Acta Psychologica 212 (2021) 103191

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

From simple agents to information sources: Readers' differential processing of story characters as a function of story consistency

Gaston Saux [a,*], Nicolas Vibert [b], Julien Dampuré [c], Debora I. Burin [d], M. Anne Britt [e], Jean-François Rouet [b]

[a] Pontifical Catholic University of Argentina — National Scientific and Technical Research Council (CONICET), Av. Alicia Moreau de Justo 1500, Edif. San José, 2do piso (1107), Buenos Aires, Argentina

[b] Centre de Recherches sur la Cognition et l'Apprentissage, CNRS, Université de Poitiers, Université de Tours, MSHS - Bâtiment A5, 5, rue Théodore Lefebvre, TSA 21103, 86073 Poitiers cedex 9, France

[c] University of La Laguna - University of La Sabana, Facultad de Psicología, Campus del Puente Común, Km. 7 Autopista Norte de Bogotá, Chía, Cundinamarca, Colombia

[d] University of Buenos Aires – National Scientific and Technical Research Council (CONICET), Lavalle 2353 (1052), Buenos Aires, Argentina

[e] Northern Illinois University, office 363, 100 Normal Rd. DeKalb, IL 60115, USA

ARTICLE INFO

Keywords:
Memory
Information source
Text comprehension
Eye-tracking

ABSTRACT

The study examined how readers integrate information from and about multiple information sources into a memory representation. In two experiments, college students read brief news reports containing two critical statements, each attributed to a source character. In half of the texts, the statements were consistent with each other, in the other half they were discrepant. Each story also featured a non-source character (who made no statement). The hypothesis was that discrepant statements, as compared to consistent statements, would promote distinct attention and memory only for the source characters. Experiment 1 used short interviews to assess participants' ability to recognize the source of one of the statements after reading. Experiment 2 used eye-tracking to collect data during reading and during a source-content recognition task after reading. As predicted, discrepancies only enhanced memory of, and attention to source-related segments of the texts. Discrepancies also enhanced the link between the two source characters in memory as opposed to the non-source character, as indicated by the participants' justifications (Experiment 1) and their visual inspection of the recognition items (Experiment 2). The results are interpreted within current theories of text comprehension and document literacy.

1. Introduction

Accessing written information has never been easier than in the twenty-first century. With the problem of accessibility fading away, the new literacy challenges for digital readers include how to select a piece of information, determine its pertinence to one's goals, evaluate its quality, and integrate it into one's knowledge representations (e.g., Bråten et al., 2018). Within this context, there has been a growing interest in understanding lay readers' heuristic ability to "source" (Wineburg, 1991) as part of their comprehension strategies. Although not new in essence, sourcing skills have become critical to understanding how readers construct coherence and validate information in complex but frequent digital reading scenarios, such as when having to make sense out of multiple perspectives or having to find reliable advice from information that comes in a variety of media and forms (List et al., 2020). They have also become fundamental to our understanding of how readers use (e.g., share, question, incorporate into an argument) the information they can access (e.g., Goldman, 2004). The construct of a source is broad and multifaceted and can be defined in multiple ways (Rouet & Britt, 2014). The term source is sometimes used to refer broadly to a document, that is, an informational corpus clearly delimitated from other corpora. A source can also be defined more specifically as any information, embedded or provided outside the document, about the identity, purpose or context of the people or institutions who produced or made the information available (Braasch et al., 2018). In this paper we use the term "source" in the latter way. For instance, a story character issuing some information as part of the story is the source of that information. We refer to sourcing as paying attention to, encoding,

* Corresponding author at: CIPP-UCA, Edif. San José, Av. Alicia Moreau de Justo 1500, 2do piso (1107), Buenos Aires, Argentina.
E-mail address: gaston_saux@uca.edu.ar (G. Saux).

https://doi.org/10.1016/j.actpsy.2020.103191
Received 21 January 2020; Received in revised form 29 September 2020; Accepted 3 October 2020
Available online 2 November 2020

and using information about information sources, i.e., who says what. In this restricted sense, sourcing is a way for readers to coherently integrate multiple discourse contents by indexing them as multiple viewpoints and, eventually, to discard inaccurate or misleading information by weighing and evaluating it on the basis of who endorses it (Pilditch et al., 2020; Stadtler & Bromme, 2014).

The present study examines the cognitive processes whereby readers incorporate information from and about multiple sources into an integrated memory representation and how they later retrieve this information. We start with a brief review of current theories of source comprehension, and we then focus on readers' representation of story characters as sources. In the second part of the article, we present two empirical studies in which we investigated readers' integration of source information through verbal protocols (Experiment 1) and eye-tracking data (Experiment 2).

1.1. Representing information from multiple sources and the D-ISC hypothesis

The Documents Model Framework (Britt et al., 1999; Britt & Rouet, 2012; Perfetti et al., 1999) describes the possible mental representations and processes underlying the integrated comprehension of multiple texts. Consistent with general models of text comprehension (e.g., Kintsch, 1998), the Documents Model Framework explains that readers manage to integrate multiple texts by producing a distinct source-based layer within their representation of discourse, resulting in a "documents" model (Britt et al., 1999; see Fig. 1). Constructing a documents model would be particularly helpful when trying to understand complex issues involving multiple viewpoints, such as historical events (Wineburg, 1991), technology and health issues (Kammerer et al., 2016; Strømsø & Bråten, 2014), daily life recommendations (Salmerón et al., 2016) or scholarly tasks (e.g., Pérez et al., 2018).

A documents model includes a model of one or more situations (to the right in Fig. 1), which represent the state of affairs depicted in the text/s (van Dijk & Kintsch, 1983), and an intertext model (to the left and center of Fig. 1), which represents features of the sources of the contents. The intertext model also includes links that connect each source with their contents and among themselves. The Source-to-Content links integrate the semantic representation of the situation(s) with knowledge about the sources (e.g., A said X, B said Y). The Source-to-Source links integrate the sources together by means of rhetorical relations (e.g., A disagrees with B), trustworthiness evaluations (e.g., A is more knowledgeable than B; A has commercial interests as compared to B), and other types of associations (Britt et al., 1999; Perfetti et al., 1999).

Identifying the array of conditions that affect readers' construction of a documents model, and particularly their representation of the sources has been the focus of an increasing amount of research. It has been found that lay readers' encoding of "who said what" when reading multiple accounts of a single event is far from systematic (Bråten et al., 2016; Britt & Aglinskas, 2002; Bromme et al., 2015; de Pereyra et al., 2014; Saux et al., 2018; Steffens et al., 2014; Strømsø & Bråten, 2014). Therefore, researchers have invested considerable effort to identify the variables related to the reader, the text, and the situation that would increase or restrain the likelihood of constructing a documents model (e.g., Bråten et al., 2018).

One such variable is the presence of situational discrepancies across the statements made in a single text, that is, coherence breaks between two or more pieces of text information. Braasch et al. (2012) presented to readers fictitious flash news reports on various topics. Each report included two characters that functioned as sources because they were making specific statements about the same situation or topic. The compatibility of the statements was manipulated so that participants read some reports in a discrepant version (i.e., the two sources disagreed), and others in a consistent version (i.e., the two sources agreed). This manipulation impacted several online and offline measures: Participants gazed more and for longer times at sources during reading, made more references to sources in summaries, and showed better memory for who said what when confronted with discrepant rather than agreeing assertions. To explain these results, Braasch et al. proposed the Discrepancy-Induced Source Comprehension (D-ISC) hypothesis. According to the D-ISC hypothesis, semantic discrepancies trigger reader's efforts to maximize coherence: When source information is available, then tagging or linking incompatible statements to different sources (i.e., source-to-content links) is an effective strategy to restore coherence. These efforts will then translate into enhanced memory and integration of the embedded sources and their assertions, particularly when compared with consistent statements. The D-ISC effect has been replicated and extended in several other studies (see Braasch & Bråten, 2017, for an overview).

The D-ISC hypothesis is compatible with existing models related to sourcing, such as the Source Monitoring Framework (SMF, Johnson et al., 1993). The SMF examines the external cues (e.g., perceptual, spatial factors) and internal cognitive operations that allow one to distinguish between internal and external sources of experiences. The meaning of source in this framework is "the variety of characteristics that, collectively, specify the conditions under which a memory is acquired (e.g., the spatial, temporal, and social context of the event; the media and modalities through which it was perceived)" (p. 3). One of

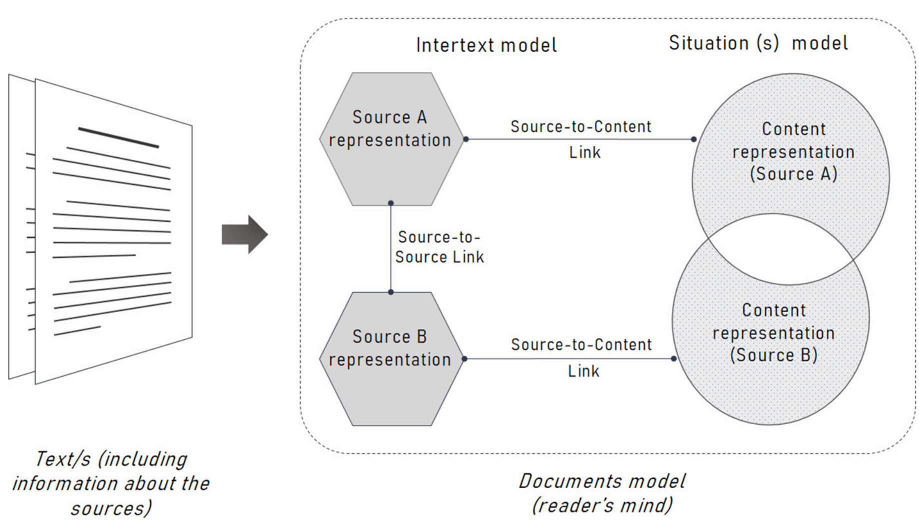

Fig. 1. Theoretical depiction of a source-content integrated mental representation, according to the Documents Model Framework (adapted from Britt & Rouet, 2012).

the conditions that will enhance memory for these characteristics at retrieval is the depth of encoding or the degree of elaboration performed on content information (Jurica & Shimamura, 1999; Long & Spooner, 2010). This aligns well with the D-ISC predictions: if discrepancies promote the integration of sources and contents during encoding, this should lead to better performance when attributing 'who said what' at retrieval, particularly when comparing it with consistent contents (Braasch & Bråten, 2017). In the SMF however, source information is typically not part of the message. It is rather experienced as part of the learning episode. In contrast, the Documents Model framework, from which the D-ISC hypothesis stems, aims at examining discourse comprehension. Thus, unlike the SMF, the Documents Model framework proposes that source nodes are conveyed through symbolic information (e.g., the name of the person issuing a statement in a story) that can be incorporated by specific mechanisms (source-to-content and source-to-source links in addition to the situation model/s) to create a documents model.

1.2. Interpreting the D-ISC effect: characters as sources vs. story agents

Given the low level of spontaneous sourcing reported in the literature, the D-ISC hypothesis has important implications from both theoretical and applied perspectives. To examine these assumptions, several controlled studies went to use short texts within repeated measures' designs (e.g., Braasch et al., 2016; de Pereyra et al., 2014; Rouet et al., 2016; Salmerón et al., 2016; Saux et al., 2017).

However, these studies involved texts that were sometimes only two sentences long (e.g., Braasch et al., 2012). This makes it difficult to determine if discrepancies really promoted the encoding of sources over other components of the mental representation, or simply made the overall information more salient. Furthermore, Braasch et al.'s (2012) study and others along the same line (e.g., de Pereyra et al., 2014; Rouet et al., 2016; Saux et al., 2017, 2018) used a particular type of source, consisting of embedded characters that produce one or more statements. Thus, the question arises whether the discrepancies enhance the representation of the characters as agents taking part in the situation, or, as predicted by the D-ISC hypothesis, as distinct sources of different statements. The latter option would entail that source characters are encoded differently in readers' mind than other characters from the text because they give information about the situation instead of just being involved in it.

