The Role of Source Credibility in the Validation of Information Depends on the Degree of (Im-)Plausibility

Andreas Wertgen, Tobias Richter, Jean-François Rouet

The Role of Source Credibility in the Validation of Information Depends on the Degree of (Im-)Plausibility

Andreas G. Wertgen[1], Tobias Richter[1] and Jean-François Rouet[2]

[1]Department of Psychology IV, University of Würzburg

[2]CNRS and University of Poitiers

Accepted for publication in the journal *Discourse Processes* (2021)

Corresponding Author:

Andreas Gabriel Wertgen

University of Würzburg, Department of Psychology IV

Röntgenring 10

97070 Würzburg, Germany

E-mail: andreas.wertgen@uni-wuerzburg.de

Telephone: +49 (0) 931 31-88736

Abstract

This study examined the role of source credibility in the validation of factual information embedded in short narratives. In a self-paced reading experiment, we tested the assumption that the degree of (im-)plausibility determines the extent that source credibility affects validation during comprehension. We used reading times of target and spillover sentences and plausibility judgments as indicators of validation. Participants read stories with a high- vs. low-credible person (expert vs. non-expert) who made plausible, somewhat implausible, or highly implausible assertions. Reading times increased and plausibility judgments varied as a function of knowledge consistency, decreasing from knowledge-consistent to implausible to knowledge-inconsistent items. Moreover, interactions of source credibility and plausibility were found for reading times of spillover sentences and plausibility judgements, indicating that source credibility and plausibility are jointly considered in validation. High-credible sources mitigated the perceived implausibility of somewhat implausible sentences but exacerbated the perceived implausibility of highly implausible information. A corresponding interactive pattern was found for the reading times of the spillover sentences. Thus, implicit and explicit indicators provided converging evidence that the modulating role of source credibility in validation depends on the degree of implausibility.

Keywords: Validation, Plausibility, Sourcing, Credibility, Text Comprehension

The Role of Source Credibility in the Validation of Information Depends on the Degree of (Im-)plausibility

A growing body of research has shown that text comprehension entails the validation of text information, that is, (implicit) judgments of its truth or falsity, as an integral component of situation model construction and updating (e.g., Cook & O'Brien, 2014; O'Brien & Cook, 2016a; 2016b; Richter et al., 2009; Singer, 2013). Yet, a dearth of research exists on conditions that affect validation. One active area of research is concerned with how world knowledge and contextual information are used in validation (e.g., Isberner & Richter, 2014; van Moort, Koornneef & van den Broek, 2018; 2020; Walsh et al., 2018; Williams et al., 2018). Source credibility is a particular type of contextual information that bears a strong conceptual relationship to the validity of information and might thus be especially relevant for validation. In particular, information on the credibility of a source might be used by the reader to decide whether the information provided by the source is believable or not. Recent research indicates an interactive relationship of source credibility and the plausibility of information. The bottom line of this research is that validation seems to rely primarily on the fit between text information and the world knowledge that readers activate during comprehension, but source credibility can exert an additional modulating influence (e.g., Braasch et al., 2012; Foy et al., 2017, Wertgen & Richter, 2020). Nevertheless, the underlying cognitive mechanisms are still unclear because the interactive patterns differ across experiments. A key to a better theoretical understanding of how exactly source information is used in validation might be to investigate the degree of (im-)plausibility of text information. In other words, the size and the direction of the modulating effect of source credibility might depend on the extent that the information is (im-)plausible. To test this assumption, we manipulated plausibility gradually by extending the experimental design used by Wertgen and Richter (2020) with an additional

intermediate level of plausibility between plausible (world-knowledge consistent) and highly implausible (world-knowledge inconsistent) information.

In the following review, we will briefly discuss research on validation during comprehension and review studies that have examined the role of source credibility in text comprehension. We will then discuss the small body of studies that have examined combined effects of plausibility and source credibility. The discussion of this research will provide the background to justify the focal assumption that the role of source credibility in validation depends on the degree of (im-)plausibility.

Validation as Implicit Assessment of Plausibility

A considerable body of research indicates that readers use their world knowledge and contextual information to routinely evaluate the plausibility of text information, a process that has been coined validation (O'Brien & Cook, 2016a; 2016b; Richter, 2015; Singer, 2013; 2019). In this context, plausibility can be defined as the "acceptability or likelihood of a situation or a sentence describing it" (Matsuki et al., 2011, p. 926). Experiments with reaction times based on the epistemic Stroop paradigm in which false, belief-inconsistent, or implausible statements slow down affirmative responses in an unrelated task provide strong evidence for routine validation with various types of linguistic and audio-visual stimuli (e.g., Gilead et al., 2019; Isberner & Richter, 2013, 2014; Piest et al., 2018; Richter et al., 2009). Further corroborating evidence for routine validation as an integral part of text comprehension stems from experiments based on a wide range of methods such as eye tracking (e.g., Matsuki et al., 2011), event-related potential data (e.g., Ferretti et al., 2008), and reading times (e.g., Cook & O'Brien, 2014; for an overview, see Isberner & Richter, 2014). For example, a typical finding from reading time experiments based on the so-called inconsistency paradigm is a slowdown for target sentences that are inconsistent with information provided earlier in the text (e.g., O'Brien et al., 1998) or that are inconsistent with world knowledge (e.g., Rapp, 2008).

The Resonance-Integration-Validation Model (RI-Val) proposed by O'Brien and Cook (2016a; 2016b) contains the assumption that resonance (activation), integration, and validation of information are three passive processes that, once started, run to completion. Text information provides cues that activate knowledge in a resonance-like process (*R*; Myers & O'Brien, 1998; O'Brien & Myers, 1999). This information is linked with content in active memory (*I*) based on conceptual overlap. These linkages are then validated against information in active memory (*Val*). Once a certain degree of coherence is matched, the reader continues reading. All three processes, resonance, integration, and validation can influence comprehension. The RI-Val Model also contains the temporal assumption that resonance, integration, and validation overlap but start successively (in a cascade-like style). Depending on readers' coherence threshold, which may vary according to their standards of coherence (van den Broek et al., 1995; 2011), validation may run to completion after the reader has moved on in the text. If that is the case, validation effects may not occur during reading the sentence whose contents are validated but during reading subsequent sentences. Actually, in reading time studies based on the inconsistency paradigm, information that is inconsistent with previous text or world knowledge slows down reading not so much on the implausible sentence but on the subsequent ("spillover") sentence (Cook & O'Brien, 2014; O'Brien & Cook, 2016a; 2016b).

Evaluation of Source Credibility

Research on source credibility usually focuses on expertise or trustworthiness as the two major dimensions of source credibility (Lombardi et al., 2014; Self, 2009). In the present research, we focused on the expertise dimension, which refers to "the extent to which a speaker is received to be capable of making correct assertions" (Pornpitakpan, 2004, p. 244). The effects of source credibility have mainly been investigated in multiple text comprehension. In multiple text reading situations, readers must integrate information from multiple texts from different perspectives on a specific topic (e.g., scientific texts about

vaccines). Ideally, readers build a mental representation that includes an integrated mental model and an intertext model of the texts (Perfetti et al., 1999). The intertext model represents source characteristics of texts (e.g., text type, author, language style) and the argumentative relationship between documents (e.g., "Text A supports Text B; opposes Text C"). Thus, source features and evaluations of source credibility based on these features can help readers make sense of multiple texts with conflicting information. Extant research shows that engaging more in processing source features, such as evaluating texts for trustworthiness based on source information, may improve multiple text comprehension (e.g., Bråten et al., 2009; Goldman et al., 2012; Wiley et al., 2009; Wineburg, 1991).

