Running Head: Understanding conflicting information on social science issues

How non-experts understand conflicting information on social science issues: The role of perceived plausibility and reading goals

Johanna Maier and Tobias Richter

University of Kassel

accepted for publication in the Journal of Media Psychology (2012)

Author note

The research reported in this article was supported by the German Research Association (Deutsche Forschungsgemeinschaft, DFG, grant RI 1100/5-1 and RI 1100/5-2). We would like to thank Sara Wippich for preparing stimulus materials and collecting data.

Corresponding author:

Johanna Maier

University of Kassel, Department of Psychology

Holländische Str. 36-38

34109 Kassel, Germany

E-mail: johanna.maier@uni-kassel.de

Telephone: +49-561 804 3582

Fax: +49-561 804 3586

Abstract

When people read Web-based science-related journalistic articles, they usually read more than one text. As a consequence, they face the task to construct a coherent mental model of the issue on the basis of conflicting information. We assume that recipients handle this task by evaluating the plausibility of information against their current understanding and prior knowledge (epistemic validation). On the one hand, information judged as implausible is often not processed further, yielding a bias towards plausible information in the mental model (plausibility bias). On the other hand, recipients can also engage in elaborative processing when they are motivated to develop a justified point of view (epistemic reading goal) which should lead to a rich mental model in contrast to the goal to memorize facts (receptive reading goal). The present study investigated the relationships of perceived plausibility and comprehension of multiple articles related to a social science topic (the PISA study) and effects of recipients' reading goal with multilevel models (items nested within recipients) on a trial-by-trial basis. As predicted, information judged as plausible was more likely integrated into recipients' mental model. This plausibility bias was independent of recipients' reading goal. However, an epistemic reading goal led to an overall stronger mental model than a receptive reading goal. Moreover, we found a positive relationship between perceived plausibility and memory for text. The present results demonstrate that recipients make sense of science communication about controversial issues by actively monitoring the plausibility of information and regulating comprehension processes according to their reading goals.

Reference Terms: comprehension, epistemic validation, plausibility, reading goal, science communication

Introduction

The World Wide Web is one of the main sources for science-related information for lay persons and scientists alike. New scientific results are produced and published at a fast pace, often prompting an instant discussion on the Internet. As a consequence, the World Wide Web offers a multitude of documents which represent different perspectives, dissimilar empirical evidence and conflicting argumentative positions on the same scientific issue. Taken together, these circumstances provide new challenges for recipients of science-related media products on the Web. Getting an adequate picture about the validity of explanatory models of climate change, for instance, requires recipients to process divergent and often conflicting information from different web sites, to evaluate the credibility and plausibility of this information and to integrate it into a coherent and well-justified point of view (Perfetti et al., 1999; Rouet, 2006). In contrast to reading science textbooks, where relationships between different explanatory models of scientific findings are made explicit and divergent perspectives are usually resolved, using the World Wide Web as source of science-related information requires recipients to resolve inconsistencies from multiple web-based articles on their own.

In this article, we will argue that recipients of science-related journal articles on the Web handle this problem by using their prior knowledge to judge the validity and plausibility of new information (*epistemic validation*, Richter, 2003; Richter, 2011), which often leads to a plausibility bias in the mental representation (Schroeder, Richter, & Hoever, 2008). Against this background, the present study investigated the relationship between plausibility judgments and comprehension outcomes constructed in studying multiple articles that contain conflicting information. Moreover, recipients search and read science-related information on the Web for different purposes and the processing of conflicting information from multiple texts is a partly strategic (goal-dependent) process (Richter & Schmid, 2010; Wiley & Voss, 1999). For these

reasons, we examined whether and to what extent recipients' reading goal affects the relationship of plausibility judgments and comprehension outcomes. In the following sections, we will discuss these issues in turn. We will then report data from a web-based experiment in which university students read multiple journal articles on an issue from the social/educational sciences, which is hotly debated in public every three years when new results are published: what conclusions (if any) can be drawn from the results of the PISA study (OECD Programme for International Student Assessment; Klieme et al., 2010; OECD, 2010)? To investigate the relationship between perceived plausibility and comprehension on different levels of representation (memory for text and the situation model, Kintsch, 1988), two types of comprehension tasks (recognition and verification) were used with different types of test items. Relationships between responses to these tasks and plausibility judgments were analyzed with a multilevel design (test items nested within persons; Raudenbush & Bryk, 2002). This multilevel approach enabled us to examine on a trial-by-trial basis how the perceived plausibility of certain pieces of information is related to the integration of the same information into recipients' mental representation of the controversial issue.

Multiple Documents Comprehension and Epistemic Validation

Reading multiple journalistic articles on science-related issues is an instance of multiple documents comprehension (Perfetti et al., 1999). Perfetti et al. (1999) proposed that in the best of cases recipients of multiple documents construct a rich documents model which consists of two parts: a referential representation of the text content attached to the respective source information and an intertext model that represents argumentative relationships – including conflicting and discrepant information – between different documents. This task requires recipients not only to comprehend, memorize and integrate information from various documents. In order to detect conflicting and discrepant information, recipients also need to evaluate the plausibility of

arguments and information presented in a given document in the light of arguments and information from other documents and their own knowledge and beliefs (*epistemic validation*, Richter, 2003).

In previous research, the validation of text information has often been conceptualized as part of a broader metacognitive activity of comprehension monitoring which refers to evaluating one's own level of understanding of texts during reading and the detection of comprehension difficulties (Westby, 2004). For example, Baker (1985, 1989) distinguishes in her theoretical account of metacomprehension between two higher-level comprehension monitoring standards that are closely related to epistemic validation: the external consistency and the internal consistency standard. Evaluating comprehension for external consistency means validating text information with regard to own prior knowledge, thus checking texts for violations of prior knowledge. Evaluating comprehension for internal consistency refers to the validation of text ideas in light of other text ideas, thus detecting contradictions and argumentative fallacies in the texts. Baker (1985) found in a study with college students that readers differ in their use of the internal and external consistency standards. Only readers with high verbal abilities consistently and spontaneously acknowledged violations of these standards, i.e., information that contradicted general world knowledge or other information given in a text.

However, other studies using more indirect methods such as reading times, eye-tracking, or event-related potentials suggest that readers routinely monitor the plausibility and internal consistency of incoming textual information, even though they might not necessarily be able to report violations of the external and internal consistency standards (e.g., Baker, 1989; Lea, 1995; Lea, Mulligan, & Walton, 2005; Richter, Schroeder, & Wöhrmann, 2009; Singer, 1993, 2006; Singer, Halldorson, Lear, & Andrusiak, 1992). For example, readers have no problem in detecting logical inconsistencies in a text when the logical relationship is signaled or the premise

information is held active in working memory (Lea, 1995; Lea et al., 2005). Recipients also validate the inferred bridging information when they draw causal and other types of bridging inferences (Singer et al., 1992; Singer, 1993). Singer (2006) used a reading time paradigm with stories that contained target sentences (e.g., The coach determined/figured that it was oranges that Ken ate) that were 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). The pattern of reading times for these sentences demonstrated that inconsistent sentences produced longer reading times. Moreover, in an event-related potentials study (ERP) with the same materials, Feretti and colleagues found an extended N400 response that occurred in implausible sentences immediately after the word that created the implausibility (Ferretti, Singer, & Patterson, 2008). In ERP studies, the N400-response describes a stereotyped electrophysiological response that is associated with semantic processing. In particular, it may be regarded as an index for integration costs associated with information which does not immediately make sense given the linguistic context of a message. The extended N400 negativity response after the word that caused the implausibility most likely indicates pragmatic processing costs associated with the implausibility. Hence, these results provide electrophysiological evidence that text verification processes occur immediately after reading a word that causes implausibility. Similar, Richter et al. (2009) showed that recipients monitor the truth value of statements even if this process interferes with an unrelated, non-semantic and non-epistemic task. In this study, participants' task was to provide orthographical judgments on half true and half false simple statements. Participants needed more time to provide the orthographical judgment for invalid sentences than for valid sentences. This result allows for the conclusion that – even though it was not required by and even interfered with the actual task – participants validated the truth and plausibility of the sentences.