Distinctly representing characters as sources as a way of handling text contradictions would thus require the construction of a documents model, (i.e., a linked representation of what is being said with who is saying it). Failing to do so would lead to a "mush" representation of the situation/s (Britt et al., 1999), with little or no indication of where the multiple pieces of information came from. When different pieces of information contradict each other, a mush model is bound to be incoherent, calling for alternate resolution strategies such as distorting or ignoring part of the information (Rouet et al., 2016).

1.3. Representing text characters as sources

The conception of discourse as a social artifact (Britt et al., 2013; Wineburg, 1991) implies that a reader can represent the agents associated with a text in different ways. When information sources are disconnected from the situation, as when a historian writes an essay about past events, the representations of sources and of the characters involved in the story are clearly separate, since they do not share features other than the author-content link. However, when information sources are embedded in the text, as when a story character issues a piece of information, one dimension of the representation (i.e., the agent as a character of the story) may overlap and become confused with the other one (i.e., the agent as an information source). This is a practical problem for some research on sourcing, which has so far assumed that characters embedded in the story can be considered as sources as long as they make one or more assertions. Research on situation model construction has provided evidence that readers can build the situation model using person-based criteria (i.e., using the protagonists of the story as main organizing concepts, see for instance Radvansky et al., 1997), and that salient characters in the story can be used to index and update the mental representation of narratives (e.g., Rapp & Kendeou, 2007; Zwaan & Radvansky, 1998). However, research on the representational status of source agents who provide information about the situation model is rather scarce.

Some prior works do indeed suggest that text characters can be distinctively represented as sources if the reader indexes them as "information providers". Graesser et al. (1999) assessed the salience in memory of characters involved in narratives, and found that a first-person narrator (who was also a character) was more salient than the other, non-narrator characters. Graesser et al. concluded that these two levels of representation can be combined or amalgamated in memory, so that the agents who are represented according to both dimensions (e.g., an agent perceived both as a character involved in the story and as a source by the reader) should be more salient than the agents represented according to one dimension only (e.g., an agent perceived only as one of the characters of the story).

Research on source memory supports this view. de Pereyra et al. (2014) reported that readers remembered text-embedded sources involved in the situation (such as a witness) better than embedded remote sources, such as someone commenting on the topic from a distance. They suggested, as Graesser et al. (1999) did, that the involvement of the sources within the story could generate a more concrete representation. Research on expository texts by means of think-aloud protocols has also provided evidence that readers pay attention to text-embedded sources, such as experts endorsing a research report, especially if the same embedded source is mentioned in more than one document, and if the documents present contradicting information (Strømsø et al., 2013; Strømsø & Bråten, 2014).

Still, the existing evidence is inconclusive regarding whether characters of a story are distinctly represented as information sources and, therefore, as something other than mere elements of a situation. Our assumption is that the representation of a text-embedded character as a source will become evident when readers are presented with contradicting statements because, following the D-ISC hypothesis (Braasch et al., 2012), this situation prompts the elaboration of who says what as a mean of coherently representing an otherwise incompatible representation of a situation.

1.4. The present study

Two experiments were conducted to test whether the previously-reported discrepancy effect stems from readers' representation of the sources of information per se with specific links to the contents of what is said (i.e., an "intertext model"; Perfetti et al., 1999). More precisely, we looked for evidence that readers of discrepant stories remember characters issuing information (i.e., source characters) in a distinct way compared to nonsource characters. We manipulated the discrepancy/consistency between the statements with the expectation that source characters should be more attended to, represented more accurately, and better retained in memory when the statements are discrepant rather than consistent. Conversely, the consistency manipulation should not affect attention or memory of the other, nonsource characters. The dependent variables combined online and offline measures. In Experiment 1 we used short interviews during a recognition task after reading, in which the participants had to recognize the source of one of the text's statements and explain their selection. The aim was to test whether discrepancy would trigger specific recollection processes when characters are presented as information sources. In Experiment 2, we used eye-tracking to record participants' fixations during reading, as in Braasch et al. (2012). Importantly, we also collected eye-tracking data during a source-content recognition task presented after reading and similar to the one used in Experiment 1. The aim was to provide new insight on the relatedness and availability of the information at the time of retrieval during the recognition task.

2. Experiment 1

The aim of the first experiment was to elicit the memory retrieval processes at work when readers try to remember the source of a piece of information (who said it) after reading a text passage. To that aim, we asked participants to read short stories and to later recall them. The stories included source characters (which according to the Documents Model Framework would be part of both the situation model and the intertext model; Britt et al., 1999), and nonsource characters (which are only part of the situation model). This enabled us to assess whether readers encoded source information and the links between sources and content (i.e., their intertext model). In addition, we asked participants to recognize the source of a statement from the story and to explain their selection. We expected that the presence of a discrepancy would result in more accurate source recognition (Braasch et al., 2012) and more discrimination between source and nonsource characters (Britt et al., 1999). Specifically, the discrepancy manipulation would influence the distribution of the two type of responses involving source characters (correct responses and other-source errors), but not of the responses involving a nonsource character (nonsource errors). We also expected that source recognition from discrepant stories would be based on retrieval as opposed to familiarity or plausibility. Conversely, we expected recognition to be based on familiarity when the stories did not include any contradiction.

2.1. Method

2.1.1. Participants

Forty-two 2nd-year Psychology students from a large Argentinian university participated for course credit. Two participants had to be excluded due to failure of the response recording material, resulting in a final sample of 40 (Age $M = 20.0$, $SD = 1.5$). All participants were native Spanish speakers and declared no reading impairment (glasses or lenses were used when necessary). Participants signed an informed consent form and were debriefed after completing the experiment.

2.1.2. Materials

Sixteen experimental, 8 filler fictitious, and two practice printed flash news reports were adapted from Braasch et al.'s (2012) original pool. Each text described events in a newspaper brief-report style on topics such as science, society, and economy (see Table 1 for an example). The original Braasch et al.'s texts were composed of a title and two sentences, with each sentence presenting an embedded source (e.g., "The policeman declared…"), namely a "source agent", making a statement about the situation (e.g., "The fire was due to a sabotage"). The adaptation consisted in adding three new sentences to each text so that each story was now composed of five sections:

1) A title;
2) an introductory sentence setting the situation;
3) a sentence introducing a third, nonsource character (i.e., someone involved in the situation but making no statement, called the "nonsource" hereafter);
4) the two sentences presenting the source agents and their utterances, which included an additive or contrasting connector depending on the experimental condition;
5) a final sentence, related to the main situation, but with no direct overlap with the sentences involving the story and source agents.

Table 1
Example of experimental text showing the Discrepancy manipulation (Experiments 1 and 2, translated into English).

#	Section	Example
1	Title	TRANSGENICS
2	Introduction	More than 300 Greenpeace activists protested yesterday against the tests of transgenic seeds that are being carried out near the village of San Antonio de Areco.
3	Nonsource + action	A farmer from the area put a green ribbon on his arm to show support for the protest.
4	Source A + statement	A correspondent of a news agency announced that [an/no] extreme protest action is foreseen.
	Source B + statement	[Likewise/conversely], the owner of an organic restaurant affirmed that the activists plan to destroy the implanted plots.
5	Final sentence	The village of San Antonio de Areco is located in the middle of the main agricultural region of the country.

Note. Discrepancy manipulation in brackets.

The sentence involving the nonsource was always placed after the introductory sentence and before the sentences involving the source agents. Because we compared the same text in consistent and discrepant versions (see Design section below), presentation order could not account for observed differences between conditions. The verbs associated with the nonsources could involve spatial displacement (e.g., arriving to the critical location), interaction with an object (e.g., putting on a piece of clothes), perceptual activities (e.g., watching a game), or passive events (e.g., being rescued from an accident), but never the production of an assertion or statement. All verbs associated with the source agents explicitly indicated a speech act (e.g., said, affirmed, declared, etc.). Only high-assertiveness verbs were used, based on the scale provided by Guerry et al. (1993). Filler texts were similar in length and style to the experimental stories, but differed in that they did not present a "two sources-two statements" structure.

For the recognition task, a separate booklet with 16 multiple-choice recognition trials (one per experimental story) was created. Each recognition trial was presented on a separate sheet and was composed of one of the statements from the text and four options to choose from for the character who made that statement. The four options were the three characters involved in the story (the two source and the nonsource characters) plus a control, distracter item, namely a potential character attuned to the general atmosphere of the story but not mentioned in the materials (e.g., "a representative of an agricultural union" in the story example in Table 1).

2.1.3. Design

A repeated measures design with one fixed factor (discrepancy) and two levels (consistent/discrepant) was used. The manipulation involved changing one word from the statement of source A and the subsequent connector to make it either consistent or discrepant with source B's statement (see Table 1, section 4). The texts were designed so that the three characters in the story could swap roles and be either the nonsource, the source agent A or the source agent B (only character names were rotated, the text structure stayed the same). This was done to control for potential influences of characters' typicality on memory accuracy. Three versions of each experimental story were then created by manipulating the roles taken by each character (the different story versions were pre-tested with small groups of students to ensure that they were perceived as plausible and coherent). In combination with the two modalities of Discrepancy (consistent/discrepant), six versions of each story were prepared. Equal groups of participants were randomly assigned to each version.

The dependent variables were collected during the recognition phase. Firstly, we recorded the final response to each trial, i.e., the character selected by the participant, which could be the correct response (the right source for the statement), the "other-source" (i.e., the incorrect source, or the source of the other statement in the story), the "nonsource" (i.e., the character from the story that did not say anything), or a "control character" (i.e., a character who was plausible but was not mentioned in the story). Secondly, we analyzed the participants' justifications of their responses in short, semi-structured interviews. This enabled us to investigate their perception of why they were choosing a certain character. We chose this task instead of other tasks that have shown to facilitate source-content integration (e.g., argumentative writing tasks) because our interest was specifically on source-content links' availability in memory.

2.1.4. Procedure

All the activities took place in individual sessions of approximately 50 min. After signing an informed consent form, participants completed the reading phase. To increase engagement in the reading task, an overarching scenario was provided to the participants (e.g., Sabatini et al., 2014). They were instructed to imagine that they had been hired by a web-based news agency and that the editor-in-chief had given them a series of brief reports, currently under consideration for publication. Participants were then told that their first task was to read these reports carefully to comprehend them. They were also told that later they would have to do other tasks that would require recalling the information they were about to read. To promote elaborative encoding, participants were asked to think of and write down an alternative title for each story after reading each text (Saux et al., 2017). Then, they were presented with a printed booklet containing one story per page, a pen, and an additional sheet of paper to write down the alternative titles. The two practice texts were used to exemplify the title production task and to familiarize participants with the style and length of the stories.

Once this was done, the reading booklet and title sheet were removed and participants completed a distracter task for approximately 15 min. This task consisted of filling in basic socio-demographic information and taking the local adaptation of the Letter–Number Sequencing subtest of WAIS III (Wechsler, 2002). Then, they were given the recognition booklet. Each recognition trial was presented on a separate page and showed the original title of the story, the statement whose source had to be recognized, and the four characters to choose from, listed randomly. The instructions for the recognition task were as follows (English translation):

> "Some of the reports you read depicted characters making assertions on a certain situation. One at a time, you will be presented now with a series of assertions extracted from the texts and four characters to choose from. Your task is to recognize the character who made the assertion. We are interested in your final responses, but also in how you justify them. Therefore, we ask you to explain your response. So, for each story we will ask you to tell us: 1) what you remember from that story; 2) which character you select and why. Your answers do not have to be long, as long as they are informative. We may ask you some additional questions to clarify your responses, when needed".

Responses were recorded with a digital voice recorder application for Android phone, with participants' consent. After completing a practice trial (corresponding to one of the practice texts) participants moved to the recognition trials. They read the title, the cued statement and the response options and gave their responses orally. Once the recognition phase was completed, participants were thanked and debriefed.