Evidence for an Interplay of Plausibility and Source Information during Comprehension

To our knowledge, only few studies have investigated the combined effects of source credibility and plausibility on validation and comprehension. The discrepancy-induced source comprehension (D-ISC) model assumes that when readers encounter discrepant (i.e., inconsistent) text information, they shift their attentional resources to sources and their characteristics, possibly in an attempt to resolve the discrepancy. Preliminary evidence for this assumption stems from the experiments by Braasch et al. (2012) and other research associated with the D-ISC model (for an overview see Braasch & Bråten, 2017). Readers showed better memory for discrepant text versions, fixated on source information more, and spent more time on source information. Although the D-ISC assumption holds that the detection of inconsistencies, which may be construed as the outcome of validation processes, intensifies sourcing during moment-by-moment processing, it does not specify how source information, in turn, affects the validation of text information. To our knowledge, Foy et al. (2017) conducted the first study to examine this question by investigating in reading time experiments with short narratives how trustworthy or untrustworthy sources affect the validation of implausible (e.g., seeing wolves in the backyard at a party) and plausible (e.g.,

seeing that it rains outside at a party) assertions. Participants were not able to determine the truth status of these assertions as they referred to events in the story world. However, the assertions were implausible (the described events unlikely) or plausible (the described events likely) according to general world knowledge. Foy et al. found that implausible assertions were read faster when they came from trustworthy sources. However, there was no effect of source credibility on plausible assertions. Their experiments show that plausibility and source credibility each affect validation with stronger effects exerted by plausibility. Wertgen and Richter (2020) followed a similar approach by collecting reading times and explicit plausibility judgments to examine joint effects of plausibility and source credibility (high vs. low-expertise sources) on validation. Unlike Foy et al., however, Wertgen and Richter used target sentences that were clearly consistent (e.g., Jupiter is the biggest planet in the Solar System) or inconsistent (e.g., The sun is the biggest planet in the Solar System) with general world knowledge and could thus be accepted or rejected based on activated knowledge. They found that credible sources slowed down the reading of world-knowledge inconsistent sentences and lowered their plausibility of world-knowledge inconsistent sentences. A possible explanation for the divergent effects of source credibility in the experiments by Foy et al. and Wertgen and Richter might lie in the degree of implausibility of information. This idea is elaborated in the following section.

The Role of Source Credibility Possibly Depends on the Degree of (Im-)Plausibility

We start from the general assumption that source credibility as contextual information and plausibility are jointly considered in validation processes. However, world knowledge dominates the validation process in most cases, whereas source credibility only modulates it, and the direction of these modulatory effects depends on the degree of implausibility. Social judgment theory (e.g., Sherif et al., 1965; Sherif & Sherif, 1967) provides a framework to conceptualize the differential effects of source credibility depending on the degree of (im-)plausibility. According to social judgment theory, judgements of belief-relevant information

occur on a continuum with latitudes of acceptance, rejection and noncommitment. Plausibility varies along such a continuum (Isberner & Richter, 2014). Analogous to the notion of latitudes of acceptance, rejection, and noncommitment in social judgment theory, we assume that the influence of source credibility depends on the degree of (im-)plausibility. When an assertion in a text is clearly consistent with accessible world knowledge, this knowledge should dominate validation, and the influence of source credibility should be minimized (comparable to the latitude of acceptance). With decreasing plausibility, the impact of source credibility should increase. Thus, source credibility should mitigate the implausibility of text information that readers cannot clearly reject or accept (comparable to the latitude of noncommitment) but only until a certain degree of implausibility is reached. If the information is clearly false (i.e., highly implausible), the effect of source credibility flips. In that case, a high-credible source creates a mismatch between source credibility and the false information, increasing the disruptive effect of the inconsistency during comprehension.

Rationale of the Present Experiment

The present research tested the assumption whether the degree of (im-)plausibility affects how source credibility is considered in validation during comprehension. To this end, we extended the experimental design used by Wertgen and Richter (2020) by adding a level of plausibility that is between extreme points of plausibility. We used sentences that were highly plausible (world-knowledge consistent), somewhat implausible and highly implausible (clearly world-knowledge inconsistent). These sentences were embedded in short stories and stated by a person described as a source with a high or low level of expertise. This method made it possible to investigate the relationship between plausibility and source credibility for validation as a continuum. We included online measures (reading times) and explicit measures (plausibility judgments) to investigate possible convergences and divergences in moment-to-moment processes during reading and more global judgments after reading (Rapp & Mensink, 2011).

For target sentences, we expected readers to process plausible sentences faster than sentences that are somewhat implausible or highly implausible. Additionally, we expected readers to process somewhat implausible sentences faster than highly implausible sentences (Hypothesis 1a). More importantly, we expected source credibility and plausibility to interact (Hypothesis 2a). Somewhat implausible sentences by high-expertise sources should lead to faster reading times compared with low-expertise sources. With increasing implausibility, we expected the pattern to flip, that is, longer reading times for highly implausible sentences asserted by a high-expertise source compared with low expertise sources. Reading times of plausible sentences should be unaffected by source credibility. Thus, we expected world knowledge to dominate validation in assertions that were close to the endpoints of the plausibility continuum and source information to exert an effect in assertions that readers cannot clearly reject or accept.

In line with the temporal assumptions of the RI-Val model (O'Brien & Cook, 2016a; 2016b), the expected effects on target sentences should also be revealed and might even be more pronounced on the subsequent (i.e., spillover) sentences. Thus, we expected reading times of spillover sentences to increase with decreasing plausibility (Hypothesis 1b). We also expected plausibility and source credibility to interact, with longer reading times for a high-expertise vs. a low-expertise source in highly implausible sentences and the reverse pattern in somewhat implausible sentences (Hypothesis 2b).

For plausibility ratings, we expected a decline in plausibility ratings from plausible sentences to somewhat implausible sentences to highly implausible sentences (Hypothesis 1c), mirroring the Hypotheses 1a and 1b. Moreover, we expected plausibility and source credibility to interact on plausibility ratings. For highly implausible sentences, a low-expertise source should lead to higher plausibility ratings than a high-expertise source, whereas the opposite pattern should occur for somewhat implausible sentences (Hypothesis 3).

Method

Participants

Ninety-nine participants with an average age of 24.40 years (*SD* = 8.14 years) participated in the experiment. Most participants were female (80%) and university students (89%). On average, the university students had completed 2.93 semesters (*SD* = 2.66). The data from six participants who spoke a first language other than German were excluded from the analyses. Thirty-two participants received study credit and 67 participants received a monetary compensation (11 Euros).

Material

The experimental materials were 36 eight-sentence short stories (number of words: *M* = 100.66, *SD* = 12.75) that were based on the materials developed by Wertgen and Richter (2020) and extended by five newly developed stories. The stories described everyday situations (e.g., vacations, restaurant visits). The third sentence described the protagonist either as a source with high or low credibility (person with high vs. low expertise in a certain field, e.g., a physics professor vs. a hairdresser apprentice making a statement about theory of relativity). The sixth (target) sentence was an assertion made by the protagonist in direct speech. This assertion was plausible (i.e., consistent with world knowledge), somewhat implausible, or highly implausible (see Table 1 for an example story) and matched the field of expertise mentioned in the description of the protagonist. The three categories of assertions in the target sentences were based on the general knowledge norms reported by Nelson and Narens (1980). Tauber et al. (2013) updated these norms and presented a table with the most frequent false responses. Based on this table, materials in the somewhat implausible condition corresponded to inaccurate statements that were provided by 6 to 65% of respondents as answers to knowledge questions with constructed responses.