These findings allow for the conclusion that recipients regularly monitor the consistency of text information with their world knowledge (and other text information) and that the cognitive processes involved here are a routine and early part of comprehension. Reading multiple journalistic articles on science-related issues should be no exception from this rule. Quite to the contrary, articles dealing with a controversial social science issue which is hotly debated in public – such as the interpretation of the PISA results – are particularly likely to provoke spontaneous (im)plausibility judgments which depend on recipients' prior beliefs, prior knowledge and their current understanding about the topic. In the next section, we will discuss how such plausibility judgments affect the comprehension of science-related articles.

The Role of Perceived Plausibility in Comprehending Multiple Documents

Provided that plausibility judgments are routinely involved in the comprehension of multiple journalistic articles on science-related topics, the question arises how recipients process information they judge as plausible or implausible during reading. Several lines of research in social as well as cognitive psychology suggest that recipients often stick to previously learned information rather than changing them when encountering information that directly contradicts or discredits the earlier presented information (e.g., Johnson & Seifert, 1994; Oostendorp, 2002; Oostendorp, Otero, & Campanario, 2001; Ross, Lepper, & Hubbard, 1975; Schroeder et al., 2008). The bottom line of these studies is that recipients tend to disregard information that seems implausible in light of their current understanding (*plausibility bias*). For instance, in an experiment by Schroeder et al. (2008) university students read expository texts that contained plausible as well as (some) implausible sentences. The main research question was how the (perceived) plausibility of information affects comprehension, i.e. the construction of a mental representation of the state of affairs described in the texts (situation model, van Dijk & Kintsch, 1983; or mental model, Johnson-Laird, 1983). A multinomial models analysis of recognition and

plausibility judgments revealed a close bi-directional relationship of validation and the construction of the situation model: Plausible information was more likely to be integrated into participants' situation model than implausible information. On the other hand, information that was already part of the situation model was more likely to be judged as plausible, regardless of its "objective" plausibility. These results suggest that recipients have a tendency to reject information that does not fit into their current situation model. As a consequence, the integration of new information into a situation model depends in part on whether it is evaluated as being plausible.

In the comprehension of multiple texts on controversial science-related issues, rejecting information perceived as implausible can be a reasonable strategy. It allows recipients to maintain a coherent situation model and, hence, an internally consistent set of beliefs about a controversial issue with relatively little cognitive effort. On the downside, such a strategy may lead to a one-sided and impoverished mental representation of controversial issues. This raises the question if recipients can adopt other, more elaborate strategies of processing conflicting information from multiple texts depending on their reading goals. This possibility is discussed next.

Reading Goals Moderate the Role of Perceived Plausibility

The pervasiveness of rejecting implausible information notwithstanding, recipients of multiple texts do not always adopt such a simply strategy of maintaining a coherent mental representation of conflicting information. In some situations, the detection of incoming information as implausible might prompt recipients to engage in further knowledge-based processing directed at resolving the inconsistency between incoming information and their current understanding (*epistemic elaboration*; for a model-based account, cf. Johnson-Laird, Girotto, & Legrenzi, 2004). In that case, the cognitive conflicts which may be detected by

moment-by-moment plausibility assessments give rise to processes that are beneficial for the comprehension of plausible as well as implausible information. For example, when recipients notice that two science-related articles argue for contradictory theoretical positions, they might search for reasons supporting position A or position B, infer boundary conditions for the validity of position A and B, or think of ways how specific predictions of the two positions might be experimentally tested against each other. All of these processes can foster the construction of a rich and well-balanced mental representation and, hence, should work against the plausibility bias. The beneficial effects of such activities were shown in a study by Blanc, Kendeou, van den Broek, and Brouillet (2008). In their study students created logical and causal connections between different explanations for one event and were able to integrate contrary explanatory approaches into one coherent mental representation of the issue.

Epistemic elaboration is strategic and demands a large degree of cognitive effort – similar to other types of elaborative processing (O'Brien, Shank, Myers, & Rayner, 1988; Richter, 2003). Considering that the plausibility bias has been observed in so many studies and that human beings are cognitive misers (Fiske & Taylor, 1991), it seems likely to assume that recipients engage in effortful epistemic elaboration only if they are motivated and able to do so. One of the most important motivational preconditions for epistemic elaboration is that recipients follow a reading goal that requires a highly realistic and accurate mental representation of the issue. Such an *epistemic reading goal* can be created with the instruction to come to an own justified point of view (Richter, 2003) or to write an argumentation (Wiley & Voss, 1999). Instead of merely focusing on the given text information, these tasks involve that recipients really understand the meaning of the text, build inferences and connect different pieces of information with one another. In line with this idea, participants in a study by Wiley and Voss (1999) performed better in an essay writing task, as well as in an inference and an analogy task, if they

were instructed to write an argumentation instead of writing a narrative, summary or explanation. This pattern of results indicates that the instruction to write an argumentation increased recipients' elaboration on and an interconnection of the texts. In contrast, a *receptive reading goal* typically requires that recipients build a strong memory for text information without paying much attention on understanding and elaboration. Thus, tasks necessitating a thoughtful evaluation of arguments presented by a text (such as an epistemic reading goal) may be expected to foster the situation model construction in reading multiple texts. Moreover, since these tasks are assumed to create an epistemic elaboration of the texts content, such a task might also reduce the plausibility bias in situation model construction.

Rationale of the Present Experiment

The first aim of the present experiment was to investigate the role of perceived plausibility of information in the comprehension of multiple science-related articles with conflicting information. The second aim was to investigate whether and to what extent recipients' reading goal affects comprehension of multiple science-related texts and, possibly, moderates the relationship of perceived plausibility and comprehension. These research questions were investigated against the background of the notion of epistemic validation outlined in the previous paragraphs. The study was based on web-based magazine articles which covered a recent controversy from educational science, namely whether the results of the PISA studies from 2000 to 2006 indicate an overall improvement in the performance of German students and, hence, an improvement of the German educational system. PISA stands for Programme for International Student Assessment. It.is a comprehensive international school achievement study directed by the OECD (Organisation for Economic Co-Operation and Development). PISA assesses students' performance in reading comprehension, mathematics, and scientific literacy of students at the end of compulsory education (grade 9) in participating OECD countries Starting in the year 2000, the

PISA study is conducted every three years. In the aftermath of the publication of the PISA 2006 results lively discussions about the conclusion the study permits emerged in the German print media (Klieme et al., 2009). At the same time, German university students cannot be expected to have comprehensive knowledge about the PISA studies. This makes this social science issue particularly suitable for investigating the relationship between plausibility judgments and comprehension outcomes when lay people read science-related texts.