2.1.5. Scoring of the retrospective interviews

The interviews were transcribed and analyzed by the first author and a research assistant based on an adaptation of McCrudden et al.'s (2016) method. Participants' justifications of their responses were concerned mostly with what they remembered from the story and how they came to recognize the selected character. Participants' justifications were coded in terms of presence or absence of three categories: retrieval, familiarity, and plausibility. Appendix A provides a summary and examples of each of these categories. The "retrieval" category coded whether participants expressed remembering the chosen character from that story. This included expressing certainty or explicitly referring to thoughts elicited during the reading phase. The "familiarity" category coded whether participants expressed being hesitant about the selected character, such as expressing that the character "sounded familiar". The "plausibility" category coded whether participants expressed using prior knowledge to help determine the likelihood of a character being present in a specific story setting. These justification categories were meant to reflect different levels of the initial encoding of an intertext model and its integration with the rest of the contents. Retrieval responses indicate a distinct representation of the source-content link. Familiarity responses indicate a less differentiated representation, but specific enough to locate the chosen character as part of the story, and plausibility responses indicate a poor overall recall and the need to access world knowledge to infer the likelihood of a character being present in the story. We expected a greater number of retrieval responses in the discrepant condition.

The frequency of each category was calculated by assigning 1 point when present and 0 when absent within each justification (see Table A.1 in the Appendix for examples and a note on scoring criteria).

The first author and the research assistant independently assessed ten randomly chosen transcripts (160 observations, representing 25% of the sample). The level of acceptance for the interrater agreement (*kappa*) was set to 0.7, corresponding to moderate to strong agreement (McHugh, 2012). Disagreements were resolved by discussion. The rest of the interviews were scored by the research assistant.

2.1.6. Statistical analyses

The analyses were run on IBM SPSS software (version 24, IBM Corp., 2016). We applied generalized mixed models (GLMM) set to logistic binary distributions to the dependent variables, i.e., accuracy and justifications. Participants gave an answer in 98% of the trials. In the remaining 2% of the observations (13 out of the 640 trials), participants reported that they preferred not to give an answer because of faulty memory. These trials were distributed across 10 participants and 8 different texts and were not further considered in the analyses.

In all cases, Discrepancy (consistent, discrepant) was used as a fixed factor. The analyses of justifications also included the Response type and its interaction with Discrepancy as additional fixed factors. To simplify the response type factor, control character errors were excluded due to their low frequency (2.9% of the total). Response type was therefore treated as a three-level factor (correct response, other-source source error, nonsource error). All models included random intercepts to account for variability across participants and texts. Odds Ratios (OR) were reported as indices of the strength of the predictor-outcome association.

2.2. Results

2.2.1. Recognition accuracy

Table 2 presents the distribution of recognition accuracy as a function of Discrepancy. Averaging across the consistent and discrepant conditions, the participants correctly recognized 45.8% of the sources. Errors in which the other source or the nonsource was selected as final response were observed in 31.9% and 22.3% of the trials, respectively. An initial mixed logistic model comparing correct vs. incorrect responses revealed a significant effect of Discrepancy, $F(1, 607) = 6.20$, p

Table 2
Distribution of recognition responses as a function of story discrepancy (Experiment 1).

Story discrepancy	Correct response	Other source errors	Nonsource errors
Consistent	123 (40.7%)	110 (36.4%)	69 (22.9%)
Discrepant	156 (50.8%)	84 (27.4%)	67 (21.8%)
Total	279 (45.8%)	194 (31.9%)	136 (22.3%)

Note. Trials in which participants did not provide an answer or chose the control character (2% and 2.9% of the observations, respectively) are not included in the percentage calculation.

= .013. The odds of correctly recognizing the source of a statement were higher in the discrepant condition than in the consistent condition, $OR = 1.50$, $CI_{0.95} = 1.09, 2.07$. Interestingly, the distribution of error types differed: When reading consistent stories, readers chose both source characters (i.e., the correct source and the other source) in similar proportions (40.7 vs. 36.4%). When reading inconsistent stories, however, readers were more likely to choose the correct source character (50.8%) over the other source (27.4%), $F(1, 471) = 7.21$, $p = .008$, $OR = 1.66$, $CI_{0.95} = 1.15, 2.40$. In other words, readers were less likely to mix-up source characters when asked to recognize the source of a conflicting statement. However, story version did not predict the distribution of nonsource errors (i.e., the character that did not say anything): Either when contrasting it with correct responses, $F(1, 413) = 1.59$, $p = .21$, or contrasting it with "other source" errors, $F(1, 328) = 1.15$, $p = .29$, participants were equally likely to choose the nonsource character after reading consistent or discrepant stories (22.9% vs. 21.8%).

2.2.2. Justification analysis

Table 3 presents the distribution of the justification categories as a function of Discrepancy and Response type. Discrepancy showed a pattern aligned with the theoretical predictions (see Retrieval and Familiarity categories, Table 3), but did not reach statistical significance, $p > .25$. Response type predicted the distribution of two types of justifications: Retrieval, $F(2, 603) = 8.87$, $p < .001$, and Familiarity, $F(2, 603) = 11.33$, $p < .001$. The odds of finding "retrieval" justifications were higher among correct responses than nonsource errors (50.1 vs. 17.6%), $OR = 3.23$, $CI_{0.95} = 1.56, 6.69$. A similar tendency was found when contrasting correct responses with other source errors (32.51%), but it did not reach significance, $p = .15$. Conversely, the odds of finding "familiarity" justifications were higher among both error types than among correct responses (other source errors: $OR = 2.32$, $CI_{0.95} = 1.07, 5.04$; nonsource errors: $OR = 3.40$, $CI_{0.95} = 1.56, 7.45$). Response type did not predict the distribution of "plausibility" justifications, $F(2, 603) = 0.28$, $p = .75$. The interaction between Discrepancy and Response type was nonsignificant for all categories, $p > .71$.

2.3. Discussion

Experiment 1 aimed at extending prior research on source-content integration. By presenting source and nonsource characters in the same text, we examined whether participants would refer to or differentially justify their memory for these two types of characters during a later source recognition task, as a function also of the discrepancy between statements made in the story.

The first hypothesis focused on accuracy and predicted that discrepancies would increase precision in recognizing the source of the cued statement (Braasch et al., 2012) and would result in more discrimination between source and nonsource characters (Britt et al., 1999), as compared to the consistent condition. In line with prior research (see Braasch & Bråten, 2017, for an overview), participants recognized the sources more accurately for discrepant than for consistent stories, suggesting that reading two opposite statements promoted the source-content integration. Furthermore, readers mistakenly selected the other source character less frequently in the discrepant than in the consistent condition. In other terms, better recognition of the source in the discrepant condition was due to a better discrimination between the two source characters. This result extends the D-ISC effect by specifically differentiating memory enhancement associated with source characters as opposed to other type of characters. Whereas precision in differentiating the two sources informs on their degree of distinctiveness *within* the intertext model, the ability to distinguish sources from the nonsource character informs on the distinctiveness *between* the intertext model and the representation of text contents that are not related to the sources (Britt et al., 1999). The present results suggest different degrees in the construction of the intertext model and its integration with the rest of the contents, so that sometimes participants may have built a distinct representation of each source and their corresponding claims. In other occasions however, their memory for sources was less distinct, but specific enough to differentiate sources from the nonsource character. In yet other occasions, participants seem to construct a "mush" memory representation, characterized by a poor differentiation between sources and text contents (Britt et al., 1999). The discrepancy manipulation only influenced the first two scenarios, in which a certain degree of source representation is assumed. Although our predictions concentrate on the D-ISC hypothesis and the DMF, other theoretical proposals have also concluded that comprehension processes tend to be just "good enough" for (Ferreira et al., 2002) or fit the standards of coherence (van den Broek et al., 2011) of the task the reader has set to perform. The higher presence of underspecified source-link representations in the consistent condition could thus reflect readers' tendency not to construct a costly representation (i.e., a documents model) if unneeded.

The second hypothesis focused on the verbal justifications made during the recognition task and predicted that source recognition from discrepant stories would be based on retrieval as opposed to familiarity or plausibility. We expected recognition to be based on retrieval when the stories included a contradiction. Although the data distribution followed these tendency (see Table 3), the variable that predicted the type of justification was response type, rather than story discrepancy: Correct answers included more "retrieval" justifications (i.e., clear recollection), whereas both error types included more "familiarity" justifications. Please note, though, that correct responses were more frequent among discrepant stories (as observed on the analysis of accuracy, see hypothesis 1), but the effect size was somewhat small.

The staggered effects observed for "retrieval" and "familiarity" justifications as a function of the type of response are also worth noting. In all cases, correct answers and nonsource errors are positioned at the extremes, with other-source errors in the middle position. Just like accuracy patterns suggest degrees of documents model construction (see hypothesis 1), this staggered effect (related to hypothesis 2) also suggests that other-source errors may represent a particular response, in which participants constructed an incomplete source representation that allowed them to access some, but not all source-related information when performing the recognition task. This interpretation has, though, a limitation: Asking participants to justify their selection may have led them to develop specific strategies to address the recognition-plus-interview protocol, which do not directly reflect the underlying memory representation. This would be particularly relevant for errors, which in our interpretation may correspond to weak source representation or even lack of source traces.

In sum, accuracy was higher for discrepant sources. Additionally, discrepant statements specifically enhanced memory of the source-related segments of the texts. Finally, the results suggest that the construction of an integrated source representation is on a continuum, with

Table 3
Distribution of the justification categories as a function of story discrepancy and response type.

Justification category	Story discrepancy	Correct source character	Other source character	Nonsource character
Retrieval	Consistent	100 (44.4%)	83 (36.9%)	42 (18.7%)
	Discrepant	133 (56.1%)	65 (27.4%)	39 (16.5%)
Familiarity	Consistent	13 (22.8%)	19 (33.3%)	25 (43.9%)
	Discrepant	17 (29.3%)	19 (32.8%)	22 (37.9%)
Plausibility	Consistent	9 (56.3%)	5 (31.2%)	2 (12.5%)
	Discrepant	3 (60%)	0 (0%)	2 (40%)
No response/ other	Consistent	1 (25%)	3 (75%)	0 (0%)
	Discrepant	3 (42.9%)	0 (0%)	4 (57.1%)
Total		279 (45.8%)	194 (31.9%)	136 (22.3%)

Note. Trials in which participants did not provide or chose the control character (2% and 2.9% of the observations, respectively) are not included in the percentage calculation.

cases in which there is no evidence of sourcing (nonsource errors), cases in which some evidence of sourcing is observed (other-source errors), and cases in which participants have a detailed representation of where the information came from (correct responses, discrepant stories' condition). Overall, these results support the view that source-related information becomes more relevant when reading discrepant multiple viewpoints. More specifically, readers seem to "tag" selectively discrepant source characters as distinct from other text information, including characters that take part in the situation but make no assertions.

3. Experiment 2

The goal of the second experiment was to further investigate source recognition processes by providing eye-tracking data collected during reading, but also during the item selection process involved in the recognition task, once the texts were not available anymore. Experiment 2 addresses three limitations found in Experiment 1. First, the design of Experiment 1 did not include data collection during the reading phase; therefore, the patterns observed in the memory task cannot be directly attributed to the level of attention assigned to sources during reading, as predicted by the D-ISC hypothesis (Braasch et al., 2012). Second, the recognition–plus-interview protocol may have influenced the final response by promoting specific response tactics, particularly in the trials in which participants were hesitant. Third, analyzing recollection parameters a posteriori from short interviews may have prevented us from capturing some processes systematically.

Studying the processes underlying high-level comprehension should ideally rely on offline as well as moment-by-moment indicators (e.g., Ferreira & Yang, 2019). Online techniques have been so far used in sourcing research mainly to study the allocation of attention during reading (Braasch et al., 2012; de Pereyra et al., 2014; Kammerer et al., 2016). To the best of our knowledge, they have not yet been used to study how the source-tagged representation is accessed after reading. This may be achieved using a recognition task similar to the one used in Experiment 1, in which readers identify the target source among several items, while their eye movements are recorded during the visual inspection of the items.