The experimental stories had an average Flesch score (Flesch, 1948, German adaptation by Amstad, 1978) of 56.46 (*SD* = 5.84) which translates to “demanding” or “fairly difficult” to read. Moreover, 36 plausible filler stories were used (20 adapted and translated

from Foy et al., 2017). The filler stories were eight sentences long and were linguistically similar to and covered topics comparable to the experimental stories. However, the filler stories contained no cues to the protagonist's expertise and no direct speech.

Norming Study

A norming study was conducted with the experimental texts to confirm that the story versions differed in perceived credibility between the two sources and in perceived plausibility between the degrees of plausibility. The 48 participants were mostly female (88%) and undergraduates from the University of Würzburg. The average age was 23.38 years (*SD* = 6.27 years). They were compensated with 5 Euros. Participants read the 36 stories in a randomized order and rated the plausibility of the assertions (1 = "very implausible" to 7 = "very plausible") and the credibility of the source (1 = "not credible at all" to 7 = "very credible") with respect to the field of expertise associated with the assertion. Presentation of story versions and the order of the two rating tasks were counterbalanced across participants. High-expertise sources received higher source credibility ratings ($M = 4.50$, $SE = 0.12$) than low-expertise sources ($M = 3.12$, $SE = 0.12$), $\beta = 0.69$, $t(1641) = 15.43$, $p < .001$, $d = 0.70$. For plausibility ratings, we found the expected monotonic decline from plausible over somewhat implausible to highly implausible assertions. Plausible assertions ($M = 5.23$, $SE = 0.11$) were judged as more plausible as somewhat implausible assertions ($M = 3.76$, $SE = 0.11$), $t(1634) = 13.61$, $p < .001$, $d = 0.77$, and somewhat implausible assertions were judged as more plausible compared with highly implausible assertions ($M = 2.47$, $SE = 0.11$), $t(1636) = 11.98$, $p < .001$, $d = 0.68$. These findings suggest that the manipulation of plausibility and source credibility was successful.

Design

The design was a 2 (source credibility: high expertise vs. low expertise) x 3 (plausibility: plausible vs. somewhat implausible vs. highly implausible) within-subjects design. Each participant read one version of every story. The assignment of stories to

experimental conditions across participants was counterbalanced across participants. Participants read the stories in a randomised order.

Procedure

For the most part, the procedure was identical to Wertgen and Richter (2020). The experiment took place on two appointments in order to mitigate fatigue and order effects, which were 4.22 (*SD* = 3.83) days apart on average. Participants read all 72 stories (experimental stories plus filler stories) on a computer screen in a self-paced fashion (sentence by sentence) at the first appointment. Participants were instructed to read the stories for comprehension and to answer questions about the story after some of the stories. A fixation cross was displayed at the location of the first word for 500 ms. Participants could advance to the next sentence by pressing a key. Practice trials were included at the beginning to familiarize participants with the self-paced reading method. Letters in all sentences except the currently read one were masked with an 'x'. After every filler story, participants responded to a yes/no comprehension question. At the second appointment, participants were given a definition of plausibility ("Plausibility describes how likely we think it is that an assertion is true or that the described situation actually took place") and instructed to read the stories again in a self-paced fashion. Participants were asked to judge the plausibility of the target sentence on a scale from 1 (= "not plausible at all") to 7 (= "very plausible)". Subsequently, participants judged for each assertion whether the assertion is true or false and their confidence in their decision on a scale from 1 (= "not confident at all") to 7 (= "very confident"). This measure was included as a manipulation check for the plausibility manipulation and will be referred to as general knowledge test. The experiment lasted 64.87 min (*SD* = 21.57) on average.

Results

Reading times and plausibility ratings were analyzed with linear mixed models with random effects (random intercepts) of participants and stories (Baayen et al., 2008). The

models were estimated with the lmer function of the R package lme4 version 1.1.-23 (Bates et al., 2015). The emmeans function in the lsmeans package (Lenth, 2016) was used for follow-up tests and to derive model-based estimates of condition means and the associated standard errors. The Type-I-Error probability was set at .05 (two-tailed) in all significance tests. All factors were effect-coded, and their main effects and the interaction were entered as fixed effects in the model. Sources with high credibility (high expertise) were coded as 1, and sources with low credibility (low expertise) were coded as -1. For plausibility, two effect-coded contrasts were constructed. In the first contrast, plausible assertions were coded as 1, highly implausible assertions were coded as -1, and somewhat implausible assertions were coded as 0. In the second contrast, plausible assertions were coded as 1, somewhat implausible assertions were coded as -1, and highly implausible assertions were coded as 0. For analyses of reading times, sentence length and the position of the story in the experiment were entered in the model as centred predictors (fixed effects). One story version was excluded from all analyses because of a programming error (15 data points overall). In the reading times, we examined processing effects on a millisecond level. Therefore, data from six non-native speakers and three participants with low performance on comprehension questions (less than 80% correct) were excluded (276 data points, or 8.3% of data of target and spillover sentences) for reading times analyses because reading times were higher on average compared to native speakers and participants with satisfying performance on comprehension questions. Reading times lower than 500ms per sentence were excluded from the analysis (6 data points, or 0.2% of data for target sentences; 8 data points, or 0.2% of data for spillover sentences). Distributions of reading times normally have a positive skewness with extreme outliers in the right tail of the distribution (e.g., Ratcliff, 1993). To account for this characteristic while not excluding too much data, we excluded reading times that deviated more than 2 *SD* from the participant mean or the item mean from the analysis (247 data points, or 7.4% of data for target sentences; 217 data points, or 6.5% of data for spillover

sentences). After data trimming, reading times were only moderately skewed (0.87 for target sentences, 0.81 for spillover sentences). The final sample for the reading time analysis consisted of 90 participants with a mean accuracy of 91.62% (*SD* = 4.91) on the comprehension questions. See Table A1 for descriptive statistics of all dependent variables.

A separate analysis of plausibility judgments without data from non-native participants and from participants with a low performance on comprehension questions (parallel to the exclusion criteria for the reading time data) elicited no substantial differences in results. Thus, we excluded no data for the analysis of plausibility judgements.

We estimated linear mixed models with all predictors (full model) and compared them to reduced models to test the fixed effects of the main effect of plausibility and the interaction effect. These tests were based on differences in deviances (which follow a χ^2 distribution) between the models for target and spillover sentences and for plausibility ratings.

Moreover, we estimated effect sizes (Cohen's *d*) for differences in condition means based on the approximate formula proposed by Westfall et al. (2014) for linear mixed models with contrast-coding and one degree-of-freedom tests (see also Judd et al., 2017). We also conducted a post-hoc sensitivity analysis of the effects based on the method proposed by Westfall et al. (2014), as implemented in the accompanying web-based app (https://jakewestfall.shinyapps.io/crossedpower/). For the sensitivity analysis, we used the smallest effect size found (*d* = -0.10) and the corresponding variance components of the random effect of participants (0.37), the random effect of stories (0.08) and the residual variance (0.55) taken from the corresponding linear mixed model. All other variance components were assumed to be 0 since the random intercept of participants and stories were the only random effects in the model. With our sample size of 90 participants and 36 stories, we estimated a post-hoc sensitivity (1-β) of .99.