Hypotheses. Epistemic validation processes, and hence the effects of perceived plausibility, are supposed to occur in the course of the situation model construction. In other words, recipients' understanding of the scientific controversy should be closely related to what they find plausible. Accordingly, we expected that information judged as plausible by a recipient should more likely be integrated into the situation model than information judged as implausible (Hypothesis 1). However, recipients' reading goal is assumed to affect the construction of the situation model as well. In particular, an epistemic reading goals is likely to foster situation model construction by increasing the amount of knowledge-based, elaborative processing of conflicting information (e.g., Wiley & Voss, 1999). Against this background, we expected recipients who follow an epistemic reading goal to construct a situation model which is richer overall compared to recipients who follow a receptive reading goal (Hypothesis 2). Moreover, we also tested whether recipients' reading goal moderates the effect of perceived plausibility on the situation model construction. Such a moderating effect should occur if an epistemic reading goal fosters elaborative processing of conflicting information. Accordingly, the plausibility bias should be weaker if participants follow an epistemic reading goal compared to participants following a receptive reading goal (Hypothesis 3). Not only the situation model, but also memory for the text itself might be related to perceived plausibility as experiments examining plausibility effects on memory for text suggest. Whereas some experiments found a stronger memory for

plausible text information than for implausible text information (Black, Freeman, & Johnson-Laird, 1986; McAllister & Anderson, 1991) other experiments found a reverse pattern (Graesser, 1981; Schank & Abelson, 1977). For this reason, we investigated the relationship between plausibility judgments and memory for text as an exploratory research question.

Methodological approach. In order to investigate whether perceived plausibility was related specifically to the situation model construction or to memory for text as well, text comprehension was assessed on different levels with a recognition/verification task. First, recognition responses to paraphrases of sentences from the text were used to assess the strength of memory for text (propositional textbase). Second, verification responses to inference items were used to assess text comprehension on the level of the situation model (Kintsch, 1988). Moreover, participants provided validation responses, e.g. plausibility judgments, for the same set of test items. By analyzing the different responses to the test items (recognition, verification, and validation) with logistic multilevel models (test items nested within participants, Raudenbush & Bryk, 2002), we were able to investigate the relationship of plausibility judgments and comprehension outcomes on a trial-by-trial basis. As an additional asset, multilevel models provide an elegant mean to control for individual response tendencies in recognition, verification, and plausibility judgments by including responses to distracter items (information unrelated to the text) as predictors on the participant level in the multilevel model. Participants greatly differ in their propensity to provide positive recognition and verification judgments. Like in any type of forced-choice task, such response tendencies contain a response bias which can compromise estimates of text memory and situation model strength if these are based solely on proportions of true-positive recognition and verification judgments (Schmalhofer & Glavanov, 1986). By including false-negative responses to distracter items as participant-level predictors in a multilevel model, the response bias can be estimated and controlled effectively. Moreover, no

strong distributional assumptions are needed which is an advantage over the signal detection analysis that is typically used to control for response bias in forced-choice tasks (MacMillan & Creelman, 2005; Schmalhofer & Glavanov, 1986).

Method

Participants

Participants were 75 undergraduates (50 women and 25 men) majoring in Psychology or Education. Their average age was 24.9 years (*SD* = 5.8) and they received study credits or a small financial reward (8 Euros/hour) for participating in the experiment.

Text Material

The experimental texts were four texts that discussed longitudinal trends in the first three PISA studies (PISA 2000, 2003, and 2006). More specifically, the texts debated whether or not the PISA studies are able to indicate that the abilities of German students as well as the German educational system have improved from 2000 to 2006. The texts were based on science-related journal articles published in reputable German magazines that are available online as well as in print (such as the weekly magazine *Die Zeit*). The texts were comparable in writing style and length (600-800 words each). In order to investigate how students deal with conflicting journalistic texts on controversial social science issues, the texts took opposing positions in the discussion about the PISA results. Thus, two of the texts argued that the data from the PISA studies demonstrate a positive trend in students' performance from 2000 to 2006, whereas the other two texts argued against this proposition. In the following, the former two texts will be referred to as pro-texts whereas the latter two texts will be referred to as contra-texts. The texts were not accompanied by source information to ensure that such information could not be used as a cue to assess the plausibility of individual text information.

Comprehension and Plausibility Measures

Text comprehension was measured with a combined recognition/verification task (modified after Schmalhofer & Glavanov, 1986). After each text, participants were presented with 30 sentences that were paraphrase, inference or distracter items (ten sentences per item type, see Table 1 for examples translated into English). Paraphrase items contained information directly presented by sentences included in the texts. Thus, responses to paraphrases were indicative for recognition memory for text. In contrast, inference items matched the content of the text but presented information that was not explicitly stated in the text. Rather, this information needed to be inferred by the participants. Responses to inference items were supposed to be indicative of situation model strength. Finally, distracter items contained information that was neither a paraphrase nor an inference but that superficially matched the broad content discussed by the texts. Responses to distracter items were used to control for response bias. Participants' task was to indicate for each test item whether it contained a paraphrase of information explicitly mentioned in the text (recognition), an inference matching the contents of the text, but not explicitly mentioned (verification), or no information related to the text. True-positive classification of inference items served as an indicator of situation model integration whereas true-positive classifications of paraphrase items served as an indicator of text memory. False-positive classifications of distracter items as paraphrases and inferences were used in order to control for participants' response tendencies (similar to Schmalhofer & Glavanov, 1986).

In a separate block of questions, participants were asked to indicate for the same set of test items whether they found the statement expressed in these items plausible or not (validation task). Again, we used participants' validation responses to the distracter items in order to control for participants' response tendency to judge items as being plausible.

Recipient Characteristics

Prior attitudes. Participants' global attitudes towards the issue of long-term trends in the

German PISA results were assessed by participants' agreement to four statements (response categories ranging from 1= *totally agree* to 5 = *totally disagree*). Two statements expressed the claims close to the two pro-texts (e.g., *The PISA studies made successful reforms of the educational system possible*) whereas the other two statements expressed claims close to the two contra-texts (e.g., *The PISA studies show that the educational system is still in need of reforms*). Despite the fact that the items expressed opposing views, the correlations between the two pro-items and the two contra-items were not negative but ranged from zero to moderately positive ($.00 < r < .28$), yielding a scale with a low internal consistency (Cronbach's $\alpha = .49$). Neither the attitude scale nor the individual items were correlated with any of the dependent variables. For this reason, the attitude measure was not included in the analysis.

Prior knowledge. Participants' prior knowledge was assessed with seven multiple choice questions (one correct answer and three distracters). In the present sample, the multiple choice test reached a poor internal consistency of .45. Moreover, participants' knowledge was overall very low as indicated by the mean item difficulty of .38, which is only slightly above the 25 percent probability of guessing the right answer. Given the low overall level and the resulting low variance of prior knowledge, the measure was not included in the analyses.

