Title Page (must include corresponding and co-author information)

Inside Document models:

The role of source attributes in integrating multiple text contents

Jean-François Rouet (1), Gaston Saux (2), Christine Ros (1)

Marc Stadtler (3), Nicolas Vibert (1), & M. Anne Britt (4)

(1) CNRS and University of Poitiers

(2) Pontifical Universidad Católica, Buenos Aires – CONICET

(3) University of Bochum

(4) Northern Illinois University

Contact author: jean-francois.rouet@univ-poitiers.fr

Author note:

Research presented in this paper was funded in part through a grant from The Region Nouvelle Aquitaine (grant CPER INSECT 2014-2020) and a Fulbright grant to the first author. The authors would like to extend their appreciation to Susan Goldman and two anonymous reviewers for their thoughtful and constructive comments on earlier drafts of this paper. Very special thanks to Susan Goldman for her detailed editing and her contribution to clarifying our terminology. The authors also thank Baptiste Berry, Lachezar Dimitrov, Mélanie Guérineau, Ioana Sandulescu, and Adonis Savarit for their assistance with data collection and analysis.

Inside Document models:

The role of source attributes in integrating multiple text contents

Abstract

Text comprehension involves the ability to understand how texts relate to the situation they describe and to each other (i.e., a Document model; Perfetti, Rouet, & Britt, 1999). Research into Document models has emphasized the role of information sources in structuring readers' mental models of situations. The present paper reviews research on source comprehension and examines new hypotheses regarding source encoding during text comprehension. Two experiments examined college students' evaluation and recognition of two embedded sources as a function of the consistency of their statements (Experiments 1 and 2) and the demands of the reading task (Experiment 2). Sources were introduced either with a knowledge feature (e.g., how the source knew what they knew) or a control feature. Readers were more likely to select the knowledgeable source as more competent and to justify their selection by referring to the knowledge feature (Experiment 1). Experiment 2 found that readers' preference for knowledgeable sources was stronger when the content statements were discrepant. Preference for knowledgeable sources was also stronger when the task focused on source knowledge, as compared to a control evaluation task or a baseline reading task. The source knowledge evaluation task distinctively enhanced participants' integration of the source designations with their respective statements, as compared to the control and baseline conditions. Results are discussed in light of current theories of single and multiple document comprehension.

Keywords: source feature, memory, evaluation, discrepancies, task

Most of what we know or believe about the world comes from what we hear or read from others (Wilson, 1983). When reading new information, readers seek to comprehend, but also to find out how good the information is. For instance, readers may ask whether the information is accurate, up to date, or unbiased. This is especially important when reading from multiple texts that may provide discrepant descriptions of a situation. One way to assess information quality is to check its source, that is, where the information comes from (Bråten, Strømsø, & Britt, 2009; Goldman, Britt, Brown, Cribb, George, et al., 2016; Rouet & Britt, 2014; Stadtler & Bromme, 2014).

Across the past three decades, research has attempted to provide theoretical accounts of the role of sources (i.e., where the information comes from: author, date, publication outlet and so forth; Perfetti, Rouet, & Britt, 1999) in text comprehension. However, there is still little consensus regarding when and how source details get integrated in readers' representation of a text. The studies presented in this paper aim to contribute to a general theory of when and how readers take information sources into account as they read short informational texts. In the first part of this paper, we provide an overview of multiple document comprehension theories. We focus on the issue of what exactly readers try to do when they evaluate the textual description of a source. In the second part of the paper, we report two recent experiments in which we have examined the interplay of task and text factors on readers' encoding of source features and source-content links when reading short news texts.

Comprehending texts: Representing who says what about a situation.

A core outcome of text comprehension is the construction of a memory representation of the situation described in the text, or situation model (Kintsch, 1998). In this perspective, a text is a coherent set of linguistic statements that describe a real or fictitious situation. In everyday reading, however, people often learn about situations through multiple pieces of text

that may come from diverse sources (e.g., different authors, dates, publication outlets and so forth). This is true for instance of scholars working with historical documents (Wineburg, 1991), business analysts reviewing articles from specialized magazines (Pirolli & Card, 1999), but also students learning science through inquiry tasks (Wiley et al., 2009), or lay people browsing the daily news (Braasch, Rouet, Vibert & Britt, 2012). In the present paper, we use the term *document* to refer to a piece of text (i.e., the discourse content) presented together with the description of where the text comes from, which we call its *source* (Rouet & Britt, 2014). We use the term *text* in a more general sense, to designate any cohesive piece of written discourse, with or without any information regarding sources. Although constructs such as text, source, and documents are complex and multifaceted (see Goldman & Brand-Gruwel, 2018, for a discussion), our definition supports a distinction between the *content* of a document (what the text says about the topic or situation) and its *source* (who wrote it, when, where and so forth).

A theoretical issue of interest is whether memory representations constructed from multiple documents differ from those that are based on a single text (e.g., a story with no source information). Early research into the comprehension of multiple documents has found that knowledgeable readers scrutinize the source of the information they read (Wineburg, 1991; Rouet, Favart, Britt & Perfetti, 1997). In doing so, readers generate attributions regarding the competence (i.e., whether a person or an organization has training, qualification or experience on a topic) and trustworthiness (i.e., whether the person or organization is neutral, benevolent, reliable with respect to the topic) of the source(s). These attributions contribute toward an assessment of the credibility of the discourse contents, which in turn may lead the reader to accept or dismiss the information. For instance, authors who appear to have little knowledge about a situation, or whose account of the situation seems tainted with bias (e.g., a conflict of interest) are less likely to be trusted than knowledgeable and/or benevolent authors (Bråten,

Strømsø & Britt, 2009; Bråten, Strømsø & Salmerón, 2010). In addition, skilled readers are more likely to cite sources that play a direct part in a situation (e.g., a character involved in a historical event) than those who are not involved (e.g., a historian writing about past events), and they remember those "embedded" sources better than sources that are physically or temporally distant from the situation they write about (de Pereyra, Britt, Braasch, & Rouet, 2014). Thus, the comprehension and evaluation of source information seems to play a significant part in understanding information from multiple documents.

Perfetti et al. (1999) provided an early description of the cognitive processes and representations involved in readers' comprehension of multiple documents. They claimed that under some conditions, readers of multiple documents build "Document models"[1] that include two components: a Situations model and an Intertext model. The Situations model (also referred to as *Integrated Mental Model* for non-narrative genres; Britt & Rouet, 2012) represents accounts of the situation as conveyed in the documents, including discrepant or conflicting assertions. The Intertext model represents the sources of information and how they relate to each other (e.g., author Y supports, corroborates or contradicts author X). Importantly, the Situations model and the Intertext model are integrated through source-content links, whereby some components of the Situations model may be "tagged" to their respective sources. Document models enable readers to understand complex and potentially conflicting situations by keeping critical portions of the model tied to their respective sources. The Document model framework was initially put forward to account for situations in which readers are comprehending multiple documents (with each document constituting a distinct artifact; Britt, Rouet, & Braasch, 2013). It extends, however, to the comprehension of a single text in which

[1] The phrase was initially used with the word document in the plural form, i.e. "Documents model". In this paper, we use the singular "Document model(s)", as the construct applies to the comprehension of single or multiple documents (de Pereyra et al., 2014). We capitalize the word document in the phrase "Document model" to differentiate the theoretical construct from the more general use of the word document.

multiple sources of information are embedded (such as the present academic paper; see also Rouet & Britt, 2014; Rouet, Britt & Potocki, 2019, for discussions).

The role of conflict in constructing Document models

When do readers construct Document models (with both sources and contents represented) as opposed to just merging the contents of multiple documents? Britt, Perfetti, Sandak and Rouet (1999) assumed that readers are more likely to "source-tag" specific claims if they contradict other claims, because source-tagging represents an effective mechanism to account for discrepant versions of a situation. By definition, readers cannot construct a single coherent mental model of a situation from documents that contradict each other. Instead, remembering that the discrepant claims come from distinct sources is a way to integrate the information into a Document model rather than a situation level. A study by Braasch et al. (2012) fully confirmed this prediction. They asked college students to read short news stories in which two characters issued either consistent or discrepant statements about a situation. Using eyetracking evidence, Braasch et al. found that readers spent more time re-inspecting textual segments that referenced source phrases (e.g., "the art critic claims that…") when the text included a discrepancy than when it did not. Participants were also more likely to recall the source of specific statements after reading discrepant stories. This discrepancy-induced source comprehension (or D-ISC) effect has been replicated and extended to younger readers, web-like documents, sources that are not embedded within the text, and other aspects of the source-content mental representation (e.g., Kammerer, Kalbfell, & Gerjets, 2016; Rouet, Le Bigot, De Pereyra, & Britt, 2016; Saux et al, 2017; Stang Lund, Bråten, Brante, & Strømsø, 2017; see also Bråten & Braasch, 2018; List, Du, Wang, & Lee, 2019, for recent discussions).