Reading Time for Target Sentences

Table 2 provides estimates and significance tests of the fixed effects in the model for the reading time of target sentences. The sentence length and the position of the story in the experiment had a significant effect on reading times. Longer target sentences led to longer reading times. More time was needed to read target sentences presented in stories appearing earlier in the course of the experiment.

As expected in Hypothesis 1a, we found a strong main effect of plausibility, $\chi^2(2) = 56.34$, $p < .001$. Participants read plausible sentences ($M = 3811$ ms, $SE = 113$ ms) faster than somewhat implausible sentences ($M = 4020$ ms, $SE = 113$ ms), $t(2670) = -4.37$. $p = < .001$, $d = -0.15$. They also read somewhat implausible sentences faster than highly implausible sentences, $t(2684) = -3.14$, $p = .002$, $d = -0.11$. However, no interaction effect of plausibility and source credibility emerged, $\chi^2(2) = 1.06$, $p = .589$ (Figure 1). Thus, Hypothesis 2a was not supported. Instead, the lack of evidence for an influence of source credibility on the reading times for the target sentence suggests a dominating role of world knowledge for initial validation processes.

Reading Time for Spillover Sentences

Table 3 provides estimates and significance tests of the fixed effect in the model for reading times of the spillover sentences. We found significant main effects of sentence length and item position. Longer spillover sentences led to longer reading times. Stories presented later in the experiment led to faster reading times.

As expected in Hypothesis 1b, the analysis revealed again a main effect of plausibility, $\chi^2(2) = 12.76$, $p = .002$. Sentences subsequent to plausible target sentences ($M = 2613$ ms, $SE = 68.40$ ms) were read significantly faster than sentences subsequent to highly implausible target sentences ($M = 2715$ ms, $SE = 68.50$ ms), $t(2675) = -3.56$, $p < .001$, $d = -0.12$. However, no significant difference was found between sentences subsequent to a plausible target sentence and sentences subsequent to a somewhat implausible target sentence, $t(2677) = -1.56$, $p = .120$.

The main effect was qualified by an interaction effect of source credibility and plausibility, $\chi^2(2) = 10.09$, $p = .006$. The pattern of the interaction is displayed in Figure 2. A somewhat implausible spillover sentence was read faster when it was combined with a high-expertise source ($M = 2614$, $SE = 71.30$) compared to a low-expertise source ($M = 2701$, $SE = 71.40$), $t(2675) = -2.14$, $p = .033$, $d = -0.10$. This pattern flipped on highly implausible sentences. That is, high-expertise sources lead to longer reading times in spillover sentences subsequent to highly implausible sentences ($M = 2763$, $SE = 71.20$) compared with low-expertise sources ($M = 2667$, $SE = 71.70$), $t(2676) = 2.35$, $p = .019$, $d = 0.11$. Thus, Hypothesis 2b, regarding the modulating role of source credibility for the validation of somewhat implausible and highly implausible information, was supported.

Plausibility Ratings

Table 4 provides estimates and significance tests of the fixed effects for the plausibility ratings. Plausibility ratings were available from 99 participants. Models that controlled for item position and the difference in days between the two appointments as centred metric predictors did not elicit substantial differences in results. Therefore, these two control variables were not included in the analyses.

As predicted in Hypothesis 1c, we found a strong main effect of plausibility, $\chi^2(2) = 1778.70$, $p < .001$. As expected, perceived plausibility declined from plausible ($M = 5.58$, $SE = 0.08$) to somewhat implausible assertions ($M = 3.62$, $SE = 0.08$), $t(3411.60) = 27.45$, $p < .001$, $d = 1.08$, and from somewhat implausible to highly implausible assertions ($M = 2.13$, $SE = 0.08$), $t(34113.38) = 20.74$, $p < .001$, $d = 0.82$. Moreover, there was a significant interaction effect, $\chi^2(2) = 25.10$, $p <. 001$ (Figure 3). In line with Hypothesis 3, a plausible statement by a high-expertise source ($M = 5.75$, $SE = 0.10$) was judged as more plausible than the same statement coming from a low-expertise source ($M = 5.40$, $SE = 0.10$), $t(3413) = 3.51$. $p = .006$, $d = 0.20$. In contrast, a high-expertise source making a highly implausible assertion ($M = 1.96$, $SE = 0.10$) lowered the plausibility compared with a low-expertise source ($M = 2.30$, SE

= 0.10), $t(3417) = -3.39$. $p = .009$, $d = -0.19$. However, Hypothesis 3 was not completely supported because there was no significant difference between somewhat implausible assertions by a high-expertise source ($M = 3.69$, $SE = 0.10$) compared with a low-expertise source ($M = 3.54$, $SE = 0.10$), $t(3413) = 1.53$. $p = .646$.

In sum, the findings for plausibility ratings show again that world knowledge is the primary source for validation, but source credibility can affect validation as well. Unexpectedly, we found no effect of source credibility on somewhat implausible sentences. Readers apparently neglected source information in the explicit judgments of somewhat implausible sentences.

Accuracy and Confidence in the General Knowledge Test

Participants recognized false and correct world-knowledge facts with an average accuracy of 81% ($SD = 39\%$). The individual accuracy ranged from 58% to 97%. Nine participants had an accuracy of less than 70%. Analyses that excluded these nine participants did not substantially change the effects relevant for the hypotheses. Therefore, data from these participants remained in the data file. Accuracies and confidence judgments differed between plausibility levels. On average, accuracy for plausible facts was 87% ($SD = 32\%$), 63% ($SD = 48\%$) for somewhat implausible facts, and 94% ($SD = 24\%$) for highly implausible facts. The confidence judgements mirrored this pattern (plausible: $M = 5.00$, $SD = 1.36$; somewhat implausible: $M = 4.25$, $SD = 1.58$; highly implausible: $M = 5.35$, $SD = 1.22$). This pattern of results suggests that the manipulation of plausibility was successful, with high accuracy and confidence for facts close to the endpoints of the plausibility continuum and only low accuracy (slightly above chance level) and lower confidence for somewhat implausible facts.

Discussion

The present experiment tested the assumption that the degree of (im-)plausibility affects the extent that source credibility is considered in validation during comprehension. We used reading times of target and spillover sentences as an implicit online-measure of

validation and plausibility ratings as an explicit offline-measure. Two major findings emerged. First, we found strong plausibility effects on reading times for target and spillover sentences and on the plausibility ratings (supporting Hypotheses 1a, 1b, and 1c). In line with numerous reading time studies based on the contradiction paradigm (e.g., Cook & Guéraud, 2014; see Cook & O'Brien, 2014, for an overview) and many other experiments on the role of plausibility in comprehension (e.g., Isberner & Richter, 2014), consistency with world-knowledge seems to have dominated validation, from the early phases of processing to explicit judgments of plausibility.

Second, in line with our main guiding assumption, we also found evidence for a modulating role of source credibility on implicit and explicit validation. We expected high-credible sources to increase the perceived plausibility of somewhat implausible text information whose veracity was difficult to determine based on participants' world knowledge but also to increase the perceived implausibility of highly implausible sentences whose falsity was easy to determine based on their world knowledge. No evidence was found for such an interaction of plausibility and source credibility on reading times for the target sentence (Hypothesis 2a). However, the expected pattern was found for the spillover sentences (Hypothesis 2b). The reading of somewhat implausible spillover sentences stated by a high-credible source was faster compared with the same sentences stated by a low-credible source. This effect flipped in spillover sentences subsequent to highly implausible sentences, which were read more slowly when the highly implausible assertions came from high-credible sources compared with low-credible sources. In sum, these results suggest that source credibility might not affect the immediate phases of processing a statement but that it takes effect after a delay – even though the simple main effects indicate that the modulating effect of source information is rather small, compared to the strong main effect of plausibility.