An important body of research has examined decision making based on visual information by means of eye-tracking evidence (Cavanagh et al., 2014; Orquin & Loose, 2013). The aim of these studies has been to determine which information people attend to when making a decision. The assumption is that only information which is fixated may be used in the decision process. In eye-tracking studies, the number of fixations made on a particular item would reflect its ability to attract participants' attention compared to the other items in the display. In contrast, the total time spent gazing at an item, i.e., the product of the number of fixations made that item and the average fixation duration, would rather be an indication of the total amount of attention/processing devoted to this particular item. In decision making processes, the total time spent gazing at an item often predicts the reader's decision to select a visual item out of multiple alternatives (Cavanagh et al., 2014; Orquin & Loose, 2013). Some studies also indicated that the words or images that are semantically related to a target attract visual attention more than other items, even at early stages of the visual search (e.g., Dampuré et al., 2014; Huettig et al., 2011). Taking the above studies as a basis, we assumed that the way participants explore the target and distractors while deciding who said something (for instance, if they fixate more the source agents than the nonsources) will inform on the way the multiple statements, their respective sources and the other characters involved in the story are organized in memory.

Thus, Experiment 2 was aimed at complementing the data from Experiment 1 by shedding some light on the processes underlying the recognition phase and to integrate it with evidence from the reading phase. As in Experiment 1, the obtained evidence will be used to infer the validity of using characters embedded within a story to study sourcing.

In accordance with the D-ISC assumption (Braasch et al., 2012), the first hypothesis was that discrepancies would affect how much attention is allocated to sources during reading. When statements are discrepant, sourcing becomes a good strategy to deal with the contradiction. Therefore, the number and duration of fixations on the source agents would increase during reading when the statements are discrepant rather than consistent. This would replicate Braasch et al.'s results in more complex stories, in which a nonsource character is involved in the depicted situation.

Similar to Experiment 1, the second hypothesis was that discrepancies would also influence participants' accuracy during the source recognition task. As a result of the increased allocation of resources to discrepant sources during reading (hypothesis 1), discrepancies would increase recognition accuracy and decrease recognition time in comparison with consistent statements (as a consequence of the construction of source-to-content links). In addition, an indicator of the clarity of recollection was collected more systematically than in Experiment 1 by using an adapted version of the RK paradigm (Tulving, 1985). Participants were asked to indicate, after each trial, whether their decision was based on a clear recollection of who said it ("Remember"), or on a familiarity feeling without having a clear recall ("Know"). Experiment 2 also included "Guess" as response option to distinguish between knowing and guessing (Eldridge et al., 2002; Gardiner et al., 1997). We expected discrepant sources to be associated with more "Remember" responses.

Finally, the third hypothesis was that, in addition to accuracy, discrepancies would also increase the integration of multiple sources in memory, as suggested by the Documents Model Framework. Therefore, we expected that the two source characters would not only be represented more distinctively than simple story agents (hypothesis 2), but would also be more linked together in memory in the discrepant condition than in the consistent condition. In contrast with Experiment 1, in which a verbal justification protocol was used, Experiment 2 collected eye movements during the recognition task. The prediction was that the two discrepant sources should attract more visual attention than the nonsource during the inspection of the items. This is because discrepant statements would promote the elaboration of a specific link between the two opposing sources in which the nonsource would not be involved (i.e., the source-to-source link). Based on the results of Experiment 1, we also predicted that this effect would be evident in correct responses and other-source errors, but not in nonsource errors.

3.1. Method

3.1.1. Participants

Thirty-six undergraduate students from a large Spanish university participated voluntarily for course credit (75% female, age $M = 20$ years, $SD = 4$). All participants had Spanish as their mother tongue and declared no reading impairment (glasses or lenses were used when necessary). Participants signed an informed consent form and were debriefed after completing the experiment. None of them took part in Experiment 1.

3.1.2. Materials

The same texts as in Experiment 1 (16 experimental stories, 8 fillers and 2 practice texts) were used. Two more experimental stories were added to increase the number of repeated measures per condition, resulting in 18 critical texts. The new texts were also adapted to have the same length and structure as the other ones, including the introductory sentence to set the situation, the sentence introducing the nonsource, two sentences presenting one source agent and her/his statement each, and the final sentence. Because Experiment 2 was implemented in the same language, but in a different country than Experiment 1, the stories were slightly modified when needed and pre-tested to account for local nuances.

For the recognition task, 18 multiple-choice recognition sets were created. As in Experiment 1, each set corresponded to one experimental story and asked the participant to identify the author of a particular statement from that text among four options (the two source agents, the nonsource, and the control character). The recognition sets were designed in slide-format in order to be presented on a computer screen. The screen was split into four quadrants, and each response option was placed at the center of each quadrant. Finally, the set also included a Remember-Know-Guess judgment (RKG) of the response. Fig. 2 shows an example of a recognition set.

3.1.3. *Design*

The design mirrored that of Experiment 1: the discrepancy between the statements A and B of each story was manipulated so that they could be consistent or discrepant. Also, as in Experiment 1, three versions of each experimental story were created by swapping the characters' roles in each version. In combination with the two modalities of Discrepancy (consistent/discrepant), six versions of each story were prepared. Equal groups of participants were randomly assigned to each version.

During the reading phase, each text section (see example in Table 1) was assigned to a separate areas of interest (AOI) which could vary from a single word to whole sentences. However, statistical comparisons were not made between different areas, but between the same area in the consistent and discrepant conditions. The first AOI corresponded to the introductory sentence. In the second sentence that introduced the nonsource, two separate AOIs were defined: one for the nonsource character itself (e.g., a farmer from the region) and another one for her/his action (e.g., putting a green ribbon on the arm to support the protest). In the two sentences that introduced the source agents, five separate AOIs were defined: one for each one of the two sources, one for each one of their respective statements, and one for the connector. The last AOI corresponded to the whole final sentence. In addition, since the D-ISC hypothesis predicts increased attention to sources once the discrepancy is detected (i.e., after reading the second statement), a "cut point" was established as in Braasch et al. (2012) to separate the first reading from second readings of previously read text (i.e., reprocessing). The cut point was defined as the first regression out of the AOI of the final sentence and into any previously fixated AOI. All fixations preceding the cut point were considered first reading, whereas all fixations following the cut point were considered second reading. The main dependent variables associated with reading were the total the total number and the total duration of fixations made on each AOI during first and second reading passes.

During the recognition trials, each set of items was divided into four AOIs, one for each one of the four possible responses. Eye movement data was collected prior to the response to investigate processes underlying memory retrieval. RKG judgments were collected once the participants had responded. Recognition accuracy was also calculated as in Experiment 1.

3.1.4. *Apparatus*

Eye movements were recorded using an Eye-Link 1000 eyetracker (SR Research Ltd., Mississauga, Ontario, Canada), which displayed the stimuli on a 21-inch monitor using a screen resolution of 768 × 1024 pixels. The eyetracking system was interfaced with a Thermaltake PC, which controlled the operation of the eyetracker and recorded the time spent on each slide and all eye movement data via dedicated Eye-Link software (DataViewer v1.11.900). The eyetracker provided gaze positions at a sampling frequency of 1000 Hz and with a precision of 0.1° of visual angle. Fixations were defined as any period in which the participant's gaze rested for 40 ms or more within a 20-pixel (about 0.7°) diameter area. Texts were presented in triple-spaced format, and all letters were in lowercase (except the first letters of sentences and proper nouns), 17-point Calibri font (i.e., approximately 0.5 cm in height, with an interval of about 2 cm between the lines). Recognition trials presented the different possible responses in similar format and size. The four items were displayed in quadrangular fashion, at approximately 5 cm from the screen center. The viewing distance was approximately 100 cm for all participants. E-Prime 2.0 software (Psychological Software Tools, Pittsburgh, PA), was coupled to the eye tracking system and was used to run the experimental script and record participants' oral responses to the recognition trials via a Marantz Professional's Solid State Recorder PMD671 (Oade Brothers Audio).

3.1.5. *Procedure*

Participants were tested individually. The materials were presented on a computer screen. The same scenario-based instructions as in Experiment 1 were used, i.e., being hired by a news agency and being assigned for reading carefully several brief news reports. Participants were also warned that they would have to perform a memory task later, based on what they would read. Each text was presented on a single slide, preceded by a 500 ms fixation cross that oriented the gaze to the starting point. As a means of ensuring careful reading, participants were asked to produce orally a summary of each story. Once they finished reading a text, they pressed the space bar, the backlight of the screen changed into a light-blue color, and they produced their summary.

The summaries were recorded via the experimental software connected to the microphone, but their content was not analyzed for this study. The texts remained visible but eye movements were not collected during the production of the summaries. The two practice texts were presented before the experimental session to introduce the reading and summary-production procedure. Once this first phase was finished, the

Slide 1	Fix. point	Slide 2
Who said that the the activists plan to destroy the implanted plots?	*	a correspondent from a news agency *[incorrect source]* a representative of an agricultural union *[control item]* a farmer from the region *[story agent]* an owner of an organic restaurant *[correct source]*

Fig. 2. Example of a recognition trial in Experiment 2. The trial corresponds to the example of text presented in Table 2. Information between brackets was not included in the actual trial.

participants completed a computerized version of the operation-word span task (OSPAN, Redick et al., 2012). The OSPAN is a working memory capacity task that test requires participants to verify larger or smaller series of simple mathematical operations while trying to recall a list of unrelated letters (performance on the OSPAN task is unrelated to our research questions and hence, they are not analyzed here. The supplementary analyses of this variable do not change the main conclusions and are included in Appendix C).

Each recognition trial included three slides (see Fig. 2): The first slide asked who made the critical statement, which was centered on the screen. The second slide displayed the four characters to choose the answer from, equidistant from the screen center. The characters' positions in the quadrangle were counterbalanced across texts and participants. Once they had taken their decision, participants pressed the spacebar again and gave their response. This caused the multiple-choice slide to disappear. A final slide asked the participants to qualify their response with a Remember-Know-Guess judgment (RKG). Participants were told to choose "Remember" if they clearly remembered that the character had said the cued statement, "Know" if they felt some familiarity but did not have a clear recall, or "Guess" if they were choosing randomly. Before the presentation of the statement and characters, a 500 ms duration fixation cross oriented participants' gaze to the center of the screen. Responses to each trial and RKG judgments were recorded orally as in the reading phase. Before the experimental trials, participants performed two practice trials using the statements and characters of the two practice texts.

3.1.6. Statistical analyses

As in Experiment 1, data were analyzed with IBM SPSS software (version 24, IBM Corp., 2016) using generalized mixed models. Linear and logistic models were applied on the continuous (i.e., eye movements and recognition times) and the categorical (recognition accuracy and RKG judgments) dependent variables, respectively. In all cases, Discrepancy (consistent, discrepant) was specified as a fixed factor. The models accounting for recognition times and RKG judgments also included fixed factors for Response type (correct response, other-source error, nonsource error) and its interaction with Discrepancy. The models accounting for eye movement during the recognition task included in addition the Fixated AOI (i.e., correct source, other-source, nonsource, control item) and its interactions with Discrepancy and Response type, when applicable, as fixed factors. All models included random intercepts to account for variability across participants and texts. The estimated coefficients or the odds ratio (for continuous and categorical outcomes respectively), standard errors (*SE*), and confidence intervals (95%) were reported as an indication of the degree of change in the outcome. Additional comparisons were conducted using paired contrasts (sequential Bonferroni) when required.

3.2. Results

3.2.1. Eye movements during reading

One participant was excluded from the analyses due to low overall quality of the eye movement recordings. Table B.1 (see Appendix) displays the descriptive statistics for the number of fixations and the total duration of fixation in each area of interest (AOI) during the first and second readings, respectively, as a function of Discrepancy. Merging across all AOIs, missing observations increased between the first reading (M = 4.5%, SD = 0.99) and the second reading (M = 37.7%, SD = 18.17), indicating that many participants did not re-read every story. The average fixation number was higher during the first than during the second reading for almost every AOI, even when comparing only the participants who did both readings, $p < .001$.