Source focusing may be triggered when the reader perceives that one source is more trustworthy than another. Kammerer et al. (2016), for example, manipulated the presence or

absence of contradictory assertions across two web documents. Each document included descriptions and logos of the sources, which indicated that one of them was more commercially biased than the other. Among other results, Kammerer et al. reported that the sample of college students who read the version with contradictions judged the commercially-biased page as less trustworthy and less convincing than those who read consistent information across documents. Likewise, Salmerón, Macedo-Rouet, and Rouet (2016) found that, when asked to evaluate multiple, conflicting messages in "Question & Answer" forum scenarios, students ranging from primary to higher education became more aware of and preferred recommendations from authors who were experts on the topic (e.g., a doctor) as opposed to authors whose competence was uncertain (e.g., a blog user under a pseudonym).

Readers' focusing on the source description when reading discrepant content has been interpreted as reflecting their attempt to develop their own interpretation of the situation (Britt et al., 1999). The Content-Source Integration Model (or CSI model, Stadtler & Bromme, 2014) explains how the source of a document may help readers draw conclusions about conflicting information. CSI proposes that, when reading contradicting information, a reader will first detect a conflict between document contents. Conflict detection likely depends on the coactivation in working memory of the discrepant propositions (Cook & O'Brien, 2014). Detection is considered an initial step and a condition for further elaborations to take place. One of these elaborations consists in "regulating" the conflict, or deciding on a way to handle the contradiction. According to the CSI model, there are several routes to regulating a conflict (Stadtler & Bromme, 2014). Readers may simply ignore the conflict (for instance by dismissing contradicting information, leaving one's mental representation free of unexplained discrepancies); they may attempt to reconcile the discrepant statements by inferentially generating an explanation for the conflict; or they may acknowledge the conflict as reflecting different points of view by integrating source features into the mental representation (Braasch

et al., 2012; Rouet et al., 2016). Conflict resolution represents a third and last type of reaction to conflicts. Here, readers can either rely on their own understanding of the subject, deciding by themselves what to believe, or they can use instead source competence to determine whom to trust. Thus, according to the CSI model, source features may play a central role in both the regulation and resolution of conflicts, at least for some strategies.

An important consequence is that, to the extent that readers use a source-based strategy to regulate and/or resolve the conflict, they should focus on the source features that are relevant to competence evaluation, as compared to other features of the same sources. However, given that other, simpler strategies are available (e.g., just ignore conflicting information), the conditions that promote source-content integration need to be specified.

The role of reading goals in source-content integration

Research has found that the reading task, usually operationalized as a situational instruction or prompt, will induce the reader to establish a reading goal, which will in turn guide processing (e.g., McCrudden, Magliano & Schraw, 2010; Rapp & McCrudden, 2018; van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011). As a consequence, readers' processing and encoding of the text depends on the general context and the precise task under which reading takes place (e.g., Cerdán & Vidal-Abarca, 2008; Kaakinen & Hyönä, 2014; McCrudden, Magliano, & Schraw, 2010; Schraw, Wade, & Kardash, 1993).

Britt, Rouet and Durik's (2018) RESOLV model assumes that readers represent the reading context and tasks in order to perform decisions regarding what to read and how to read it. They proposed that the task will shape not just the reader’s decision to attend to source features, but also their decision to elaborate on their knowledge of the sources in some way (e.g., is the author of this assertion knowledgeable? Do they have a conflict of interest?). This assumption is important because it distinguishes two ways in which the task could influence

sourcing: by indicating *when* to attend to the sources, and *how* to use the available information about the sources. Indeed, empirical evidence suggests that readers' consideration of source information depends on their generation of adequate reading goals. For example, prompts that instruct students to produce arguments after reading a series of documents on a controversial topic (including forming an opinion about the controversy) promote the inclusion of source information in the written product, as compared to prompts that instruct them to describe or summarize the same contents (see Wiley, Jaeger, & Griffin, 2018, for a review). Tasks that explicitly instruct readers to evaluate the information and how to do it seem to have even stronger effects (e.g., Britt & Angliskas, 2002; de Pereyra et al., 2014; Macedo-Rouet et al., 2019; McGrew, Smith, Breakstone, Ortega, & Wineburg, 2019; Pérez et al., 2018; Sanchez, Wiley, & Goldman, 2006; Stadtler & Bromme, 2007; Stadtler, Paul, Globoschütz, & Bromme, 2015). These studies have demonstrated that instructions that warn of risks of misinformation in digital reading environments, or encourage the examination of source features as a reliable filter, or explicitly recommend checking for sources' intentions and abilities on every accessed document, are successful in enhancing readers' online and offline integration and evaluation of this information.

Regarding the critical dimensions or source features that readers should use for their competence judgments, research has typically leaned toward a normative approach. That is, readers have been directly presented with descriptions that explicitly convey a sense of competence (e.g., a medical doctor vs. a layperson talking about a medical issue). There is less evidence that typical readers can draw elaborations from descriptive features of sources, such as their position, or activity with respect to the situation (e.g., a person who has access to evidence vs. a person who has not). Among the few studies relevant to this issue, Gottschling, Kammerer, and Gerjets (2019) reported that readers' consideration for the source information was increased, as evidenced in their participants' eye movements and responses to

comprehension questions, when readers were informed about a potential difference in trustworthiness. De Pereyra et al. (2014) also found that authors who participate in a situation are more easily remembered than those who do not. They also found a large increase in readers' memory for who said what when the participants were tasked to evaluate the sources upon reading the documents. Thus, a task prompting readers to evaluate and compare the competence of multiple authors is likely to induce readers to focus on source features. Readers would then selectively encode source features that are instrumental to deciding which source is more reliable, together with the content issued by that source.

The present research

The present research examined the assumption that some features presented as part of the description of a source are more instrumental than others in helping readers decide which source is the most reliable, when trying to comprehend a document involving conflicting content. We further assumed that readers selectively encode those instrumental features together with the contents, so as to form an integrated mental model of the situation (Britt & Rouet, 2012; Stadtler & Bromme, 2014).

A study by Saux et al. (2018) used short stories in which two characters (i.e., embedded sources) agreed or disagreed about some event or situation. The characters were introduced with a description of either how they had access to knowledge (e.g., [the policeman] "who inspected the facility") or what they looked like (e.g., [the policeman] “who held a tissue on his face”). Readers' recall of the sources' features was enhanced when the two sources disagreed, and also when the task prompted participants to evaluate the sources’ competence (Experiment 1). Unexpectedly, the competence evaluation task increased readers' memory for both types of features, including the physical description, although it was not relevant to interpreting the disagreement. However, in this study the two sources in each story came with a similar type of

features (i.e., either a physical description or a description of their access to knowledge). Therefore, the evaluation of the source features did not permit readers to clearly decide which assertion was more reliable.

The present studies were designed to extend those initial findings by examining readers' evaluation (Experiments 1 and 2) and memory (Experiment 2) of two sources issuing statements on the same situation as a function of the discrepancy between the statements, how the sources are described (or their features), and the type of reading task. In contrast with prior research, the sources came with contrasting features. One of the sources was presented with a description of how s/he could have gained knowledge of the situation (i.e., a knowledge description). The other source was presented with a description of what s/he looked like (i.e., a physical description). The main assumption was that contrasting features would let participants resolve the contradiction by assigning one of the sources a higher level of competence (Stadtler & Bromme, 2014). We predicted that both discrepant statements and a competence-evaluation task would increase readers' memory for the competent source, her/his description, and the link with the content statement. These situations were compared to conditions in which two sources made consistent statements, and conditions in which participants were asked to read to comprehend (a baseline task), or to evaluate the sources with a different criterion (i.e., assessing which source seemed older).