The temporal pattern of the reading-time effects is interpretable in the light of general assumptions about the competition of world knowledge and contextual information for

activation, integration, and validation. A growing consensus is that general world knowledge (knowledge-based validation) is the primary criterion for validation and therefore will dominate initial processing (e.g., Cook & Guéraud, 2005), but contextual information can also influence validation and comprehension (text-based validation). Note, however, that in some instances source information can be the primary criterion for assessing the plausibility or believability of information as evidence from studies on multiple text comprehension in novices vs. experts suggests (e.g., Rouet et al., 1997; Wineburg, 1991). A growing body of research sheds light on the relative importance of both types of information. For example, evidence from eye movements (van Moort, Koornneef & van den Broek, 2020) and neuroimaging (van Moort, Jolles, et al., 2020) indicates distinct time courses for knowledge-based and text-based validation. Among other findings, van Moort, Kornneef and van den Broek (2020) found stronger disruptive effects of inconsistencies based on world knowledge compared to contextual contradictions. As such, the strong effect of plausibility on target sentences found in the present experiment might be interpreted as further evidence for the dominating role of world knowledge in the initial validation of information. The pattern of effects is also in line with the temporal assumptions of the RI-Val model by which validation overlaps with integration but starts and runs to completion later, possibly after readers have moved on to the next sentence (Cook & O'Brien, 2014; O'Brien & Cook, 2016a; 2016b).

The direction of the effect of source credibility on the spillover sentences also differed between somewhat implausible and highly implausible sentences (as predicted in Hypothesis 2b). In somewhat implausible sentences, source credibility is informative because validation cannot lead to a conclusive outcome based on world knowledge alone. Therefore, high-credible sources can affect validation by mitigating the disruptive effect of implausible information during reading. Highly implausible sentences, in contrast, can be validated based only on world knowledge. In such sentences, a high-credible source is at odds with the outcome of the knowledge-based validation process, increasing to the disruption of the

reading process. The differential effects of source credibility for somewhat implausible and highly implausible sentences can be explained by assuming that plausibility forms a continuum that is structured by latitudes of rejection, noncommitment and acceptance (as described in social judgment theory, Sherif et al., 1965; Sherif & Sherif, 1967). Based on a similar conception of plausibility as a continuum, Hinze et al. (2014, Experiment 2) have compared accurate statements, inaccurate but plausible statements, and inaccurate and implausible statements, and collected readers' cognitive responses to these statements via think-aloud data. They showed that readers were more skeptical and less likely to accept inaccurate statements that were implausible, as compared to inaccurate statements that were plausible. According to the theoretical framework that our study is based on, whether a piece of textual information falls in the latitudes of rejection, noncommitment, or acceptance, is determined initially by the world-knowledge-dominated process of validation. Source information is most informative in the "gray" area of noncommitment but provides no additional information when the outcome of the validation process falls within the latitude of acceptance. This theoretical perspective not only accommodates the pattern of effects found in the present study but also integrates the seemingly divergent findings by Foy et al. (2017) and Wertgen and Richter (2020).

Of note, we found an influence of source information on the processing of somewhat implausible or highly implausible information, but not in the processing of plausible information. This pattern is also consistent with the D-ISC assumption (e.g., Braasch et al., 2012) and associated research (e.g., de Pereyra et al., 2014; Rouet et al., 2016; Saux et al., 2018), according to which attention to source information is triggered by inconsistent information.

Implicit validation processes are assumed to feed into explicit plausibility judgments (e.g., Schroeder et al., 2008). Therefore, the effect of high-credible sources on the moment-by-moment processing of somewhat implausible information should be mirrored in the

explicit plausibility ratings. We found an interaction effect of source credibility and plausibility on explict plausibility ratings, which was in line with the predictions (Hypothesis 3). Participants judged plausible sentences as even more plausible coming from a high-credible source. In contrast, high-credible sources lowered the plausibility of highly implausible sentences compared with low-credible sources. However, no significant effect of source credibility was found on the perceived plausibility of somewhat implausible sentences, although the pattern matched the one predicted by Hypothesis 3 descriptively. It is difficult to explain why the reading times and the plausibility ratings diverge at this point. Generally speaking, plausibility ratings are global judgements and more strategic compared with measurements of moment-by-moment processing; they partly rest on different (and more variable) psychological processes. These features might account for differences found between online indicators of reading processes and offline indicators of reading outcomes (Rapp & Mensink, 2011). Moreover, it must be kept in mind that the simple main effects of source credibility in the spillover reading times of the somewhat implausible information were significant but rather small. Thus, the impact of source information on online validation processes might have been too weak to carry over to the explicit plausibility judgments. Further research that elucidates the processes and the kind of information involved in the plausibility judgments is needed to clarify this point.

The present experiment raises interesting questions regarding the relationships between source information, validation, and readers' tendency to pick up false information from (fictional) stories (i.e., misinformation effects; Marsh & Fazio, 2006; Rapp, 2008). Although our study was not designed to study the influence of inaccurate information on readers' beliefs, we propose that validation may protect readers from misinformation effects. However, research shows that this protection is far from perfect (for overviews see Isberner & Richter, 2014; Singer, 2019), in part because readers do not always possess the required prior knowledge to tell accurate from inaccurate statements. Source information might be an

additional cue that modulates the accuracy of validation and, hence, the likelihood of misinformation effects to occur. For instance, a high-credible source in a narrative might exacerbate misinformation effects compared with a low-credible source. Note, that this interpretation is mostly speculative at this point and that further research, based on longer narratives and including plausible inaccuracies, is needed.

The core processes of comprehension, activation, validation, and integration are closely intertwined and jointly influence the mental representation that is constructed during reading (e.g. the RI-Val model, O'Brien & Cook, 2016a; 2016b). For example, activation interacts with validation as only information that is currently active can be used in validation. In the present study, we designed the experimental stories in a way that the activation of source information was highly likely by placing the source information close to the critical information in the target sentence. Moreover, we described the source in a way that readers could easily infer whether the source had a high or low credibility. Future research might focus on the interplay of validation and activation and how variations in factors that affect the activation of source information might also affect whether, how, and when source information is used in the validation process.

The results of the present experiment are consistent and make sense theoretically, but they need to be interpreted with its limitations in mind. One limitation is that we used short narratives developed by the experimenter with a schematic structure to enhance experimental control, and participants were required to read numerous stories successively. We cannot determine whether this relatively artificial reading situation might have induced specific strategies that altered the results. It would be worthwhile to replicate the basic finding of differential effects of source credibility for highly implausible and somewhat implausible sentences with various settings and task contexts. For instance, Sparks and Rapp (2011) and de Pereyra et al. (2014) found that source-focusing instructions impacted readers' attention to and memory for source information, respectively. Moreover, we used the paradigm of self-

paced reading (moving window) which poses certain restrictions on the reader. For example, participants could not revisit previously read text. Although studies have shown that text comprehension is only marginally impaired by a self-paced reading paradigm with linear reading (e.g., Chung-Fat-Yim et al., 2016), using eye tracking as a more naturalistic paradigm, which allows readers to regress to earlier sentences, would be a fruitful next step of research. Finally, the present research is also limited because we examined only three levels of plausibility, two levels of source credibility, and only one type of source credibility (expertise). A more comprehensive understanding of the interplay of plausibility and source credibility in validation could be gained by including a broader range of degrees of plausibility and credibility and other types of source credibility (such as trustworthiness).