The discrepancy manipulation did not have any significant impact on either the number or the total duration of fixations during the first reading, $p > .09$. During the second reading, however, Discrepancy affected the reading pattern on Source B's AOI [Fixation number: $F(1, 280) = 4.55$, $p = .034$; Fixation duration: $F(1, 280) = 3.78$, $p = .052$]. Compared to consistent statements, discrepant statements increased the number of fixations by 0.97 ($SE = 0.45$, $CI_{0.95} = 0.08, 1.86$) and the fixations' duration by 228.64 ms ($SE = 117.63$, $CI_{0.95} = -2.9, 460.19$) on this particular AOI. Note though that CIs are wide in both cases, so we cannot exclude that the effects are different than reported. No other significant impact of Discrepancy was found during second reading.

3.2.2. Recognition times, accuracy and RKG judgments

Table 4 presents the frequency of the different response types to the recognition trials, the recognition times and the RKG judgments as a function of Discrepancy (18 trials per subject, $n = 648$). The proportion of correct responses went from 48.0% of the trials in the consistent condition to 64.4% of the trials in the discrepant condition. Errors in which the incorrect source or the nonsource were chosen represented 28.3% and 12.2% of the responses, respectively. Control item errors represented only 3.3% of the total (21 trials). Because more than half of control item errors were associated to guess judgments, we decided to exclude them from further analyses. Thus, as in Experiment 1, Response type was treated as a three-level factor (correct response, other-source error, nonsource error).

Discrepancy did not have any significant impact on recognition times, $F(1, 574) = 0.03$, $p = .85$. However, a significant effect of Response type, $F(2, 592) = 17.14$, $p < .001$, and a significant Response type × Discrepancy interaction, $F(2, 583) = 3.76$, $p = .024$, were found. Compared to correct responses, recognition times increased by 2.71 s in nonsource errors ($SE = 0.60$, $CI_{0.95} = 1.53, 3.89$), and by 2.13 s in other-source errors ($SE = 0.47$, $CI_{0.95} = 1.22, 3.05$). The difference between the two error types did not reach statistical significance, $p = .98$. The paired contrasts indicated that the interaction with Discrepancy was such that, for correct responses only, participants were faster to respond in the discrepant than in the consistent condition, $p = .005$ (contrast estimation = −968.37, $SE = 344.19$, $CI_{0.95} = -1644.40, -292.34$), (see Table 4).

For recognition accuracy, the model revealed a significant effect of Discrepancy, $F(1, 646) = 18.91$, $p < .001$. The odds of producing correct responses were about twice as high in the discrepant than in the consistent condition, $OR = 2.04$, $CI_{0.95} = 1.48, 2.82$.

For the analysis of RKG judgments, Discrepancy and Response type were included as predictors into the model. A first analysis contrasted "remember" vs. "know" judgments. Both Response type, $F(2, 527) = 12.48$, $p < .001$, and Discrepancy, $F(1, 527) = 9.83$, $p = .002$, significantly predicted the degree of recollection: The odds of producing "remember" judgments were greater for correct responses than for other-source errors, $OR = 2.71$, $CI_{0.95} = 1.77, 4.14$, and nonsource errors, $OR = 2.61$, $CI_{0.95} = 1.35, 5.03$. They were also greater in the discrepant than in the consistent condition, $OR = 1.79$, $CI_{0.95} = 1.24, 2.59$. A second analysis contrasted "remember" vs. "guess" judgments. Again, Response type significantly predicted the distribution, $F(2, 374) = 13.61$, $p < .001$: The odds of getting "remember" judgments were higher for correct responses than for other-source errors, $OR = 3.17$, $CI_{0.95} = 1.57, 6.38$, $p = .001$, and nonsource errors, $OR = 8.17$, $CI_{0.95} = 3.64, 18.86$, $p < .001$. The effect of Discrepancy, on the other hand, showed a similar but marginal trend, $F(1, 374) = 3.39$, $p = .07$.

To sum up, participants were faster, more accurate, and felt more confident in their responses when they gave the correct answer, but also when the texts included discrepant rather than consistent statements.

3.2.3. Eye movements during the recognition task

Table B.2 (see Appendix B) displays the eye movement data obtained during the recognition task as a function of the presence of discrepancies and of response type. Eye movement recordings indicated that participants spent in average 4.4 s ($SD = 1.8$) reading the slide with the target statement (Fig. 2, slide 1). The average duration of participants' fixations ($M = 235$ ms, $SD = 46$) on this slide was typical of silent reading (Rayner, 2009), suggesting that participants were not anticipating their

Table 4
Participants' response and recollection judgment in the recognition task as a function of Discrepancy (Experiment 2).

		Frequency (%)		Recognition time – mean (SD)	
		Consistent	Discrepant	Consistent	Discrepant
Final response to the recognition task	Correct response	154 (48.0%)	208 (64.4%)	6.0 (3.2)	5.1 (3.6)
	Other-source error	111 (34.6%)	71 (22.0%)	7.2 (4.2)	7.2 (4.0)
	Nonsource error	42 (13.0%)	37 (11.4%)	6.6 (3.3)	7.8 (5.1)
	Control item error	14 (4.4%)	7 (2.2%)	7.4 (4.0)	7.8 (4.7)
Recollection judgment	Remember	117 (36.1%)	171 (52.8%)	–	–
	Know	146 (45.1%)	106 (32.8%)	–	–
	Guess	61 (18.8%)	46 (14.2%)	–	–

Note. Recognition times are reported in seconds. Trials in which participants chose the control character (3.3% of the observations) are not included in the calculation.

responses before seeing the multiple-choice slide. Discrepancy, Response type, and their interaction had no significant impact on the total number or duration of fixations on slide 1, $p > .26$.

Then, we analyzed participants' visual inspection of the items on the next slide, which displayed the four items to choose from. The initial analyses included Discrepancy (consistent, discrepant), Response type (correct response, other-source error, nonsource error), the Fixated AOI (correct source, other-source, nonsource, and control item) and their interactions as fixed factors. Response type [Fixation number: $F(2,2346) = 23.44$, $p < .001$; Fixation duration: $F(2,2346) = 25.24$, $p < .001$] and the Fixated AOI [Fixation number: $F(3,2346) = 18.07$, $p < .001$; Fixation duration: $F(3,2346) = 15.36$, $p < .001$] affected eye movements. These effects were qualified by two interactions: Discrepancy × Response type [Fixation number: $F(2,2346) = 3.15$, $p = .04$; Fixation duration: $F(2,2346) = 5.12$, $p = .006$] and Item × Response type [Fixation number: $F(6,2346) = 6.67$, $p < .001$; Fixation duration: $F(6,2346) = 6.56$, $p < .001$]. Finally, all these effects were qualified by a three-way interaction between the three fixed factors, both for the total number of fixations, $F(6,2346) = 2.39$, $p = .026$, and for the total fixation duration, $F(6,2347) = 2.83$, $p = .009$. To interpret the three-way interaction, follow-up analyses for each type of response were conducted separately, because the goal was to compare eye movements within response types, and not across response types. The mean number of fixations per AOI as a function story discrepancy and response type are shown in Fig. 3 and Table B.2 (Appendix).

For correct responses, the analyses showed a significant effect of Discrepancy [Fixation number: $F(1,1352) = 7.61$, $p = .006$; Fixation duration: $F(1,1352) = 10.79$, $p = .001$]. Participants fixated more ($b =$

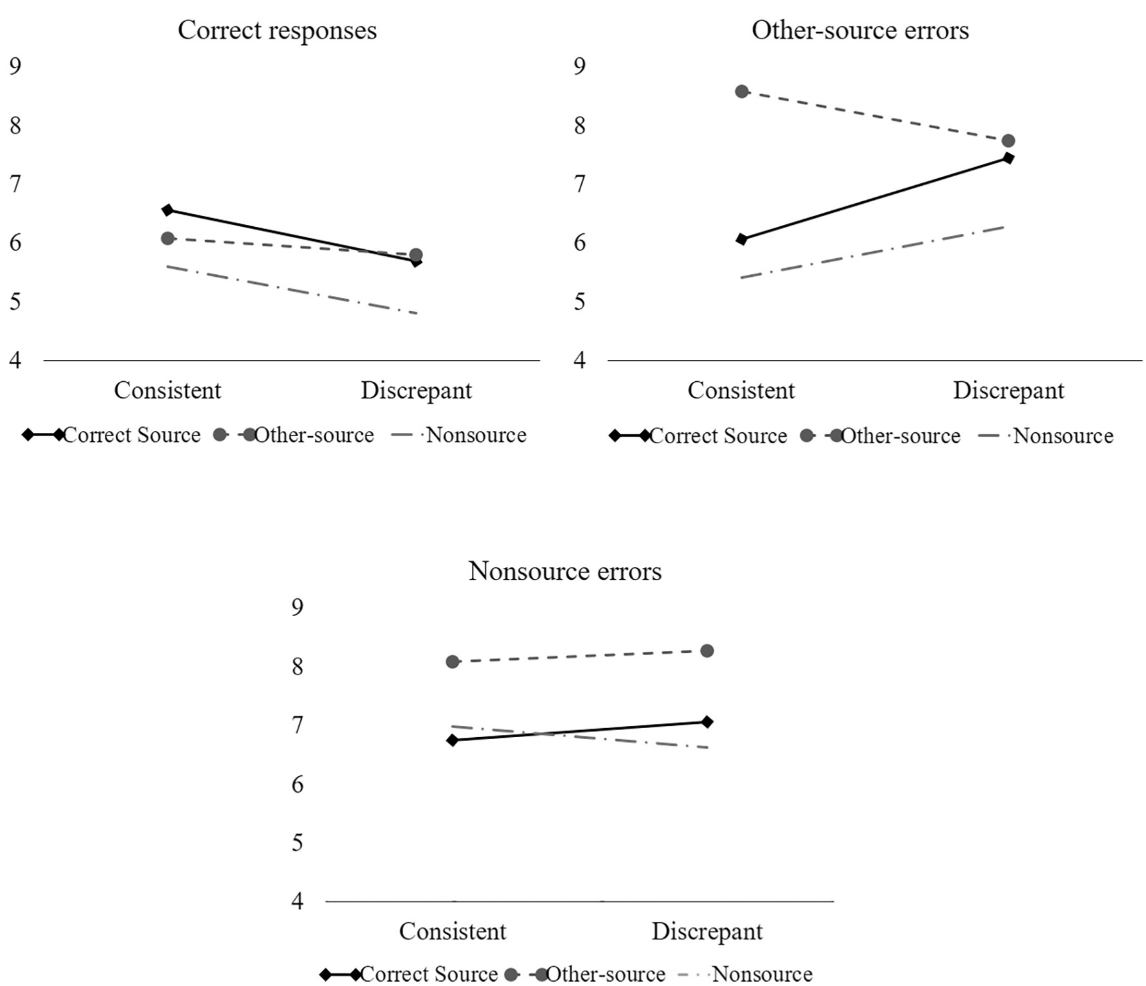

Fig. 3. Mean number of fixations obtained for correct responses, other-source errors and nonsource errors as a function of the Fixated AOI (correct source, other-source, nonsource) and Discrepancy, Experiment 2.

0.73, $SE = 0.40$, $CI_{0.95} = -0.06, 1.53$) and for more time ($b = 233.95$, $SE = 97.05$, $CI_{0.95} = 43.57, 424.3$) the AOIs in the consistent than in the discrepant condition. The analyses also showed a significant effect of the Fixated AOI [Fixation number: $F(3,1352) = 6.76$, $p < .001$; Fixation duration: $F(3,1352) = F(3,1352) = 7.19$, $p < .001$]. Paired comparisons revealed that participants fixated more and for more time on the two source items (both the correct and incorrect ones) than the nonsource item, $p < .027$. Averaging the two source items, the estimated increase as compared to the nonsource item was of 0.91 fixations ($SE = 0.38$, $CI_{0.95} = 0.15, 1.67$) and of 270.81 ms ($SE = 92.36$, $CI_{0.95} = 88.91, 472.71$). No significant difference was detected between the sources, nor between the nonsource and the control item. The Discrepancy × Fixated AOI interaction did not yield any significant result, $F(3,1352) = 0.54$ $p > .66$.