Experiment 1

Experiment 1 aimed at verifying whether college-level readers identify and use knowledge descriptions when assessing source competence. Although a description such as "the [source] who inspected the facility" implicitly suggests that the source has some knowledge of the situation, the degree to which these descriptions influence readers' attribution of

competence, compared to, for instance, the designation of the source (e.g., "the policeman"), or the plausibility of their assertions, is an empirical question. Consistent with prior studies (e.g., Braasch et al., 2012) we used short stories in which two sources were designated by their occupations (e.g., "the policeman", "the fireman"). The occupations were relevant to the situation (e.g., a fire in a warehouse). However, an additional description or "source feature" was provided for each source. For one of the sources, the feature described how the source had gained knowledge on the situation (e.g., "the policeman/fireman who examined the installation"); for the other source, the feature described a physical detail or accessory (e.g., "the policeman/fireman who held a tissue on his face"). We tested the prediction that, in response to the evaluation prompt, participants would select the source with the knowledge feature more often than the source with the physical feature, especially when statements were discrepant rather than consistent. In addition, we expected the participants to quote the knowledge feature as a main justification for selecting the most knowledgeable source.

Method

Participants

Thirty-six first- and second-year students at a mid-size French university participated for course credit. Data from three participants had to be removed either because they were non-native French speakers (n=1) or because they failed to complete the procedure (n=2). The resulting sample (n=33) included twenty-six female participants (79%). The participants' average age was 19.5 years (range: 17-25). All participants signed an informed consent and were debriefed after the activity.

Materials and design

The main experimental materials comprised sixteen short, fictitious newsflash reports adapted from Braasch et al. (2012). Each story involved two characters issuing statements about

a situation (e.g., a fire in a warehouse). An example of a story with the manipulations is shown in Table 1. The characters were labelled according to their profession or occupation (e.g., "the policeman" and “the fireman”, see Table 1). Importantly, character labels were chosen so that both characters were plausible with respect to the topic or situation. In addition, each character was associated with an additional description that was manipulated to be a “physical” or a “knowledge” feature. Physical features described what the character looked like, such as body details, clothing or posture. Knowledge features described how the character gained knowledge about the situation, such as her/his position, access to evidence, or participation in an investigation (see Table 1, features 1 and 2). The statement issued by the first source was manipulated so that it could be consistent or discrepant with the statement issued by the second source (see Table 1, statements 1 and 2). Additionally, source features were switched across source designations to create two new versions, resulting in four different versions of each story. Switching attributes permitted the measurement of the effect of the type of feature regardless of the source designations.

(Table 1 about here)

The materials were assembled in booklets containing 8 consistent and 8 discrepant stories. The critical stories were mixed up with 6 filler stories that included no features for the sources and a slightly different structure. The presentation order of the stories was randomized across participants. In addition, two practice stories were placed at the beginning of the booklet in order to demonstrate the task. The first practice story involved consistent accounts of a situation, whereas the second story involved contradictory accounts.

Each story was presented together with two multiple-choice questions. The first question (Q1) asked about the critical fact in the story contents. The purpose of the first question was to make sure that participants were actually perceiving the contradiction between the statements, when present (Stadtler & Bromme, 2014). In the example in Table 1, Q1 was "What

is the cause of the fire?" Participants had to choose among three options, in this case: “A breakdown of the electric circuit / A malevolent individual / It is uncertain". We assumed that the selection of the “uncertain” response would be higher when the stories were discrepant. In contrast, we expected participants to select the definite response (i.e., the fact supported by both sources) when the stories were consistent.

The second question (Q2) asked participants to choose which of the sources had the most knowledge about the situation. Three answer options were provided, including the two source designations and "uncertain". In addition, the participants were given a blank space to write a short justification of their answer to Q2.

In addition, a standardized reading fluency test (Lefavrais, 1968) was included in order to make sure that participants were in the normal range. The test consists in a list of 486 words presented on three A4 pages. Half of the words are names of animal species (e.g., a rat, a wolf, a fly). The task is to read all the words one by one and to highlight all the animal names. The reading fluency score is the total number of animal names highlighted in exactly three minutes, minus the number of words mistakenly highlighted.

Procedure

Participants were run in groups of 2 to 8 in a quiet meeting room over the course of a single, 50-minute session. After signing a consent form, the participants received a story booklet and were given instructions as follows:

"In this experiment, we ask you to imagine that you are an intern in a news agency. You are asked to check the readability[2] of the news messages that will be released on the agency's Web site. You will read a series of short texts that cover the news in the scientific, social or everyday life events. Please read each text carefully and answer the two questions that are displayed below the text. The first question is factual, you have to answer as a function of what

[2] The instructions are adapted from French. The original word used in the instructions was "lisibilité", which corresponds to a broad sense of readability that includes the coherence and cohesiveness of a text.

the text says. The second question calls upon your personal judgment. You are asked to tell which of the characters has the most knowledge of the topic. You have to decide for yourself based on the information provided in the text. Then you have to justify your choice in a single short sentence or even a simple phrase. We will go through two practice trials for you to understand the task."

The experimenter then read out loud the first practice story and the first question (i.e., a question about the critical fact in the story). As stated above, the experimenter explained that the participants were to answer according to the information stated in the text and invited them to select an answer. Then the experimenter discussed the three options as well as the expected answer (in this case, a definite answer since the two sources provided consistent accounts). The experimenter then read out loud the second question (i.e., a question about the source characters' knowledge). The experimenter reiterated that, unlike the first question, the answer was a matter of personal interpretation, but that the participants were to use the information provided in the text to form an opinion. The experimenter then let the participants select an answer and write a short justification for their selection. No feedback was provided.

The same procedure was repeated for the second example (i.e., a discrepant story). The experimenter explained that "uncertain" was the expected response to the first question because the two sources provided conflicting accounts of the situation. The experimenter emphasized that "uncertain" was the expected answer regardless of the participants" own opinion or interpretation of the situation. The experimenter then reiterated that the answer to the second question (i.e., which source has the most knowledge) was a matter of personal interpretation. Then the experimenter let the participants answer the question and write their justification. The experimenter invited clarification questions about the procedure and informed the participants that they had about 40 minutes to go through the rest of the stories.

The participants then went through the stories at their own pace. The experimenter indicated when half of the allotted time was elapsed and when only 5 minutes were left. However, the experimenter allowed a few extra minutes in the rare cases where participants were not quite done within 40 minutes. After all participants had finished reading the stories and answering the questions, the experimenter collected the booklets.

The participants then took the standardized reading test. Finally, the experimenter thanked and debriefed the participants.

Results

All participants' reading fluency scores were in the normal range for undergraduates (M = 106, SD = 16). As regards performance on the main reading task, the data were analyzed using IBM SPSS 24.0 software. To examine the distribution of the responses to Q1 and Q2 as a function of story consistency, we estimated generalized linear mixed models (GLMM). Because the responses were categories, the model type was set to a binomial logistic regression (which assumes a binominal distribution and a generalized logit relationship with the linear model). In all cases, story consistency (consistent, discrepant) was entered as the fixed factor and random intercepts for participants and texts were specified. Effect sizes for significant effects are reported in terms of odds ratio (OR) and their confidence intervals, calculated by exponentiating the fixed coefficients.

Comprehension of the situation (Q1)

Participants provided an answer to Q1 in 98% of the trials (only one participant left one trial blank). The percentage of correct answers was 86.4% for consistent stories (SD = 11.4) and 76.5% for discrepant stories (SD = 32.7). Thus, participants correctly understood the stories and/or noted the discrepancy in a vast majority of cases. The logistic mixed model showed that story consistency significantly predicted response accuracy $F(1,526) = 9.26$, $p = .002$. The odds

of selecting the expected answer when reading consistent stories (i.e., the fact supported by both sources) was about twice the odds of selecting the expected answer when reading discrepant stories (i.e., “it is uncertain”), *OR* = 2.11, *CI*.95= 1.30, 3.41. In other words, participants provided the expected answer more frequently in the consistent than in the discrepant stories. In some cases, participants selected the statement they agreed with instead of acknowledging the discrepancy (on readers' tendency to resolve discrepancies by siding with one interpretation, see Rouet et al., 2016).

Source knowledge evaluation (Q2)

Table 2 presents the respective frequency of each type of response to the source knowledge evaluation question as a function of story consistency.