To conclude, the present experiment yielded three important insights. First, we present further evidence for validation during text comprehension as found by numerous studies with the contradiction paradigm (e.g., O'Brien et al., 1998). Second, we provide further support for the general assumption that source credibility as contextual information and plausibility are jointly considered in validation but that source information might be considered after a slight delay. Third and most importantly, the assumption was supported that the role of source credibility depends on the degree of (im-)plausibility. The extent that source credibility affects validation seems to depend on the outcome of initial, knowledge-based validation processes that determine the degree of (im-)plausibility.

Author Note

We would like to thank Manuel Klein for his help in collecting data. The experimental texts, data files, and R-scripts for the full analyses are available at https://osf.io/w9htv/. The reported experiment was not preregistered. The authors report no conflict of interest.

References

Amstad, T. (1978). *Wie verständlich sind unsere Zeitungen?* [How understandable are our newspapers?] Unpublished dissertation, University of Zurich, Switzerland.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). c. *Journal of Memory and Language, 59* (4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

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. https://doi.org/10.3758/s13421-011-0160-6

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28. https://doi.org/10.1598/RRQ.44.1.1

Chung-Fat-Yim, A., Peterson, J. B., & Mar, R. A. (2017). Validating self-paced sentence-by-sentence reading: Story comprehension, recall, and narrative transportation. *Reading and Writing*, *30*(4), 857–869. https://doi.org/10.1007/s11145-016-9704-2

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), 365–378. https://doi.org/10.1080/0163853X.2005.9651683

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, *51*(1–2), 26–49. https://doi.org/10.1080/0163853X.2013.855107

de Pereyra, G., Britt, M. A., Braasch, J. L. G., & Rouet, J.-F. (2014). Reader's memory for information sources in simple news stories: Effects of text and task features. *Journal of Cognitive Psychology*, *26*(2), 187–204. https://doi.org/10.1080/20445911.2013.879152

Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition, 108*(3), 881–888. https://doi.org/10.1016/j.cognition.2008.06.002

Flesch, R. F. A. (1948). New readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233. https://doi.org/10.1037/h0057532

Foy, J. E., LoCasto, P. C., Briner, S. W., & Dyar, S. (2017). "Would a madman have been so wise as this?" The effects of source credibility and message credibility on validation. *Memory & Cognition, 45*(2), 281–295. https://doi.org/10.3758/s13421-016-0656-1

Gilead, M., Sela, M., & Maril, A. (2019). That's my truth: Evidence for involuntary opinion confirmation. *Social Psychological and Personality Science, 10*(3), 393–401. https://doi.org/10.1177/1948550618762300

Goldman, S. R., Braasch, J. L., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*, *47*(4), 356–381. https://doi.org/10.1002/RRQ.027

Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: Plausibility effects on memory for misinformation. *Memory & Cognition, 42*(2), 305–324. https://doi.org/10.3758/s13421-013-0359-9

Isberner, M.-B., & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, *142*(1), 15–22. https://doi.org/10.1016/j.actpsy.2012.10.003

Isberner, M.-B., & Richter, T. (2014). 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: MIT Press.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*, 601-625. https://doi.org/10.1146/annurev-psych-122414-033702

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lombardi, D., Seyranian, V., & Sinatra, G. M. (2014). Source effects and plausibility judgments when reading about climate change. *Discourse Processes, 51*(1–2), 75–92. https://doi.org/10.1080/0163853X.2013.855049

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. https://doi.org/10.1037/a0022964

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition, 34*(5), 1140-1149. https://doi.org/10.3758/BF03193260

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*(2–3), 131–157. https://doi.org/10.1080/01638539809545042

O'Brien, E. J., & Cook, A. E. (2016a). Coherence threshold and the continuity of processing: The RI-Val model of comprehension. *Discourse Processes, 53*(5–6), 326–338. https://doi.org/10.1080/0163853X.2015.1123341

O'Brien, E. J., & Cook, A. E. (2016b). Separating the activation, integration, and validation components of reading. *Psychology of Learning and Motivation* (Vol. 65, pp. 249–276). New York: Academic Press. https://doi.org/10.1016/bs.plm.2016.03.004

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. 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. https://doi.org/10.1037/0278-7393.24.5.1200

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

Piest, B. A., Isberner, M. B., & Richter, T. (2018). Don't believe everything you hear: Routine validation of audiovisual information in children and adults. *Memory & Cognition, 46*(6), 849-863. https://doi.org/10.3758/s13421-018-0807-7

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology, 34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x.

Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition, 36*(3)**,** 688–701. https://doi.org/10.3758/MC.36.3.688

Rapp, D. N., & Mensink, M. C. (2011). Focusing effects from online and offline reading tasks. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 141–164). Charlotte, NC: Information Age Publishing.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Richter, T. (2015). Validation and comprehension of text information: Two sides of the same coin. *Discourse Processes, 52*(5–6), 337–352. https://doi.org/10.1080/0163853X.2015.1025665

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. https://doi.org/10.1037/a0014038

Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, *15*(1), 85–106. https://doi.org/10.1207/s1532690xci1501_3

Rouet, J.-F., Le Bigot, L., de Pereyra, G., & Britt, M. A. (2016). Whose story is this? Discrepancy triggers readers' attention to source information in short narratives. *Reading and Writing*, *29*(8), 1549–1570. https://doi.org/10.1007/s11145-016-9625-0

Saux, G., Ros, C., Britt, M. A., Stadtler, M., Burin, D. I., & Rouet, J.-F. (2018). Readers' selective recall of source features as a function of claim discrepancy and task demands. *Discourse Processes*, *55*(5–6), 525–544. https://doi.org/10.1080/0163853X.2018.1463722

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. https://doi.org/10.1016/j.jml.2008.05.001

Self, C. (2009). Credibility. In D. W. Stacks, & M. B. Salwen (Eds.), *An integrated approach to communication theory and research* (2nd ed., pp. 435–456). New York: Routledge. https://doi.org/10.4324/9780203710753

Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). *Attitude and attitude change: The social judgement-involvement approach*. Philadelphia, PA: Saunders.

Sherif, C. W., & Sherif, M. (Eds.). (1967). *Attitude, ego-involvement, and change*. New York, NY: Wiley.