For other-source errors, the analyses revealed a main effect of the Fixated AOI [Fixation number: $F(3,697) = 26.47$, $p < .001$; Fixation duration: $F(3,697) = 22.46$, $p = .018$], qualified by an interaction with Discrepancy [Fixation number: $F(3,697) = 3.40$, $p = .018$; Fixation duration: $F(3,697) = 4.43$, $p = .004$]. The post hoc comparisons revealed that, in the discrepant condition, the number of fixations made on both sources was the highest as opposed to the control item, which was fixated the least, $p < .001$ (contrast estimations: Correct source vs. Control item = 3.05, $SE = 0.72$, $CI_{0.95} = 1.18, 4.92$; Incorrect source vs. Control item = 3.17, $SE = 0.72$, $CI_{0.95} = 1.27, 5.07$). Importantly, the number of fixations made on both sources was similar ($p = .86$). The inspection of the non-source item was placed in between the two sources and the control item, tending to differentiate from them, but without reaching statistical significance $p > .056$. In contrast, in the consistent condition there was a large difference between the numbers of fixations made on the two sources ($p < .001$), with participants fixating more on the incorrect source than on any of the other three AOIs, $p < .001$ (contrast estimations: Incorrect Source vs. Correct source = 2.87, $SE = 0.56$, $CI_{0.95} = 1.46, 4.29$; Incorrect source vs Nonsource = 3.55, $SE = 0.56$, $CI_{0.95} = 2.10, 5.01$; Incorrect Source vs. Control item = 4.74, $SE = 0.56$, $CI_{0.95} = 3.25, 6.23$). Data relative to the duration of fixations mirrored the AOI × Discrepancy interaction, but the post hoc analyses showed less discrimination between the two sources and the nonsource item in the discrepant condition, $p > .60$.

Finally, for nonsource errors, no significant effect of Discrepancy, the Fixated AOI, or their interaction was found for either the total number of fixations, $p > .18$, or their total duration, $p > .08$.

3.3. Discussion

The second experiment aimed at extending the results of the first experiment by providing eye movement data collected during reading, but also during the recognition task.

The first hypothesis predicted that discrepant statements would increase attention to sources during reading. In line with this prediction, the number and duration of fixations was higher for the second source of the stories when statements were discrepant. Interestingly, this was the only AOI affected by the manipulation during reading, thus suggesting a certain specificity of the discrepancy-induced reading processes. The effect replicates Braasch et al.'s (2012) findings using more complex stories, in which additional contents (a nonsource character who is also involved in the depicted situation, an introductory sentence, and a final sentence) were added.

The second hypothesis predicted that discrepancies would increase participants' accuracy and decrease recognition time. In accordance with this prediction, we found first, that participants were better at recognizing the sources of discrepant than consistent statements, and second that participants recognized the sources faster in the discrepant condition, particularly when making the correct choice in the recognition trial (i.e., discrepancies promoted the construction of source-to-content links). Consistent with this pattern, participants also felt more confident of their responses after reading discrepant stories. Overall, this set of results provides support for the second hypothesis and complements well the results of Experiment 1.

Finally, the third hypothesis was that discrepancies would not only promote the distinctiveness of the source nodes in memory, as reflected by the recognition accuracy, but would also promote a stronger link between the two source nodes, as suggested by the Documents Model Framework (e.g., Perfetti et al., 1999). To test this prediction, we examined participants' visual inspection of the items before producing a response in the recognition trials. Our assumption was that similar fixation patterns on the two source items would indicate more integration of these two nodes in memory (i.e., the source-to-source link). Interestingly, this happened when participants gave correct answers, and in the discrepant condition when participants made other-source errors, suggesting that, in those particular situations, the two sources were more "tied" together in participants' memory. Particularly in the other-source type of error, in which an incomplete underlying source-content representation is assumed, the discrepant condition seems to have turned the two source characters into *better* but *similar* contenders in memory, as compared to the other response options. We believe that, above and beyond accuracy, this particular pattern provides evidence that discrepancies induce some separation of the source nodes from the rest of the text contents, but also some integration between them (e.g., Britt & Rouet, 2012). 'Who said what' tasks have been used before to examine the creation of the source-to-content link (e.g., Braasch et al., 2012; Rouet et al., 2016). An important contribution of Experiment 2 is that, by incorporating eye tracking into the recognition task, both the source-to-content and the source-to-source links were examined in the same task. Of note is that evidence of the construction of a source-to-source link was observed during recognition (i.e., similar fixations for the two source items in discrepant condition) rather than during reading (i.e., no 'tying' of the two discrepant source areas was observed during the reading phase). In line with prior research (Saux et al., 2017), this pattern suggests that producing source-to-source links may be more costly than producing source-to-content links, and that the D-ISC effect on the former may be observed after additional elaboration by the reader.

4. General discussion

The two experiments from this study had two main goals. The first goal was to provide evidence on whether characters who produce discrepant informative statements in the same text are distinctly represented by the reader from other text contents. The second goal was to expand research on memory for discrepant sources by including additional indicators of memory availability and integration beyond accuracy. To reach the first goal, both studies included two types of characters in the material: those who made statements (i.e., the sources) and those who were just involved in the situation depicted in the text, but made no statement (i.e., the nonsource characters). To reach the second goal, both studies collected data during the recognition task beyond the participant's final response, using verbal protocols in Experiment 1 and eye-tracking evidence in Experiment 2, and considered differences between the different types of possible answers (correct responses and different error types).

Overall, the findings of both experiments were consistent with the theoretical predictions of the documents model framework (Perfetti et al., 1999). In Experiment 1, the presence of discrepancies promoted memory distinctiveness for the source targeted in the recognition task and for the other source, but not for the non-source, which remained unaffected by the manipulation. The verbal protocols from Experiment 1 also suggested that distinct cognitive processes were associated with the different incorrect responses to the recognition task, indicating more similarity between correct responses and other-source errors than between the two error types.

To further explain why both source characters seemed similarly salient in some situations, eye-tracking was used in Experiment 2 as a measure of participants' attention to the three characters during reading and during the recognition trial. Participants were more attentive to sources when reading discrepant statements and, as in Experiment 1, correct responses to recognition trials were made faster, with stronger accuracy feelings than errors, and occurred more frequently after having read discrepant statements.

Importantly, the degree to which participants visually inspected the items differed according to response types and statement discrepancy. When making correct responses and other-source errors, participants fixated the two sources to the same extent, and more than the other response options. In other-source errors, this was so after having read discrepant statements. Conversely, nonsource errors remained unaffected by the manipulation.

In sum, the combined results from both experiments indicate that the impact of discrepancies is specific to source as opposed to nonsource characters, as originally proposed by documents model framework but not directly tested in earlier experiments (e.g., Braasch et al., 2012). This claim is supported by the selective memory enhancement for source-related segments of the text, complemented by a lack of effect on other text segments or characters, as observed in both experiments. The results also suggest that discrepancies promote not only a distinct, but also an integrated representation of multiple sources, in which conflicting information providers seem to be more tied together in memory than when the statements are consistent. Finally, the results also suggest that this effect is incremental rather than categorical, with some cases in which readers construct a fuzzy documents model, which is incomplete but still more specific than other cases, in which an indistinct (untagged) representation should be assumed.

4.1. Limitations and future research

Several aspects of the present experiments limit the generalization of the results. First, it could be argued that character types (sources and nonsource) differ both as information providers (i.e., the nonsource never produces a statement) and in their importance in the story (i.e., in terms of text structure, the sources present more connections with the rest of the story than the nonsource). Our hypotheses (in both Experiments), however, were not merely aimed at identifying differences in the processing of character types, but were always conditional to statement discrepancy. In other words, our claim is that sources become instrumental when there are coherence breaks in what characters say. Because we compared the same texts in consistent and discrepant versions, the processing and representational differences observed among character types as a function of the manipulation cannot be attributed only to their causal importance in text structure. Future research is needed to examine whether story discrepancies promote attention to any source embedded in the story or only to those sources that provide contested information. This could be examined by including characters who provide utterances unrelated to the conflicting claims. Recent research has shown that discrepant statements can selectively favor the processing of source features that are useful to deal with the discrepancy, as compared to other, less relevant source features (Gottschling et al., 2019; Saux et al., 2018). Although this was not a specific goal of this study, we believe that this line of work complements well the results reported here.

Another point to consider is the nature of the discrepancies in the stories. In most of our texts (14 out of 18), discrepancies were causal in nature and concerned factual statements (causal-factual discrepancies). In the remaining four stories, discrepancies were either evaluative (i.e., different assessments of the same situation) or predictive (i.e., different estimations of what would happen). As a control, we have re-run all the analyses for Experiment 1 and Experiment 2 (except for eye-tracking data) with only the texts presenting causal-factual discrepancies. The effects reported in the paper do not change, $p < .05$. However, it is worth noting that the D-ISC hypothesis predicts that source-to-content links would be most likely created when the contradiction makes it difficult to integrate the dissimilar perspectives into the same mental model, thus making the sources valuable tools for restoring coherence (Rouet et al., 2016). Whether different types of discrepancies (differences in aesthetical appreciations, ideological positions, etc.) differ in their influence on the representation of embedded sources and other characters is not clear to us, but it would be worthwhile examining in the future.

The influence of text discrepancies was systematic across the two experiments. However, other factors could also play a role. For example, our study does not explain why readers would build sometimes clearer and sometimes fuzzier representations that would lead them to correctly identify the source in the first case and to confuse the two sources in the other case. Readers use different means to achieve coherence in the presence of factual inconsistencies. The instrumental use of source information is one way, but readers can also ignore part of the information, distort it, or try to reconcile it (Stadtler & Bromme, 2014). For instance, employing experimental materials similar to the ones from this study, Rouet et al. (2016) reported that their participants sometimes used hedging and other tactics to resolve the contradictions when asked to summarize the texts. One particular factor that has been proven to influence readers' sourcing is the task. For example, readers' amount of attention and source use increases when they are instructed at the outset to do so (for a review, see Wiley et al., 2018). Recent studies have reported simultaneous but independent effects of text discrepancies and prompting tasks on source memory tasks, with no interaction between them, with the size of the task effect being generally stronger (Rouet et al., 2020; Saux et al., 2018). However, the exact interplay between the presence of text discrepancies and the reader's task model in prompting sourcing is still not clear and should be further explored. On a similar line, we cannot fully discount an influence of perceived plausibility and prior beliefs about the sources or content, even after switching the role of the characters within each story across participants. Several recent reviews have attempted to unravel the set of factors that impact sourcing (Braasch et al., 2018; Bråten et al., 2018; List & Alexander, 2017). Whereas the presence of discrepancies in a text is one of them, other text, reader, and situation-related variables should be taken into consideration when generalizing the present results.

Another limitation relates to the potential interactions of the current results with evaluative processes. As stated in the introduction, this paper focuses on source memory. However, the importance of sourcing also relates to the ability to use source information to evaluate the reliability of what is being read (Putnam & Phelps, 2017). The relationship between epistemic validation processes, such as source evaluation, and other comprehension processes, such as the construction of a multi-level representation of discourse, may be closer than traditionally assumed (e.g., O`Brien & Cook, 2016; Richter & Maier, 2017). So far, evaluation processes seem to support rather than undermine the D-ISC effect (e.g., Bråten et al., 2016; Kammerer et al., 2016). However, the exact impact of source evaluation on source memory remains an open question.

4.2. Implications

This study has implications both for research and applied contexts. From a research perspective, it contributes to the specification of mechanisms to account for the creation of source-to-content links during reading as part of the reading experience (Braasch et al., 2012; Britt et al., 1999; Perfetti et al., 1999). Although basic assumptions regarding the construction of a source representation in memory have been articulated (e.g., Britt et al., 1999; Britt & Rouet, 2012), the encoding and retrieval processes underlying a documents model under different task conditions remain to be specified. Our results contribute by proposing a specific and incremental nature to the effect attributed to text discrepancies. At the same time, they highlight the need to further develop the study of memory representations that underlie the comprehension of multiple texts.