Procedure

Participants started by reading the four experimental texts which were presented paragraph by paragraph on a computer screen. Half of the participants read the texts with the goal to memorize as much factual information from the text as possible (*receptive reading goal*), whereas the other half read the text with the goal to develop an own point of view on the topic covered by the texts (*epistemic reading goal*). In order to ensure that participants maintained the reading goal throughout the experiment, participants in the epistemic reading goal condition were prompted to provide ratings whether or not the argumentation in the paragraph was internally

consistent and whether or not the arguments were consistent with what participants already knew about the topic. These prompts were used to focus recipients on epistemic elaboration by a) judging the intertext-consistency of the paragraph and a) judging the relationship of the text information with their prior knowledge. In contrast, participants in the receptive reading goal condition provided ratings of how well they were able to comprehend and memorize the information in each paragraph. These prompts were used to focus participants in the receptive reading goal instruction on the accumulation of facts. Comprehension judgments for the test items were measured directly after each text was presented. Participants were asked to judge whether the test item contained a paraphrase of a sentence from the text, an inference matching the contents of the text, or no information related to the text (recognition/verification task). Subsequently participants provided plausibility judgments for the same set of test items (validation task). Responses were given by pressing one of two response keys (marked green and red for *yes* and *no*, respectively). Finally, participants were asked to recall what instruction they had been given for reading the texts in the beginning of the experiment. Seventy-two percent of the participants were able to explicitly recall their reading goal instruction. The rest of the participants failed to recall the reading instruction correctly because they answered the question inaccurately (20 %) or provided insufficient responses such as “clear” (8 %). At the end, participants were thanked and debriefed.

Design

The design was a multilevel design because the dependent variables (recognition, verification, and validation responses) were located on the level of test items (level 1) and the independent variables were located either on the level of the test items (level 1) or on the level of participants (level 2). The structure of the multilevel model is explained in more detail in the section *Multilevel Model*.

Results

Multilevel Model

We used multilevel models (generalized linear mixed models) to analyze responses to the recognition/verification task (McCulloch & Searle, 2001; Raudenbush & Bryk, 2002; Richter, 2006; Richter & Naumann, 2002). Multilevel models are the method of choice here because the data conform to a multilevel structure: Responses to test items (level 1) were nested within each participant (level 2). Due to the fact that a measured variable (plausibility judgments) was included as predictor variable, the design was not fully balanced, necessitating a model which accounts for test items as well as participants as sources of error variance. For inference and paraphrase test items, two separate two-level logistic regression model were estimated (hierarchical non-linear model with a logit link function, cf. Barr, 2008). In the model of the inference items, the true positive classification of inferences as inferences vs. all other classifications was used as outcome variable. The probability of true-positive classifications was interpreted as an indicator of situation model integration. For the paraphrase items, the true positive classifications as paraphrases vs. all other classifications served as outcome variable. This variable was interpreted as an indicator of the memory for text.

Parameters were estimated using a restricted maximum likelihood in combination with generalized least-squares estimates. Fixed effects were estimated with a population-average model with robust standard errors (computed with HLM 6.08, Raudenbush, Bryk, & Congdon, 2009). In all models, responses to the comprehension question were used as binary outcome variables and parameter estimation was based on a Bernoulli-sampling model with a logit-link function (Raudenbush & Bryk, 2002, ch. 10).[1]

Level 1 model. On the item level, the main predictor was participant's plausibility judgment for each test item (contrast-coded: -1 = implausible vs. 1 = plausible). Including this

predictor allowed us to investigate the relationship between recipients' recognition or verification responses to the test items as outcome variable and their plausibility judgments for the same test items as predictor variable. In addition, the position of the text which the item referred to within the experimental session (ranging from 1 to 4, grand-mean centered) and the argumentative stance of the text (contrast-coded; -1=contra vs. 1 = pro) were included as control variables. The resulting model for level 1 (test items) was the following:

$$\ln\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_{0ij} + \beta_{1ij}\, Plausibility_{ij} + \beta_{2ij}\, Position_{ij} + \beta_{3ij}\, Stance_{ij} + r_{ij,}$$

In this model, the probabilities of a true-positive recognition or verification response by participant *j* to item *i* (given as logg-odds/logits) serve as the outcome variable (for the logic of using the logit-link function here see Raudenbush & Bryk, 2002, ch. 10). The intercept β_{0ij} represents the average log-odds for participant j after controlling for plausibility, text position and text stance. The coefficient β_{1ij} represents the slope of plausibility on the log-odds after controlling for text position and stance. Similarly, β_{2ij} specifies the slope of text position and β_{3ij} the slope of text stance while controlling for the remaining level 1 variables. Finally, r_{ij} is the level 1 error term.

Level 2 model. The level 2 model incorporated between-participants variability of the level 1 intercept and the level 1 coefficients. Four submodels were specified, one for the intercept and one for each of the slopes of the level 1 predictors:

$\beta_{0ij} = \gamma_{00} + \gamma_{01}\, Reading\ Goal_j + \gamma_{02}\, Responses\ Distracters_j + \gamma_{03}\, Plausibility\ Distracters_j +$

u_{0j} (intercept model)

$\beta_{1ij} = \gamma_{10} + \gamma_{11}\, Reading\ Goal_j + \gamma_{12}\, Responses\ Distracters_j + \gamma_{13}\, Plausibility\ Distracters_j +$

u_{1j} (model for the slope of plausibility)

$\beta_{2ij} = \gamma_{20} + \gamma_{21}\ Reading\ Goal_j + \gamma_{22}\ Responses\ Distracters_j + \gamma_{23}\ Plausibility\ Distracters_j + u_{2j}$ (model for the slope of position)

$\beta_{3ij} = \gamma_{30} + \gamma_{31}\ Reading\ Goal_j + \gamma_{32}\ Responses\ Distracters_j + \gamma_{33}\ Plausibility\ Distracters_j + u_{3j}$ (model for the slope of text stance)

Participants' reading goal (contrast-coded: -1 = receptive reading goal vs. 1 = epistemic reading goal), as well as the mean proportion of positive recognition/verification responses to the distracter items in the recognition and verification task (*Responses Distracters*, z-standardized) and the validation task (*Plausibility Distracters*, z-standardized) were included as predictors in all level 2 models. Moreover, all level 2 models incorporated an error term, allowing the intercept and the slopes to vary randomly between participants (random coefficients model). Including participants' reading goal into the intercept model allowed us to estimate the main effect γ_{01} of this variable on recognition/verification responses. Including participants' reading goal in the submodel of the slope of plausibility allowed us to estimate the cross-level interaction γ_{11} between reading goal and plausibility. In other words, this coefficient captures the interaction of plausibility and reading goal, i.e. the extent to which participants' reading goal moderated the effects of perceived plausibility on recognition/verification responses. The mean proportion of false-positive recognition/verification responses to the distracter items in the recognition/verification task and the mean proportion of positive responses ("plausible" responses) to the distracter items the validation task were included to control for participants' response tendencies in these tasks.

