empirical evidence) suggest that validation is a “by-product” of comprehension when relevant knowledge is easily accessible, because under these circumstances, comprehension and validation of information rely on partly the same processes (cp. Wyer & Radvansky, 1999). I concur with Singer’s (2006) proposal that the reported failures of validation should be seen as diagnostic of how validation (and comprehension) function, rather than as evidence for a general gullibility on the part of the reader.

To summarize, the results of the studies reported in this thesis, in conjunction with the reviewed research, support the conclusion that language has a built-in plausibility monitoring

mechanism that, in general, alerts readers to inconsistencies with their knowledge or prior text portions and thereby generally promotes accurate and stable representations. This process is likely to aid comprehension processes such as syntactic analysis and reference resolution by checking whether the established analysis is actually plausible with regard to world knowledge. These findings speak against two-step models that assume a strict separation of comprehension and validation in distinct, non-overlapping stages of processing. Rather, some analysis of real-world truth or plausibility appears to be carried out in the earliest stages of comprehension, although this analysis may often be incomplete. Epistemic monitoring should therefore be explicitly taken into account by theories of text comprehension, and its underlying mechanisms, as well as its relation to reading outcomes and characteristics of the reader and the text, need to be further explored by future research.

References

- Appel, M., & Richter, T. (2010). Transportation and need for affect in narrative persuasion: A mediated moderation model. *Media Psychology*, 13(2), 101–135.
- Baker, L. (1989). Metacognition, comprehension monitoring, and the adult reader. *Educational Psychology Review*, 1(1), 3–38.
- Baker, L., & Wagner, J. L. (1987). Evaluating information for truthfulness: The effects of logical subordination. *Memory & Cognition*, 15(3), 247–255.
- Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21(4), 477–487.
- Bohan, J. T. (2008). *Depth of processing and semantic anomalies* (Doctoral dissertation). University of Glasgow, Glasgow.
- Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory*, 18(6), 670–678.
- Braasch, J. L. G., Rouet, J.-F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, 40(3), 450–465.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67(2), 135–144.
- Bromme, R., Kienhues, D., & Porsch, T. (2010). Who knows what and who can we believe? Epistemological beliefs are beliefs about knowledge (mostly) attained from others. In L. D. Bendixen, & F. C. Feucht (Eds.), *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice* (pp. 163–193). Cambridge: Cambridge University Press.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1), 45–73.
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3(3), 472–517.
- Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science*, 30(1), 95–120.

- Cook, A. E., & O'Brien, E. J. (2013). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2013.855107
- Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 540–551.
- Ferretti, T. R., Singer, M., & Harwood, J. (2013). Processes of discourse integration: Evidence from event-related brain potentials. *Discourse Processes*, 50(3), 165–186.
- Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition*, 108(3), 881–888.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46(2), 107–119.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59(4), 601–613.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65(2), 221–233.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10(6), 597–602.
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 7, Speech Acts* (pp. 41–58). New York: Academic Press.
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 801–811.
- Hakala, C. M., & O'Brien, E. J. (1995). Strategies for resolving coherence breaks in reading. *Discourse Processes*, 20(2), 167–185.
- Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-differences perspective. *Memory & Cognition*, 29(3), 449–461.
- Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2013). Pilgrims sailing the Titanic: Plausibility effects on memory for misinformation. *Memory & Cognition*. Advance online publication. doi:10.3758/s13421-013-0359-9
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, New York: Oxford University Press.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

- Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, 111(3), 640–661.
- Kamas, E. N., Reder, I. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition*, 24(6), 687–699.
- Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology*, 88(2), 260–271.
- Kaup, B., & Foss, D. J. (2005). Detecting and processing inconsistencies in narrative comprehension. In D. T. Rosen (Ed.), *Trends in experimental psychology research* (pp. 67–90). Hauppauge, NY: Nova Science Publishers.
- Kendeou, P. (2013). Validation and comprehension: An integrated overview. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2013.855874
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 854–865.
- Kennedy, A., Murray, W., & Boissiere, C. (2004). Parafoveal pragmatics revisited. *European Journal of Cognitive Psychology*, 16(1-2), 128–153.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240.
- Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*, 37(2), 165–186.
- Lombardi, D., Seyranian, V., & Sinatra, G. M. (2013). Source effects and plausibility judgments when reading about climate change. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2013.855049
- Maier, J. (2013). *Learning scientific information from multiple texts: The role of epistemic validation* (Doctoral dissertation). University of Kassel, Kassel.
- Maier, J., & Richter, T. (2013a). How nonexperts understand conflicting information on social science issues: The role of perceived plausibility and reading goals. *Journal of Media Psychology: Theories, Methods, and Applications*, 25(1), 14–26.
- Maier, J., & Richter, T. (2013b). Text-belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, 31(2), 151–175.

- Marsh, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments, & Computers*, 36(4), 650–655.
- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, 3(6), 1417–1429.
- Murray, W. S. (1998). Parafoveal pragmatics. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 181–199). Oxford, England: Elsevier.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26(2&3), 131–157.
- Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, 19(3), 338–368.
- Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, 3(4), 229–235.
- Park, H., & Reder, L. M. (2004). Moses illusion: Implication for human cognition. In R. F. Pohl (Ed.), *Cognitive Illusions* (pp. 275–291). Hove, UK: Psychology Press.
- Pickering, M. J., & Traxler, M. J. (2000). Parsing and incremental understanding during reading. In M. W. Crocker, M. J. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 238–258). Cambridge, UK, New York: Cambridge University Press.
- Piest, B. (2013). *Ich glaube nicht, was du da sagst: Validierung audio-visuell vermittelter Informationen erfolgt bei starkem Hintergrundwissen automatisch* (Unpublished bachelor thesis). University of Kassel, Kassel.
- Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, 89(3), 250–280.
- Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: The Moses illusion revisited. In A. Graesser & G. Bower, (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 233–258). New York: Academic Press.
- Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses Illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, 30(4), 385–406.
- Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen*. Lengerich: Pabst.
- Richter, T. (2011). Cognitive flexibility and epistemic validation in learning from multiple texts. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility* (pp. 125–140). Dordrecht: Springer Netherlands.

- Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning*, 5(1), 47–65.
- Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96(3), 538–558.
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1–14). New Haven, CT: Yale University Press.
- Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26(2&3), 159–190.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382–386.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498–504.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59(3), 237–255.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54(4), 574–591.
- Singer, M., & Doering, J. C. (2013). Exploring individual differences in language validation. *Discourse Processes*. Advance online publication. doi:10.1080/0163853X.2013.855534
- Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition and Instruction*, 31(2), 130–150.
- Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6), 1162–1169.
- Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and Instruction*, 18(6), 513–527.

- Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, 11(5), 882–888.
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47(1), 69–90.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35(3), 454–475.
- van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, 20(9), 1092–1099.
- van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. J. Schraw (Eds.), *Text relevance and learning from text*. (pp. 123–139). Greenwich, CT: Information Age Publishing.
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39(2&3), 299–316.
- van den Broek, P., Risdén, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Hillsdale, NJ: Erlbaum.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1998). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Hove, UK: Psychology Press.
- van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, 74(1), 35–46.
- van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. *Language and Cognitive Processes*, 5(2), 105–113.
- Walsh, C. R., & Johnson-Laird, P. N. (2009). Changing your mind. *Memory & Cognition*, 37(5), 624–631.

- Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, 8(6), 647–653.
- Wittwer, J., & Ihme, N. (2013). Reading skill moderates the impact of semantic similarity and causal specificity on the coherence of explanations. *Discourse Processes*. Advance online publication. doi: 10.1080/0163853X.2013.855577
- Wyer, R. S., & Radvansky, G. A. (1999). The comprehension and validation of social information. *Psychological Review*, 106(1), 89–118.
- Zabrocky, K., Moore, D., & Schultz, N. R. (1987). Evaluation of comprehension in young and old adults. *Developmental Psychology*, 23(1), 39–43.

Universität Kassel, Fachbereich Humanwissenschaften
Erklärung zum Eigenanteil (Dissertationen im Promotionsfach Psychologie)

Erklärung zum Eigenanteil an den veröffentlichten oder zur Veröffentlichung vorgesehenen wissenschaftlichen Schriften innerhalb meiner Dissertationsschrift gemäß § 5a Abs. 4 Satz 1 der Allgemeinen Bestimmungen für Promotionen an der Universität Kassel vom 13. Juni 2012:

I. Allgemeine Angaben

Isberner, Maj-Britt

Institut für Psychologie, Universität Kassel

Thema der Dissertation: The role of epistemic monitoring in language comprehension

II. Nummerierte Aufstellung der eingereichten Schriften:

1. Isberner, M.-B. & Richter, T. (in press). Comprehension and validation: Separable stages of information processing? A case for epistemic monitoring in language comprehension. In D. N. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Boston, MA: MIT Press.
2. Isberner, M.-B. & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, 142, 15-22.
3. Isberner, M.-B. & Richter, T. (in press). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes*.
4. Isberner, M.-B., Richter, T., & Kaakinen, J. (2013). Epistemic modality in sentence comprehension: Effects of epistemic adverbs on eye movements. Manuscript submitted for publication to *Discourse Processes* (acknowledgment of receipt: 1.11.2013).

III. Darlegung des eigenen Anteils an diesen Schriften:

zu Nr. 1: Anteile an der Entwicklung der Konzeption, Literaturrecherche, Erstellen des Manuskriptes, Beweisführung:

15% Prof. Dr. Tobias Richter, 85% Dipl.-Psych. Maj-Britt Isberner

zu Nr. 2 und 3: Anteile an der Entwicklung der Konzeption, Literaturrecherche, Methodenentwicklung, Versuchsdesign, Datenerhebung, Datenauswertung, Ergebnisdiskussion, Erstellen des Manuskriptes, Programmierung, Beweisführung:

15% Prof. Dr. Tobias Richter, 85% Dipl.-Psych. Maj-Britt Isberner

• ERKLÄRUNG ZUM EIGENANTEIL •

Zu Nr. 4: Anteile an der Entwicklung der Konzeption, Literaturrecherche, Methodenentwicklung, Versuchsdesign, Datenerhebung, Datenauswertung, Ergebnisdiskussion, Erstellen des Manuskriptes, Programmierung, Beweisführung: 40% Prof. Dr. Tobias Richter, 40% Dipl.-Psych. Maj-Britt Isberner, 20% Dr. Johanna Kaakinen

IV. Anschriften der Mitautoren:

Prof. Dr. Tobias Richter: tobias.richter@uni-kassel.de

Dr. Johanna Kaakinen: johkaa@utu.fi

Datum, Unterschrift der Antragstellerin:

.....
Ort, Datum

.....
Unterschrift

Ich bestätige die von Dipl.-Psych. Maj-Britt Isberner unter Pkt. III abgegebene Erklärung:

Prof. Dr. Tobias Richter:
Ort, Datum

.....
Unterschrift

Dr. Johanna Kaakinen:
Ort, Datum

.....
Unterschrift

Eidesstattliche Versicherung und Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.

.....
Ort, Datum

.....
Unterschrift

• EIDESSTATTLICHE VERSICHERUNG UND ERKLÄRUNG •

Acknowledgements

I would like to thank a number of people who, directly or indirectly, have had an influence on this thesis.

First of all, I am grateful to my advisor, Prof. Dr. Tobias Richter, for his unrelenting support and guidance in the preparation of this thesis. More generally, he has fostered my academic development by encouraging independence while at the same time never failing to offer support and feedback when it was needed. I would also like to express my gratitude to the other members of my doctoral committee: To Prof. Dr. Barbara Kaup, for having welcomed me so warmly into her group in Tübingen and for her valuable comments and suggestions on parts of the research presented in this thesis, and to Prof. Dr. Mirjam Ebersbach and Prof. Dr. Holden Härtl, for their interest in this work.

During my graduate studies, I have benefited from the doctoral support programs of the Special Priority Program “Science and the General Public” funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) and of the Research Initiative “Forschungsinitiative Sprachdiagnostik und Sprachförderung” funded by the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung). I am grateful to the organizers and coordinators of both of these programs, and for all the stimulating exchanges at the various meetings and workshops.

A heart-felt thank you also goes to my colleagues and collaborators along the way for many helpful and inspiring discussions; in particular, to Dr. Johanna Kaakinen, for introducing me to eye-tracking and helping me with technical questions; to David Nitz, for his assistance, technical and otherwise, in the early stages of this thesis; and to Julia Knoepke and Isabel Lindner, who never ceased to support me with both their expertise and friendship.

I would also like to thank our student assistants, in particular Irina Fix, Theresa Hahn, and Benjamin Piest, who assisted me with quite a number of tedious tasks at different stages of the studies.

Last but not least, I am thankful to a number of people who have been an invaluable support throughout my life:

To (Dr.!) Benjamin Ernst, for taking the time to read and give me feedback on a draft of this thesis, and for never failing to brighten up the daily grind with his incomparable humor.

To my friends, in particular Anna Fuhrmann, Daniel Göhring, Ricarda Moll, and Helena Zanter, for their incessant encouragement, in particular during the final stages of this thesis.

To my family – my parents, Gabriele Kern-Isberner and Klaus Isberner, my siblings, Silja and Malte, and my grandparents – for their inspiration and their faith in me.

And finally, to Marc Grosjean, for his truly unconditional love and support.

• ACKNOWLEDGEMENTS •

Appendix

Table of Contents

Appendix A: Stimulus Material for Study 1	3
Experimental Items	4
Filler Items	12
Appendix B: Stimulus Material for Study 2	25
Experimental Items	26
Filler Items	31
Appendix C: Stimulus Material for Study 3	35
Experimental Items for Experiment 1	36
Experimental Items for Experiment 2	47

Appendix A

Stimulus Material for Study 1

Experimental Items

Table A1

Experimental Items Used in Study 1 (Experiments 1 and 2)

No.	Context		Target Word			
			Correct		Misspelled (Exp. 1)	
	Plausible	Implausible	Predictable	Non-predictable	Predictable	Non-predictable
1	Sebastian ist Florist. Er verkauft	Sebastian ist Automechaniker. Er verkauft	Blumen	Pflanzen	Bluhmen	Pflansen
2	Nicolas hat schlechte Augen. Darum braucht er eine	Nicolas kann schlecht hören. Darum braucht er eine	Brille	Sehhilfe	brille	Sehilfe
3	Heute ist Freitag. Morgen ist	Heute ist Mittwoch. Morgen ist	Samstag	Wochenende	Samstak	Wochennende
4	Anna ist Augenoptikerin. Sie verkauft	Anna ist Versicherungskauffrau. Sie verkauft	Brillen	Kontaktlinsen	Brilen	Kontacktlinsen
5	Simon ist übergewichtig. Der Arzt empfiehlt ihm eine	Simon ist untergewichtig. Der Arzt empfiehlt ihm eine	Diät	Fastenkur	Diet	Fastenkuhr
6	Juliane hat wenig geschlafen. Jetzt ist sie	Juliane hat ausgeschlafen. Jetzt ist sie	müde	erschöpft	mühde	erschöbft

7	Dominik hat eine Katze. Sie fängt	Dominik hat eine Kröte. Sie fängt	Mäuse	Vögel	Mäuse	Vögel
8	Oliver ist ein verantwortungsvoller Radfahrer. Er trägt immer einen	Oliver ist ein gutgekleideter Manager. Er trägt immer einen	Helm	Kopfschutz	Helm	Kopfschutz
9	Sarah ist Erzieherin. Sie arbeitet nämlich gerne mit	Sarah ist Gärtnerin. Sie arbeitet nämlich gerne mit	Kindern	Menschen	kindern	menschen
10	John ist unzufrieden. Darum möchte er sein Leben	John ist zufrieden. Darum möchte er sein Leben	ändern	verbessern	Ändern	verbässern
11	Holger ist Workaholic. Er ist süchtig nach	Holger ist Alkoholiker. Er ist süchtig nach	Arbeit	Erfolg	Abeit	Erfolk
12	Wolfgang hat Migräne. Deswegen leidet er an	Wolfgang hat Hautausschlag. Deswegen leidet er an	Kopfschmerzen	Übelkeit	Kopfschmertzen	Übelkeidt
13	Stefan ist gesellig. Er hat viele	Stefan ist Einzelgänger. Er hat viele	Freunde	Bekannte	Fräunde	Bekannte
14	Tobias backt sich eine Pizza. Darauf streut er	Tobias backt sich eine Waffel. Darauf streut er	Käse	Gewürze	Khäse	Gewürrze
15	Marion hat einen Hund. Den Nachbarn stört sein häufiges	Marion hat einen Wellensittich. Den Nachbarn stört sein häufiges	Bellen	Winseln	Bällen	Winnseln