(Table 2 about here)

The participants selected the source with the knowledge description in 57.4% of the cases overall. The participants found that which source had the most knowledge was uncertain in 26.9% of the cases, and they selected the source with the physical description in only 15.7% of the cases. As regards the manipulation of story consistency, the selection of the source with a knowledge description did not significantly differ across versions (consistent: 55.7%; discrepant: 59.1%; $F(1,526) = 0.60$, $p = .44$). However, the percentage of "uncertain" responses was larger in consistent than in discrepant stories (31.8% vs. 22.0%, respectively, $F(1,526) = 20.16$, $p = .008$). The odds of selecting "uncertain" over the other two options were higher in the consistent than in the discrepant version, *OR* = 1.78, *CI*.95= 1.17, 2.73.

Participants' preference for the source with a knowledge feature was confirmed by the content analysis of their justifications. Participants' justifications for their selection of one of the two sources (i.e., responses 1 or 2, 72.6% of the total responses to Q2) were categorized as (a) based on the source feature, either alone or in combination with other criteria, such as "the policeman *examined the facility*, but they do not say how the fireman came to his conclusion"

(our emphasis on the source feature); (b) based on the designation of the source, such as "'the fireman has experience with fires, that's their job, their can tell whether it is accidental or criminal"; or (c) based on another characteristic of the source (most often a characteristic too vague to be assigned to either the occupation or the feature, e.g. "s/he has expertise"). Based on prior research using a similar methodology (Rouet et al., 2016), two of the co-authors scored 38 of the justifications (10%) independently, with a very high level of agreement (Cohen's k = .83). Disagreements were resolved through discussion. One of the scorers scored the rest of the justifications. Fifty eight percent of the justifications included a reference to a source feature; 19% were solely based on the occupation; and 18% were too vague to be assigned to a category. Finally, 5% of the responses did not have a justification. A further examination of the justifications based on source features showed that they only involved *knowledge* features. In other words, the physical features were never used to support source knowledge attribution.

Discussion

The goal of Experiment 1 was to examine the impact of source features on participants' evaluation of the competence of the source in stories involving consistent or discrepant statements. More precisely, we wanted to find out if a description of how the source gained knowledge about the situation would make the source seem more knowledgeable than a source described by their physical aspect. Upon reading, participants were asked to answer a comprehension question about the critical fact in the story (Q1), and to decide which of the two sources had more knowledge about the situation (Q2). Responses to Q1 confirmed that the stories were easy to understand and that the discrepancy was noted in the majority of cases, although students sometimes failed to acknowledge that the situation was uncertain. Moreover, the source bearing the knowledge feature was most often selected as the most knowledgeable (Q2), although the participants refrained from making a definite attribution in more than a third

of the trials. Most of the participants' justifications mentioned directly or implicitly the knowledge description as a basis for their decision. Our prediction that this trend would be stronger when the sources' statements were discrepant did not reach statistical significance. It is possible that the evaluation task increased readers' attention to the source features even for consistent stories, thus reducing the effect of discrepancy in promoting source-content integration (Stadtler & Bromme, 2014).

Participants' failure to acknowledge the situation as uncertain was not unexpected. Prior research (e.g., Hakala & O'Brien, 1995; Rouet et al., 2016) has found that readers sometimes try to resolve inconsistencies by distorting the facts or generating elaborative inferences. Participants may also have decided to side up with the source they found the most knowledgeable. If so, participants should show better memory for the source with the knowledge feature, as a consequence of using this piece of information to resolve the discrepancy (Stadtler & Bromme, 2014). This prediction is further explored in Experiment 2.

In sum, participants' responses to the comprehension question (Q1), and the source knowledge evaluation question (Q2) strongly suggest that they identified and took that information into account in a majority of the trials. It is possible, though, that the explicit instruction to evaluate the sources' knowledgeability overshadowed the influence of story discrepancy on their attention to sources (Braasch et al., 2012). The non-significant trend for discrepancy to increase the participants' preference for the source with the knowledge feature warrants a replication with a larger sample. Furthermore, as suggested above, if knowledge features are used at the time of reading in order to resolve the conflict, they should be more likely to be remembered later. Such a memory effect would add support to the view that readers of multiple and discrepant accounts integrate source and content information into a Document model (Perfetti, Rouet, & Britt, 1999).

Experiment 2

Experiment 2 was aimed at replicating and extending our findings concerning readers' use of source descriptions to comprehend discrepant documents. Similar to Experiment 1, our first hypothesis was that, when two sources make discrepant assertions, readers would focus on the source features that allow them to decide whom to believe, as compared to when the sources agree with each other (Stadtler & Bromme, 2014). Consequently, when asked which source is most knowledgeable, participants would select the source with the knowledge feature more often than the source with the physical feature. Additionally, we hypothesized that the focusing on source features at the time of reading would impact readers' memory for the document. Consequently, participants in Experiment 2 also took a recognition test in which they were prompted to recognize the source designations and features, the statements, and the links between these components.

To assess the impact of the task setting on readers' identification and memory for source features, Experiment 2 included three different task conditions. Participants in all three conditions read short stories and immediately answered a question about the critical fact in the story. This was done to ensure that participants would read each story attentively and detect the discrepancy when it was present. In addition, participants in the "knowledge evaluation " condition were asked which source had the most knowledge of the situation. This condition was similar to Experiment 1 except that participants were not asked to justify their selection. Participants in the "age evaluation" condition were asked which source seemed older. These two tasks required readers to focus on the source characters, but differed in the importance of knowledgeability. We also included a "baseline condition" in which participants only answered the question about the critical fact.

Based on the RESOLV model (Britt et al., 2018), our prediction was that participants would select the source with a knowledge feature more often under the competence evaluation task than under the age evaluation task. We also predicted that both source evaluation conditions would increase source memory, as compared to the baseline condition. More specifically, we expected that readers in the evaluation conditions would better recognize the components of the source descriptions, irrespective of the specific evaluation task (i.e., a simple task relevance effect; Rapp & McCrudden, 2018). Finally, we reasoned that the knowledge evaluation task would prompt readers to resolve the conflict by assessing the competence of the sources and by linking the sources to their respective statements (i.e., forming a Document model). Therefore, we expected the knowledge evaluation condition to further increase readers' memory for the links between the source designations and their respective features, on the one hand (e.g., "it was the policeman who inspected the facility"), and between the source designations and their respective statements (e.g., "it was the policeman who claimed that the fire was criminal").

Method

Participants

One-hundred and twelve first- and second-year undergraduate students from the same university as in Experiment 1 participated voluntarily for course credit (81% female, Age M = 20, SD = 1.8). All participants signed an informed consent and were debriefed after the activity. None of them participated in Experiment 1.

Materials and design

Materials included the same 16 critical texts as in Experiment 1, with a within-subject manipulation of story consistency (8 consistent and 8 discrepant stories per participant; see example in Table 1). In addition, participants were randomly assigned to one of three task conditions. In the knowledge evaluation condition (n=39), participants were asked to read each

story and to answer two questions. Question 1 asked about the key fact in the story, whereas Question 2 asked which of the two source characters was the most knowledgeable. In the age evaluation condition (n=41), Question 2 was replaced with a question asking which of the source characters seemed older. Finally, in the base-line condition (n=32) only Question 1 was asked (no explicit source evaluation). The questions were presented immediately below each story, allowing participants to re-read before selecting an answer.

Because previous efforts using a cued source recall task generated a high level of unattempted trials (e.g., Saux et al., 2018), in the present experiment we probed readers' memory using a recognition task. The recognition task was created by displaying the sources, the features, and the statements from each critical story among a set of distractors (see Figure 1). The task consisted of (a) recognizing the two characters among a set of four (including the two actual characters plus two plausible distractors; see Figure 1, central column); (b) linking each character with its respective feature among a set of four (including the two target features plus two distractors; Figure 1, left), and (c) linking each character to their respective content statement among a set of four (including the two actual statements, the alternate statement corresponding to the alternate version of the story, and a distractor; Figure 1, right).

(Figure 1 about here)

Finally, two intermediate tasks were used between the reading phase and the recognition phase, so as to decrease the impact of short-term memory on recognition. One of the tasks was a standardized reading fluency test (Lefavrais, 1967), which was used to verify that the participants had a comparable reading level across experimental conditions. The other was an unrelated comprehension task meant to divert readers' attention from the stories they had read. Completion of the intermediate tasks took approximately 15 minutes.