All hypothesis tests were based on type-I-error probability of .05. Descriptive statistics and intercorrelations of all variables are provided in Table 2. In addition to parameter estimates (see Table 3 for the fixed effects and Table 4 for the variance components), we will report

predicted (conditional) probabilities of yes-responses (back-transformed from the logit-link model with estimated standard errors) for each model. We start by reporting the parameter estimates of the model for inference items as this model is central to the question of whether and to what extent there is a plausibility bias on the situation model level when participants read multiple articles with conflicting information. We will then report the estimates of the model for the paraphrase items in order to examine whether a similar bias occurs on the level of the memory for text.

Situation Model Strength: Multilevel Model for Inference Items

In the model for the inference items, true-positive classifications of inferences in the recognition/verification task served as an indicator of situation model integration (for parameter estimates, see Table 3, left-hand columns). In line with Hypothesis 1, there was a strong positive effect of plausibility in this model (parameter γ_{10}), indicating that inference items judged as plausible were much more likely to be correctly identified as inferences matching the text content (P = .50, SE = .02) compared to inference items judged as implausible (P = .33, SE = .02; Figure 1a). This effect was moderated by the individual proportion of positive validation responses to distracter items (parameter γ_{13}): The higher a participant scored on this variable, the weaker was the effect of plausibility on true-positive classifications of inferences for this participant. Given that the individual proportion of positive validation responses reflects participants' general tendency to judge information as plausible, the negative moderator effect of this variable suggests that the plausibility effect cannot be accounted for by response tendencies at the time of testing. Rather, the plausibility effect seems to be an effect which is due to processes going on during reading.

In line with Hypothesis 2, reading goal exerted a positive main effect (parameter γ_{01}) on the true-positive classification of inferences: Overall, participants following an epistemic reading goal were more likely to correctly identify inferences (P = .44, SE = .02) compared to participants following a receptive reading goal (P = .38, SE = .02; Figure 1a). Notably, the effect of reading goal occurred over and above a positive main effect of the individual proportion of positive verification responses to the distracter items (γ_{02}). Thus, the effect of reading goal cannot be explained by participants' response tendencies at the time of testing.

Finally, the cross-level interaction of plausibility and participants' reading goal (parameter γ_{11}) predicted by Hypothesis 3 was not significant. Thus, contrary to our assumptions, the effects of perceived plausibility and reading goal on the situation model were completely additive.

In sum, consistent with Hypothesis 1, information perceived as plausible was also more likely to be integrated into participants' situation model of the controversial issue. Consistent with Hypothesis 2, an epistemic reading goal yielded a stronger situation model of the issue compared to a receptive reading goal. However, Hypothesis 3 predicting that the plausibility effect would be weakened by an epistemic reading goal could not be supported.

Memory for Text: Multilevel Model for Paraphrase Items

For the paraphrase items, we estimated a model with the true positive classifications of paraphrases as information explicitly mentioned in the text in the recognition/verification task as outcome variable (Table 3, right-hand columns). In this model, again a strong positive effect of plausibility emerged (indicated by parameter γ_{10}). Paraphrase items judged as plausible had a considerably higher chance of being correctly recognized as information provided by the text (P = .55, SE = .03) compared to paraphrase items judged as implausible (P = .23, SE = .02; Figure

1b). For the paraphrase items we found neither a significant main effect of participant's reading goal (parameter γ_{01}) nor a significant cross-level interaction of reading goal and plausibility (parameter γ_{11}).

Discussion

The present experiment investigated how plausibility judgments and recipients' reading goals are related to comprehension outcomes when laypeople read journalistic articles on a controversial social science topic. Most importantly, results revealed a strong relationship between perceived plausibility of text information and situation model strength, e.g. recipients mental representation: inferences perceived as plausible by recipients were more likely to be integrated in the situation model compared to inferences judged as implausible. Besides this plausibility bias, which also occurred for the memory for information explicitly provided in the text, we also found a main effect of recipients' reading goal for situation model strength. Recipients who were instructed to develop an own point of view on the text topic (epistemic reading goal) built a stronger situation model than recipients who were instructed to memorize as many facts as possible (receptive reading goal). However, we failed to find the expected interaction between situation model construction and reading goal.

These results are broadly in line with the proposed theoretical framework of epistemic validation according to which plausibility judgments play an important role when recipients comprehend multiple science-related texts with conflicting information (Richter, 2011). More precisely, the positive relationship between perceived plausibility of text information and situation model integration supports the idea that recipients verify text information against their world knowledge and, as one consequence, tend to reject information perceived as implausible (Schroeder et al., 2008). In comprehending multiple journalistic articles on controversial issues,

this mode of dealing with implausible information is a simple way to maintain a coherent (albeit not necessarily balanced) point of view even though conflicting information is processed. A similar mechanism might underlie failures to update a situation model when new information is presented which is at odds with information presented earlier. For example, failures to update the situation model are quite common in unfolding news reports when initial information turns out to be false and is corrected by later information (e.g., Blanc et al., 2008; Johnson & Seifert, 1994). In addition, there is evidence that the likelihood of situation model updating depends on whether the new information is consistent with recipients' prior beliefs (Lewandowsky, Stritzke, Oberauer, & Morales, 2005). From this perspective, the research reported here contributes to the confirmation bias (my-side bias), i.e. recipients' tendency to prefer information which supports their prior beliefs over belief-inconsistent information or their tendency to re-interpreted belief-inconsistent information so that it becomes consistent with their prior beliefs (Nickerson, 1998). However, the present study goes beyond existing research on the confirmation bias in two respects. First, given that recipients in the present study were rather ambivalent towards the global claims raised in the texts, the present results suggest that perceived plausibility is used as an heuristic for the selection of text information for further processing even in the absence of strong global attitudes (which were in the focus of previous research; see, Baron, 1995, for a representative example). Second, our study showed on a trial-by-trial basis that perceived plausibility is strongly and positively related to both comprehension (situation model strength) as well as memory for text.

The plausibility effect on memory for text is in line with prior experiments (Black et al., 1986; McAllister & Anderson, 1991) which found a positive relationship between plausibility judgments and memory for text. Paraphrases of sentences from the text had a higher likelihood of being recognized as coming from the text when they were perceived as plausible. One possible

explanation of the fact that a positive rather than a negative relationship of perceived plausibility with memory for text was obtained is that participants' prior knowledge regarding the PISA studies was quite low. As a consequence, participants might have lacked the schematic knowledge structures which are necessary if unusual or implausible information is to receive a privileged status in the memory representation of the text (as proposed by the schema-pointer-plus-tag hypothesis, Schank & Abelson, 1977). Further research should explore this possibility by including prior knowledge as a moderator of plausibility effects on the levels of the situation model and memory for text.

The present experiment can be placed in the wider theoretical context of research on the comprehension of multiple documents on science-related issues as they are typical for informal Web-based learning about science topics (e.g., Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti et al., 1999; Stadtler, Scharrer, & Bromme 2011). In terms of the seminal framework of multiple documents comprehension proposed by Britt et al. (1999), the present research focused on aspects of content integration – to what extent and how recipients build a coherent situation model on the basis of conflicting information from multiple texts– rather than source selection, tagging and separation. Our findings suggest that the perceived plausibility of information is used by recipients as a cue as to what information they should and should not include in their situation model of the scientific controversy. In prior research on multiple texts and Web-based science communication, the focus has been on recipients' evaluation of source credibility and its impact on comprehension (e.g., Bråten, Strømsø, & Britt, 2009; for a review see Metzger, 2007). This research suggests that recipients sometimes use external cues such as source reputation or endorsement-based credibility judgments as credibility cues to validate Web-based information (Metzger, Flanagin, & Medders, 2010). Our research expands this research insofar as it demonstrates that recipients also seem to rely on their subjective plausibility judgments as one

efficient mean to select information for further processing. We would assume that the perceived plausibility of information should become especially important if credibility cues such as site appearance or source reputation are not available (Metzger et al., 2010), which was the case in the present experiment.