• APPENDIX A • STIMULUS MATERIAL FOR STUDY 1 •

16	Theo ist Arzt. Er arbeitet im	Theo ist Anwalt. Er arbeitet im	Krankenhaus	Kreißsaal	Kranckenhaus	Kraißsaal
17	Martins Freundin ist schwanger. Deshalb geht er mit ihr zum	Martins Freundin ist depressiv. Deshalb geht er mit ihr zum	Frauenarzt	Gynäkologen	Frauenartzt	Gynekologen
18	Selma liebt Tiere. Darum hat sie eine	Selma liebt Kinder. Darum hat sie eine	Katze	Schlange	Katzse	schlange
19	Christian findet sein Zimmer zu dunkel. Darum kauft er eine	Christian findet sein Zimmer zu hell. Darum kauft er eine	Lampe	Deckenleuchte	Lammpe	deckenleuchte
20	Ben hat keine Freunde. Er fühlt sich	Ben hat viele Freunde. Er fühlt sich	einsam	allein	einsahm	Allein
21	Andreas denkt immer nur an sich. Er ist	Andreas denkt immer an andere. Er ist	egoistisch	selbstsüchtig	egoißtisch	selbssüchtig
22	Maike hat Geburtstag. Darum bekommt sie	Maike hat Hunger. Darum bekommt sie	Geschenke	Besuch	Geschencke	Behsuch
23	Bruno arbeitet als Bäcker. Er verkauft frische	Bruno arbeitet als Metzger. Er verkauft frische	Brötchen	Bienenstiche	Bröhtchen	Binenstiche
24	Stephi mag Süßigkeiten. Am liebsten isst sie	Stephi mag keine Süßigkeiten. Am liebsten isst sie	Schokolade	Weingummis	Schockolade	Weingumis

25	Marc bestellt sich einen Kaffee. Darein gießt er	Marc bestellt sich einen Sekt. Darein gießt er	Milch	Sahne	milch	Saane
26	Peter besorgt einen Blumenstrauß für seine Freundin. Er kauft	Peter besorgt Getränke für eine Feier. Er kauft	Rosen	Nelken	Rohsen	Nelcken
27	Gaby wäscht ihre Haare. Dazu benutzt sie	Gaby wäscht ihre Kleidung. Dazu benutzt sie	Shampoo	Spülung	Shampoo	Spühlung
28	Leonie hat Gürtelrose. Sie geht zum	Leonie hat Angststörungen. Sie geht zum	Hautarzt	Dermatologen	Hautarzt	Dermathologen
29	Anton ist Metzger. Er verkauft	Anton ist Florist. Er verkauft	Fleisch	Würste	Flaisch	Würrste
30	Christine hat großen Durst. Darum bestellt sie sich ein	Christine hat großen Hunger. Darum bestellt sie sich ein	Wasser	Getränk	wasser	Getrenk
31	David verkauft seinen Computer. Er braucht nämlich	David kauft sich einen Computer. Er braucht nämlich	Geld	Platz	Geldt	Plats
32	Charlie ist Angler. Heute fing er einen großen	Charlie ist Insektensammler. Heute fing er einen großen	Fisch	Lachs	fisch	Lacks
33	Marc und Gerhard trinken Kaffee. Dazu essen sie	Marc und Gerhard trinken Bier. Dazu essen sie	Kuchen	Torte	Kuchn	Thorte

34	Pascal steht nicht gerne im Mittelpunkt. Er ist sehr	Pascal steht gerne im Mittelpunkt. Er ist sehr	schüchtern	zurückhaltend	Schüchtern	zurückhaltent
35	Mara will sich operieren lassen. Dafür geht sie zum	Mara will sich malen lassen. Dafür geht sie zum	Arzt	Chirurgen	Artzt	Chiruhrgen
36	Karl gibt ungern Geld aus. Er ist sehr	Karl gibt gerne Geld aus. Er ist sehr	geizig	sparsam	geitzig	spahrsam
37	Rainer ist Bauer. Er fährt einen	Rainer ist Rennfahrer. Er fährt einen	Traktor	Mähdrescher	Tracktor	Mähdräsker
38	Pablo studiert Medizin. Er möchte nämlich gerne Menschen	Pablo studiert Medienwissenschaften. Er möchte nämlich gerne Menschen	helfen	heilen	hälfen	hailen
39	Erik ist alt. Darum hat er viele	Erik ist jung. Darum hat er viele	Falten	Erfahrungen	Fallten	Ervahrungen
40	Christoph hat schon viel erlebt. Darum ist er sehr	Christoph hat noch nicht viel erlebt. Darum ist er sehr	erfahren	weise	erfaren	weise
41	Sandra hatte einen tollen Bühnenauftritt. Dafür erntet sie	Sandra hatte einen schlechten Bühnenauftritt. Dafür erntet sie	Applaus	Beifall	Aplaus	Beiffall
42	Peter ist Gebrauchtwagenhändler. Er verkauft	Peter ist Lebensmittelhändler. Er verkauft	Autos	Fahrzeuge	Auttos	Fahrzüge

43	Frank hat einen Wasserrohrbruch. Er ruft den	Frank hat einen Schienbeinbruch. Er ruft den	Klempner	Handwerker	Klempner	Hantwerker
44	Der kleine Stefan bekommt in seiner Rechenarbeit eine Eins. Sein Vater ist	Der kleine Stefan bekommt in seiner Rechenarbeit eine Fünf. Sein Vater ist	stolz	zufrieden	stollz	zufriden
45	Olli arbeitet in einer Pommeshütte. Er verkauft gerade eine	Olli arbeitet in einer Boutique. Er verkauft gerade eine	Wurst	Frikadelle	Wurßt	Frickadelle
46	Sascha ist Journalist. Er schreibt	Sascha ist Komponist. Er schreibt	Artikel	Kolumne	Atikel	Kolummne
47	Miriam spielt Badminton. Morgen hat sie ein wichtiges	Miriam spielt Keyboard. Morgen hat sie ein wichtiges	Spiel	Turnier	Spiehl	Tunier
48	Michael hat Sarah belogen. Darum bittet er sie jetzt um	Michael hat Sarah geholfen. Darum bittet er sie jetzt um	Verzeihung	Vergebung	Verzeiung	Vergehung
49	Karsten hat einen Hasen überfahren. Darum fühlt er sich	Karsten hat einen Preis gewonnen. Darum fühlt er sich	schuldig	schlecht	schulldig	Schlecht
50	Michael ist Schriftsteller. Er schreibt	Michael ist Sänger. Er schreibt	Romane	Bücher	Rohmane	bücher

51	Thomas möchte Musical-Darsteller werden. Dafür lernt er	Thomas möchte Bäcker werden. Dafür lernt er	singen	tanzen	sinngen	tannzen
52	Daniel ist Dompteur. Sein Beruf ist sehr	Daniel ist Lehrer. Sein Beruf ist sehr	gefährlich	riskant	gefehrlich	risskant
53	Julian ist erfolglos. Das macht ihn	Julian ist erfolgreich. Das macht ihn	traurig	betrübt	traurik	betrüpt
54	Maria putzt regelmäßig ihre Zähne. Deswegen hat sie keine	Maria putzt regelmäßig ihre Wohnung. Deswegen hat sie keine	Karies	Löcher	Kharies	löcher
55	Victor ist elegant gekleidet. Er trägt eine	Victor geht schwimmen. Er trägt eine	Krawatte	Fliege	Kravatte	Flihge
56	Elisa macht Nudeln. Dazu benutzt sie	Elisa macht Hausaufgaben. Dazu benutzt sie	Wasser	Salz	Wassar	Saltz
57	Julia hat den Zug doch noch verpasst. Jetzt ist sie	Julia hat den Zug doch noch bekommen. Jetzt ist sie	sauer	verzweifelt	Sauer	verzwaifelt
58	Susanne isst gerne Geflügel. Am liebsten mag sie	Susanne ist Vegetarierin. Am liebsten mag sie	Hähnchen	Pute	Hänchen	Puhte
59	Julia liest gerne Sachbücher. Darum ist sie sehr	Julia liest gerne Comics. Darum ist sie sehr	gebildet	klug	Gebildet	kluhg

60	Nora spielt in einer erfolgreichen Fußballmannschaft. Sie ist nämlich sehr	Nora spielt in einer erfolgreichen Band. Sie ist nämlich sehr	sportlich	athletisch	spohrtlich	atletisch
61	Maren ist beliebt. Sie ist nämlich sehr	Maren ist unbeliebt. Sie ist nämlich sehr	nett	hilfsbereit	Nett	hilfsbereidt
62	Anne geht in eine Eisdiele. Sie bestellt sich einen	Anne geht in einen Weinkeller. Sie bestellt sich einen	Eisbecher	Milchshake	Eisbächer	Milchshacke
63	Lisa geht in den Park. Sie füttert die	Lisa steht am Aquarium. Sie füttert die	Tauben	Schwäne	tauben	Schwähne
64	Saskia macht anderen gerne Geschenke. Sie ist sehr	Saskia macht anderen ungern Geschenke. Sie ist sehr	großzügig	freigebig	groszügig	freigehbig

Note. Crossing each of the two versions of the context sentence (plausible vs. implausible) with each version of the target word (Experiment 1: predictable/orthographically correct vs. predictable/misspelled vs. non-predictable/orthographically correct vs. non-predictable/misspelled; Experiment 2: predictable vs. non-predictable) resulted in the eight different versions of each item in Experiment 1 (Context X Predictability X Orthographical Correctness) and the four different versions of each item in Experiment 2 (Context X Predictability) that were counterbalanced across participants via item lists.

Filler Items

Table A2

Filler Items Used in Study 1 (Experiments 1 and 2)

No.	Filler Item	Plausibility	Required Response	Position of Target Word	Misspelled Target Word (Exp. 1)
1	Inka ist Kunsthändlerin. Sie verkauft Gemälde.	plausible	negative	5	verkaufft
2	Jan ist Architekt. Er entwirft Gebäude.	plausible	negative	3	Architeckt
3	Jakob ist unglücklich. Er hat Liebeskummer.	plausible	negative	3	unglücklich
4	Simon ist Jazzmusiker. Er spielt Trompete.	plausible	positive	5	spielt
5	Martin ist krank. Er hat Schmerzen.	plausible	positive	3	krank
6	Udo ist Imker. Er züchtet Bienen.	plausible	positive	5	züchtet
7	Georg ist Feuerwehrmann. Er ist sehr mutig.	plausible	negative	3	Feuerwehrman
8	Andreas ist Zivildienstleistender. Er arbeitet im Altenheim.	plausible	negative	3	Ziwildienstleistender
9	Daniela fliegt nicht gerne. Sie hat Höhenangst.	plausible	negative	2	fliehgt
10	Christoph ist Naturforscher. Er macht eine Expedition.	plausible	positive	5	macht

11	Daniel spielt im Streichorchester. Er ist Violinist.	plausible	positive	2	spielt
12	Markus ist Informatiker. Er kann gut programmieren.	plausible	positive	6	gut
13	Marc ist Wissenschaftler. Er ist sehr intelligent.	plausible	negative	3	Wissenschaftler
14	Caroline redet viel. Sie ist sehr extravertiert.	plausible	negative	2	rehtet
15	Nina liebt Nervenkitzel. Darum macht sie Bungee-Jumping.	plausible	negative	4	Darumm
16	Manuela hat einen Papagei. Er kann sprechen.	plausible	positive	4	Papagei
17	Boris liebt Extremsport. Er ist sehr abenteuerlustig.	plausible	positive	6	sehr
18	Karsten ist Moderator. Er arbeitet beim Radio.	plausible	positive	5	arbeitet
19	Marcus ist Tutor. Er kann gut erklären.	plausible	negative	4	Err
20	Ivan ist Dirigent. Er leitet ein Orchester.	plausible	negative	3	Dirigendt
21	Johannes ist Künstler. Er ist sehr kreativ.	plausible	negative	5	istt
22	Christian fährt Motorrad. Er trägt eine Lederjacke.	plausible	positive	2	fährt
23	Rudolf ist gläubig. Er macht eine Pilgerfahrt.	plausible	positive	6	eine
24	Jasna ist Pharmazeutin. Sie entwickelt neue Medikamente.	plausible	positive	3	Pharmazeutin

25	Gregor ist Wirt. Er besitzt ein Restaurant.	plausible	negative	5	besitzt
26	Benni lacht gerne. Er ist sehr humorvoll.	plausible	negative	6	ser
27	Martha schreibt eine Mathematiklausur. Sie löst Rechenaufgaben.	plausible	negative	2	schreibt
28	Jan wird bald Vater. Darum kauft er Babykleidung.	plausible	positive	2	wird
29	Peter muss zum Kieferorthopäden. Er braucht eine Zahnspange.	plausible	positive	4	Kieferorthopäden
30	Norbert sitzt im Gefängnis. Er ist ein Krimineller.	plausible	positive	7	ein
31	Deborah gewinnt ein Tanzturnier. Sie bekommt eine Medaille.	plausible	negative	2	gewint
32	Marcel ist Archäologe. Er interessiert sich für Geschichte.	plausible	negative	5	interessirt
33	Felix mag keinen Fußball. Er spielt lieber Tennis.	plausible	negative	3	kainen
34	Philipp ist Sänger. Er hat eine schöne Stimme.	plausible	positive	7	schöne
35	Amelie besteht ihre Meisterprüfung. Sie erhält eine Urkunde.	plausible	positive	6	erhält
36	David arbeitet diszipliniert. Deswegen ist er sehr produktiv.	plausible	positive	3	diszipliniert
37	Caroline geht zum Karnevalszug. Sie trägt ein Kostüm.	plausible	negative	4	Karnävalszug

38	Bastian trägt einen Ring. Er ist nämlich verlobt.	plausible	negative	7	nähmlich
39	Miriam möchte Model werden. Sie ist sehr fotogen.	plausible	negative	4	werrden
40	Sascha hat eine Schreibblockade. Darum ist er frustriert.	plausible	positive	5	Darum
41	Rabea wandert durch ein Moor. Sie trägt Gummistiefel.	plausible	positive	3	durch
42	Alexander spielt in einer Eishockeymannschaft. Er ist Torhüter.	plausible	positive	5	Eishockeymannschaft
43	Angela süßt ihren Tee. Dazu benutzt sie Honig.	plausible	negative	5	Dazu
44	Gabriele liebt Strategiespiele. Darum spielt sie gerne Schach.	plausible	negative	7	gerne
45	Yannis lernt Vokabeln. Dafür macht er sich Karteikarten.	plausible	negative	3	Vokabelln
46	Jürgen arbeitet bei der Bahn. Er ist Schaffner.	plausible	positive	2	arbeitet
47	Sabine arbeitet in einer Bücherei. Sie ist Bibliothekarin.	plausible	positive	5	Bücherei
48	Anne liebt alte Gemälde. Darum studiert sie Kunstgeschichte.	plausible	positive	6	studiert
49	Laura trägt eine Brille. Sie ist nämlich kurzsichtig.	plausible	negative	2	träkt
50	Anna muss eine Rede halten. Sie hat Lampenfieber.	plausible	negative	7	hadt
51	Justina backt einen Kuchen. Dazu benutzt sie Eier.	plausible	negative	4	Kuchn

52	Jörg ist Doktorand. Er arbeitet an seiner Promotion.	plausible	positive	3	Doktorand
53	Mara klettert auf einen Baum. Sie pflückt Kirschen.	plausible	positive	7	pflückt
54	Klaus ist erfolgreich. Er ist nämlich sehr zielstrebig.	plausible	positive	6	nämlich
55	Tobias ist Sportjournalist. Er berichtet über ein Golfturnier.	plausible	negative	3	Sportjournalist
56	Monika trinkt keinen Alkohol. Sie ist nämlich schwanger.	plausible	negative	4	Alkohohl
57	Anja ist eine zuverlässige Mitarbeiterin. Sie ist immer pünktlich.	plausible	negative	5	Mitabeiterin
58	Sandra geht an den Strand. Sie trägt einen Bikini.	plausible	positive	4	den
59	Simon liegt im Krankenhaus. Er bekommt jeden Tag Besuch.	plausible	positive	8	Tag
60	Claire geht zu einem Ball. Sie trägt ein Abendkleid.	plausible	positive	6	Sie
61	David geht gerne auf Parties. Er ist sehr kontaktfreudig.	plausible	negative	4	auff
62	Morgen ist Weihnachten. Sandra und ihre Freundin backen Plätzchen.	plausible	negative	8	bakken
63	Thomas parkt im Halteverbot. Darum bekommt er einen Strafzettel.	plausible	negative	2	parckt
64	Arnold ist Fußballfan. Er geht jedes Wochenende ins Stadion.	plausible	positive	6	jedes