Procedure

The students participated in a single session of approximately one hour. They were run in groups of 2 to 8 in a quiet meeting room, by one or two experimenters from a group of four. They were asked to sign a consent form and were explained that they were to participate in a study about the comprehension of news stories. The participants were given the story comprehension task according to the same procedure as in Experiment 1, except that participants were given only 25 minutes (instead of 40) to complete the whole set of stories, because they did not have to provide written justifications. The experimenter walked the participants through the two practice trials, invited clarification questions and invited the participants to go through the rest of the stories at their own pace. The experimenter indicated when half of the allotted time was elapsed and when only 3 minutes were left. However, the experimenter allowed a few extra minutes in the rare cases where participants were not quite done within 25 minutes. After all participants had finished reading the stories and answering the questions (i.e., the task resolution phase), the experimenter collected the booklets.

After a short break, the participants received two intermediate tasks, for a total duration of about 15 minutes. Finally, they received the recognition booklet. The experimenter showed an example, which corresponded to the first practice story. The experimenter explained how to complete the item, and invited clarification questions. S/he emphasized that the participants were expected to fill in as many items as possible, but that they should not circle any item or draw any link randomly. A time limit of 10 minutes was loosely set for task completion. Then the participants went through the 16 critical stories one at a time at their own pace. After they had finished the source recognition task (i.e., the memory phase), the participants were thanked and debriefed.

Results

An ANOVA on the reading fluency test found no difference between the three groups, $F(2,109) = 1.92$, $p = .16$. Participants' performance was in the expected range for undergraduate students

(Lefavrais, 1967). As in Experiment 1, the main analyses consisted of estimating GLMM with random intercepts for participants and texts using IBM SPSS software (v24). Responses to the content comprehension and source evaluation questions (Q1 and Q2, respectively) were analyzed with logistic mixed models. The models included story consistency (consistent, discrepant), type of task (competence evaluation, age evaluation, baseline condition), and their interaction as fixed factors. In addition to story consistency and type of task, the models for recognition accuracy also included type of feature (knowledge description, physical description), all possible two-way interactions and the three-way interaction (story consistency X type of task X type of feature) as fixed factors. As in Experiment 1, effect size is reported in terms of odds ratio (OR) calculated by exponentiating the fixed coefficients when estimating logistic models, and in terms of fixed coefficients when estimating linear models.

Task resolution phase

Comprehension of the situation (Q1). Table 3 presents the descriptive statistics for the responses to Q1 as a function of story consistency and task type. The logistic mixed model showed that story consistency significantly predicted the number of correct (i.e. expected) answers to the factual question, $F(1,1786) = 16.24$, $p < .001$. Similar to Experiment 1, the percentage of correct answers was slightly higher in the consistent (88.9% on average across all task conditions) than in the discrepant condition (82.1%; $OR = 1.58$, $CI_{.95}$= [0.90, 2.75]). Interestingly, this effect was qualified by an interaction with the type of task, $F(2,1786) = 11.41$, $p < .001$. Contrary to the baseline condition, in the other two conditions the odds of selecting the correct answer was lower for discrepant stories (baseline-discrepant vs. competence evaluation-discrepant: OR = 0.21, CI.95= [0.10, 0.44]; baseline-discrepant vs. age evaluation-discrepant: OR = 0.19, CI.95= [0.09, 0.41]. Overall, however, the percentage of correct answers was very high (85.5%), suggesting that both consistent and discrepant stories were easy to understand.

Source evaluation (Q2). Descriptive statistics for the responses to Q2 as a function of the manipulations are presented in Table 4 (the baseline group did not receive Q2). As in Experiment 1, most of the participants either selected the source with the knowledge attribute as having the most knowledge, or chose not to decide ("uncertain"). The mean percentage of selection of the "uncertain" option was rather stable across conditions (between 41.2 and 48.4%). Taking this into consideration, and to simplify the interpretation of the categorical outcomes, the inferential analyses were run on the selection of the "source with knowledge feature" and the "source with physical description" items only.

The logistic mixed model revealed that task type, $F(1,64) = 21.42$, $p < .001$, and story consistency, $F(1,705) = 4.89$, $p = .027$, predicted the selection of the source options. As expected, the odds of selecting the source with a knowledge description were twice higher in the knowledge evaluation task than in the age evaluation task ($OR = 2.23$, $CI_{.95} = 1.36, 3.65$). In the latter condition, the two sources were about as likely to be selected regardless of the type of features they came with. Selection of the source with the knowledge attribute was also more likely overall when reading discrepant than consistent stories ($OR = 1.55$, $CI_{.95} = 1.01, 2.40$). This was expected in the knowledge evaluation task condition, since discrepancy is expected to push readers to scrutinize source competence more closely, but not in the age evaluation task, in which the knowledge feature is irrelevant. However, the Statement Consistency X Type of Task interaction failed to reach significance, $F(1,705) = 0.22$, $p = .64$.

In total, as in Experiment 1 participants in the knowledge evaluation task often - though far from systematically - noted the knowledge feature and used it to designate the source with the most knowledge of the situation. In both conditions however, participants were often reluctant to designate one source as more knowledgeable than the other based solely on the knowledge feature.

Recognition phase

Data analysis focused on participants' construction of an integrated Document model. Therefore, we considered readers' ability to correctly link different pieces of information as the main performance indicator (e.g., Britt et al., 1999). However, extending previous research (e.g., Braasch et al., 2012; Rouet et al., 2016) two different link indexes were calculated. On one hand, the source-feature integration index represented the recognition accuracy for the link between the source designation and her/his feature (i.e., the lines between the center and left columns in Figure 1). On the other hand, the source-content integration index represented the recognition accuracy for the link between the source designation and the content stated by that source (or Source-Content link, i.e., Figure 1, line connecting the center and right columns).

Analyses of recognition accuracy were conducted on the original binary values obtained from the responses. However, for clarity purposes, we report the values as percentages.

(Figure 2 about here)

Source-feature integration. Figure 2a shows the effect of task condition and discrepancy on the recognition of the link between source designations and features. On average, performance was 54.7% (chance level: 6.3%). As predicted however, recognition accuracy was higher in the two source evaluation conditions (62.9% and 59.1% in the knowledge evaluation and age evaluation conditions, respectively), than in the baseline condition (42.1%; $F(2, 102) = 9.41$, $p < .001$). The competence and age evaluation groups outperformed the baseline group [competence evaluation task vs. baseline: *OR* = 2.28, *CI*.95= 1.28, 4.06; age evaluation task vs. baseline: *OR* = 1.65, *CI*.95= 0.93, 2.92]. The odds of recognizing the source-feature link in the two evaluation groups were similar, *OR* = 0.73, *CI*.95= 0.42, 1.24. Furthermore, story consistency tended to, but it did not significantly affect source-feature integration, $F(1, 3572) = 3.07$, $p = .080$, and it did not interact with task conditions, $F(2, 3572) = 0.05$, $p = .95$. In short, source-feature integration simply increased when the instructions made source details relevant

to the reading task. Type of feature and its interactions with the task and story consistency did not yield significant effects, $p > .1$.

Source-content integration. Figure 2b shows the effect of task condition and discrepancy on the recognition of source-content links. Source-content link recognition was 50.8% on average. There was a significant effect of the task condition, $F(2, 101) = 11.32$, $p < .001$. Recognition was highest in the knowledge evaluation condition (59.4%), followed by the age evaluation condition (52.1%) and the baseline condition (41.1%). Both evaluation groups showed better recognition than the baseline group (age evaluation: $OR = 1.72$, $CI.95 = 1.18, 2.51$; knowledge evaluation: $OR = 2.20$, $CI.95 = 1.50, 3.24$). In addition, the knowledge evaluation group outperformed the age evaluation group, $OR = 1.52$, $CI.95 = 1.07, 2.16$. There was also a significant effect of story discrepancy: Source links were better recognized in discrepant as opposed to consistent stories, $F(1, 3578) = 8.14$, $p = .004$. There was no interaction between task condition and story discrepancy, $F(2, 3578) = 0.53$, $p = .59$. Thus, both the knowledge evaluation task and the presence of a discrepancy specifically promoted the integration of source designations with what was said and how the sources were described. Again, the effects of feature type and its interactions with the task and story consistency failed to reach significance, $p > .59$.