Recipients' mental representation of the scientific issue (their situation model) was not only related to the perceived plausibility of text information, but also to recipients' reading goal. In line with previous research (Richter, 2003; Wiley & Voss, 1999) recipients following an epistemic reading goal created a stronger situation model compared to recipients following a receptive reading goal. However, we did not find the expected interaction between plausibility judgments and situation model construction. Descriptively, the pattern of results was in line with the assumption that an epistemic reading goal should attenuate the plausibility bias but the corresponding interaction effect was not significant. One likely cause of the lack of an interaction effect might be the strong main effect of perceived plausibility. In the presence of a main effect, the power for establishing an ordinal interaction effect is low because main effect and interaction effect overlap in terms of explained variance (e.g., Bobko, 1986). As a consequence, ordinal interactions are often difficult to establish even if they represent true population effects. Moreover, there was an overall high level of perceived plausibility, resulting in a restricted variance of this predictor. This might have rendered the occurrence of an interaction between plausibility judgments and reading goal even less likely. Nevertheless, the main effect of reading goal suggests that recipients can strategically control their comprehension of multiple articles that contain conflicting information. Similar effects of reading goals on knowledge-based inferences have been found for reading narrative and expository texts (Narvaez, van den Broek, & Ruiz, 1999). In the study by Narvaez et al. (1999), participants who were instructed to read for study generated more evaluations of the text and identified more knowledge-based coherence breaks

during reading compared to participants who were instructed to read for entertainment. Thus, participants reading purpose lead to an adjustment in the inferential activities the readers engaged in. One possibility, which should be investigated in experiments with process-oriented measures, is that an epistemic reading goal specifically fosters elaborative processing, which itself facilitates the construction of a rich and well-developed situation model.

One limitation of the present experiment is that plausibility judgments have been assessed after recipients had read the texts. This procedure was necessary to maintain a naturalistic reading situation which mimics informal learning with science-related web-based articles as closely as possible. In contrast, the task to provide plausibility judgments for individual sentences during reading would have distorted the normal reading process extensively. Due to this procedure, our data only supports a correlative relationship of subjective plausibility and comprehension although the multilevel design used in our study is suitable to rule out certain alternative explanations in terms of response tendencies at the time of testing. Theoretically, it seems reasonable to assume that a bi-directional causal relationship underlies the relationship between perceived plausibility and comprehension (as Schroeder et al., 2008, have established it for learning with expository texts): The more plausible some piece of information appears to the reader, the more likely is it that it will be integrated into the current situation model. On the other hand, the easier it is to integrate some piece of information into the current situation model, the more plausible it will appear to the reader. The assumption of such a bi-directional relationship fits well into the theoretical framework of epistemic validation. Nevertheless, further studies should investigate the causal relationship between validation processes and comprehension outcomes in learning with multiple texts more directly, even at the risk that this investigation has to be accomplished with a rather artificial learning situation.

Moreover, we assessed perceived plausibility with forced-choice yes-/no-questions rather than rating scales with multiple levels. In this way, participants had to decide whether they perceive a statement as plausible or as implausible and were prevented from havering. While this approach has certain methodological advantages, it might also create an over-simplified picture because subjective plausibility may be regarded as a continuous dimension. Evidently, fine-grained plausibility differences and non-linear relationships of perceived plausibility with comprehension and memory for text cannot be captured with dichotomous plausibility judgments. For this reason, future studies should also use more sensitive measures such as rating scales to investigate how information neither perceived as plausible nor as implausible affects the comprehension of multiple texts.

In addition, our data does not readily provide an answer to the question, from whence participants derived their plausibility judgments. In the framework of epistemic validation proposed in the present study, we would assume that even if participants had low prior knowledge about the topic, they still used their fragmentary, subjective and maybe even objectively wrong prior knowledge as background for plausibility judgments. In addition, recipients could also base their plausibility judgments on fine-grained beliefs, as well as their current understanding they construct during reading. For instance, participants might not know whether or not German students achieved more or less points in the PISA study from 2006 compared to the results of 2003, but they might have beliefs about the overall quality of the German educational system. These beliefs are loosely associated with the discussion about the PISA studies and hence, recipients might use these beliefs to judge corresponding text information as plausible or implausible. Further studies on this issue should take a closer look at what kind of knowledge and beliefs participants access during plausibility judgments. Think-

aloud protocols or a direct manipulation of recipients' knowledge and/or beliefs might serve as a good starting point for such an investigation.

Another question that is not answered by the present study is whether the present results generalize to topics outside the social/educational sciences. Does it make a difference, for example, whether recipients are told that they read a text on an issue from the social/educational or from the natural sciences? One might suspect that the "hardness" of scientific results commonly associated with the natural sciences might cause recipients to abstain from making plausibility judgments. However, there are natural science issues which are intensely debated in public such as the risks and benefits of nuclear energy, the causes of climate change, or health topics such as the benefits of regular vaccinations. In recent experiments, we have found comparable plausibility effects for such topics, which seems to suggest that the plausibility effect is a general phenomenon (Maier & Richter, 2012).

In many societies, the World Wide Web has become the standard source for science-related information. Our results and the theoretical perspective of epistemic validation advocated here indicate how recipients can deal with the new challenges evoked by the Word Wide Web, namely building an understanding of a scientific issue on the basis of multiple articles that contain conflicting information. Plausibility judgments as regular part of comprehension can explain how recipients come to a (one-sided) understanding of controversial scientific issues with minimized cognitive effort. Moreover, fostering epistemic elaboration with an epistemic reading goal can be one effective educational mean to support recipients to become more "digitally literate" in understanding Web-based scientific communication.

References

Baker, L. (1985). Differences in the standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly, 20*, 297-313.

Baker, L. (1989). Metacognition, comprehension monitoring and the adult reader. *Educational Psychology Review, 1*, 3-38.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 7*, 221-235.

Barr, D.J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Black, A., Freeman, P., & Johnson-Laird, P.N. (1986). Plausibility and the comprehension of text. *British Journal of Psychology, 77*, 51-62.

Blanc, N., Kendeou, P., van den Broek, P., & Brouillet, D. (2008).Updating situation models during reading of news reports: Evidence from empirical data and simulations. *Discourse Processes, 45*, 103-121. doi:10.1080/01638530701792784

Bobko, P. (1986). A solution to some dilemmas when testing hypotheses about ordinal interactions. *Journal of Applied Psychology, 71*, 323-326.

Bråten, I. Strømsø, H., & Britt, A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly, 44*, 6-28.

Britt, M.A., Perfetti, C.A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S.R. Goldman, A.C. Graesser & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209-233). Mawah, NJ: Erlbaum.

Ferretti, T.R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition*, *108*, 881-888. doi:10.1016/j.cognition.2008.06.002

Fiske, S. T., & Taylor, S. E. (1991). Social cognition (2nd ed.). New York: McGraw-Hill.