65	Rabea interessiert sich für Ökonomie. Darum studiert sie Volkswirtschaftslehre.	plausible	positive	2	interessiert
66	Guido ist Tierliebhaber. Darum engagiert er sich für Artenschutz.	plausible	positive	8	für
67	Anja arbeitet in einem Reitstall. Sie liebt nämlich Pferde.	plausible	negative	7	libt
68	Sascha ist Vertrauenslehrer. Er behandelt seine Schüler sehr respektvoll.	plausible	negative	5	behandelt
69	Jürgen liebt die Natur. Darum geht er gerne wandern.	plausible	negative	4	Nathur
70	Christine renoviert ihre Wohnung. Dafür kauft sie neue Tapeten.	plausible	positive	2	renoviert
71	David ist ein guter Projektleiter. Er ist sehr verantwortungsbewusst.	plausible	positive	6	Er
72	Cora kocht sich einen Kamillentee. Sie ist nämlich erkältet.	plausible	positive	5	Kamillentee
73	Marilyns kleiner Sohn hat Geburtstag. Sie kauft ihm ein Feuerwehrauto.	plausible	negative	8	iehm
74	Nina möchte ein Bild aufhängen. Dazu benutzt sie einen Hammer.	plausible	negative	4	Bildt
75	Bernd richtet sein Badezimmer ein. Er kauft sich eine Dusche.	plausible	negative	9	einne

76	Gerhard renoviert sein Arbeitszimmer. Er kauft sich einen neuen Schreibtisch.	plausible	positive	6	kauft
77	Gregor hat seine Uhr verloren. Darum geht er zum Fundbüro.	plausible	positive	9	zum
78	Andreas hat zwei Kinder. Er geht mit ihnen häufig zum Spielplatz.	plausible	positive	7	mit
79	Lukas bringt sein Fahrrad zur Reparatur. Er braucht eine neue Lampe.	plausible	negative	10	näue
80	Heike kann sich gut in die Lage anderer versetzen. Sie ist sehr empathisch.	plausible	negative	8	anderar
81	Karl-Heinz ist Tierschützer. Er liebt Jagdsport.	implausible	negative	3	Tierschüttzer
82	Felix ist Mode-Designer. Er entwirft Baupläne.	implausible	positive	5	entwirft
83	Oskar ist Literaturkritiker. Er rezensiert Kinofilme.	implausible	positive	3	Literaturkritiker
84	Günther ist Winzer. Er erntet Bananen.	implausible	positive	2	ist
85	Tommy ist Krankenpfleger. Seine Arbeitskleidung ist orange.	implausible	negative	5	arbeitskleidung
86	Max arbeitet beim Militär. Er ist Pazifist.	implausible	negative	4	Millitär
87	Elliott ist ein Alligator. Er hat Fell.	implausible	negative	4	Alligathor

88	Isabella hat zwei Geschwister. Sie ist Einzelkind.	implausible	positive	3	zwei
89	Melanie backt Brötchen. Dazu benutzt sie Schuhcreme.	implausible	positive	6	sie
90	Nick gibt ein Konzert. Er ist Pokerspieler.	implausible	positive	4	Konzert
91	Konstantin prahlt gerne. Er ist sehr bescheiden.	implausible	negative	6	Sehr
92	Walter ist Pfarrer. Er trägt eine Uniform.	implausible	negative	5	tregt
93	Yvonne trägt Ohrringe. Sie bestehen aus Pappe.	implausible	negative	3	Ohringe
94	Es ist Mittag. Richard betrachtet die Sterne.	implausible	positive	3	Mittag
95	Celine macht eine Süßspeise. Darüber streut sie Knoblauch.	implausible	positive	6	streut
96	Marian öffnet das Fenster. Ihm ist nämlich kalt.	implausible	positive	5	Ihm
97	Matthias hat keine Haare. Er trägt einen Pferdeschwanz.	implausible	negative	4	Hahre
98	Mark duscht immer lange. Er ist sehr umweltbewusst.	implausible	negative	3	immar
99	Sandra hat ein Bewerbungsgespräch. Sie trägt einen Jogginganzug.	implausible	negative	2	hatt
100	Lars möchte ein Haustier. Er ist nämlich Allergiker.	implausible	positive	2	möchte

101	Uwe liebt Herausforderungen. Er ist nämlich sehr ängstlich.	implausible	positive	4	Er
102	Jasmin mag den Sommer. Sie liebt nämlich Schnee.	implausible	positive	6	liebt
103	Jens muss noch Autofahren. Darum trinkt er Alkohol.	implausible	negative	6	trinkt
104	Jessica ist häufig müde. Sie schläft nämlich gut.	implausible	negative	3	heufig
105	Robin ist nicht vertrauenswürdig. Deswegen ist er Kassenwart.	implausible	negative	7	Er
106	Martina ist Spanischlehrerin. Darum kann sie gut Skifahren.	implausible	positive	3	Spanischlehrerin
107	Hubert geht wandern. Dafür packt er einen Koffer.	implausible	positive	5	packt
108	Sonja macht eine Bergtour. Sie trägt eine Schwimmweste.	implausible	positive	7	eine
109	Kirsten mag keinen Sport. Sie ist nämlich kitzelig.	implausible	negative	4	sport
110	Lea wandert auf einem Gletscher. Sie trägt Stöckelschuhe.	implausible	negative	5	Gletscher
111	Ferdinand ist müde. Darum geht er ins Kino.	implausible	negative	7	inz
112	Veronika ist verheiratet. Darum trägt sie einen Schal.	implausible	positive	4	Darum
113	Martin ist ein Mönch. Er lebt im Internat.	implausible	positive	6	lebt
114	Felix ist beliebt. Er wird von den anderen gehänselt.	implausible	positive	3	beliebt

115	David hat kalte Füße. Darum strickt er sich Handschuhe.	implausible	negative	6	strikt
116	Martina mag dunkle Kleidung. Darum trägt sie gerne gelb.	implausible	negative	4	Kleidung
117	Isabelle möchte schlafen. Darum schaltet sie das Licht an.	implausible	negative	5	schaltet
118	Pauline möchte eine Vorspeise. Sie bestellt sich einen Vanille-pudding.	implausible	positive	6	bestellt
119	Rainer geht zum Raubtiergehege. Dort beobachtet er die Giraffen.	implausible	positive	8	die
120	Franziska hat gute Laune. Das Wetter ist nämlich schlecht.	implausible	positive	6	Wetter
121	Joschua lässt sich leicht beeinflussen. Er ist sehr willensstark.	implausible	negative	2	lässt
122	Sigfried möchte die Prüfung bestehen. Darum lernt er wenig.	implausible	negative	4	Prüfung
123	Michaela studiert Medizin. Dabei lernt sie viel über Politik.	implausible	negative	7	fiel
124	Sabrina fährt mit dem Aufzug. Er ist nämlich defekt.	implausible	positive	5	Aufzug
125	Paul ekelt sich vor Blut. Deswegen ist er Veterinär.	implausible	positive	6	Deswegen
126	Draußen ist es kalt. Darum trägt Hannah eine Shorts.	implausible	positive	4	kalt
127	Michaels Lieblingsfach ist Deutsch. Er kann nämlich gut rechnen.	implausible	negative	2	Lieblingsfach

128	Karl misst seine Körpertemperatur. Dazu benutzt er ein Skalpell.	implausible	negative	4	Körpertemperatur
129	Wolfgang liebt Motorsport. Deswegen träumt er von einem Pony.	implausible	negative	7	vonn
130	Franz geht zu einer Versteigerung. Dort kauft er Mehl.	implausible	positive	5	Versteigerung
131	Beate möchte zum Flughafen. Darum nimmt sie ein Bad.	implausible	positive	6	nimmt
132	Therese trocknet ihre Haare. Dazu benutzt sie einen Dosenöffner.	implausible	positive	8	einen
133	Victoria liebt Camping. Darum kauft sie sich ein Regal.	implausible	negative	7	Sich
134	Laurenz liebt Karaoke. Er ist nämlich ein guter Schwimmer.	implausible	negative	8	guhter
135	Dennis hat Angst vor Mäusen. Darum kauft er eine Alarmanlage.	implausible	negative	3	Ankst
136	Lasse ist bei seinen Mitarbeitern beliebt. Er behandelt sie herablassend.	implausible	positive	8	behandelt
137	Christian liebt italienische Küche. Darum fährt er häufig nach Moskau.	implausible	positive	3	italienische
138	Wiebke ist sehr naturverbunden. Darum schläft sie gerne im Hotel.	implausible	positive	4	naturverbunden
139	Christoph wird nicht gerne nass. Deswegen trägt er eine Uhr.	implausible	negative	5	nas

140	Ulrike trägt einen dicken Pullover. Draußen ist es nämlich warm.	implausible	negative	8	ess
141	Beatrix wärmt ihr Essen auf. Dazu benutzt sie den Kühlschrank.	implausible	negative	7	Benutzt
142	Susanna kocht heiße Schokolade. Sie füllt sie in eine Wärmflasche.	implausible	positive	4	Schokolade
143	Friedrich trägt einen Gürtel. Seine Hose ist nämlich zu kurz.	implausible	positive	9	zu
144	Mike sieht gerne die Nachrichten. Darum hat er einen Papagei.	implausible	positive	2	sieht
145	Alice liegt in der Sonne. Davon bekommt sie einen Schnupfen.	implausible	negative	6	Dafon
146	Nicolas spielt gerne Bowling. Deswegen hat er einen eigenen Schläger.	implausible	negative	5	deswegen
147	Patrick ist ein Glückspilz. Bei Spielen ist er häufig der Verlierer.	implausible	negative	4	Glückspils
148	Cassandra möchte Benzin sparen. Darum fährt sie meistens mit dem Auto.	implausible	positive	8	meistens
149	Gisela hat sich das Bein gebrochen. Deswegen geht sie zum Sport.	implausible	positive	5	Bein
150	Sandra arbeitet ehrenamtlich beim Roten Kreuz. Sie ist nämlich sehr eigennützig.	implausible	positive	10	sehr

151	Herbert kann seine Wohnungstür nicht öffnen. Er hat nämlich keinen Führerschein.	implausible	negative	7	Ehr
152	Pavel kauft sich ein Teleskop. Er interessiert sich nämlich für Golf.	implausible	negative	9	nemlich
153	Leo ist auf öffentliche Verkehrsmittel angewiesen. Er kann nämlich nicht schwimmen.	implausible	negative	4	öffentliche
154	Jonathan räumt regelmäßig sein Zimmer auf. Darum ist es sehr chaotisch.	implausible	positive	3	regelmäßig
155	Kevin macht nicht gerne den Abwasch. Darum kauft er sich eine Kaffeemaschine.	implausible	positive	11	eine
156	Jennifer versteht sich nicht mit ihren Eltern. Sie sind nämlich sehr verständnisvoll.	implausible	positive	7	Eltern
157	Jenna hat sich am Handgelenk verletzt. Darum kann sie jetzt nicht laufen.	implausible	negative	10	jezt
158	Josef bohrt ein Loch in die Wand. Dazu benutzt er einen Fön.	implausible	negative	2	bort
159	Manus interessiert sich für die asiatische Kultur. Darum reist er dieses Jahr nach Frankreich.	implausible	positive	12	Jahr
160	Gary möchte nicht mehr mit dem Bus zur Arbeit fahren. Darum lernt er jonglieren.	implausible	positive	7	Bus

Appendix B

Stimulus Material for Study 2

Experimental Items

Table B1

Experimental Items Used in Study 2, Taken from the Item Pool of Richter, Schroeder, and Wöhrmann (2009, Experiment 4)

No.	Word 1	Word 2	Word 3	Knowledge	Validity	Position of Target Word
1	Bibliotheken	haben	Bücher	high	valid	3
2	Vier	ist eine	Zahl	high	valid	3
3	Hunde	haben	Pfoten	high	valid	3
4	Fußbälle	sind	rund	high	valid	3
5	Cognac	enthält	Alkohol	high	valid	3
6	Parfüm	enthält	Duftstoffe	high	valid	3
7	Jeans	haben	Nähte	high	valid	3
8	Schwarzfahren	ist	verboten	high	valid	3
9	Wasser	ist	nass	high	valid	3
10	Handys	haben	Akkus	high	valid	3
11	Feuerwehrautos	sind	rot	high	valid	3
12	Seuchen	sind	gefährlich	high	valid	3
13	Gold	ist	wertvoll	high	valid	3
14	Disketten	sind	Datenträger	high	valid	3
15	Hauptsätze	enthalten	Wörter	high	valid	3
16	Bücher	haben	Seiten	high	valid	3

• APPENDIX B • STIMULUS MATERIAL FOR STUDY 2 •

17	Blitze	sind	hell	high	valid	3
18	Diamanten	sind	teuer	high	valid	3
19	Schwimmen	ist	gesund	high	valid	3
20	Kartoffeln	sind	nahrhaft	high	valid	3
21	Tulpen	sind	farbig	high	valid	3
22	Kissen	sind	weich	high	valid	3
23	Schokolade	enthält	Kalorien	high	valid	3
24	Cabrios	haben	Motoren	high	valid	3
25	Jalousien	sind	Musikinstrumente	high	invalid	3
26	Elefanten	sind	Zweibeiner	high	invalid	3
27	Flöten	haben	Saiten	high	invalid	3
28	Schach	ist ein	Ballspiel	high	invalid	3
29	Sonnenbrand	ist	hautfreundlich	high	invalid	3
30	Anschnallgurte	sind	unnötig	high	invalid	3
31	Rockkonzerte	sind	leise	high	invalid	3
32	Sauerbraten	ist	vegetarisch	high	invalid	3
33	Wolken	sind	quadratisch	high	invalid	3
34	Computer	haben	Emotionen	high	invalid	3
35	Glühlampen	sind	weich	high	invalid	3
36	Säuglinge	haben	Zähne	high	invalid	3
37	Sportplätze	haben	Kinos	high	invalid	3
38	Kirschtorte	ist	orange	high	invalid	3

• APPENDIX B • STIMULUS MATERIAL FOR STUDY 2 •

39	Locher	sind	Küchengeräte	high	invalid	3
40	Kredite	sind	zinslos	high	invalid	3
41	Theaterbesuche	sind	kostenlos	high	invalid	3
42	Zwieback	ist	flüssig	high	invalid	3
43	Horrorfilme	sind	entspannend	high	invalid	3
44	Falschgeld	ist	legal	high	invalid	3
45	Kontaktlinsen	sind	eckig	high	invalid	3
46	Sportwagen	sind	umweltfreundlich	high	invalid	3
47	Tannen	sind	Laubbäume	high	invalid	3
48	Fledermäuse	sind	tagaktiv	high	invalid	3
49	Dampfmaschinen	haben	Kreuzköpfe	low	valid	3
50	Mamen	sind	Bonsaigewächse	low	valid	3
51	Wollnashörner	sind	ausgestorben	low	valid	3
52	Boykott	ist ein	Eponym	low	valid	3
53	Grünlinge	sind	giftig	low	valid	3
54	Vokativ	ist ein	Kasus	low	valid	3
55	Gewölle	enthalten	Knochen	low	valid	3
56	Superfluid	ist ein	Aggregatzustand	low	valid	3
57	Bronze	ist eine	Legierung	low	valid	3
58	Ostkreta	ist	waldlos	low	valid	3
59	Trockeneis	enthält	Kohlendioxid	low	valid	3
60	Rosmarin	ist ein	Lippenblütler	low	valid	3