Recognition accuracy as a function of the response to the evaluation task (Q2)

Given the high percentage of trials in which participants could not decide which source was the most knowledgeable or older, we further explored the potential relationship between participants' source selection and their performance in the subsequent recognition task. These analyses excluded the baseline group who did not perform any source evaluation task during reading. The sub-sample was thus composed of 80 participants. A global link recognition score was calculated for each trial (i.e., the sum of both link integration indexes per text, max = 4). Trials in which the participants selected the source with the knowledge attribute in the

evaluation task had the highest recognition scores (Accuracy %: $M = 66.6$, $SD = 33.5$), followed by the cases in which the source with the physical attribute was selected ($M = 59.8$, $SD = 33.3$), followed by the cases in which the “uncertain" option was selected ($M = 51.2$, $SD = 35.5$). A mixed-model analysis revealed a significant effect of the response to the evaluation on the total recognition score, $F(2, 1232) = 18.89$, $p < .001$. Recognition was better when either of the two sources were selected by comparison with the "uncertain" option (adj. Bonferroni: $p < .002$). The estimated coefficients indicated that, compared to the “uncertain” option, the scores increased by.37 point when selecting the source with the physical attribute (SE = 0.10, CI.95= 0.17, 0.57) and by .5 point when selecting the source with the knowledge attribute (SE = 0.09, CI.95= 0.37, 0.71). No significant differences were observed between the two source options; $p = .093$).

Discussion

Experiment 2 had two goals. The first goal was to replicate and extend the findings from Experiment 1 regarding the effect of a knowledge attribute on readers' assessment of source competence; the second goal was to examine the effects of task conditions and story consistency on readers' integration of source and content information.

As in Experiment 1, participants understood the story and noted the discrepancy when present. Also consistent with Experiment 1, participants in the source evaluation condition more often selected the source with the knowledge feature as more knowledgeable. In addition, preference for the knowledgeable source was significantly increased when participants read discrepant as opposed to consistent stories. This confirms the trend observed in Experiment 1, and adds support to the role of discrepancy in readers' attention to source features (e.g., Braasch et al., 2012; Gottschling et al., 2019; Kammerer et al., 2016; Saux et al., 2018; Stadtler & Bromme, 2014). Upon noticing a discrepancy that cannot be interpreted based on prior

knowledge, readers scrutinize the sources in search for credibility cues (e.g., information about what the source knows and how they know what they know).

Participants' preference for the more knowledgeable source extended, albeit modestly, to a condition in which they were asked which of the sources seemed to be older. This unpredicted finding suggests that knowledge features may be interpreted as a cue to experience and maturity, which in turn might make the source look older. At this point, it is worth keeping in mind that the source features were assigned to either one of the two source designations across versions of the stories, therefore the effect of the source feature is independent of any other textual influence. Nevertheless, knowledgeable sources were selected more often in the knowledge evaluation condition, suggesting that the task manipulation was effective in indicating how source information was to be used during reading (Rouet et al., 2017). More generally, the data suggest that some source cues (vs. the source description as a whole) may become more or less relevant depending on the kind of goal readers pursue.

We also hypothesized that our manipulations would impact participants' memory representation of the documents. More specifically, we predicted that both evaluation tasks would enhance readers' memory for source details (source-feature integration), whereas the knowledge evaluation task would have a more specific impact on source-content links, especially when the story was discrepant. Indeed, both source-focusing tasks (i.e., knowledge and age evaluations) enhanced source node recognition as compared to the baseline group. Additional verifications found that the baseline group's memory for what was said in the stories was just as accurate as the other two groups (i.e., about 80%). This rules out the possibility that the baseline group would have simply paid less attention to the documents at the time of reading, since they only had to answer one question instead of two. The finding that both source-focusing instructions improved memory for sources is congruent with other evidence of how prompts and tasks affect reading (McCrudden et al., 2010; de Pereyra et al., 2014).

Experiment 2 also found that the competence evaluation task distinctively increased readers' recognition of the source-content links. This adds an important specification to previous studies of multiple source comprehension. Unlike the other tasks, the source knowledge evaluation task uniquely required using source descriptions to assess the reliability of the contents. Consistent with the later point, source-content recognition was also higher for discrepant than from consistent stories, suggesting that a process of resolution whereby readers integrated sources and contents into a Document model (Perfetti et al., 1999) was prompted by the presence of a conflict (Stadtler & Bromme, 2014). Overall, this last set of results aligns well with our initial predictions regarding recognition performance, and they support the idea that readers will selectively elaborate on the sources when they need to form an integrated representation of sources and statements, known as an intertext model (Britt et al., 1999).

Finally, the set of analyses linking participants' responses to the evaluation question with their recognition performance revealed that the readers who selected a source at the evaluation phase (instead of "uncertain") showed higher accuracy during recognition. A tentative interpretation, based on the CSI model (Stadtler & Bromme, 2014), is that the more readers reached a resolution of the conflict by siding with one of the sources, the stronger its influence on a later, source-related memory task. However, this last claim remains fragile as participants' responses to the evaluation question cannot be interpreted unambiguously. Selecting a source in response to a question does not necessarily mean that participants sided up with the source, but rather that they determined who seemed the most knowledgeable / oldest, depending on the condition. In other words, Experiment 2 was not designed to study readers' memory for their own conclusion from the story, though that issue opens interesting research perspectives.

General discussion and conclusions

This paper examined the view that text comprehension involves building a coherent representation not just of what was said in the document, but also of who said what and why. When a document includes gaps, inconsistencies or conflicts, readers can integrate contents with sources as a means to organize their incoherent situation model (Perfetti et al., 1999). This, however, involves an effortful process, which is more likely to take place when the task context makes it highly relevant. In the present study, we maximized the contextual relevance of evaluating the sources (i.e., characters issuing information) by asking readers explicit questions about their competence (Experiments 1 and 2) or their apparent age (Experiment 2). The experiments found that college-level readers can interpret source attributes as indicating whether a source is knowledgeable with respect to the situation. The experiments also added evidence that the kind of processing readers engage in during the task has an impact on their memory for the document contents. Consistent with Stadtler and Bromme's (2014) Content-Source Integration model (CSI), readers' recognition of source details and source-content links is the highest when they engage in an evaluative process, weighing the sources in order to determine which one is more believable. As proposed in the CSI model, readers might additionally want to decide about the validity of the competing claims. Although we did not directly examine which position readers sided with, it is likely that they would use their attributions of source competence to substantiate their validity judgments (Stadtler & Bromme, 2014). In addition, the task effects observed in Experiment 2 fit the prediction that readers selectively devote attentional resources to text information as a function of their perception of the demands of the reading task and context, a core prediction of the RESOLV model (Britt et al., 2018). More generally, these experiments support the view that in certain circumstances readers construct Document models integrating source and content information in their mental model of the document (Perfetti, Rouet & Britt, 1999).

From a practical point of view, our study also suggests that prompting sourcing effectively supports source evaluation. Whereas previous research has mainly compared the presence of sourcing prompts with no prompts (e.g., Kammerer et al., 2016) or with content-oriented prompts (e.g., Stadtler & Bromme, 2008), the present study compared tasks that oriented readers toward different aspects of the source. Our recognition task (Experiment 2) indicated that readers selectively memorized those source aspects that were relevant to their particular task. This finding points to the potential of prompts as a parsimonious way of guiding students' attention to source information.

These experiments have a number of limitations that restrict the scope of our findings and should be addressed in future research. First, our materials consisted of a series of short texts organized in a rather standardized way. Even though the booklets included some filler items with a different structure, we cannot rule out the possibility that participants learned the structure of the critical items and constructed specific strategies to address the task. The findings reported in this paper clearly call for an extension to longer documents with more complex, less predictable structures. Related to this limitation, we did not independently assess the respective salience or distinctiveness of "physical" and "knowledge" source attributes. Even though we did not reason about absolute differences between these two types of attributes *per se*, we cannot rule out that intrinsic, uncontrolled differences may explain at least in part the observed interactions between the attributes and other factors. Furthermore, the materials dealt with rather familiar topics and situations. It is unclear whether the findings would replicate with more complex topics or less clear-cut discrepancies. Theoretical discussions (e.g., Bromme & Goldman, 2014) and empirical studies suggest that a lack of familiarity with the contents would prompt lay readers to focus even more on source details as peripheral cues to credibility (Lucassen & Schraagen, 2013). However, this is an issue left for further research.