Graesser, A.C. (1981). *Prose comprehension beyond the word*. New York: Springer.

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: Toward a cognitive science of language, inference and consciousness*. Boston, MA: Harvard University Press.

Johnson-Laird, P.N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*, 640-661.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Eds.) (2010). *PISA 2009: Bilanz nach einem Jahrzehnt* [PISA 2009: Balance after one decade]. Münster, Germany: Waxmann.

Lea, R.B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 18-26.

Lea, R.B., Mulligan, E.J., & Walton, J.L. (2005). Accessing distant premise information: How memory feeds reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 387-395.

Lewandowsky, S., Stritzke, W.G.K., Oberauer, K. & Morales, M. (2005). Memory for fact, fiction, and misinformation. *Psychological Science, 16*, 190-195.

Maier, J., & Richter, T. (2012). Plausibility effects in the comprehension of controversial science texts. *Manuscript submitted to publication*.

McAllister, P.O., & Anderson, A. (1991). Conservatism and the comprehension of implausible text. *European Journal of Social Psychology*, *21*, 147–164. doi:10.1002/ejsp.2420210205

McCulloch, C.E., & Searle, S.R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide*. (2nd ed.). Mahwah, NJ: Erlbaum.

Metzger, M.J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*, 2078-2091.

Metzger, M.J., Flanagin, A.J., & Medders, R.B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*, 413-439.

Narvaez, D., van den Broek, P., & Ruiz, A.B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology, 91*, 488-496.

Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many disguises. *Review of General Psychology, 2*, 175-220.

O'Brien, E.J., Shank, D., Myers, J.L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 410-420.

OECD (2010). *PISA 2009 Results: What students know and can do: Student performance in reading, mathematics and science* (*Vol. I)*. Paris: OECD.

Oostendorp, H. van (2002). Updating mental representation during reading scientific text. In J. Otero, J. Leon, & A. Graesser (Eds.), *The psychology of science text comprehension* (pp. 309-332). Mahwah, NJ: Erlbaum.

Oostendorp, H. van, Otero, J., & Campanario, J.M. (2001). Conditions of updating during reading. In M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary studies* (pp. 55-76). Philadelphia, PA: John Benjamins.

Perfetti, C.A., Rouet, J.F. & Britt, M.A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). Mawah, NJ: Erlbaum.

Raudenbush, S. W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Raudenbush, S.W., Bryk, A.S., & Congdon, R.T. (2009). *HLM 6: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen* [Epistemic validation in text comprehension]. Lengerich, Germany: Pabst Science Publishers.

Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes, 41*, 221-250.

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. & Naumann, J. (2002). Mehrebenenanalysen mit hierarchisch-linearen Modellen [Multilevel analysis with hierarchical linear models]. *Zeitschrift für Medienpsychologie, 14* (N.F. 2), 155-159.

Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning, 5*, 47-65. doi:10.1007/s11409-009-9038-4

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.

Rouet, J.F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mawah, NJ: Erlbaum.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25,* 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*, 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.

Stadtler, M., Scharrer, L., & Bromme, R. (2011). How reading goals and rhetorical signals influence recipients' recognition of intertextual conflicts. In L. Carlson, C. Hoelscher, & T.F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1346 -1351). Austin, TX: Cognitive Science Society.

Van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Westby, C. (2004). A language perspective on executive functioning, metacognition, and self-regulation in reading. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.). *Handbook of language and literacy: Development and disorders* (pp. 398–427). New York: The Guilford Press.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 301-311. doi:10.1037/0022-0663.91.2.301

Footnotes

[1] In addition to the results reported here, we estimated a multilevel model with only the subsample of 54 participants who were able to recall their reading goal instruction explicitly and correctly. The pattern of significant results was identical (with slightly more pronounced parameter estimates) to the results for the complete sample, so only the latter are reported here.

Table 1:

Examples of Test Items used in the Recognition/Verification and Validation Tasks

Sentence type	
Original text part	Focus of the PISA study in the year 2000 has been on reading and text comprehension, 2003 on mathematics and for the first time on natural sciences. From the point of view of the OECD, conclusions about the performance trend for the basis competencies reading and text comprehension are available at the official release of the PISA study on December 4th – at this time point these competencies have been tested three times. The same will hold for mathematics after the year 2009, for the natural sciences after 2012. The German PISA coordinator Manfred Prenzel, however, has a different view on this matter. He already spotted an increase in the abilities of German students in math and science after the second PISA round in 2003.
Paraphrase	The German PISA coordinator noticed improvements in the math and science abilities of German school children since 2003.
Inference	The German coordinator has a much more positive view on the changes in the German school system compared to the officer of the OECD.
Distracter	The construct of reading competence that has been assessed in the PISA study in 2003 comprised abilities in text comprehension as well as reflective skills.

Table 2:

Descriptive Statistics and Intercorrelations of Predictor Variables, Covariates, and Criterion Variables

			Correlations							
	M	*SD*	1	2	3	4	5	6	7	8
1 Reading Goal (contrast-coded, -1 = receptive vs. 1 = epistemic) [a]	-0.07	1.00	1							
2 Text Stance (contrast-coded, -1 = contra vs. 1 = pro) [a]	0.00	1.00	.00	1						
3 Position of the Text (centered) [a]	0.00	1.12	-.00	.02	1					
4 Plausibility (contrast-coded, -1 = implausible vs. 1 = plausible) [b]	0.56	0.83	-.01	.03	-.03	1				
5 Responses Distracters [b]	0.13	0.11	-.03	.00	.00	.08***	1			
6 Plausibility Distracters [b]	0.49	0.17	.02	.00	.00	.05**	.16***	1		
7 Situation Model [b]	0.45	0.50	.05**	-.07***	-.01	.15***	.07***	.01	1	
8 Memory for Text [b]	0.49	0.50	-.03	.06**	-.02	.25**	-.04*	.04*	.40***	1

Note. Responses Distracters: Proportion of false positive responses to distracters in the recognition/verification tasks. Plausibility Distracters: Proportion of positive responses to distracters in the validation task. Situation Model: Proportion of true positive classification of inferences in the verification task. Memory for Text: Proportion of true positive classifications of paraphrases as text information directly stated in the text.

[a]$N = 75$ (person level). [b]$N = 3000$ (item level).