• APPENDIX B • STIMULUS MATERIAL FOR STUDY 2 •

61	Rezitative	sind	Sprechgesänge	low	valid	3
62	Autoreifen	enthalten	Kohlenstoff	low	valid	3
63	Benzol	ist	kloppfest	low	valid	3
64	Erdrotation	bewirkt	Polverflachung	low	valid	3
65	Dachse	sind	Marder	low	valid	3
66	Bären	sind	Allesfresser	low	valid	3
67	Kamele	sind	Paarhufer	low	valid	3
68	Goldregen	ist	giftig	low	valid	3
69	Bor	ist ein	Element	low	valid	3
70	Sperrholz	enthält	Furnierlagen	low	valid	3
71	Kiefernharz	enthält	Terpentin	low	valid	3
72	Markstücke	enthalten	Nickel	low	valid	3
73	Protonenzerfall	bewirkt	Elektronenemissionen	low	invalid	3
74	Arabien	ist	föderalistisch	low	invalid	3
75	Rohharz	enthält	Säuren	low	invalid	3
76	Grottenolme	haben	Finger	low	invalid	3
77	Kronkorken	enthalten	Polypropylen	low	invalid	3
78	Pintos	haben	Maulkorbpflicht	low	invalid	3
79	Exportbier	ist	obergärig	low	invalid	3
80	Schwertransporte	sind	maßgerecht	low	invalid	3
81	Blitzlichtpulver	enthält	Natrium	low	invalid	3
82	Flugzeugrümpfe	enthalten	Rohrgerüste	low	invalid	3

• APPENDIX B • STIMULUS MATERIAL FOR STUDY 2 •

83	Meereisblumen	sind	Pflanzen	low	invalid	3
84	Brom	ist ein	Halbmetall	low	invalid	3
85	Nachtschattengewächse	sind	einkeimblättrig	low	invalid	3
86	Grauhörnchen	sind	artengeschützt	low	invalid	3
87	Grüntee	ist	fermentiert	low	invalid	3
88	Glückskatzen	sind	zweifarbig	low	invalid	3
89	Gummibärchen	sind	fettfrei	low	invalid	3
90	Hummeln	haben	Stacheln	low	invalid	3
91	Tiergattungen	enthalten	Familien	low	invalid	3
92	Franziskanerkutten	sind	schwarz	low	invalid	3
93	Sekundenkleber	sind	Haftklebstoffe	low	invalid	3
94	Curry	enthält	Salz	low	invalid	3
95	Orkane	bewirken	Tsunamis	low	invalid	3
96	Meisen	sind	Zugvögel	low	invalid	3

Filler Items

Table B2

Filler Items Used in Study 2, Taken from the Item Pool of Richter, Schroeder, and Wöhrmann (2009, Experiment 4)

No.	Word 1	Word 2	Word 3	Knowledge	Validity	Position of Target Word
1	Zigaretten	sind	ungesund	high	valid	1
2	Segelschiffe	haben	Masten	high	valid	1
3	Bonsais	sind	klein	high	valid	2
4	Spülmittel	ist	fettlösend	high	valid	2
5	Ozeane	enthalten	Salz	high	valid	1
6	Kaffee	enthält	Koffein	high	valid	1
7	Kohlenfeuer	bewirkt	Hitze	high	valid	2
8	Starkstrom	ist	tödlich	high	valid	2
9	Wasser	enthält	Sauerstoff	high	valid	1
10	Cowboystiefel	enthalten	Leder	high	valid	1
11	Karate	ist ein	Kampfsport	high	valid	2
12	Rosen	haben	Blätter	high	valid	2
13	Gürtel	haben	Schnallen	high	valid	1
14	Berge	sind	hoch	high	valid	1
15	Kühe	sind	Wiederkäuer	high	valid	2
16	Müdigkeit	bewirkt	Gähnen	high	valid	2

• APPENDIX B • STIMULUS MATERIAL FOR STUDY 2 •

17	Boskop	ist eine	Birnensorte	high	invalid	1
18	Kalkstein	ist	schwarz	high	invalid	1
19	Piranhas	haben	Lungen	high	invalid	2
20	Schallplatten	enthalten	Silber	high	invalid	2
21	Schachbretter	sind	einfarbig	high	invalid	1
22	Schnee	ist	heiß	high	invalid	1
23	Kopien	sind	einzigartig	high	invalid	2
24	Vorhänge	sind	Kleidungsstücke	high	invalid	2
25	Prüfungen	sind	angenehm	high	invalid	1
26	Eisbären	sind	Haustiere	high	invalid	1
27	Kamillentee	ist	kalorienreich	high	invalid	2
28	Realschüler	haben	Abitur	high	invalid	2
29	Autobahnen	haben	Ampeln	high	invalid	1
30	Holz	ist ein	Metall	high	invalid	1
31	Kameras	sind	Zeitmesser	high	invalid	2
32	Schnupfen	bewirkt	Übelkeit	high	invalid	2
33	Jurten	sind	Zelte	low	valid	1
34	Violinen	haben	Bassbalken	low	valid	1
35	Landauer	sind	Kutschen	low	valid	2
36	Krypton	ist ein	Edelgas	low	valid	2
37	Gelbwurst	enthält	Safran	low	valid	1
38	Essigsäurebakterien	sind	genießbar	low	valid	1

• APPENDIX B • STIMULUS MATERIAL FOR STUDY 2 •

39	Talk	ist ein	Mineral	low	valid	2
40	Transistoren	sind	Halbleiter	low	valid	2
41	Rinder	sind	farbenblind	low	valid	1
42	Mais	ist ein	Süßgras	low	valid	1
43	Wein	enthält	Glycerin	low	valid	2
44	Heringseier	sind	klebrig	low	valid	2
45	Bologna	hat eine	U-Bahn	low	invalid	1
46	Simbabwe	hat	Rechtsverkehr	low	invalid	1
47	Walisisch	hat einen	Hauptdialekt	low	invalid	2
48	Adagios	sind	Tänze	low	invalid	2
49	Schirmladen	sind	Kommoden	low	invalid	1
50	Muscheln	haben	Hörorgane	low	invalid	1
51	Spitzwegerich	ist	harntreibend	low	invalid	2
52	Handkäse	enthält	Rohmilch	low	invalid	2
53	Lurche	sind	Wirbellose	low	invalid	1
54	Diesel	ist	Schweröl	low	invalid	1
55	Indigo	ist ein	Gewürz	low	invalid	2
56	Pinguine	sind	Nestbauer	low	invalid	2

Appendix C

Stimulus Material for Study 3

Experimental Items for Experiment 1

Table C1

Experimental Items Used in Study 3 (Experiment 1)

No.	Context Sentence: Constraint		Target Sentence: Marker			Target Word
	High	Low	No Marker	High Certainty	Low Certainty	
1	Kristiina haluaa tuoksua hyvältä treffeillä	Kristiina pu-keutuu men-näkseen teatteriin	Hän laittaa hajuvettä korvansa taakse	Hän tietenkin laittaa hajuvettä korvansa taakse	Hän kenties laittaa hajuvettä korvansa taakse	hajuvettä
2	Julius on vä-risokea	Juliuksella on vaikeuksia näön kanssa	Hän ei pysty erottamaan punaista ja vihreää	Hän ei tietenkään pysty erottamaan punaista ja vihreää	Hän ei kaiketi pysty erottamaan punaista ja vihreää	punaista
3	Julia nukkui huonosti viime yönä	Julialla oli tapaa-minen eilen illalla	Hän on nyt väsynyt ja hänen on vaikea keskittyä	Totta kai hän on nyt väsynyt ja hänen on vaikea keskittyä	Ehkä hän on nyt väsynyt ja hänen on vaikea keskittyä	väsynyt
4	Martin tyt-töystävä synny-ttää	Martin tyt-töystävällä on vakava ongelma	Martti lähtee hänen kanssaan sairaalaan pitääkseen tyt-töystävästään hyvää huolta	Totta kai Martti lähtee hänen kanssaan sairaalaan pitääkseen tyttöystävästään hyvää huolta	Kenties Martti lähtee hänen kanssaan sairaalaan pitääkseen tyt-töystävästään hyvää huolta	sairaalaan
5	Törmänen pa-keni murhattu-aan vaimonsa kylmäverisesti	Törmänen on joutunut kerras-saan omituiseen tilanteeseen	Suomen kaikki poliisit yrittävät nyt jäljittää häntä	Tietenkin Suomen kaikki poliisit yrittävät nyt jäljittää häntä	Kenties Suomen kaikki poliisit yrittävät nyt jäljit-tää häntä	poliisit

6	Katariina otti hiuksiinsa uuden permanentin	Katariina kävi kampaajalla	Nyt hänellä on kiharat hiukset	Nyt hänellä on tietenkin kiharat hiukset	Nyt hänellä on kaiketi kiharat hiukset	kiharat
7	Seppo on floristi	Sepolla on kiosk	Hän myy kukkia pienessä liikkeessään, joka sijaitsee keskussairaalan aulassa	Hän tietenkin myy kukkia pienessä liikkeessään, joka sijaitsee keskussairaalan aulassa	Hän kaiketi myy kukkia pienessä liikkeessään, joka sijaitsee keskussairaalan aulassa	kukkia
8	Juha menee nyt nukkumaan	Juha on pieni poika	Hänellä on päällään pyjama, jossa on sinisiä tähtiä	Hänellä on tietenkin päällään pyjama, jossa on sinisiä tähtiä	Hänellä on ehkä päällään pyjama, jossa on sinisiä tähtiä	pyjama
9	Pekka on töissä leipomossa	Pekka on töissä kaupassa	Hän myy tuoretta leipää ja muita leivonnaisia	Hän tietysti myy tuoretta leipää ja muita leivonnaisia	Hän kaiketi myy tuoretta leipää ja muita leivonnaisia	leipää
10	Mika on kirjailija	Mika on varsinainen sanaseppo	Hän kirjoittaa parhaillaan kirjaa Suomen suosituimmista vitseistä	Hän kirjoittaa tietenkin parhaillaan kirjaa Suomen suosituimmista vitseistä	Hän kirjoittaa kenties parhaillaan kirjaa Suomen suosituimmista vitseistä	kirjaa
11	Julie tuli Suomeen suoraan Lontoosta	Julie tuli Suomeen suoraan Afrikasta	Hänen äidinkielenä on englantia, jota hän puhuu erikoisella korostuksella	Hänen äidinkielenä on tietysti englantia, jota hän puhuu erikoisella korostuksella	Hänen äidinkielenä on kaiketi englantia, jota hän puhuu erikoisella korostuksella	englanti
12	Markku ja Tero juovat kahvia	Markku ja Tero juovat colaa	He syövät pullaa ja keksejä	He syövät tietysti pullaa ja keksejä	He syövät ehkä pullaa ja keksejä	pullaa
13	Annella on optikkoliike kaupungin keskustassa	Anne työskentelee pienessä liikkeessä kaupungin keskustassa	Hän myy silmälaseja ja suurennuslaseja	Hän myy tietysti silmälaseja ja suurennuslaseja	Hän myy kaiketi silmälaseja ja suurennuslaseja	silmälaseja

14	Markku ei tykkää olla huomion keskipisteenä	Markku on parikymppinen nuori mies	Hän on tosi ujo eikä puhu paljon	Hän on tietysti tosi ujo eikä puhu paljon	Hän on kenties tosi ujo eikä puhu paljon	ujo
15	Timi on korviaan myöten veloissa	Timi on onneton	Hänellä ei ole rahaa maksaa ensi kuun vuokraa	Hänellä ei tietenkään ole rahaa maksaa ensi kuun vuokraa	Hänellä ei ehkä ole rahaa maksaa ensi kuun vuokraa	rahaa
16	Johanna pesee hiuksiaan	Johanna on pesulalla	Hän käyttää shampoota ja hoitoainetta	Hän käyttää tietenkin shampoota ja hoitoainetta	Hän käyttää kaiketi shampoota ja hoitoainetta	shampoota
17	Maarit tilaa kupin kahvia	Maarit leipoo kakkua	Hän laittaa siihen sokeria ja maitoa	Hän tietenkin laittaa siihen sokeria ja maitoa	Hän kaiketi laittaa siihen sokeria ja maitoa	sokeria
18	Matti polki uuden reittiennäytksen mennessään töihin	Matti käveli aamulla töihin	Hän oli saapuessaan hikinen ja hengästynyt	Hän oli saapuessaan tietenkin hikinen ja hengästynyt	Hän oli saapuessaan kenties hikinen ja hengästynyt	hikinen
19	Veikon vaimo on jo pitkään sanonut, että Veikon pitäisi leikkauttaa hiuksensa	Veikon vaimo on jo pitkään sanonut, että Veikon pitäisi huolehtia itsestään paremmin	Veikko meneekin työpäivän jälkeen parturiin siistimään tukkansa	Tietysti Veikko meneekin työpäivän jälkeen parturiin siistimään tukkansa	Ehkä Veikko meneekin työpäivän jälkeen parturiin siistimään tukkansa	parturiin
20	Heikki käynnisti juuri tietokoneensa	Heikki istahti pöytänsä ääreen	Hän lukee sähköpostinsa ja käy sitten töiden kimppuun	Hän lukee tietenkin sähköpostinsa ja käy sitten töiden kimppuun	Hän lukee kenties sähköpostinsa ja käy sitten töiden kimppuun	sähköpostinsa
21	Roope on raskaansarjan nyrkkeilijä	Roope on hermostunut	Hänellä on huomenna tärkeä ottelu Helsingin MM-kisoissa	Hänellä on huomenna tietenkin tärkeä ottelu Helsingin MM-kisoissa	Hänellä on huomenna kenties tärkeä ottelu Helsingin MM-kisoissa	ottelu

22	Holger istuu lempipubissaan	Holger on syömässä ravintolassa	Hän juo olutta ja tilaa sen kanssa pizzan	Hän juo tietysti olutta ja tilaa sen kanssa pizzan	Hän juo kenties olutta ja tilaa sen kanssa pizzan	olutta
23	Valtteriin asunnossa puhkesi vesiputki	Valtterilla on asunnossaan paha ongelma	Hän soittaa putkimiehelle ja pyytää tätä apuun	Hän soittaa tietysti putkimiehelle ja pyytää tätä apuun	Hän soittaa kaiketi putkimiehelle ja pyytää tätä apuun	putkimiehelle
24	Jouluaatto valkeni aurinkoisena ja lumisena	Aamu valkeni aurinkoisena ja lumisena	Isä haki kuusen sisälle ja lapset koristelivat sen kauniiksi	Tietenkin isä haki kuusen sisälle ja lapset koristelivat sen kauniiksi	Kenties isä haki kuusen sisälle ja lapset koristelivat sen kauniiksi	kuusen
25	Simo on yli-painoinen	Simo tuntee olonsa heikkovointiseksi	Hänen lääkärintensä suosittelee laihduttamista, jotta hänen olonsa paranisi	Hänen lääkärintensä tietysti suosittelee laihduttamista, jotta hänen olonsa paranisi	Hänen lääkärintensä kaiketi suosittelee laihduttamista, jotta hänen olonsa paranisi	laihduttamista
26	Tänään on Maikun syntymäpäivä	Maikku potee kipeänä sairaalassa	Hän saa paljon lahjoja ja kukkasia	Tietenkin hän saa paljon lahjoja ja kukkasia	Ehkä hän saa paljon lahjoja ja kukkasia	lahjoja
27	Hannes treenaa päivittäin kuntosalilla	Hannes on töissä rakennustyömaalla	Hänellä on isot lihakset, joita hän esittelee mielellään	Tietenkin hänellä on isot lihakset, joita hän esittelee mielellään	Kenties hänellä on isot lihakset, joita hän esittelee mielellään	lihakset
28	Ossi on vastuuntuntoinen pyöräilijä	Ossi on luonteeltaan varovainen	Hän käyttää aina kypärää pyöräillessään	Tietenkin hän käyttää aina kypärää pyöräillessään	Kenties hän käyttää aina kypärää pyöräillessään	kypärää
29	Pertillä ei ole yhtään ystävää	Pertillä on ongelma	Hän on yksinäinen ja kaipaisi lomamatkoilleen seuraa	Hän on tietysti yksinäinen ja kaipaisi lomamatkoilleen seuraa	Hän on kaiketi yksinäinen ja kaipaisi lomamatkoilleen seuraa	yksinäinen
30	Taneli ja Ville menevät elokuviin	Taneli ja Ville menevät ruoka-kauppaan	He ostavat popcornia ja limsaa	He ostavat tietenkin popcornia ja limsaa	He ostavat ehkä popcornia ja limsaa	popcornia