As another limitation, our assessment of text comprehension was limited to a simple factual question (during reading) and to the assessment of content and source recognition after a short delay. Although participants answered the comprehension question quite accurately, this question alone hardly makes for a test of whether they constructed an accurate model of the situation. For instance, we have no information regarding whether participants were able to draw any conclusion regarding the situation, or whether they would be able to use the information in any application or transfer task. A more complete assessment of content comprehension might increase the probability to observe differential effects of source features on students' resolution of the discrepancy. Finally, our interpretations are limited in the absence of process data (e.g., reading time and attention to source features during reading). Although prior research suggests that the construction of a Document model takes place online (e.g., Braasch et al., 2012), future experiments should investigate whether readers devote additional attention to the details of a source description upon realizing that the story includes uncertain or discrepant assertions.

Despite these limitations, our study adds support to the view that integration, as it applies to text comprehension, encompasses both the contents and the sources of textual information (Britt & Rouet, 2012; Perfetti et al., 1999). The tagging of specific contents to their respective sources combined with the evaluation of a source's competence are powerful mechanisms to enable readers to restore some coherence upon reading discrepant or conflicting accounts of a situation (Britt et al., 1999; Britt et al., 2013; Stadtler & Bromme, 2014). However, the integration of sources and contents is a sophisticated and effortful comprehension strategy, whose acquisition needs to be supported through adequate instructional means (Britt et al., 2018). In a world of abundant but changing and sometimes unreliable information, this would certainly be a worthwhile educational endeavor.

References

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,* 450-465.

Bråten, I., & Braasch, J.L. (2018). The role of conflict in multiple source use. In J.L.G. Braasch, I. Bråten, & M.T. McCrudden (Eds.) *Handbook of multiple source use* (pp. 184-201). New York: Routledge.

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

Bråten, I., Strømsø, H., & Salmerón, L. (2010). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction, 21*, 180-192.

Britt, M.A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*, 485-522.

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). Mahwah, NJ: Lawrence Erlbaum Associates.

Britt, M.A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M.J. Lawson and J.R. Kirby (Eds.) *The Quality of Learning* (pp. 276-314). New York, NY: Cambridge University Press.

Britt, M.A., Rouet, J.-F., & Braasch, J.L.G. (2013). Documents as entities: Extending the situation model theory of comprehension. In M.A. Britt, S.R. Goldman, and J.-F. Rouet (Eds.), *Reading: From Words to Multiple Texts* (pp. 160-179). New York: Routledge.

Britt, M.A., Rouet, J.-F., & Durik, A. (2018). *Literacy beyond text comprehension: a theory of purposeful reading*. New York: Routledge.

Bromme, R., & Goldman, S.R. (2014). The public's bounded understanding of science. *Educational Psychologist, 49*(2), 59-69.

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*, 209-222.

Cook, A.E., & O'Brien, E.J. (2014). Knowledge activation, integration, and validation during narrative text comprehension, *Discourse Processes, 5(1-2)*, 26-49.

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*, *24*(2), 187-204.

Goldman, S.R., & Brand-Gruwel, S. (2018). Learning from multiple sources in a digital society. In F. Fischer, C.E. Hmelo-Silver, S.R. Goldman, & P. Reimann (Eds.) *International handbook of the learning sciences* (pp. 86-95). New York: Routledge.

Goldman, S.R., Britt, M.A., Brown, W., Cribb, G., George, M., Greenleaf, C., ... & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, *51*(2), 219-246.

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.

Hakala, C.M., & O'Brien, E.J. (1995). Strategies for resolving coherence breaks in reading. *Discourse Processes*, *20*, 167–185.

Kaakinen, J.K., & Hyönä, J. (2014). Task relevance induces momentary changes in the functional visual field during reading. *Psychological Science, 25*(2), 626-632.

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.

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

Lefavrais, P. (1968). *La pipe et le rat [the pipe and the rat].* Issy-les-Moulineaux, France : EAP.

List, A., Du, H., Wang, Y., & Lee, H.Y. (2019). Toward a typology of integration: Examining the documents model framework. *Contemporary Educational Psychology, 58*, 228-242.

Lucassen, T., & Schraagen, J.M. (2013). The influence of source cues and topic familiarity on credibility evaluation. *Computers in Human Behavior, 29*(4), 1387-1392.

Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmerón, L., & Rouet, J.F. (2019). How good is this page? Benefits and limits of prompting on teenagers' assessment of Web information quality. *Reading Research Quarterly, 54*(3), 299-321.

McCrudden, M.T., Magliano, J., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, *35*, 229-241.

McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students' web savvy: An intervention study. *British Journal of Educational Psychology, 89*(3), 485-500.

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J.F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, *58*, 53-64.

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

Pirolli, P., & Card, S.K. (1999). Information Foraging. *Psychological Review, 106*, 643-675.

Rapp, D. N., & McCrudden, M. T. (2018). Relevance before, during, and after discourse experiences. In J. Strassheim & H. Nasu (Eds.) *Relevance and Irrelevance: Theories, Factors and Challenges* (pp. 141-160). Berlin: De Gruyter.

Rouet, J.-F., & Britt, M.A. (2014). Learning from Multiple Documents. In *R.E. Mayer (Ed.) Cambridge Handbook of Multimedia Learning, 2nd Edition* (pp. 813-841). Cambridge, MA: Cambridge University Press.

Rouet, J.-F., Britt, M.A., & Potocki, A. (2019). Multiple-text comprehension. In J. Dunlosky & K. Rawson (Eds.) *Cambridge Handbook of Cognition and Education* (pp. 356-380). Cambridge, MA: Cambridge University Press.

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

Rouet, J.-F., Le Bigot, L., de Pereyra, G., & Britt, M.A. (2016). Whose story is this? Discrepancy triggers readers' attention to source information in short narratives. *Reading and Writing*, *29*(8), 1549-1570.

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.

Sanchez, C.A., Wiley, J., & Goldman, S.R. (2006). Teaching students to evaluate source reliability during internet research tasks. In S.A. Barab, K. Hay, & D.T. Hickey (Eds.)

Proceedings of the 7th International Conference of the Learning Sciences (pp. 662–666). Bloomington, IN: International Society of the Learning Sciences.

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 and Cognition, 45*, 151-167.

Saux, G., Ros, C., Britt, M.A., Stadtler, M., Burin, D., & Rouet, J.-F. (2018). Readers' selective recall of source features as a function of claim discrepancy and task demands. *Discourse Processes. 55*(5-6), 525-544.

Schraw, G., Wade, S.E., & Kardash, C.A. (1993). Interactive effects of text-based and task-based importance on learning from text. *Journal of Educational Psychology*, *85*, 652–661.

Stadtler, M. & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer Supported Collaborative Learning, 2*, 191-210.

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. Braasch (Eds.), *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences* (pp. 379-402). Cambridge, MA: MIT Press.

Stadtler, M., Paul, J.M., Globoschütz, S., & Bromme, R. (2015). Watch out!-An instruction raising students' epistemic vigilance augments their sourcing activities. In D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, & P.P. Maglio (Eds.) *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2278-2283). Austin, TX: Cognitive Science Society.

Stang Lund, E., Bråten, I., Brante, E.W., & Strømsø, H.I. (2017). Memory for textual conflicts predicts sourcing when adolescents read multiple expository texts. *Reading Psychology, 38*(4), 417-437.

van den Broek, P., Bohn-Gettler, C.M., Kendeou, P., Carlson, S., & White, M.J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M.T. McCrudden, J.P. Magliano, & G. Schraw (Eds.), *Text Relevance and Learning from Text* (pp. 123-140). Greenwich, CT: Information Age Publishing.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source Evaluation, Comprehension, and Learning in Internet Science Inquiry Tasks. *American Educational Research Journal, 46*, 1060-1106.

Wiley, J., Jaeger, A.J., & Griffin, T.D. (2018). Effects of instructional conditions on comprehension from multiple sources in history and science. In J.L. Braasch, I. Bråten, & M.T. McCrudden (Eds.), *Handbook of Multiple Source Use* (pp. 341-361). New York: Routledge.

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport, CT: Greenwood Press.

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*, 73-87.

Table 1

Example of a story used in experiments 1 and 2, with the manipulation of source features and content consistency.