Singer, M. (2013). Validation in reading comprehension. *Current Directions in Psychological Science*, *22*(5), 361–366. https://doi.org/10.1177/0963721413495236

Singer, M. (2019). Challenges in validation and comprehension. *Discourse Processes*, *56*(5–6), 465–483. https://doi.org/10.1080/0163853X.2019.1598167

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 230–247. https://doi.org/10.1037/a0021331

van den Broek, P., Bohn-Gettler, C.M., 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. Schraw (Eds.), *Text Relevance and Learning from Text* (pp. 123-140). Greenwich, CT: Information Age Publishing.

van den Broek, P., Risden, K., Husbye-Hartmann, E., (1995). The role of readers' standards of coherence in the generation of inferences during reading. In R.F. Lorch, Jr., & E.J. O'Brien (Eds.), *Sources of coherence in text comprehension* (pp. 353-373). Hillsdale, NJ: Erlbaum.

van Moort, M. L., Koornneef, A., & van den Broek, P. W. (2018). Validation: Knowledge- and text-based monitoring during reading. *Discourse Processes, 55*(5–6), 480–496. https://doi.org/10.1080/0163853X.2018.1426319

van Moort, M. L., Koornneef, A., & van den Broek, P. W. (2020). Differentiating text-based and knowledge-based validation processes during reading: Evidence from eye movements. *Discourse Processes,* 1–20. https://doi.org/10.1080/0163853X.2020.1727683

van Moort, M. L., Jolles, D. D., Koornneef, A., & van den Broek, P. (2020). What you read versus what you know: Neural correlates of accessing context information and background knowledge in constructing a mental representation during reading. *Journal of Experimental Psychology: General*, *149*(11), 2084-2101. https://doi.org/10.1037/xge0000764

Walsh, E. K., Cook, A. E., & O'Brien, E. J. (2018). Processing real-world violations embedded within a fantasy-world narrative. *Quarterly Journal of Experimental Psychology, 71*(11), 2282–2294. https://doi.org/10.1177/1747021817740836

Wertgen, A. G., & Richter, T (2020). Source credibility modulates the validation of implausible information. *Memory & Cognition, 48*(8), 1359–1375. https://doi.org/10.3758/s13421-020-01067-9

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020-2045. https://doi.org/10.1037/xge0000014

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106. https://doi.org/10.3102/0002831209333183

Williams, C. R., Cook, A. E., & O'Brien, E. J. (2018). Validating semantic illusions: Competition between context and general world knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1414-1429. https://doi.org/10.1037/xlm0000526

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87. https://doi.org/10.1037/0022-0663.83.1.73

Appendix

Table A1

Means and Standard Deviations of the Dependent Variables by Experimental Condition

Plausibility	Source Credibility	Target Sentence (ms)	Spillover Sentence (ms)	Plausibility Ratings
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
Plausible	High Expertise	3798 (1379)	2594 (928)	5.75 (1.65)
	Low Expertise	3738 (1323)	2576 (939)	5.40 (1.61)
Somewhat Implausible	High Expertise	3995 (1616)	2570 (891)	3.70 (2.08)
	Low Expertise	4050 (1560)	2704 (974)	3.54 (1.94)
Highly Implausible	High Expertise	3983 (1473)	2777 (1030)	1.96 (1.61)
	Low Expertise	4150 (1556)	2606 (946)	2.29 (1.71)

Note. Plausibility ratings range from 1 (= “not plausible at all”) to 7 (= “very plausible”).

Table 1

Sample Experimental Story

Introduction:
Today was Aaron's big day: he was a candidate on the TV show 'Who wants to be a millionaire?'. It was his first time on live television.
Expertise
Low expertise:
Aaron did not have a lot of general knowledge and he was only on the show because his friends had applied in his name and against his will.

High expertise:
Aaron was very knowledgeable in various domains; as such, he liked to watch as much quiz shows as possible.
Continuation:
All his friends were in the audience. The show host was about to read the possible answers to his question as Aaron interrupted him.
Assertion
Plausible assertion: 'I know the answer without having to choose from the possible answers, watt is the measurement of electric power,' Aaron said confident of victory.
Somewhat implausible assertion: 'I know the answer without having to choose from the possible answers: ampere is the measurement of electric power,' Aaron said confident of victory.
Highly implausible assertion: 'I know the answer without having to choose from the possible answers: kilogram is the measurement of electric power,' Aaron said confident of victory.
Spillover: Before the right answer was revealed, the TV station decided, it was time for a commercial break.
Ending: Aaron couldn't stand the tension.

Table 2

Estimated Coefficients, Standard Errors, Degrees of Freedom, and t Values for the Linear Mixed Model of the Reading Times of the Target Sentence.

	Est.	*SE*	*df*	*t*	*p*
(Intercept)	4001.89	103.49	106.37	38.66	< .001
Length of Sentence	726.44	47.59	88.13	15.27	< .001
Position	-238.40	20.19	2673.15	-11.81	< .001
Source Credibility	-30.01	19.60	2648.27	-1.53	.126
Plausibility Contrast 1	-171.32	28.05	2667.80	-6.11	< .001
Plausibility Contrast 2	-18.66	27.80	2672.73	-0.67	.502
Source Credibility x Contrast 1	24.77	28.00	2652.21	0.89	.376
Source Credibility x Contrast 2	0.00	27.66	2649.42	0.00	1.00

Note. Source Credibility (contrast coded: high expertise = 1, low expertise = -1). Plausibility (effect coded: contrast 1: plausible = 1, somewhat implausible = 0, highly implausible = -1; contrast 2: plausible = 1, somewhat implausible = -1, highly implausible = 0).

Table 3

Estimated Coefficients, Standard Errors, Degrees of Freedom, and t Values for the Linear Mixed Model of the Reading Times of the Spillover Sentence.

	Est.	*SE*	*df*	*t*	*p*
(Intercept)	2661.62	66.40	111.59	40.09	< .001
Length of Sentence	476.51	38.85	36.00	12.26	< .001
Position	-139.32	12.03	2691.93	-11.58	< .001
Source Credibility	3.46	11.72	2673.15	0.30	.768
Plausibility Contrast 1	-53.36	16.60	2673.87	-3.22	.001
Plausibility Contrast 2	4.41	16.55	2676.52	0.27	.80
Source Credibility x Contrast 1	-44.48	16.64	2674.56	-2.67	.008
Source Credibility x Contrast 2	46.79	16.57	2673.61	2.82	.005

Note. Source Credibility (contrast coded: high expertise = 1, low expertise = -1). Plausibility (effect coded: contrast 1: plausible = 1, somewhat implausible = 0, highly implausible = -1; contrast 2: plausible = 1, somewhat implausible = -1, highly implausible = 0).

Table 4

Estimated Coefficients, Standard Errors, Degrees of Freedom, and t Values for the Linear Mixed Model of the Plausibility Ratings.

	Est.	*SE*	*df*	*t*	*p*
(Intercept)	3.77	0.07	62.51	54.23	< .001
Source Credibility	0.03	0.03	3413.92	0.94	.350
Plausibility Contrast 1	1.64	0.04	3414.54	39.70	< .001
Plausibility Contrast 2	0.16	0.04	3411.93	3.83	< .001
Source Credibility x Contrast 1	0.20	0.04	3415.51	4.82	< .001
Source Credibility x Contrast 2	-0.05	0.05	3414.26	-1.21	.228

Note. Source Credibility (contrast coded: high expertise = 1, low expertise = -1). Plausibility (effect coded: contrast 1: plausible = 1, somewhat implausible = 0, highly implausible = -1; contrast 2: plausible = 1, somewhat implausible = -1, highly implausible = 0).