31	Topi tekee pizzaa	Topi viimeistelee ruoka-annosta	Hän laittaa sen päälle juustoa ja mausteita	Tietenkin hän laittaa sen päälle juustoa ja mausteita	Kenties hän laittaa sen päälle juustoa ja mausteita	juustoa
32	Leolla on puhjennut paha ihottuma	Leo haluaa jäädä tänään kotiin	Hän menee lääkäriin saadakseen lääkärin-todistuksen	Hän menee tietenkin lääkäriin saadakseen lääkärintodistuksen	Hän menee kenties lääkäriin saadakseen lääkärin-todistuksen	lääkäriin
33	Eevalla on huono näkö	Eeva haluaa saada ajokortin	Hän tarvitsee silmälasit voidakseen ajaa autoa	Hän tarvitsee tietenkin silmälasit voidakseen ajaa autoa	Hän tarvitsee ehkä silmälasit voidakseen ajaa autoa	silmälasit
34	Saara on opettaja	Saara on töissä kaupungin palveluksessa	Hän on tekemisissä lasten ja nuorten kanssa päivittäin	Hän on tietenkin tekemisissä lasten ja nuorten kanssa päivittäin	Hän on kaiketi tekemisissä lasten ja nuorten kanssa päivittäin	lasten
35	Lailalla oli eilen syntymäpäivä	Lailan täti oli tullut heille kylään	Lailan äiti oli leiponut kakun, jonka päällä oli nonparelleja	Tietenkin Lailan äiti oli leiponut kakun, jonka päällä oli nonparelleja	Ehkä Lailan äiti oli leiponut kakun, jonka päällä oli nonparelleja	kakun
36	Maria rakastaa kukkia ja koristekasveja	Maria on erittäin varakas	Hänellä on iso puutarha, jossa hän kasvattaa ruusuja	Tietenkin hänellä on iso puutarha, jossa hän kasvattaa ruusuja	Ehkä hänellä on iso puutarha, jossa hän kasvat- taa ruusuja	puutarha
37	Raine on maanviljelijä	Raine on pikku-poika	Hän ajelee traktorilla joka päivä	Tietysti hän ajelee traktorilla joka päivä	Kenties hän ajelee traktorilla joka päivä	traktorilla
38	Esko on omistautunut isä	Esko ei malta odottaa vapaapäiväänsä	Hänellä on pieni poika, jonka kanssa hänestä on kiva viettää aikaa	Hänellä on tietenkin pieni poika, jonka kanssa hänestä on kiva viettää aikaa	Hänellä on ehkä pieni poika, jonka kanssa hänestä on kiva viettää aikaa	poika
39	Katariinalla on työn alla historiallinen romaani	Katariina kerää ideoita työtään varten	Hän on kuuluisa kirjailija, joka on suosittu ympäri maailmaa	Hän on tietenkin kuuluisa kirjailija, joka on suosittu ympäri maailmaa	Hän on kaiketi kuuluisa kirjailija, joka on suosittu ympäri maailmaa	kirjailija

40	Harri tekee koko ajan paljon töitä	Harrilla on ongelma	Hän on työnarkomaani eikä hänellä ole vapaa-ajan harrastuksia	Hän on tietysti työnarkomaani eikä hänellä ole vapaa-ajan harrastuksia	Hän on kaiketi työnarkomaani eikä hänellä ole vapaa-ajan harrastuksia	työnarkomaani
41	Villellä on paha migreeni	Villellä on paha flunssa	Hän kärsii kovasta päänsärystä ja pysyy sängyssä koko päivän	Hän kärsii tietysti kovasta päänsärystä ja pysyy sängyssä koko päivän	Hän kärsii kaiketi kovasta päänsärystä ja pysyy sängyssä koko päivän	päänsärystä
42	Reiskan asunnosta katkaistiin sähköt	Reiska osti talon kaupungin laidalta	Talossa on illan tullen pimeää ja kylmää	Talossa on illan tullen tietenkin pimeää ja kylmää	Talossa on illan tullen kenties pimeää ja kylmää	pimeää
43	Jarno on toiminut merikapteenina jo 20 vuotta	Jarno on ollut kokkina jo 20 vuotta	Hän on töissä laivalla, joka seilaa Karibialla	Hän on tietenkin töissä laivalla, joka seilaa Karibialla	Hän on kenties töissä laivalla, joka seilaa Karibialla	laivalla
44	Eero elää erittäin terveellisesti	Eero asuu maaseudulla	Hän syö paljon vihanneksia ja hedelmiä omasta puutarhastaan	Hän syö tietenkin paljon vihanneksia ja hedelmiä omasta puutarhastaan	Hän syö kenties paljon vihanneksia ja hedelmiä omasta puutarhastaan	vihanneksia
45	Viktor on puukeutunut tyylikkäästi	Viktor näyttää erilaiselta kuin tavallisesti	Hänellä on kaulassaan kravatti ja päällään tumma puku	Hänellä on tietenkin kaulassaan kravatti ja päällään tumma puku	Hänellä on kenties kaulassaan kravatti ja päällään tumma puku	kravatti
46	Pentti on lähdössä kauppaan, vaikka ulkona tulee vettä kaatamalla	Pentti on lähdössä kauppaan	Hän ottaa mukaansa sateenvarjon ja kauppakassin	Hän ottaa tietenkin mukaansa sateenvarjon ja kauppakassin	Hän ottaa kenties mukaansa sateenvarjon ja kauppakassin	sateenvarjon
47	Esa-Pekka on sinfoniaorkesterin kapellimestari	Esa-Pekka on sivistynyt mies	Hän tietää paljon musiikista ja erityisesti sävellyksistä	Hän tietysti tietää paljon musiikista ja erityisesti sävellyksistä	Hän kaiketi tietää paljon musiikista ja erityisesti sävellyksistä	musiikista

48	Kirsi rakastaa klassista musiikkia	Kirsi rakastaa korkeakulttuuria	Hän käy usein konserteissa ja oopperassa	Hän käy tietenkin usein konserteissa ja oopperassa	Hän käy kenties usein konserteissa ja oopperassa	konserteissa
49	Martti kärsii araknofobiasta	Martilla ei ole helppo elämä	Hän pelkää hämähäkkejä ja hyönteisiä niin paljon, että ei suostu lähtemään juhannukseksi maalle	Hän tietenkin pelkää hämähäkkejä ja hyönteisiä niin paljon, että ei suostu lähtemään juhannukseksi maalle	Hän kaiketi pelkää hämähäkkejä ja hyönteisiä niin paljon, että ei suostu lähtemään juhannukseksi maalle	hämähäkkejä
50	Kalle on sairaanhoitaja	Kalle on hallintohommissa	Hän on töissä sairaalassa kaupungin laidalla	Hän on tietysti töissä sairaalassa kaupungin laidalla	Hän on ehkä töissä sairaalassa kaupungin laidalla	sairaalassa
51	Sirkalla on upea ääni	Sirkka on monella tapaa lahjakas	Hän on laulaja tanssiorkesterissa	Hän on tietysti laulaja tanssiorkesterissa	Hän on kenties laulaja tanssiorkesterissa	laulaja
52	Tommi voitti lotossa	Tommi on tehnyt kovasti töitä jo monta vuotta	Nyt hän on rikas ja rakennuttaa upean talon	Nyt hän on tietysti rikas ja rakennuttaa upean talon	Nyt hän on kaiketi rikas ja rakennuttaa upean talon	rikas
53	Aarne haluaa mennä Tiinan kanssa naimisiin	Aarne todella pitää Tiinasta	Aarne aikoo vihdoinkin kosia Tiinaa, sillä he ovat olleet yhdessä jo vuosikausia	Totta kai Aarne aikoo vihdoinkin kosia Tiinaa, sillä he ovat olleet yhdessä jo vuosikausia	Ehkä Aarne aikoo vihdoinkin kosia Tiinaa, sillä he ovat olleet yhdessä jo vuosikausia	kosia
54	Tiina on liikinäköinen	Tiina on muotitietoinen	Hän käyttää silmälaseja joka päivä	Hän tietenkin käyttää silmälaseja joka päivä	Hän kaiketi käyttää silmälaseja joka päivä	silmälaseja
55	Marjaana on menossa polvileikkaukseen	Marjaana on lähdössä tropiikkiin lomamatkalle	Hän tapaa lääkärin kysyäkseen, kuinka voisi valmistautua siihen	Hän tietenkin tapaa lääkärin kysyäkseen, kuinka voisi valmistautua siihen	Hän ehkä tapaa lääkärin kysyäkseen, kuinka voisi valmistautua siihen	lääkärin

56	Teron on pakko muuttaa	Tero sai juuri potkut työpäikastaan	Hän etsii nyt asuntoa, johon hänellä olisi varaa	Hän tietysti etsii nyt asuntoa, johon hänellä olisi varaa	Hän ehkä etsii nyt asuntoa, johon hänellä olisi varaa	asuntoa
57	Mikko valehteli Ullalle ja nyt Mikkoa harmittaa	Mikko ei ole nähnyt Ullaa pitkään aikaan	Mikko pyytää anteeksi ja kutsuu Ullan illallisille	Mikko pyytää tietysti anteeksi ja kutsuu Ullan illallisille	Mikko pyytää ehkä anteeksi ja kutsuu Ullan illallisille	anteeksi
58	Rauli rakastaa taidetta	Rauli ei jaksa istua kotona vapaa-aikanaan	Hän käy usein museoissa ja keskustelee taiteesta ystäviensä kanssa	Hän käy tietenkin usein museoissa ja keskustelee taiteesta ystäviensä kanssa	Hän käy kaiketi usein museoissa ja keskustelee taiteesta ystäviensä kanssa	museoissa
59	Maria pesee hampaansa säännöllisesti	Maria pitää hyvää huolta terveydestään	Hänellä ei ole reikiä eikä ientulehdusta	Hänellä ei tietenkään ole reikiä eikä ientulehdusta	Hänellä ei kaiketi ole reikiä eikä ientulehdusta	reikiä
60	Piia aikoo mapittaa laskut siistiin järjestykseen	Piia aikoo siivota kaappinsa	Hän ostaa kansioita ja niihin välilehtiä	Hän ostaa tietenkin kansioita ja niihin välilehtiä	Hän ostaa kenties kansioita ja niihin välilehtiä	kansioita
61	Pikku Antti sai kympin matematiikasta	Pikku Antti soittaa pianoa	Hänen isänsä on erittäin ylpeä taitavasta pojastaan	Hänen isänsä on tietysti erittäin ylpeä taitavasta pojastaan	Hänen isänsä on kaiketi erittäin ylpeä taitavasta pojastaan	ylpeä
62	Jari odottaa tärkeää soittoa	Jari on rakastunut	Hän tuijottaa jatkuvasti puhelintaan ja toivoo, että se soisi	Hän tietenkin tuijottaa jatkuvasti puhelintaan ja toivoo, että se soisi	Hän kenties tuijottaa jatkuvasti puhelintaan ja toivoo, että se soisi	puhelintaan
63	Pertsan on odottanut Pirkkoa treffeille jo melkein tunnin	Pertsan on menossa Pirkon kanssa treffeille	Pirkko on myöhässä ja Pertsaa alkaa jo hermostuttua	Pirkko on tietysti myöhässä ja Pertsaa alkaa jo hermostuttua	Pirkko on kaiketi myöhässä ja Pertsaa alkaa jo hermostuttua	myöhässä

64	Antti ajattelee aina vain omaa napaansa	Antti on hieman erikoinen ihminen	Hän on itsekäs ja vieläpä ylpeä siitä	Hän on tietysti itsekäs ja vieläpä ylpeä siitä	Hän on kaiketi itsekäs ja vieläpä ylpeä siitä	itseks
65	Elina on töissä Korkeasaarella	Elinalla on hyvä työpaikka	Hän huolehtii eläimistä ja niiden aitauksista	Hän tietenkin huolehtii eläimistä ja niiden aitauksista	Hän kaiketi huolehtii eläimistä ja niiden aitauksista	eläimistä
66	Hetalla oli eilen tosi hyvä keikka	Heta valmistui luokkansa parhaana oppilaana	Hän sai mahtavat aplodit ja paljon kehuja	Hän sai tietenkin mahtavat aplodit ja paljon kehuja	Hän sai kaiketi mahtavat aplodit ja paljon kehuja	aplodit
67	Karoliina opiskelee ATK:ta	Karoliina on kiinnostunut monista asioista	Hän tietää paljon tietokoneista ja ohjelmoinnista	Hän tietysti tietää paljon tietokoneista ja ohjelmoinnista	Hän kenties tietää paljon tietokoneista ja ohjelmoinnista	tietokoneista
68	Kari ei pidä rahan tuhlaamisesta	Kari on tarkka joissakin asioissa	Hän on todella pihi eikä koskaan osta mitään normaalilla hinnalla	Hän on tietysti todella pihi eikä koskaan osta mitään normaalilla hinnalla	Hän on kaiketi todella pihi eikä koskaan osta mitään normaalilla hinnalla	pihi
69	Teppo on myynyt autonsa	Teppo soitti isälleen	Hän tarvitsi rahaa Italian matkaa varten	Hän tarvitsi tietenkin rahaa Italian matkaa varten	Hän tarvitsi kenties rahaa Italian matkaa varten	rahaa
70	Antonilla on karjatila	Antonilla on maatila	Hän kasvattaa lehmiä ja lampaista	Tietenkin hän kasvattaa lehmiä ja lampaista	Kenties hän kasvattaa lehmiä ja lampaista	lehmiä
71	Anneli laittaa toppatakin päällensä	Anneli lähtee aamulla reippaasti töihin	Ulkona on kylmä eikä Anneli halua vilustua	Tietenkin ulkona on kylmä eikä Anneli halua vilustua	Ehkä ulkona on kylmä eikä Anneli halua vilustua	kylmä
72	Tomi söi äidin paistamaa pannukakkua	Tomi söi jäätelöä	Hän laittoi sen päälle hilloa ja kermavaahtoa	Tietenkin hän laittoi sen päälle hilloa ja kermavaahtoa	Ehkä hän laittoi sen päälle hilloa ja kermavaahtoa	hilloa

73	Daniel on leijonankesyttäjä	Daniel on töissä sirkuksessa	Hänen työnsä on vaarallista ja jännittävää	Hänen työnsä on tietenkin vaarallista ja jännittävää	Hänen työnsä on kaiketi vaarallista ja jännittävää	vaarallista
74	Salme rakastaa eläimiä	Salme rakastaa luontoa	Hänellä on kissoja, jotka on otettu löytöeläinkodista	Hänellä on tietenkin kissoja, jotka on otettu löytöeläinkodista	Hänellä on kenties kissoja, jotka on otettu löytöeläinkodista	kissoja
75	Tirkkoset ovat hätää kärsimässä hevosensa takia	Tirkkoset ovat hätää kärsimässä kotieläimensä takia	Se syö hirveät määrät heinää, eikä Tirkkosten rahat riitä hevosen ylläpitoon	Se syö tietysti hirveät määrät heinää, eikä Tirkkosten rahat riitä hevosen ylläpitoon	Se syö kenties hirveät määrät heinää, eikä Tirkkosten rahat riitä hevosen ylläpitoon	heinää
76	Tuomaksella on kissa	Tuomaksella on tarantellahämähäkki	Se pyydystää hiiriä ja syö ne	Tietenkin se pyydystää hiiriä ja syö ne	Ehkä se pyydystää hiiriä ja syö ne	hiiriä
77	Erkki on jo hyvin iäkäs	Erkki on kokenut näyttelijä	Hänen kasvonsa ovat ryppyiset ja ilmeikkäät	Hänen kasvonsa ovat tietysti ryppyiset ja ilmeikkäät	Hänen kasvonsa ovat kenties ryppyiset ja ilmeikkäät	ryppyiset
78	Kristianin mielestä hänen huoneensa on liian pimeä iltaisin	Kristian aikoo sisustaa huoneensa uudelleen	Hän haluaa ostaa lampun saadakseen huoneeseen lisää valoa	Hän haluaa tietysti ostaa lampun saadakseen huoneeseen lisää valoa	Hän haluaa kenties ostaa lampun saadakseen huoneeseen lisää valoa	lampun
79	Tero on erittäin sosiaalinen	Tero on kuuluisa	Hänellä on paljon ystäviä ja tuttavias	Hänellä on tietysti paljon ystäviä ja tuttavias	Hänellä on kaiketi paljon ystäviä ja tuttavias	ystäviä
80	Sanna soittaa uudessa bändissä	Sanna ja hänen ystävänsä ovat erittäin lahjakkaita	Heillä on pian ensimmäinen keikka klubilla	Tietenkin heillä on pian ensimmäinen keikka klubilla	Ehkä heillä on pian ensimmäinen keikka klubilla	keikka

81	Marion on vannoutunut vegaani	Marion on aktiivisesti mukana ympäristöliikkeessä	Hän periaatteidensa mukaisesti kieltäytyy lihasta ja tehotuotetuista kasviksista	Totta kai hän periaatteidensa mukaisesti kieltäytyy lihasta ja tehotuotetuista kasviksista	Kenties hän periaatteidensa mukaisesti kieltäytyy lihasta ja tehotuotetuista kasviksista	lihasta
82	Petri haluaa antaa tytöstävälleen kukkia ystävänpäivänä	Petri haluaa yllättää työstävänsä viikonloppuna	Hän ostaa ruusuja kukkakaupasta	Tietenkin hän ostaa ruusuja kukkakaupasta	Kenties hän ostaa ruusuja kukkakaupasta	ruusuja
83	Eläintarhaan on tuotu uusi apina	Eläintarhaan on tuotu uusi elefantti	Se syö banaaneja, joita hoitaja laittaa hänen häkkiinsä	Se tietysti syö banaaneja, joita hoitaja laittaa sen häkkiin	Se kaikei syö banaaneja, joita hoitaja laittaa sen häkkiin	banaaneja
84	Roope on innokas lukija	Roope on akateemisesti koulutettu	Hänellä on paljon kirjoja kirjahyllyssään	Hänellä on tietenkin paljon kirjoja kirjahyllyssään	Hänellä on kenties paljon kirjoja kirjahyllyssään	kirjoja

Note. Crossing each of the two contexts (high constraint vs. low constraint) with each of the three versions of the target sentence (no marker vs. high certainty vs. low certainty) resulted in the six different versions of each item that were counterbalanced across participants via six item lists.