#	Section	Example
1	General introduction	Last night a violent fire ravaged a warehouse in the industrial district.
2	Source introduction	A policeman and a journalist came to the scene the next morning.
3	Source 1 - [Feature 1] - [Statement 1 / Alternate statement 1]	The policeman, who held a tissue on his face, declared that the fire was due to a *sabotage / breakdown* of the electric circuit.
4	Source 2 - [Feature 2] - Statement 2.	The fireman, after inspecting the facility, concluded that the fire was set by a malevolent individual.

Note. Original version in French. The manipulation of story consistency is marked in italics. The feature of the sources are underlined.

Table 2.

Mean (and SD) percentage of to the source evaluation question (Q2) as a function of statement consistency (Experiment 1).

	Response category		
Story version	*Source w/ knowledge feature*	*Source w/ physical feature*	*Uncertain*
Consistent	55.7 (25.7)	12.5 (13.3)	31.8 (24.6)
Discrepant	59.1 (20.1)	18.9 (16.6)	22.0 (18.2)

Table 3.

Mean (and SD) percentage of correct answers to the factual question (Q1) as a function of statement consistency and task type (Experiment 2).

Q1 (factual question)	*Consistent stories*	*Discrepant stories*
Knowledge evaluation task	90.4 (12.9)	77.9 (27.9)
Age evaluation task	91.5 (15.2)	78.9 (26.6)
Baseline task	84.8 (18.8)	89.4 (26.8)

Note. In all three task conditions, the expected answer for consistent stories was the statement the two sources agreed upon; whereas the expected answer for discrepant stories was “uncertain”.

Table 4.

Mean (and SD) percentage of correct answers to the source evaluation question (Q2) as a function of statement consistency and task type (Experiment 2).

		Response category		
Task condition	*Story version*	*Source w/ knowledge feature*	*Source w/ physical feature*	*Uncertain*
Competence evaluation task	*Consistent*	38.8 (25.7)	17.8 (16.7)	43.4 (28.0)
	Discrepant	43.8 (23.2)	15.0 (17.7)	41.2 (26.3)
Age evaluation task	*Consistent*	24.4 (18.9)	27.2 (23.1)	48.4 (33.1)
	Discrepant	31.6 (23.8)	23.1 (19.9)	45.3 (35.9)

Note: The baseline condition is not represented in this table because participants were not asked any question about the sources.

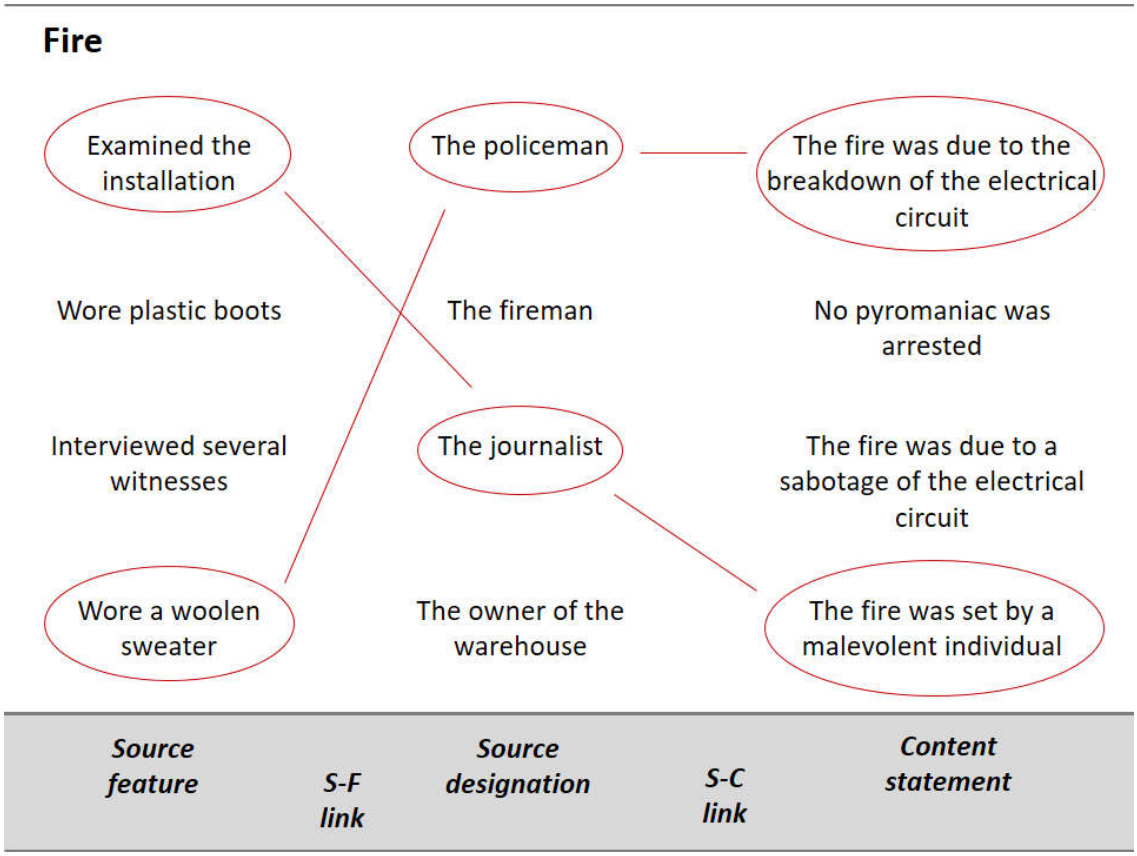

Figure 1. Example of a trial from the source recognition task. Participants were first asked to mark the actual sources included in the story ("source designation", central column). Then they were asked to draw a link between each source and their respective feature ("source feature" and "S-F link", left-hand side column). Finally, they were asked to draw a link between the source and their respective content statement ("Content statement" and "S-C link", right-hand side column).

Note. The ellipses and lines exemplify the optimal response to the "Fire" story example (discrepant version, Table 1). The lower line that identifies the recognition task components (highlighted in grey) was not included in the actual recognition sheet.

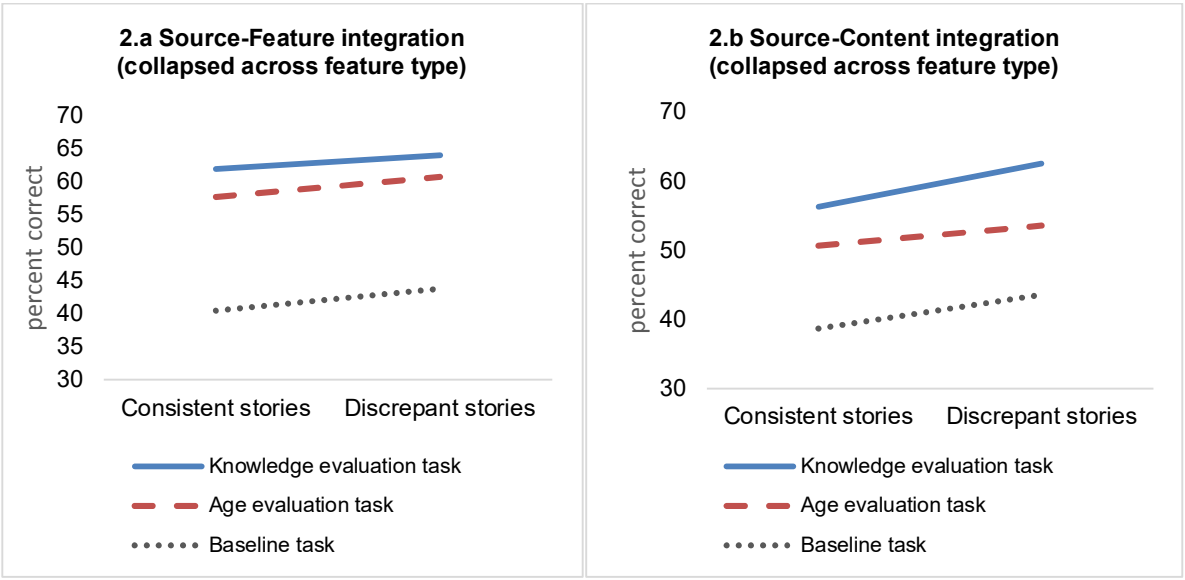

Figure 2. Recognition of source nodes (i.e., designation and feature; 2a) and source-content link (2b) as a function of task condition and story discrepancy.