Another contribution is the novel use of eye-tracking during post-reading tasks in Experiment 2. To the best of our knowledge, this is the first study to examine source-tagged text representations by means of this methodology in an offline (i.e., post-reading) memory task. Combined with the use of verbal protocols (Experiment 1), the eye tracking methodology allowed us to examine source memory beyond accuracy indicators, and proved effective at detecting rather subtle nuances associated with the memory task. Salmerón et al. (2018) have pointed out several limitations in the use of eye tracking to the study of sourcing, including lack of consistency between eye movements' data and other measures, such as post-reading interviews. However, the authors also made clear that the problem lies mostly in the absence of a connection between the data and theoretical models of sourcing. In the present study, the combined interpretation of the results was possible by keeping the data in tight association with the specific theoretical predictions derived from the D-ISC hypothesis (Braasch et al., 2012) and the Documents Model Framework (e.g., Britt et al., 1999).

From an applied perspective, the results may be of interest to programs oriented to develop sourcing strategies in young or struggling readers. The Internet tends to confront readers with complex reading scenarios, including contradictions across multiple documents, formal differences, and variations in trustworthiness (Salmerón et al., 2018). However, the notion of source and the idea that knowledge is not equally distributed can be traced back to young elementary students (Stadtler et al., 2018). The simplified scenario used in this research (i.e., short texts that include embedded source and nonsource characters) could thus serve as a basis for developing or training specific sourcing heuristic skills at initial levels in a concentrated, intensive way. In fact, focused instructional interventions using short texts have proven useful to promote sourcing among teenagers (e.g., Pérez et al., 2018). We believe this could be a fruitful path to foster basic functional document comprehension required by the digital era.

CRediT authorship contribution statement

Gaston Saux: Conceptualization, Formal analysis, Investigation, Writing – Original Draft.
Nicolas Vibert: Conceptualization, Writing – Review & Editing.
Julien Dampuré: Investigation, Resources.
Debora I. Burin: Formal analysis, Writing – Review & Editing.
M. Anne Britt: Conceptualization, Writing – Review & Editing.
Jean-François Rouet: Conceptualization, Writing – Review & Editing, Project administration.

Funding

This work was supported by the Agencia Nacional de Promoción Científica y Tecnológica from Argentina [PICTO ANPCyT-UCA 2017-0068]; and the Agence Nationale pour la Recherche from France [ANR-12-cord-0028-02 "ARCSYS"].

Declaration of competing interest

None.

Appendix A

Example table and note on scoring criteria corresponding to Section 2.1.5 "Scoring of the retrospective interviews" (Experiment 1).

Table A.1
Categories, quotes, and scoring examples for the content analysis of the short interviews (recognition task, Experiment 1).

Category	Quote examples
• Retrieval (expressing certainty when identifying that a character was in the story; references to thoughts about the character elicited during the reading phase)	P. 28, text nr. 7: "It's the [*correct source*], I don't need to see the options (…)". P. 25, text nr. 3: "It's the first option [*nonsource*] (…) because I remember I wrote [*ref. to the title produced during reading*] as the new title, that's why I remember". P. 11, text nr. 8: "Oh, this one I remember. In this case I'm sure it was the [*correct source*]".
• Familiarity (expressing a feeling of familiarity when identifying that a character was in the story, with no clear recollection, expressing hesitation)	P. 1, text nr. 15: "(…) I discard [*ref. to the correct source*]… There was someone that contradicted this… I think it was the [*control item*], but I'm not very sure, I don't feel it very strongly". P. 5, text nr. 14: "(…) I think the [*nonsource*] was in the story, but I don't know if she said this… (…) OK, I'll go with the [*nonsource*], but I'm not sure, it only sounds familiar" P. 39, text nr.9: "(…) I do not remember, I focused more on the phrase than on the person… emm… I have the impression that the [*incorrect source*] said this, but I'm not completely sure".
• Plausibility (using world prior knowledge during retrieval to infer the plausibility of a certain character for a specific setting)	P. 12, text nr. 4: "Oh, this story was about [*ref. to the setting of the story*] (…) I think the [*nonsource*] must have been there, if it was in the [*ref. to the setting of the story*]". *P.* 22, text nr. 10: "I'm not really sure, I don't remember this one (…) I don't know. I'll say the [*correct source*] because I relate that the [*correct source*] should know about this stuff". P. 34, text nr. 8: "(…) I remember there were, like, several sentences… I don't know why but I remember there was a lot of people (…) oh no, I'm not sure… so…, I say the [*nonsource*], but no one rings a bell".

Note on scoring criteria (Recognition Task, Experiment 1).

The level of acceptance for the interrater agreement (*kappa*) was set to 0.7, which corresponds to moderate to strong agreement (McHugh, 2012). Initially, the "Retrieval" and "Familiarity" sub-categories did not surpass the level of acceptance. A detailed inspection revealed that participants tended to use an idiomatic regional expression ("*me suena*"; literally "it sounds to me") to refer to any type of recall, even when they also expressed feeling certain about their memory. This created confusion in the categorization. A more conservative criterion was then agreed, according to which an utterance would be coded as "Retrieval" if participants made explicit mention of feeling certain about their recall, regardless of the use of the idiomatic expression. The protocols were reanalyzed after applying this new criterion, resulting in acceptable levels of interrater agreement ($k > 0.78$).

Appendix B

Tables with descriptive statistics for eye movements during reading (Table B.1) and during the recognition task (Table B.2), corresponding to Section 3.2 (Results, Experiment 2).

Table B.1
Means of the number and total duration of fixations (in seconds) made in each AOI during first and second readings of the texts, as a function of Discrepancy (Experiment 2).

Text section	AOI	DV	First reading		Second reading	
			Consistent Mean (SD)	Discrepant Mean (SD)	Consistent Mean (SD)	Discrepant Mean (SD)
1	Introductory sentence	Fix count	25.5 (12.9)	24.4 (12.8)	17.5 (15.9)	17.3 (17.2)
		Fix duration	5.9 (3.3)	5.5 (3.0)	3.9 (3.7)	4.0 (4.4)
2	Story agent	Fix count	7.6 (4.8)	7.1 (3.9)	5.6 (4.7)	5.7 (5.6)
		Fix duration	1.6 (1.2)	1.5 (1.0)	1.2 (1.0)	1.3 (1.9)
	Action (story agent)	Fix count	11.8 (5.6)	11.5 (5.7)	6.5 (7.0)	7.7 (7.6)
		Fix duration	2.6 (1.4)	2.5 (1.4)	1.2 (1.2)	1.5 (1.6)
3	Source A	Fix count	6.9 (6.3)	6.1 (3.8)	2.7 (4.5)	3.1 (5.2)
		Fix duration	1.4 (1.0)	1.4 (0.9)	1.3 (1.2)	1.4 (1.4)
	Utterance (Source A)	Fix count	9.1 (5.0)	9.2 (4.4)	7.5 (7.6)	8.1 (8.0)
		Fix duration	2.0 (1.3)	2.0 (1.0)	1.6 (1.5)	1.7 (1.7)
	Connector	Fix count	2.7 (1.9)	2.4 (1.7)	1.3 (2.5)	1.4 (2.4)
		Fix duration	0.6 (0.5)	0.6 (0.6)	0.7 (0.8)	0.8 (0.8)
	Source B	Fix count	5.5 (2.9)	5.07 (2.5)	**3.8 (3.5)**	**4.8 (4.3)**
		Fix duration	1.2 (0.9)	1.1 (0.7)	**0.8 (0.9)**	**1.4 (1.1)**
	Utterance (Source B)	Fix count	10.6 (5.6)	10.1 (4.2)	7.8 (7.5)	7.3 (7.6)
		Fix duration	2.4 (1.8)	2.3 (1.4)	1.5 (1.7)	1.6 (2.0)

Note. Highlights in bold indicate significant differences between consistent and discrepant conditions ($p < .05$).

Table B.2
Mean total numbers and durations (in seconds) of fixations made in each AOI during the recognition task as a function of Response type and Discrepancy (Experiment 2).

AOI		Correct responses		"Incorrect source" errors		"Story agent" errors	
		Consistent	Discrepant	Consistent	Discrepant	Consistent	Discrepant
Correct source	Fix count	6.6 (4.1)	5.7 (4.1)	6.1 (4.8)	7.4 (5.3)	6.7 (3.6)	7.1 (3.7)
	Fix duration	1.5 (1.0)	1.3 (1.0)	1.5 (1.3)	1.8 (1.5)	1.4 (0.9)	1.6 (0.8)
Incorrect source	Fix count	6.1 (4.0)	5.8 (5.1)	8.6 (5.9)	7.7 (5.3)	8.1 (4.9)	8.3 (7.1)
	Fix duration	1.4 (1.0)	1.3 (1.2)	2.2 (1.6)	1.8 (1.3)	1.8 (1.3)	2.1 (2.2)
Story agent	Fix count	5.6 (3.9)	4.8 (3.9)	5.4 (3.5)	6.3 (4.0)	7.0 (4.3)	6.6 (6.4)
	Fix duration	1.2 (1.0)	1.0 (0.9)	1.3 (0.9)	1.3 (1.0)	1.6 (1.0)	1.6 (1.7)
Control item	Fix count	5.3 (3.4)	4.8 (4.0)	4.4 (3.2)	4.5 (2.4)	6.7 (4.6)	8.0 (6.4)
	Fix duration	1.2 (0.8)	1.1 (0.9)	1.1 (0.8)	1.0 (0.5)	1.5 (1.0)	1.9 (1.6)

Appendix C

Experiment 2 data analysis including performance in the OSPAN task as covariable.

The analyses reported in this section replicate the analyses reported in the main body of the paper, but adding participants' performance in the operation-word span task (OSPAN, Unsworth et al., 2005) as a covariable.

Description of the automated OSPAN task

The OSPAN task is a widely used measure of working memory capacity (WMC) that has been associated to fluid abilities and controlled or executive attention (Unsworth & Engle, 2005). The OSPAN automated version used as distracter task in Experiment 2 is divided into sets of three to seven items (Redick et al., 2012). Each item requires the participant to solve a math operation and decide whether a suggested answer is correct or not. Then, a letter appears on the screen for 1000 ms. After completing a set, the participant must recall in correct order the presented letters by clicking on a letter grid. For the current analyses, we used the absolute method to score performance (i.e., the sum of all perfectly recalled sets; $M = 31.09$, $SD = 14.37$).

Statistical analyses

Model estimation mirrored the approach reported in the main body of the paper (please refer to the Results' section of Experiment 2 for specs). The only differences with the analyses reported here are the inclusion of the OSPAN scores as a continuous fixed predictor and the application of the Satterthwaite approximation to fit the degrees of freedom. Initial models included all possible interactions. However, non-significant interaction terms involving the continuous predictor were removed from the final models, since failure to remove these can lead to incorrect conclusions (Engqvist, 2005). In case of significance, the effect (main or interaction) of the continuous predictor was analyzed by inspecting the slope of the estimated pendent.

Results

Eye movements during reading

During second reading passes, a significant OSPAN × Discrepancy interaction was found for source B's AOI, only on the number of fixations, F(1,

261) = 4.98, $p = .027$. The higher the score, the higher the number of fixations for this area in discrepant trials ($B = 0.07$). A similar, marginal tendency was also found on the AOI containing the statement of source B [fixation duration: $F(1, 293) = 3.61$, $p = .058$; fixation number: $F(1, 292) = 3.84$, $p = .051$]. The higher the score, the more participants fixated on this statement in discrepant as compared to consistent trials (fixation duration: $B = 21.46$; fixation number: $B = 0.09$).

Recognition times, accuracy and RKG judgments

Regarding recognition times, the general pattern was the same as in the original analysis: Whereas correct responses were overall faster than errors, correct responses in discrepant trials were, in addition, significantly faster than in consistent trials, $p < .005$. OSPAN scores also predicted recognition times, so that the higher the score, the faster the response, $F(1, 34) = 11.39$, $p = .002$, $B = -62{,}38$. The interaction between OSPAN scores and the fixed predictors were non-significant.