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

Table 3:

Estimates for Fixed Effects in the Logistic Multilevel Model for the Responses to the Inference (Left Columns) and Paraphrase Items (Right Columns)

	Responses to Inference Items		Responses to Paraphrase Items	
Parameter	*Estimate (SE)*	*t* (71)	*Estimate (SE)*	*t* (71)
Intercept (γ_{00})	-0.37 (0.05)	-7.20***	-0.49 (0.08)	-6.34***
Level 1 (Item Level)				
Reading Goal (γ_{01})	0.14 (0.05)	2.87**	-0.03 (0.08)	-0.41
Responses Distracters (γ_{02})	0.16 (0.06)	2.54*	-0.20 (0.09)	-2.16*
Plausibility Distracters (γ_{03})	0.07 (0.05)	1.34	0.20 (0.09)	2.35*
Level 2 (Person Level)				
Plausibility (Main Effect, γ_{10})	0.35 (0.06)	5.90***	0.71 (0.06)	12.86***
Plausibility X Reading Goal (γ_{11})	-0.06 (0.06)	-0.98	-0.04 (0.05)	-0.74
Plausibility X Responses Distracters (γ_{12})	-0.05 (0.05)	-0.94	0.09 (0.06)	1.5
Plausibility X Plausibility Distracters (γ_{13})	-0.15 (0.06)	-2.66*	-0.14 (0.06)	-2.37*

Position (Main Effect, γ_{20})	0.00 (0.03)	0.09	-0.06 (0.03)	-1.89
Position X Reading goal (γ_{21})	-0.01 (0.03)	-0.18	-0.01 (0.03)	-0.37
Position X Responses Distracters (γ_{22})	-0.03 (0.04)	-0.75	-0.06 (0.03)	-2.00*
Position X Plausibility Distracters (γ_{23})	-0.02 (0.04)	-0.66	-0.03 (0.03)	-1.03
Stance (Main Effect, γ_{30})	-0.16 (0.04)	-4.27***	0.12 (0.04)	2.89**
Stance X Reading Goal (γ_{31})	-0.00 (0.04)	-0.05	0.06 (0.04)	1.48
Stance X Responses Distracters (γ_{32})	-0.06 (0.03)	-2.13*	0.01 (0.04)	0.34
Stance X Plausibility Distracters (γ_{33})	-0.02 (0.04)	-0.36	-0.01 (0.05)	-0.18

Note. Reading Goal: Contrast-coded, -1 = receptive reading goal vs. 1 = epistemic reading goal. Responses Distracters: Proportion of false positive responses to distracters in the recognition/verification task, z-standardized. Plausibility Distracters: Proportion of positive responses to distracters in the validation task, z-standardized. Plausibility: Contrast-coded, -1 = implausible vs.1 = plausible. Position: Position of the text in the experimental session, centered. Stance: Argumentative stance of the text, contrast-coded, -1 = contra (PISA does not show an improvement of German students) vs. 1 = pro (PISA shows an improvement of German students).

* $p = .05$, ** $p = .01$, *** $p = .001$ (two-tailed).

Table 4:

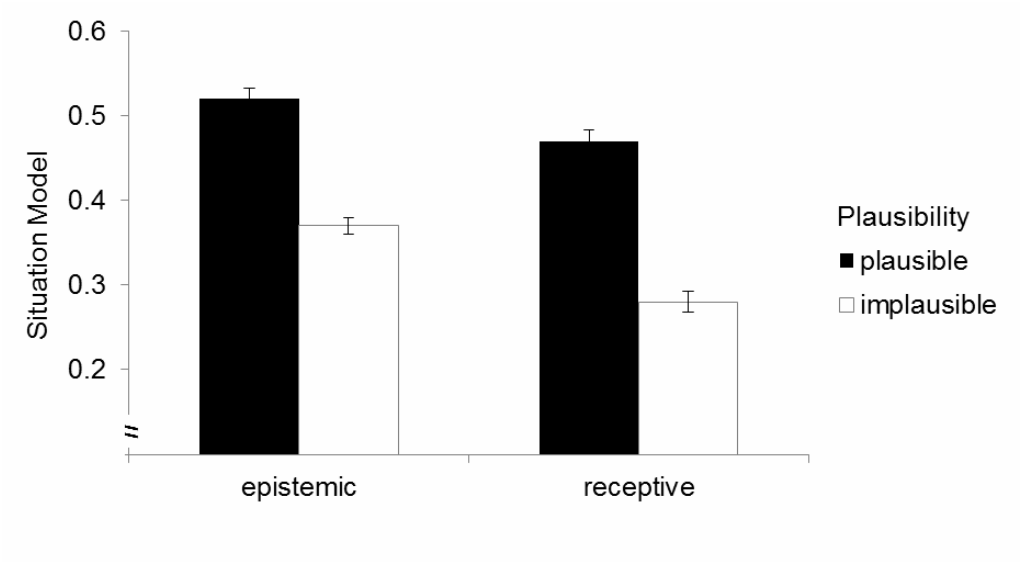
Estimates for Variance Components in the Logistic Multilevel Model for the Responses to the Inference (left) and Paraphrase Test Items (right)

Parameter	Responses to Inference Items		Responses to Paraphrase Items	
	Estimate	χ^2 (71)	Estimate	χ^2 (70)
Intercept	0.05	94.10*	0.43	158.70***
Plausibility	0.13	119.97***	0.08	81.11
Position	0.01	89.04	0.02	72.65
Stance	0.00	75.68	0.06	101.40**

Note. Plausibility: Contrast-coded, -1 = implausible vs. 1 = plausible. Position: Position of the text in the experimental session, centered. Stance: Argumentative stance of the text, contrast-coded, -1 = contra (PISA does not show an improvement of German students) vs. 1 = pro (PISA shows an improvement of German students).

* p = .05, ** p = .01, *** p = .001 (two-tailed).

a)



b)

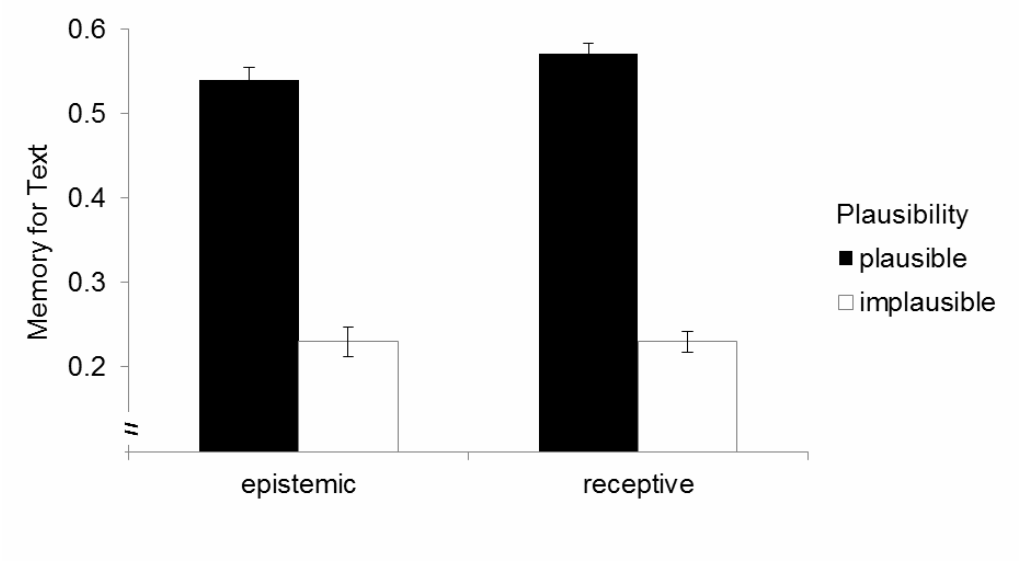


Figure 1. Effects of perceived plausibility and reading goal (epistemic vs. receptive) on (a) situation model strength (predicted conditional probabilities of true-positive classifications of inferences, back-transformed from the logit-link model, with estimated standard errors) and (b) memory for text (predicted conditional probabilities of true-positive classifications of paraphrases, back-transformed from the logit-link model, with estimated standard errors).