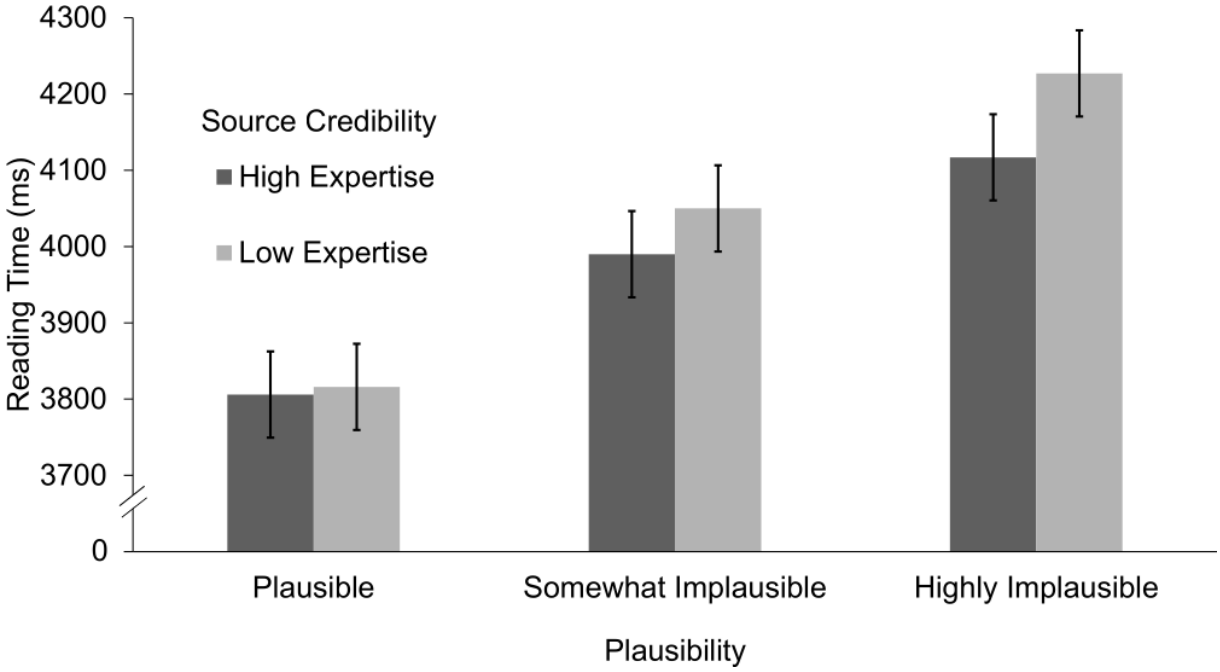

Figure 1. Mean reading times (with standard errors) on target sentences by experimental condition.

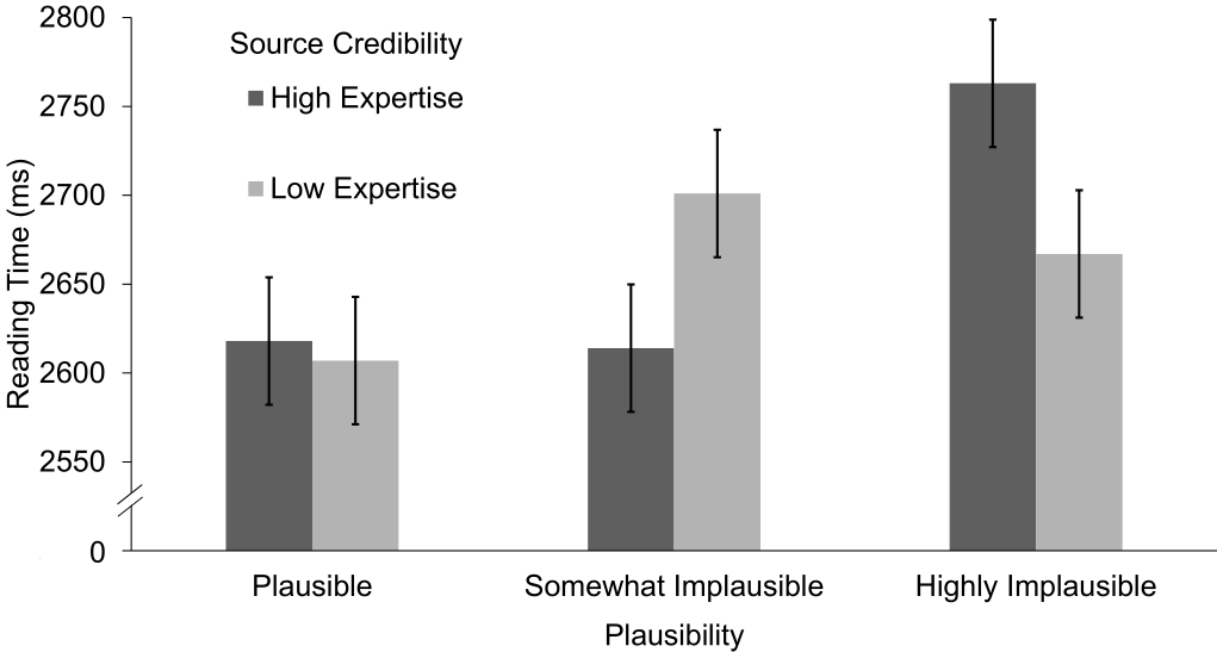

Figure 2. Mean reading times (with standard errors) on spillover sentences by experimental condition.

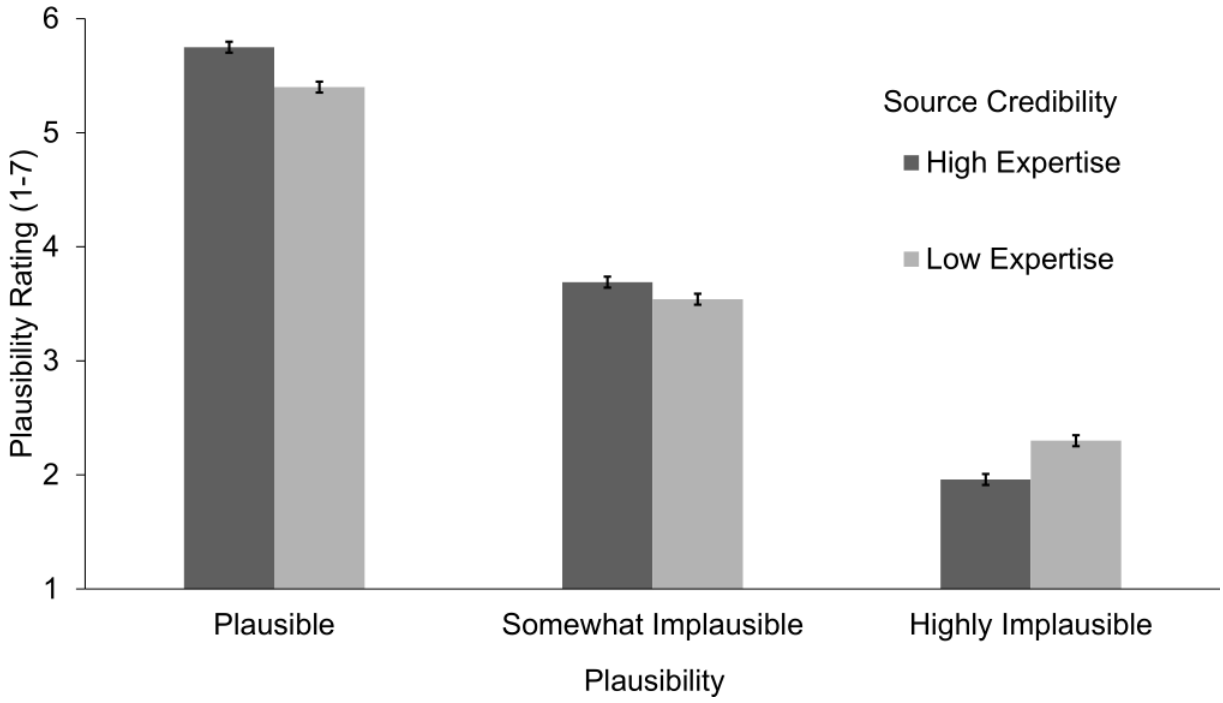

Figure 3. Mean plausibility ratings (with standard errors) of the target sentence by experimental condition.