Regarding recognition accuracy, the general pattern was the same as in the original analysis: discrepant trials presented more correct responses than consistent trials, $p < .001$. In addition, the OSPAN scores also predicted accuracy, so that for one point increase in the score, the odds of correctly recognizing the source increased by 1.02, $F(1, 34) = 5.48$, $p = .025$. The Discrepancy by OSPAN interaction was nonsignificant.

Regarding RKG judgments, no significant effects of the OSPAN scores or its interactions with the fixed predictors were found.

Eye movements during the recognition task

As in the original analyses, each response type was examined separately.

Regarding correct responses, the pattern was the same as in the original analysis: participants fixated more and for more time the AOIs in the consistent than in the discrepant condition. Also, in both conditions, they fixated more and for more time on the two source items than the nonsource item, $p < .01$. In addition, OSPAN scores also predicted eye movements in correct responses, so that fixation number and duration decreased as the score increased [fixation duration: $F(1, 32) = 14.11$, $p = .001$, $B = -15.05$; fixation number: $F(1, 30) = 16.29$, $p < .001$, $B = -0.05$]. The OSPAN interactions with the fixed predictors were nonsignificant.

Regarding incorrect source errors, as in the original analysis, the Discrepancy manipulation was qualified by an interaction with the Type of Area so that, in the discrepant condition, both sources were fixated the most but similarly, whereas in the consistent condition there was a large difference between the numbers of fixations made on the two sources, with participants fixating more on the incorrect source than on any of the other three AOIs, $p < .05$. In addition, the Discrepancy manipulation was also qualified by an interaction with OSPAN scores [fixation duration: $F(1, 686) = 3.66$, $p = .056$; fixation number: $F(1, 685) = 6.32$, $p = .012$]. The higher the score, the less participants fixated on the items in the consistent as compared to discrepant trials (fixation duration: $B = -11.44$; fixation number: $B = -0.06$).

Regarding nonsource errors, a main effect of OSPAN scores was observed, only for fixations' duration. Overall, the higher the score, the faster the duration of the fixations when producing this type of error, $F(1,38) = 4.98$, $p = .032$, $B = -19.18$.

Summary and conclusions

Overall, the inclusion of the OSPAN scores as a continuous predictor into the models did not modify the effects and interactions reported in the main body of the paper and therefore do not affect the general conclusions of the paper. However, the OSPAN scores did predict performance, sometimes in interaction with the Discrepancy manipulation. Taken as a whole, these new effects suggest that high WMC may facilitate the construction of a documents model during reading (as indicated by the OSPAN × Discrepancy interaction when inspecting source B and her/his statement during second reading) and its controlled retrieval during a source memory task (as indicated by the OSPAN main effects on recognition time and accuracy and its interaction with Discrepancy on the inspection of the AOIs during recognition).

Although a detailed examination of these additional effects fall outside the scope of the present study, we hope they will stimulate future research. We believe they show the need of a more systematic examination of the role of WMC differences on different aspects of the construction and strategic use of a documents model.

References

Braasch, J. L. G., & Bråten, I. (2017). The Discrepancy-Induced Source Comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist, 52*(3), 167–181. https://doi.org/10.1007/978-981-10-4331-4

Braasch, J. L. G., Bråten, I., & McCrudden, M. T. (Eds.). (2018). *Handbook of multiple source use*. Routledge. https://doi.org/10.4324/9781315627496.

Braasch, J. L. G., McCabe, R. M., & Daniel, F. (2016). Content integration across multiple documents reduces memory for sources. *Reading and Writing: An Interdisciplinary Journal, 29*(8), 1571–1598. https://doi.org/10.1007/s11145-015-9609-5

Braasch, J. L. G., Rouet, J.-F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory and Cognition, 40*(3), 450–465. https://doi.org/10.3758/s13421-011-0160-6

Bråten, I., Stadtler, M., & Salmerón, L. (2018). The role of sourcing in discourse comprehension. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Handbook of discourse processes* (2nd ed., pp. 141–166). Routledge.

Bråten, I., Strømsø, H. I., & Andreassen, R. (2016). Sourcing in professional education: Do text factors make any difference? *Reading and Writing: An Interdisciplinary Journal, 29* (8), 1599–1628. https://doi.org/10.1007/s11145-015-9611-y

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485–522. https://doi.org/10.1207/S1532690XCI2004_2

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). Lawrence Erlbaum Associates Publishers.

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M. J. Lawson, & J. R. Kirby (Eds.), *The quality of learning: Dispositions, instruction, and mental structures*. Cambridge University Press.

Britt, M. A., Rouet, J.-F., & Braasch, J. L. G. (2013). Documents experienced as entities: Extending the situation model theory of comprehension. In M. A. Britt, S. R. Goldman, & J.-F. Rouet (Eds.), *Reading: from words to multiple texts* (pp. 160–179). Routledge.

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. 123–139). IAP Information Age Publishing.

Bromme, R., Scharrer, L., Stadtler, M., Hömberg, J., & Torspecken, R. (2015). Is it believable when it's scientific? How scientific discourse style influences laypeople's resolution of conflicts. *Journal of Research in Science Teaching, 52*(1), 36–57. https://doi.org/10.1002/tea.21172

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General, 143*(4), 1476–1488. https://doi.org/10.1037/a0035813

Dampuré, J., Ros, C., Rouet, J.-F., & Vibert, N. (2014). Task-dependent sensitization of perceptual and semantic processing during visual search for words. *Journal of Cognitive Psychology, 26*, 530–549. https://doi.org/10.1080/20445911.2014.907576

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

Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review, 9*(1), 139145. https://doi.org/10.3758/BF03196270

Engqvist, L. (2005). The mistreatment of covariate interaction terms in linear model analyses of behavioural and evolutionary ecology studies. *Animal Behaviour, 70*(4), 967–971. https://doi.org/10.1016/j.anbehav.2005.01.016

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes, 56*(7), 485–495. https://doi.org/10.1080/0163853X.2019.1591885

Gardiner, J. M., Richardson-Klavehn, A., & Ramponi, C. (1997). On reporting recollective experiences and "direct access to memory systems". *Psychological Science, 8*(5), 391–394. https://doi.org/10.1111/j.1467-9280.1997.tb00431.x

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Ferris, & D. M. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 313–347). Information Age Publishing.

Gottschling, S., Kammerer, Y., & Gerjets, P. (2019). Readers' processing and use of source information as a function of its usefulness to explain conflicting scientific claims. *Discourse Processes, 56*(5–6), 429–446. https://doi.org/10.1080/0163853x.2019.1610305

Graesser, A. C., Bowers, C., Olde, B., & Pomeroy, V. (1999). Who said what? Source memory for narrator and character agents in literary short stories. *Journal of Educational Psychology, 91*(2), 284–300. https://doi.org/10.1037/0022-0663.91.2.284

Guerry, M., Catelain, A., & Caron, J. (1993). La compréhension de marqueurs modaux: verbes d'attitude propositionelle et adverbes. *L'Année Psychologique, 93*(2), 201–225.

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

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28.

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition, 27*(4), 648–656.

Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes, 53*, 430–456. https://doi.org/10.1080/0163853X.2016.1169968

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

List, A., & Alexander, P. A. (2017). Analyzing and integrating models of multiple text comprehension. *Educational Psychologist, 52*(3), 143–147. https://doi.org/10.1080/00461520.2017.1328309

List, A., Van Meter, P., Lombardi, D., & Kendeou, P. (2020). Loggers and conservationists: Navigating the multiple resource forest through the trees. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (pp. 1–14). Routledge.

Long, D. L., & Spooner, A. (2010). Placing a text in context. *Psychonomic Bulletin & Review, 17*, 237–242.

McCrudden, M. T., Stenseth, T., Bråten, I., & Strømsø, H. I. (2016). The effects of topic familiarity, author expertise, and content relevance on Norwegian students' document selection: A mixed methods study. *Journal of Educational Psychology, 108*(2), 147–162. https://doi.org/10.1037/edu0000057

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/bm.2012.031

O'Brien, E. J., & Cook, A. E. (2016). 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

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica, 144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

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

Pérez, A. M., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J. M., Salmerón, L., & Rouet, J. (2018). *Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions*. https://doi.org/10.1016/j.learninstruc.2018.04.006

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). Lawrence Erlbaum Associates Publishers.

Pilditch, T. D., Madsen, J. K., & Custers, R. (2020). False prophets and Cassandra's curse: The role of credibility in belief updating. *Acta Psychologica, 202*, 102956. https://doi.org/10.1016/j.actpsy.2019.102956

Putnam, A. L., & Phelps, R. J. (2017). The citation effect: In-text citations moderately increase belief in trivia claims. *Acta Psychologica, 179*, 114–123. https://doi.org/10.1016/j.actpsy.2017.07.010

Radvansky, G. A., Wyer, R. S., Jr., Curiel, J. M., & Lutz, M. F. (1997). Situation models and abstract ownership relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1233–1246. https://doi.org/10.1037/0278-7393.23.5.1233

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition, 35*(8), 2019–2032. https://doi.org/10.3758/BF03192934

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist, 52*(3), 148–166. https://doi.org/10.1080/00461520.2017.1322968

Rouet, J.-F., & Britt, M. A. (2014). Learning from multiple documents. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (2nd ed., pp. 813–841). Cambridge: MA, Cambridge University Press.

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: An Interdisciplinary Journal, 29*(8), 1549–1570. https://doi.org/10.1007/s11145-016-9625-0

Rouet, J.-F., Saux, G., Ros, C., Stadtler, M., Vibert, N., & Britt, M. A. (2020). Inside document models: The role of source attributes in integrating multiple text contents. *Discourse Processes*. https://doi.org/10.1080/0163853X.2020.1750246

Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 36–43. https://doi.org/10.1111/ldrp.12028

Salmerón, L., Gil, L., & Bråten, I. (2018). Using eye-tracking to assess sourcing during multiple document reading: A critical analysis. *Frontline Learning Research, 6*, 105–122. https://doi.org/10.14786/flr.v6i3.368

Salmerón, L., Macedo-Rouet, M., & Rouet, J.-F. (2016). Multiple viewpoints increase students' attention to source features in social question and answer forum messages. *Journal of the Association for Information Science and Technology., 67*, 2404–2419. https://doi.org/10.1002/asi.23585

Saux, G., Britt, M. A., Le Bigot, L., Vibert, N., Burin, D., & Rouet, J. F. (2017). Conflicting but close: Readers' integration of information sources as a function of their disagreement. *Memory & Cognition, 45*(1), 151–167. https://doi.org/10.3758/s13421-016-0644-5

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*. https://doi.org/10.1080/0163853X.2018.1463722

Stadtler, M., & Bromme, R. (2014). The content-source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp, & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). MIT Press.

Stadtler, M., Bromme, R., & Rouet, J.-F. (2018). Learning from multiple documents: How can we foster multiple document literacy skills in a sustainable way? In E. Manalo, Y. Uesaka, & C. A. Chinn (Eds.), *Routledge research in achievement and gifted education. Promoting spontaneous use of learning and reasoning strategies* (pp. 46–61). Routledge.

Steffens, B., Britt, M. A., Braasch, J. L. G., Strømsø, H. I., & Bråten, I. (2014). Memory for scientific arguments and their sources: Claim-evidence consistency matters. *Discourse Processes, 51*(1–2), 117–142. https://doi.org/10.1080/0163853X.2013.855868

Strømsø, H. I., & Bråten, I. (2014). Students' sourcing while reading and writing from multiple documents. *Nordic Journal of Digital Literacy, 9*, 92–111.

Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction, 31*(2), 176–203. https://doi.org/10.1080/07370008.2013.769994

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12. https://doi.org/10.1037/h0080017

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/BF03192720

Wechsler, D. (2002). *WAIS – III. Escala de inteligencia para adultos de Wechsler* (3rd. ed.). Paidós.

Wiley, J., Jaeger, A. J., & Griffin, T. D. (2018). Effects of instructional conditions on comprehension from multiple sources in history and science. In *Handbook of multiple source use* (pp. 341–361). Routledge.

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

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162