Experimental Items for Experiment 2

Table C2

Experimental Items Used in Study 3 (Experiment 2)

No.	Context Sentence:		Target Sentence:				Target Word	Final Sentence
	Plausibility		Focus of Marker					
	Plausible	Implausible	No Marker	Phrase (epistemic/ neutral Marker)	Sentence (epistemic/ neutral Marker)			
1	William hält sich gerne an Orten auf, an denen viel los ist	William möchte mit dem Auto verreisen	Er kauft sich am Bahnhof eine Fahrkarte und eine Zeitung	Er kauft sich am Bahnhof womöglich/ gerade eine Fahrkarte und eine Zeitung	Womöglich/ Gerade kauft er sich am Bahnhof eine Fahrkarte und eine Zeitung	Fahrkarte	Sein Auto ist mal wieder kaputt	
2	Sebastian ist Kioskbesitzer	Sebastian ist Metzger	Er verkauft Schnittblumen in seinem Laden	Er verkauft vielleicht/ jetzt Schnittblumen in seinem Laden	Vielleicht/ Jetzt verkauft er Schnittblumen in seinem Laden	Schnittblumen	Davon verspricht er sich einen höheren Umsatz	
3	Tina ist sehr modebewusst	Tina kann gut sehen	Sie trägt eine Brille beim Autofahren	Sie trägt womöglich/ heute eine Brille beim Autofahren	Womöglich/ Heute trägt sie eine Brille beim Autofahren	Brille	Die Brille schützt ihre Augen vor dem grellen Sonnenlicht	
4	Marc hat nette Nachbarn	Marc hat seine Haustiere verschenkt	Er bringt ihnen Tierfutter vom Einkaufen mit	Er bringt ihnen eventuell/ morgen Tierfutter vom Einkaufen mit	Eventuell/ Morgen bringt er ihnen Tierfutter vom Einkaufen mit	Tierfutter	Er fühlt sich ihnen sehr verbunden	

5	Nicolas will den Führerschein machen	Nicolas hat gute Augen	Er benötigt eine Brille, um scharf zu sehen	Er benötigt möglicherweise/ jetzt eine Brille, um scharf zu sehen	Möglicherweise/ Jetzt benötigt er eine Brille, um scharf zu sehen	Brille	Er hat heute Morgen starke Medikamente eingenommen
6	Helmut ist ein lebenslustiger Mensch	Helmut trinkt nur ungern Bier und Schnaps	Er war betrunken in der Kneipe	Er war vielleicht/ gestern betrunken in der Kneipe	Vielleicht/ Gestern war er betrunken in der Kneipe	betrunken	Sein Kumpel hat ihn zu dem Kneipenbesuch überredet
7	Anna arbeitet in einem Kaufhaus	Anna arbeitet in einem Spielzeugladen	Sie verkauft Brillen und Lupen	Sie verkauft womöglich/ heute Brillen und Lupen	Womöglich/ Heute verkauft sie Brillen und Lupen	Brillen	Kinder können sich damit als Detektiv verkleiden
8	Simon fühlt sich nicht wohl	Simon ist untergewichtig	Sein Arzt empfiehlt ihm eine Diät mit speziellen Nahrungsmitteln	Sein Arzt empfiehlt ihm vielleicht/ nun eine Diät mit speziellen Nahrungsmitteln	Vielleicht/ Nun empfiehlt ihm sein Arzt eine Diät mit speziellen Nahrungsmitteln	Diät	Das soll seinen Stoffwechsel wieder in Ordnung bringen
9	Juliane hatte gestern eine Verabredung	Juliane hat sehr viel geschlafen	Sie ist müde und kann sich nicht konzentrieren	Sie ist möglicherweise/ jetzt müde und kann sich nicht konzentrieren	Möglicherweise/ Jetzt ist sie müde und kann sich nicht konzentrieren	müde	Ihr Kreislauf kommt nicht richtig in Schwung
10	Dominik hat eine Eule	Dominik hat eine Hündin	Sie fängt Mäuse, die im Haus herumlaufen	Sie fängt möglicherweise/ gegenwärtig Mäuse, die im Haus herumlaufen	Möglicherweise/ Gegenwärtig fängt sie Mäuse, die im Haus herumlaufen	Mäuse	Sie spielt gerne mit kleinen Tieren

11	Alex ist ein Elefant	Alex ist ein Löwe	Er frisst die Bananen, die der Zoowärter ihm in den Käfig stellt	Er frisst vielleicht/ heute die Bananen, die der Zoowärter ihm in den Käfig stellt	Vielleicht/ Heute frisst er die Bananen, die der Zoowärter ihm in den Käfig stellt	Bananen	Er hat einen ausgefallenen Geschmack
12	Oliver ist ein vorsichtiger Mensch	Oliver mag keine Kopfbedeckungen	Er trägt einen Helm auf dem Fahrrad	Er trägt vielleicht/ nun einen Helm auf dem Fahrrad	Vielleicht/ Nun trägt er einen Helm auf dem Fahrrad	Helm	Dies ist Vorschrift und dient der Gesundheit
13	Robert ist ein Akademiker	Robert liest nicht gerne	Er hat viele Bücher bei sich zuhause	Er hat womöglich/ mittlerweile viele Bücher bei sich zuhause	Womöglich/ Mittlerweile hat er viele Bücher bei sich zuhause	Bücher	Er benötigt sie für seine Arbeit
14	Sarah ist bei der Stadt angestellt	Sarah ist Zoowärterin	Sie kümmert sich jeden Tag um Kinder und Jugendliche	Sie kümmert sich möglicherweise/ mittlerweile jeden Tag um Kinder und Jugendliche	Möglicherweise/ Mittlerweile kümmert sie sich jeden Tag um Kinder und Jugendliche	Kinder	Sie wollte immer schon gerne etwas Soziales machen
15	Johns Freundin ist in Urlaub gefahren	John versteht sich gut mit seiner Freundin	Er ist unglücklich und weint den ganzen Tag	Er ist möglicherweise/ momentan unglücklich und weint den ganzen Tag	Möglicherweise/ Momentan ist er unglücklich und weint den ganzen Tag	unglücklich	Er vermisst sie nämlich gerade sehr
16	Holger hat ein psychisches Problem	Holger mag seinen Beruf nicht	Er ist süchtig nach Arbeit und hat wenig andere Hobbies	Er ist womöglich/ derzeit süchtig nach Arbeit und hat wenig andere Hobbies	Womöglich/ Derzeit ist er süchtig nach Arbeit und hat wenig andere Hobbies	Arbeit	Er muss sich immer anderen beweisen

17	Wolfgang hat häufig eine Erkältung	Wolfgang treibt manchmal Sport	Er leidet dann an schlimmen Kopfschmerzen und bleibt den ganzen Tag im Bett	Er leidet dann möglicherweise/ morgens an schlimmen Kopfschmerzen und bleibt den ganzen Tag im Bett	Möglicherweise/ Morgens leidet er dann an schlimmen Kopfschmerzen und bleibt den ganzen Tag im Bett	Kopfschmerzen	Seine Gesundheit ist sehr instabil
18	Stefan ist ein berühmter Mensch	Stefan ist ein schüchterner Mensch	Er geht oft auf Partys und hat viele Freunde	Er geht möglicherweise/ mittlerweile oft auf Partys und hat viele Freunde	Möglicherweise/ Mittlerweile geht er oft auf Partys und hat viele Freunde	Partys	Kontakte sind gut für seine Karriere
19	Tobias rollt den Teig aus	Tobias backt Kuchen	Er streut darauf Käse und viele Gewürze	Er streut darauf eventuell/ nun Käse und viele Gewürze	Eventuell/ Nun streut er darauf Käse und viele Gewürze	Käse	Er liebt besondere Speisen
20	Marion hat einen Dackel	Marion hat einen Papagei	Den Nachbarn störte sein häufiges Bellen und Winseln	Den Nachbarn störte womöglich/ gestern sein häufiges Bellen und Winseln	Womöglich/ Gestern störte den Nachbarn sein häufiges Bellen und Winseln	Bellen	Das hat ihm Marion beigebracht
21	Theo ist Physiotherapeut	Theo ist Künstler	Er arbeitet im Krankenhaus am Ende der Stadt	Er arbeitet eventuell/ morgens im Krankenhaus am Ende der Stadt	Eventuell/ Morgens arbeitet er im Krankenhaus am Ende der Stadt	Krankenhaus	Er hatte sich auf den Job beworben, um seine Schulden zu bezahlen
22	Martins Freundin macht Leistungssport	Martins Freundin fühlt sich gesund	Er begleitet sie zum Arzt für einen Gesundheitscheck	Er begleitet sie vielleicht/ später zum Arzt für einen Gesundheitscheck	Vielleicht/ Später begleitet er sie zum Arzt für einen Gesundheitscheck	Arzt	Beide wollen sicher sein, dass ihr wirklich nichts fehlt

23	Selma liebt die Natur	Selma ist allergisch gegen Tierhaare	Sie hat eine Katze als Haustier	Sie hat vielleicht/ mittlerweile eine Katze als Haustier	Vielleicht/ Mittlerweile hat sie eine Katze als Haustier	Katze	Sie wollte gerne ein Haustier, das sie im Freien halten kann
24	Christian richtet sein Zimmer neu ein	Christian ist mit der Beleuchtung in seinem Zimmer sehr zufrieden	Er will sich eine neue Lampe kaufen	Er will sich eventuell/ jetzt eine neue Lampe kaufen	Eventuell/ Jetzt will er sich eine neue Lampe kaufen	Lampe	Seine Schwester wünscht sich von ihm etwas Selbstgemachtes
25	Ben sitzt oft still im Park	Ben ist ein berühmter Schauspieler	Er fühlt sich sehr einsam und möchte etwas daran ändern	Er fühlt sich womöglich/ momentan sehr einsam und möchte etwas daran ändern	Womöglich/ Momentan fühlt er sich sehr einsam und möchte etwas daran ändern	einsam	Er hätte gerne jemanden, dem er vertrauen kann
26	Andreas ist ein unbeliebter Mensch	Andreas ist ein netter Kerl	Er ist ein Egoist und stößt andere gelegentlich vor den Kopf	Er ist möglicherweise/ mittlerweile ein Egoist und stößt andere gelegentlich vor den Kopf	Möglicherweise/ Mittlerweile ist er ein Egoist und stößt andere gelegentlich vor den Kopf	Egoist	Er kann aber auch über seinen Schatten springen
27	Maike ist ein beliebtes Mädchen	Maike ist gerade von ihrem Freund verlassen worden	Sie bekommt eine Menge Geschenke und Blumen	Sie bekommt vielleicht/ heute eine Menge Geschenke und Blumen	Vielleicht/ Heute bekommt sie eine Menge Geschenke und Blumen	Geschenke	Es ist nämlich Valentinstag, und sie hat viele Verehrer

28	Bruno arbeitet im Lebensmittel- laden	Bruno arbeitet in einem Getränke- laden	Er verkauft viele frische Brötchen und andere Backwaren	Er verkauft womöglich/ morgens viele frische Brötchen und andere Backwaren	Womöglich/ Morgens verkauft er viele frische Brötchen und andere Backwaren	Brötchen	Die Kunden kaufen diese Produkte gerne
29	Tina macht sich schick für einen Theaterbesuch	Tina möchte heute alleine zu Hause bleiben	Sie benutzt Parfüm und Haarspray	Sie benutzt eventuell/ gerade Parfüm und Haarspray	Eventuell/ Gerade benutzt sie Parfüm und Haarspray	Parfüm	Sie liebt es, sich herauszuputzen
30	Marc bestellt sich im Restaurant einen Tee	Marc bestellt sich im Restaurant eine Suppe	Dort hinein gibt er einen Löffel Zucker	Er gibt möglicherweise/ gerade einen Löffel Zucker hinein	Möglicherweise/ Gerade gibt er einen Löffel Zucker hinein	Löffel	Er mag es gerne süß
31	Peter möchte seine Freundin am Wochenende überraschen	Peter möchte sich von seiner Freundin trennen	Er kauft ihr Rosen in einem kleinen Laden	Er kauft ihr womöglich/ morgen Rosen in einem kleinen Laden	Womöglich/ Morgen kauft er ihr Rosen in einem kleinen Laden	Rosen	Er möchte ihr damit seine Dankbarkeit für die gemeinsame Zeit ausdrücken
32	Klaus ist kulturell sehr interessiert	Klaus war noch nie im Kino	Er liebt Filme und weiß sehr viel darüber	Er liebt möglicherweise/ mittlerweile Filme und weiß sehr viel darüber	Möglicherweise/ Mittlerweile liebt er Filme und weiß sehr viel darüber	Filme	Er hat eine große DVD-Sammlung zuhause
33	Gaby steht unter der Dusche	Gaby schrubbt den Fußboden in der Küche	Dabei benutzt sie das Shampoo von ihrem Freund	Sie benutzt dabei eventuell/ heute das Shampoo von ihrem Freund	Eventuell/ Heute benutzt sie dabei das Shampoo von ihrem Freund	Shampoo	Sie hat gerade nichts anderes zur Hand
34	Leonie ist schlecht drauf	Leonie hat Lust auf einen Ausflug	Sie geht zu ihrem Hausarzt in der Stadt	Sie geht möglicherweise/ gerade zu ihrem Hausarzt in der Stadt	Möglicherweise/ Gerade geht sie zu ihrem Hausarzt in der Stadt	Hausarzt	Dieser stellt Atteste aus, ohne viele Fragen zu stellen

35	Anton ist Künstler	Anton ist Kioskbesitzer	Er verkauft Statuen aus Ton	Er verkauft möglicherweise/ momentan Statuen aus Ton	Möglicherweise/ Momentan verkauft er Statuen aus Ton	Statuen	Abends und am Wochenende verbringt er viel Zeit mit Töpfern
36	Ralf bleibt in seiner Freizeit nicht gerne zuhause	Ralf hält nichts von Kunst	Er geht ins Museum in seiner Stadt	Er geht womöglich/ bald ins Museum in seiner Stadt	Womöglich/ Bald geht er ins Museum in seiner Stadt	Museum	Dort arbeitet eine Frau, die er gerne kennenlernen möchte
37	Christine sitzt im Café und liest ein Buch	Christine trinkt gerne Kakao mit Sahne	Sie bestellt sich ein Wasser mit Kohlensäure	Sie bestellt sich eventuell/ heute ein Wasser mit Kohlensäure	Eventuell/ Heute bestellt sie sich ein Wasser mit Kohlensäure	Wasser	Sie macht gerade eine Diät
38	David spricht mit seinem Vater	David hat einen gut bezahlten Job	Er braucht Geld für eine Reise nach Italien zu seiner kranken Mutter	Er braucht möglicherweise/ jetzt Geld für eine Reise nach Italien zu seiner kranken Mutter	Möglicherweise/ Jetzt braucht er Geld für eine Reise nach Italien zu seiner kranken Mutter	Geld	Er bezahlt momentan eine Hypothek ab
39	Charlie ist sehr geduldig	Charlie ist Vegetarier	Er fängt einen großen Fisch beim Angeln	Er fängt vielleicht/ gerade einen großen Fisch beim Angeln	Vielleicht/ Gerade fängt er einen großen Fisch beim Angeln	Fisch	Diesen möchte er seinem Opa schenken
40	Marc und Gerhard unterhalten sich bei einer Cola	Marc und Gerhard unterhalten sich im Fitnessstudio	Dabei essen sie Kuchen und Plätzchen	Sie essen dabei möglicherweise/ gerade Kuchen und Plätzchen	Möglicherweise/ Gerade essen sie dabei Kuchen und Plätzchen	Kuchen	Heute ist Weihnachtsfeier
41	Stefanie ist eine begabte junge Frau	Stefanie ist schüchtern	Sie hat bald ihren ersten Auftritt auf der Bühne	Sie hat eventuell/ nun bald ihren ersten Auftritt auf der Bühne	Eventuell/ Nun hat sie bald ihren ersten Auftritt auf der Bühne	Auftritt	Ihr Gesangslehrer hat sie dazu gedrängt

42	Pascal ist viel zuhause	Pascal hat viele Freunde	Er ist sehr schüchtern und redet nicht viel	Er ist vielleicht/ derzeit schüchtern und redet nicht viel	Vielleicht/ Derzeit ist er schüchtern und redet nicht viel	schüchtern	Er kann anderen aber gut zuhören
43	Mara möchte in die Tropen reisen und braucht Rat	Mara möchte ein Picknick machen und braucht Rat	Sie geht zum Arzt in ihrem Viertel	Sie geht vielleicht/ heute zum Arzt in ihrem Viertel	Vielleicht/ Heute geht sie zum Arzt in ihrem Viertel	Arzt	Sie ist gegen viele Pflanzen allergisch
44	Karl kleidet sich immer sehr schlicht	Karl kauft oft Geschenke für seine Freunde	Er ist sehr geizig und sparsam	Er ist vielleicht/ derzeit sehr geizig und sparsam	Vielleicht/ Derzeit ist er sehr geizig und sparsam	geizig	Dies trifft aber nur zu, wenn es um seine eigene Person geht
45	Rainer wohnt auf dem Land	Rainer hat keinen Führerschein	Er fährt einen Traktor seines Vaters bei der Ernte	Er fährt möglicherweise/ gerade einen Traktor seines Vaters bei der Ernte	Möglicherweise/ Gerade fährt er einen Traktor seines Vaters bei der Ernte	Traktor	Sein Vater hat ihm dies früh beigebracht
46	Gabriel arbeitet in einem interessanten Beruf	Gabriel arbeitet gerne mit Menschen	Sein Bereich sind die Raubtiere und ihr Gehege	Sein Bereich sind vielleicht/ momentan die Raubtiere und ihr Gehege	Vielleicht/ Momentan sind sein Bereich die Raubtiere und ihr Gehege	Raubtiere	Diese ziehen jedes Jahr viele kleine und große Besucher an
47	Pablo studiert an der Hochschule	Pablo studiert Architektur	Er möchte Menschen helfen und sie gesund machen	Er möchte Menschen womöglich/ später helfen und sie gesund machen	Womöglich/ Später möchte er Menschen helfen und sie gesund machen	helfen	Er will sie in Fragen des gesunden Wohnens beraten
48	Erik ist ein freundlich aussehender Mann	Erik ist ein junger Mann	Im Gesicht hat er zahlreiche Falten und Grübchen	Er hat im Gesicht möglicherweise/ mittlerweile zahlreiche Falten und Grübchen	Möglicherweise/ Mittlerweile hat er im Gesicht zahlreiche Falten und Grübchen	Falten	Er lacht oft und gerne

49	Christoph ist bei seinen Freunden sehr beliebt	Christoph ist erst 16	Er ist in vielen Dingen sehr erfahren und weiß Bescheid	Er ist in vielen Dingen womöglich/ mittlerweile sehr erfahren und weiß Bescheid	Womöglich/ Mittlerweile ist er in vielen Dingen sehr erfahren und weiß Bescheid	erfahren	Er hat mit seinen Eltern eine Weltreise gemacht
50	Sandra hat eine Klassenarbeit zurückbekommen	Sandra hat eine schlechte Note für ihre Klassenarbeit bekommen	Sie bekommt von ihren Eltern viel Lob und Zuwendung	Sie bekommt von ihren Eltern vielleicht/ nun viel Lob und Zuwendung	Nun/ Vielleicht bekommt sie von ihren Eltern viel Lob und Zuwendung	Lob	Sie hat sich bei der Klassenarbeit sehr angestrengt
51	Peter ist Spielwarenhändler	Peters Kunden sind Kinder	Er verkauft unter anderem Autos und Motorräder	Er verkauft unter anderem eventuell/ gegenwärtig Autos und Motorräder	Eventuell/ Gegenwärtig verkauft er unter anderem Autos und Motorräder	Autos	Diese sind als Spielzeug sehr beliebt
52	In Linas Haus ist ein Problem aufgetreten	In Linas Haus muss der Parkettfußboden erneuert werden	Sie ruft den Klempner aus dem Nachbarort an	Sie ruft möglicherweise/ morgen den Klempner aus dem Nachbarort an	Möglicherweise/ Morgen ruft sie den Klempner aus dem Nachbarort an	Klempner	Sie vermutet einen Rohrbruch als Ursache
53	Der kleine Stefan macht seine Mathematik-Hausaufgaben	Der kleine Stefan bekommt in seiner Rechenarbeit eine Vier	Sein Vater ist sehr stolz auf ihn und kauft ihm ein Eis	Sein Vater ist womöglich/ jetzt sehr stolz auf ihn und kauft ihm ein Eis	Womöglich/ Jetzt ist sein Vater sehr stolz auf ihn und kauft ihm ein Eis	stolz	Stefan hat bis vor Kurzem immer Fünfen und Sechsen geschrieben
54	Fabian lebt auf dem Land	Fabian mag am liebsten Fleisch	Er isst viel Obst und Gemüse aus seinem Garten	Er isst vielleicht/ gerade viel Obst und Gemüse aus seinem Garten	Vielleicht/ Gerade isst er viel Obst und Gemüse aus seinem Garten	Obst	Das kostet ihn nichts

55	Sabine fährt zum Einkaufen in die Stadt	Sabine geht zu ihrer Nachbarin	Sie möchte eine neue Frisur und eine neue Haarfarbe	Sie möchte eventuell/ morgen eine neue Frisur und eine neue Haarfarbe	Eventuell/ Morgen möchte sie eine neue Frisur und eine neue Haarfarbe	Frisur	Sie möchte sich dazu erst einmal beraten lassen
56	Alexis hat gerade Abitur gemacht	Alexis muss für eine Prüfung lernen	Er starrt auf sein Telefon und hofft, dass es klingelt	Er starrt womöglich/ jetzt auf sein Telefon und hofft, dass es klingelt	Womöglich/ Jetzt starrt er auf sein Telefon und hofft, dass es klingelt	Telefon	Er wartet auf eine Rückmeldung zu einer Bewerbung
57	Olli geht zu einem Bauernmarkt	Olli ist Vegetarier	Er kauft eine Bratwurst mit Ketchup	Er kauft möglicherweise/ heute eine Bratwurst mit Ketchup	Möglicherweise/ Heute kauft er eine Bratwurst mit Ketchup	Bratwurst	Er möchte seinem Sohn eine Freude machen, der Bratwürste liebt
58	Sascha ist talentiert	Sascha ist Legastheniker	Er schreibt Artikel für eine Wochenzeitung	Er schreibt vielleicht/ gerade Artikel für eine Wochenzeitung	Vielleicht/ Gerade schreibt er Artikel für eine Wochenzeitung	Artikel	Er diktiert seine Texte auf Band
59	Miriam ist verärgert	Miriam ist verletzt	Sie hat ein wichtiges Spiel für ihren Verein	Sie hat eventuell/ morgen ein wichtiges Spiel für ihren Verein	Eventuell/ Morgen hat sie ein wichtiges Spiel für ihren Verein	Spiel	Sie muss jedoch wegen eines Knöchelbruchs aussetzen
60	Sabrina liebt die Kunst	Sabrina macht sich nicht viel aus Musik	Sie geht gerne ins Konzert und in die Oper	Sie geht eventuell/ mittlerweile gerne ins Konzert und in die Oper	Eventuell/ Mittlerweile geht sie gerne ins Konzert und in die Oper	Konzert	Bei ihren Freundinnen ist dies ein beliebter Treffpunkt

61	Michael hat Sarah seit langer Zeit nicht gesehen	Michael hat Sarah eine große Freude gemacht	Er bittet sie um Verzeihung und hofft auf ihr Verständnis	Er bittet sie möglicherweise/ nun um Verzeihung und hofft auf ihr Verständnis	Möglicherweise/ Nun bittet er sie um Verzeihung und hofft auf ihr Verständnis	Verzeihung	Sie ist normalerweise sehr nachsichtig
62	Karsten ist mit dem Auto sehr schnell gefahren	Karsten ist ein sehr rücksichtsvoller Autofahrer	Er fühlt sich schuldig und drosselt die Geschwindigkeit	Er fühlt sich möglicherweise/ jetzt schuldig und drosselt die Geschwindigkeit	Möglicherweise/ Jetzt fühlt er sich schuldig und drosselt die Geschwindigkeit	schuldig	Er hat aus Versehen ein Tier überfahren
63	Linus kann sich gut ausdrücken	Linus ist Schreiner	Er schreibt Romane und Kurzgeschichten für eine Zeitschrift	Er schreibt womöglich/ mittlerweile Romane und Kurzgeschichten für eine Zeitschrift	Womöglich/ Mittlerweile schreibt er Romane und Kurzgeschichten für eine Zeitschrift	Romane	Er möchte sich etwas dazu verdienen
64	Franz hat seine Rechnung nicht bezahlt	Franz hat immer das Licht und alle Geräte an	In dieser Nacht ist sein Haus dunkel und das Telefon geht nicht	In dieser Nacht ist sein Haus eventuell/ nun dunkel und das Telefon geht nicht	Eventuell/ Nun ist in dieser Nacht sein Haus dunkel und das Telefon geht nicht	dunkel	Die Hauptsicherung ist möglicherweise abgeschaltet
65	Thomas möchte Lehrer werden	Thomas möchte Makler werden	Vor der Aufnahmeprüfung lernt er Singen und Klavierspielen	Er lernt vor der Aufnahmeprüfung möglicherweise/ nun Singen und Klavierspielen	Möglicherweise/ Nun lernt er vor der Aufnahmeprüfung Singen und Klavierspielen	Singen	Man hat ihm einen kreativen Ausgleich zur Entspannung empfohlen
66	Daniel arbeitet im Zirkus	Daniel langweilt sich oft	Sein Beruf war sehr gefährlich und spannend	Sein Beruf war vielleicht/ früher sehr gefährlich und spannend	Vielleicht/ Früher war sein Beruf sehr gefährlich und spannend	gefährlich	Für ihn ist er bereits Routine

67	Julian zieht weg in eine andere Stadt	Julian trifft in seiner neuen Heimat viele neue Leute	Das macht ihn traurig und mutlos	Das macht ihn womöglich/ gerade traurig und mutlos	Womöglich/ Gerade macht ihn das traurig und mutlos	traurig	Er vermisst seine alten Freunde
68	Maria achtet sehr auf ihre Gesundheit	Maria hat Zahnschmerzen	Sie hat keine Karies und auch keine Parodontose	Sie hat womöglich/ momentan keine Karies und auch keine Parodontose	Womöglich/ Momentan hat sie keine Karies und auch keine Parodontose	Karies	Sie hatte nur schon immer sehr empfindliche Zähne
69	Manuel ist ein kleiner Junge	Manuel ist Anwalt	Er trägt einen Schlafanzug mit kleinen Sternen	Er trägt möglicherweise/ derzeit einen Schlafanzug mit kleinen Sternen	Möglicherweise/ Derzeit trägt er einen Schlafanzug mit kleinen Sternen	Schlafanzug	Dies war schon immer sein liebster Schlafanzug
70	Victor hat heute einen wichtigen Termin	Victor mag bequeme Kleidung	Er trägt eine Krawatte und einen schwarzen Anzug	Er trägt vielleicht/ heute eine Krawatte und einen schwarzen Anzug	Vielleicht/ Heute trägt er eine Krawatte und einen schwarzen Anzug	Krawatte	Er ist auf einem Empfang eingeladen
71	Johann freut sich auf seine Freizeit	Johann ist oft traurig	Er hat einen Sohn, mit dem er gerne spielt	Er hat vielleicht/ jetzt einen Sohn, mit dem er gerne spielt	Vielleicht/ Jetzt hat er einen Sohn, mit dem er gerne spielt	Sohn	Sein Job lässt ihm aber wenig Zeit für seine Familie
72	Judith hat nachmittags häufig Verabredungen	Judith hat meistens Langeweile	Sie trifft sich mit den anderen Kindern aus dem Kindergarten	Sie trifft sich eventuell/ derzeit mit den anderen Kindern aus dem Kindergarten	Eventuell/ Derzeit trifft sie sich mit den anderen Kindern aus dem Kindergarten	Kindern	Dann vergeht ihre Langeweile
73	Christian wird nie langweilig	Christian ist Grobmotoriker	Er strickt einen Schal aus grüner Wolle	Er strickt möglicherweise/ momentan einen Schal aus grüner Wolle	Möglicherweise/ Momentan strickt er einen Schal aus grüner Wolle	strickt	Seine Freundin drängt ihn dazu

74	Sanne war gestern beim Friseur	Sanne hat von Natur aus glattes Haar	Sie hat viele Locken auf dem Kopf	Sie hat vielleicht/ nun viele Locken auf dem Kopf	Vielleicht/ Nun hat sie viele Locken auf dem Kopf	Locken	Ihr Friseur hatte ihr dazu geraten
75	Carmen hat eine Menge Bega- bungen	Carmen hat eine krächzende Stimme	Sie arbeitet als Sängerin in einer Band	Sie arbeitet womög- lich/ mittlerweile als Sängerin in einer Band	Womöglich/ Mitt- lerweile arbeitet sie als Sängerin in einer Band	Sängerin	Ihre ungewöhnli- che Stimme ist sehr gefragt
76	Laila hat bei einem Wettbe- werb gewonnen	Laila muss auf ihr Gewicht achten	Ihre Mutter hat ihr einen Kuchen mit vielen Kerzen gebacken	Ihre Mutter hat ihr eventuell/ heute einen Kuchen mit vielen Kerzen gebacken	Eventuell/ Heute hat ihre Mutter ihr einen Kuchen mit vielen Kerzen gebacken	Kuchen	Laila hat es sich so sehr gewünscht
77	Finn ist sehr empfindlich	Finn hat wenig Ahnung von Musik	Er hasst schiefe Töne in Musikstücken	Er hasst vielleicht/ mittlerweile schiefe Töne in Musikstücken	Vielleicht/ Mittler- weile hasst er schiefe Töne in Musikstücken	schiefe	Er mag keine experimentelle Musik
78	Paula ist Künstlerin	Paula ist vier Jahre alt und malt gerne	Eines ihrer Werke hängt sogar im Museum in der Abteilung für moderne Kunst	Eines ihrer Werke hängt sogar mög- licherweise/ bald im Museum in der Abteilung für mo- derne Kunst	Möglicherweise/ Bald hängt sogar eines ihrer Werke im Museum in der Abteilung für mo- derne Kunst	Museum	Ihre Gemälde sind außerge- wöhnlich schön
79	Christiane hört gerne Radio	Christiane ist taub	Sie mag laute Musik aus ihrer Stereoanlage	Sie mag vielleicht/ derzeit laute Musik aus ihrer Stereoanla- ge	Vielleicht/ Derzeit mag sie laute Musik aus ihrer Stereoanla- ge	laute	Sie kann die Vibration spüren

80	Rüdiger hat es nicht einfach	Rüdiger ist Fernsehmoderator	Er fürchtet sich vor Räumen mit vielen Menschen	Er fürchtet sich eventuell/ gegenwärtig vor Räumen mit vielen Menschen	Eventuell/ Gegenwärtig fürchtet er sich vor Räumen mit vielen Menschen	Räumen	Er bekommt dann oft Herzrasen
81	Tom liebt guten Weißwein	Tom liebt süße Kekse	Dazu isst er gerne Tomaten und Basilikum	Er isst dazu möglicherweise/ mittlerweile gerne Tomaten und Basilikum	Möglicherweise/ Mittlerweile isst er dazu gerne Tomaten und Basilikum	Tomaten	Er liebt diese Geschmackskombination
82	Gudrun ist sehr gebildet	Gudrun hat keine feste Anstellung	Sie arbeitet an einer Schule in ihrer Stadt	Sie arbeitet eventuell/ momentan an einer Schule in ihrer Stadt	Eventuell/ Momentan arbeitet sie an einer Schule in ihrer Stadt	Schule	Sie hilft Kindern in der Freizeit bei den Hausaufgaben
83	Franz-Egon ist sehr fleißig	Franz-Egon ist ein nettes Kind	Er ist in seiner Klasse ein Außenseiter mit wenig Freunden	Er ist in seiner Klasse vielleicht/ momentan ein Außenseiter mit wenig Freunden	Vielleicht/ Momentan ist er in seiner Klasse ein Außenseiter mit wenig Freunden	Außenseiter	Er schreibt sehr gute Noten
84	Inken geht einkaufen	Inken hat heute schon viel Wasser getrunken	Sie hat großen Durst und braucht eine Erfrischung	Sie hat womöglich/ jetzt großen Durst und braucht eine Erfrischung	Womöglich/ Jetzt hat sie großen Durst und braucht eine Erfrischung	Durst	Sie hat ausgiebig Sport gemacht
85	Michaela hat gleich ein wichtiges Vorstellungsgespräch	Michaela hat seit gestern keinen Schnupfen mehr	Sie putzt sich mit einem Taschentuch die Nase	Sie putzt sich möglicherweise/ gerade mit einem Taschentuch die Nase	Möglicherweise/ Gerade putzt sie sich mit einem Taschentuch die Nase	Taschentuch	Sie hat sie aus Versehen mit Lippenstift beschmiert

86	Dirk sucht einen neuen Job	Dirk möchte mit Kindern arbeiten	Er geht zum Friseur um die Ecke	Er geht eventuell/ heute zum Friseur um die Ecke	Eventuell/ Heute geht er zum Friseur um die Ecke	Friseur	Er möchte bei seinem Bewerbungsgespräch in einem Kindergarten einen guten Eindruck machen
87	Svenja mag Tiere	Svenja ist Busfahrerin	Sie arbeitet auf einem Bauernhof am Ende des Ortes	Sie arbeitet vielleicht/ mittlerweile auf einem Bauernhof am Ende des Ortes	Vielleicht/ Mittlerweile arbeitet sie auf einem Bauernhof am Ende des Ortes	Bauernhof	Sie holt immer die Feriengäste vom Bahnhof ab, die Ferien auf dem Bauernhof machen
88	Mikaella hat ein altes Hobby wiederentdeckt	Mikaella studiert Physik	Sie lernt dabei viel über Musik und Instrumente	Sie lernt dabei womöglich/ nun viel über Musik und Instrumente	Womöglich/ Nun lernt sie dabei viel über Musik und Instrumente	Musik	Ihr Spezialgebiet ist die Akustik
89	Hilmar ist seit 20 Jahren Koch	Hilmar ist Anwalt	Er hat auf einem Schiff in der Karibik angeheuert	Er hat möglicherweise/ gerade auf einem Schiff in der Karibik angeheuert	Möglicherweise/ Gerade hat er auf einem Schiff in der Karibik angeheuert	Schiff	Er wollte mal etwas anderes machen
90	Max interessiert sich für Zoos	Max hat Angst vor den Hunden seiner Nachbarn	Jeden Abend füttert er die Tiere und sieht nach dem Rechten	Er füttert eventuell/ nun jeden Abend die Tiere und sieht nach dem Rechten	Eventuell/ Nun füttert er jeden Abend die Tiere und sieht nach dem Rechten	Tiere	Er wurde darum gebeten
91	Adam mag frische Luft	Adam ist Chemiker	Er ist im Wald, um Bäume zu fällen	Er ist vielleicht/ gerade im Wald, um Bäume zu fällen	Vielleicht/ Gerade ist er im Wald, um Bäume zu fällen	Wald	Dies macht er gerne in seiner Freizeit

92	Julius hat eine Sehschwäche	Julius ist voll verkehrstüchtig	Er verwechselt Rot und Grün beim Sehtest	Er verwechselt möglicherweise/ abends Rot und Grün beim Sehtest	Möglicherweise/ Abends verwechselt er Rot und Grün beim Sehtest	Rot	Das behindert ihn aber nicht im Straßenverkehr
93	Mirka ist sehr fleißig	Mirka ist schreibfaul	Sie schreibt Briefe und verschickt sie	Sie schreibt womöglich/ momentan Briefe und verschickt sie	Womöglich/ Momentan schreibt sie Briefe und verschickt sie	Briefe	Ihre Chefin hat sie darum gebeten
94	Till ruft seinen Vater an	Till wohnt noch bei seinen Eltern	Am Ende des Monats hat er kein Geld für die Miete	Er hat am Ende des Monats eventuell/ nun kein Geld für die Miete	Eventuell/ Nun hat er am Ende des Monats kein Geld für die Miete	Geld	Er gibt immer alles sofort aus
95	Sven hat einen anstrengenden Job	Sven ist sehr gebildet	Er arbeitet auf einer Baustelle in seiner Stadt	Er arbeitet vielleicht/ morgens auf einer Baustelle in seiner Stadt	Vielleicht/ Morgens arbeitet er auf einer Baustelle in seiner Stadt	Baustelle	Er koordiniert und überwacht Bauarbeiten
96	Marion ist gesundheitsbewusst	Marion ist Metzgerin	Sie ernährt sich schon seit mehreren Wochen vegetarisch und verzichtet auf Fleisch	Sie ernährt sich möglicherweise/ jetzt vegetarisch und verzichtet auf Fleisch	Möglicherweise/ Jetzt ernährt sie sich vegetarisch und verzichtet auf Fleisch	vegetarisch	Sie möchte ihr Cholesterin niedrig halten
97	Annette nimmt Nahrungsergänzungsmittel	Annette treibt viel Sport	Sie ist schwer krank und hofft auf eine Genesung	Sie ist womöglich/ momentan schwer krank und hofft auf eine Genesung	Womöglich/ Momentan ist sie schwer krank und hofft auf eine Genesung	krank	Dafür unternimmt sie alles, was sie für sinnvoll hält

98	Arno mag Tina sehr gerne	Arno und Tina streiten viel	Er will ihr einen Antrag machen	Er will ihr eventuell/ nun einen Antrag machen	Eventuell/ Nun will er ihr einen Antrag machen	Antrag	Er möchte sich mit ihr versöhnen
99	Thomas ist gefeuert worden	Thomas besitzt ein Haus in der Stadt	Er sucht dringend eine Wohnung, die er finanzieren kann	Er sucht vielleicht/ derzeit dringend eine Wohnung, die er finanzieren kann	Vielleicht/ Derzeit sucht er dringend eine Wohnung, die er finanzieren kann	Wohnung	Er will aufs Land ziehen
100	Martina ist in einer heiklen Lage	Martina ist Richterin	Sie wird von der Polizei gesucht	Sie wird möglicherweise/ gegenwärtig von der Polizei gesucht	Möglicherweise/ Gegenwärtig wird sie von der Polizei gesucht	Polizei	Sie hat eine Straftat begangen
101	Georg musiziert jeden Tag	Georg ist unmusikalisch	Er spielt als Pianist in einer Band	Er spielt womöglich/ momentan als Pianist in einer Band	Womöglich/ Momentan spielt er als Pianist in einer Band	Pianist	Er hält sich für ein großes Talent
102	Hannes ist Bauarbeiter	Hannes ist Schreibkraft	Er hat viele Muskeln, auf die er sehr stolz ist	Er hat vielleicht/ nun viele Muskeln, auf die er sehr stolz ist	Vielleicht/ Nun hat er viele Muskeln, auf die er sehr stolz ist	Muskeln	Er trainiert in jeder freien Minute
103	Lauras Leben ist nicht sehr abwechslungsreich	Laura ist von Natur aus faul	Sie verbringt viel Zeit mit Lernen und sieht ihre Freunde kaum noch	Sie verbringt womöglich/ gerade viel Zeit mit Lernen und sieht ihre Freunde kaum noch	Womöglich/ Gerade verbringt sie viel Zeit mit Lernen und sieht ihre Freunde kaum noch	Lernen	Sie will die Abschlussprüfung unbedingt bestehen
104	Jens kommt von der Arbeit nach Hause	Jens ist ein sehr aktiver Mensch	Er legt sich aufs Sofa und hört Musik zum Entspannen	Er legt sich aufs Sofa und hört möglicherweise/ nun Musik zum Entspannen	Möglicherweise/ Nun legt er sich aufs Sofa und hört Musik zum Entspannen	Musik	Er braucht eine Pause

105	Huan ist zweisprachig aufgewachsen	Huans Vater ist Chinese	Ihre Muttersprache ist Englisch mit einem leichten Akzent	Ihre Muttersprache ist womöglich/ längst Englisch mit einem leichten Akzent	Womöglich/ Längst ist ihre Muttersprache Englisch mit einem leichten Akzent	Englisch	Ihre Eltern haben ihr diese Sprache von klein auf beigebracht
106	Holger besucht eine Grillparty	Holger mag keinen Alkohol	Er trinkt ein Bier und isst ein Würstchen	Er trinkt vielleicht/ heute ein Bier und isst ein Würstchen	Vielleicht/ Heute trinkt er ein Bier und isst ein Würstchen	Bier	Sein Freund hat ihn dazu eingeladen
107	Lars sieht aus dem Fenster den Schnee fallen	Lars genießt das warme Wetter	Es ist der erste Advent und bald ist Weihnachten	Es ist eventuell/ nun der erste Advent und bald ist Weihnachten	Eventuell/ Nun ist es der erste Advent und bald ist Weihnachten	Advent	Bis dahin ist Lars aus seinem Urlaub zurück
108	Anna macht Pläne für die nächsten zwei Wochen	Anna hat die nächste Zeit viel zu tun	Sie fährt in den Urlaub, um sich zu erholen	Sie fährt möglicherweise/ morgen in den Urlaub, um sich zu erholen	Möglicherweise/ Morgen fährt sie in den Urlaub, um sich zu erholen	Urlaub	Das kann sie am besten, wenn sie viel unternimmt
109	Lisa hält sich zwei Stunden in der Küche auf	Lisa kauft Ananas und Eis	Bald ist Weihnachten und es gibt noch viel vorzubereiten	Bald ist womöglich/ nun Weihnachten und es gibt noch viel vorzubereiten	Womöglich/ Nun ist bald Weihnachten und es gibt noch viel vorzubereiten	Weihnachten	Sie will das Weihnachtsessen diesmal anders gestalten
110	Antonia zieht eine Jacke an	Antonia packt ihre Strandsachen ein	Draußen regnet es und sie möchte sich nicht erkälten	Es regnet vielleicht/ heute draußen und sie möchte sich nicht erkälten	Vielleicht/ Heute regnet es draußen und sie möchte sich nicht erkälten	regnet	Sie ist durch den Wechsel von Sonne und Regen schon angeschlagen

111	Astrid hat sehr zugenommen	Die Ärzte hatten Astrid gesagt, dass sie nicht schwanger werden kann	Sie erwartet ein Baby, das sie sich lange gewünscht hat	Sie erwartet womöglich/ bald ein Baby, das sie sich lange gewünscht hat	Womöglich/ Bald erwartet sie ein Baby, das sie sich lange gewünscht hat	Baby	Sie hat nicht mehr damit gerechnet
112	Tom arbeitet seit Jahren sehr viel	Tom verdient in seinem Job nur sehr wenig	Er ist reich und kann sich viel leisten	Er ist möglicherweise/ jetzt reich und kann sich viel leisten	Möglicherweise/ Jetzt ist er reich und kann sich viel leisten	reich	Er hat im Lotto gewonnen
113	Tina hat gerade ihren Schulabschluss gemacht	Tina hat Flugangst	Sie möchte Stewardess werden und viel herumkommen	Sie möchte eventuell/ derzeit Stewardess werden und viel herumkommen	Eventuell/ Derzeit möchte sie Stewardess werden und viel herumkommen	Stewardess	Sie versucht gerade, ihre Flugangst zu überwinden
114	Katharina sammelt Ideen	Katharina ist Anwältin	Sie ist eine berühmte Schriftstellerin, deren Romane sehr beliebt sind	Sie ist womöglich/ mittlerweile eine berühmte Schriftstellerin, deren Romane sehr beliebt sind	Womöglich/ Mittlerweile ist sie eine berühmte Schriftstellerin, deren Romane sehr beliebt sind	Schriftstellerin	Sie schreibt überzeugend über Mordfälle
115	Horst sitzt am Schreibtisch	Horst singt ein Lied	Er schreibt auf der Tastatur und benutzt die Maus	Er schreibt womöglich/ nun auf der Tastatur und benutzt die Maus	Womöglich/ Nun schreibt er auf der Tastatur und benutzt die Maus	Tastatur	Er schreibt gerade einen Liedtext
116	Benjamin öffnet die Tür	Benjamin ist gerne und häufig drinnen	Er braucht frische Luft zum Atmen	Er braucht möglicherweise/ gerade frische Luft zum Atmen	Möglicherweise/ Gerade braucht er frische Luft zum Atmen	Luft	Er hält sich schon den ganzen Tag im Haus auf

117	Ulla interessiert sich für viele Dinge	Ulla ist Bäuerin	Sie ist Computerexpertin und kennt sich mit Hard- und Software aus	Sie ist womöglich/ mittlerweile Computerexpertin und kennt sich mit Hard- und Software aus	Womöglich/ Mittlerweile ist sie Computerexpertin und kennt sich mit Hard- und Software aus	Computerexpertin	Sie macht die gesamte Buchhaltung und Organisation am Computer
118	Maria ist reich	Maria hat keinen grünen Daumen	Sie hat einen Garten mit vielen schönen Pflanzen und Blumen	Sie hat vielleicht/ derzeit einen Garten mit vielen schönen Pflanzen und Blumen	Vielleicht/ Derzeit hat sie einen Garten mit vielen schönen Pflanzen und Blumen	Garten	Ihr Mann interessiert sich sehr für Botanik
119	Pia möchte Ordnung schaffen	Pia ist sehr chaotisch	Sie kauft sich einen Ordner und Büroklammern	Sie kauft sich eventuell/ morgen einen Ordner und Büroklammern	Eventuell/ Morgen kauft sie sich einen Ordner und Büroklammern	Ordner	Sie braucht einen sicheren Ort für ihre wichtigsten Unterlagen
120	Walter fährt zum Baumarkt	Walter ist spät dran für einen wichtigen Termin	Er steuert eine Tankstelle an, um dort einen Zwischenstopp zu machen	Er steuert möglicherweise/ gerade eine Tankstelle an, um dort einen Zwischenstopp zu machen	Möglicherweise/ Gerade steuert er eine Tankstelle an, um dort einen Zwischenstopp zu machen	Tankstelle	Er muss tanken, weil er sonst sein Ziel nicht erreichen kann

Note. Crossing each of the two contexts (plausible vs. implausible) with each of the five versions of the target sentence (no marker vs. epistemic marker with focus on phrase vs. epistemic marker with focus on sentence vs. neutral marker with focus on phrase vs. neutral marker with focus on sentence) resulted in the ten different versions of each item that were counterbalanced across participants via ten item lists.