Causal Inferences During EFL Reading of Expository Texts:
Effects of Two Kinds of Familiarity and L2 Reading Proficiency

Masaya HOSODA
Graduate School, University of Tsukuba

Abstract

The present study examined effects of two kinds of familiarity (inference ideas and text content) and L2 reading proficiency on EFL readers’ on-line (i.e., during-reading) causal inference generation from expository texts. Fifty-five Japanese university students participated in the experiment and read texts that varied in familiarity: Familiar (FA), Partially Unfamiliar (PU), and Unfamiliar (UF). The combined measure of reading times for target sentences and response times for inference questions revealed that low-proficiency readers failed to make on-line causal inferences regardless of familiarity condition, but high-proficiency readers made the inferences during the reading of FA and PU texts. In addition, error rates for inference questions suggest that low-proficiency readers generated the inferences when answering a question, that is, when the task required them to do so. Together, these findings reveal the conditions in which EFL readers make causal inferences during expository reading. Pedagogical implications are discussed in terms of the interplay between the reader and text.

1. Introduction

Successful text comprehension requires that readers make inferences to understand texts’ implicit meaning. For example, upon encountering the sentence, *Miso is recommended for keeping the body warm because it contains a rich amount of salt*, the reader should recognize the causal relation by the causal connective *because*: being rich in salt makes Miso suitable for keeping the body warm. However, the deep-level understanding of this causal relation further requires an unstated causal explanation pertaining to why being rich in salt is good for keeping the body warm. That is, the reader has to infer *Salt keeps the body warm* to fully understand the causal relation in this sentence. Such inferences are called causal inferences because they link text ideas through causal relations (Singer, Harkness, & Stewart, 1997; Singer & O’Connell, 2003; Noordman, Vonk, & Kempff, 1992). Making causal inferences is particularly important for understanding expository texts, which aim “to communicate information so that the reader might learn something” (Weaver & Kintsch, 1991, p. 230). As the above example shows, causal
inference generation allows the reader to glean causal explanations for events described in the text, which is crucial for deep expository comprehension. This is largely because texts often provide information through causal relations (León & Peñalba, 2002). Thus, making causal inferences enables the reader to go beyond understanding of the text per se, allowing to learning new knowledge from the text (Singer & O’Connell, 2003). This is particularly important in educational settings, as expository texts play a central role in conveying novel concepts or ideas to students (Weaver & Kintsch, 1991).

Despite the importance of causal inferences in understanding and learning from expository texts, there is limited research in this area, especially for English as a Foreign Language (EFL) learners. For example, an important but unresolved question is under what circumstances EFL learners automatically make causal inferences during expository reading (i.e., on-line). Hosoda (2014) found that although EFL readers made causal inferences by engaging in a post-reading task, they failed to do so on-line even if their L2 reading proficiency was high. However, past studies are limited because they did not sufficiently control information familiarity, that is, the extent to which one has relevant knowledge. Numerous studies have found that information familiarity has a critical impact on inference generation and reading processes (e.g., McNamara & O’Reilly, 2009; Miller & Keenan, 2011; Myers & O’Brien, 1998), but past inference studies have not fully considered this issue. Specifically, in the context of inference generation from expository texts, we can assume that at least two kinds of familiarity matters—familiarity of inferences themselves, and that of text content from which inferences are derived. Nevertheless, prior L1 and L2 research has only considered one type of familiarity (e.g., Noordman et al., 1992; Singer et al., 1997), and no study has manipulated both kinds of familiarity simultaneously. Moreover, although research has shown that the nature of inferences made by L2 readers depends on their L2 reading proficiency (e.g., Yoshida, 2003), few studies have addressed a possible relationship between L2 reading proficiency and these two kinds of familiarity.

These issues inspired the present study, in which EFL readers’ on-line causal inference generation from expository texts was investigated with respect to two kinds of familiarity (inference ideas and text content) and L2 reading proficiency. The present study would be of pedagogical importance because it may provide a theoretical basis for reading instruction that accommodates individual differences in reading proficiency and text characteristics.

1.1 Inference Familiarity and Inference Generation: Two Types of Inferences

When the inference to be made is familiar, it is assumed that readers have pertinent background knowledge to that idea. In such a case, causal inference processing can be accounted for by the validation model, which was developed to explain how causal relations are understood in discourse processes (e.g., Singer, Halldorson, Lear, & Andrusiak, 1992). Specifically, the model posits that in the course of causal comprehension, one activates background knowledge pertinent to the causal relation through inference generation to assess the validity of the causal
relation. For example, to coherently understand the statement, *Dorothy poured the bucket of water on the bonfire. The fire went out*, readers should activate the knowledge, *Water extinguishes fire*, which mediates the causal relation between the first event (pouring water on the fire) and the second (the fire went out), to justify the causality. Indeed, researchers have found empirical evidence for this process in various studies with a conclusion that causal comprehension inherently involves the activation of relevant knowledge associated with inference generation (Singer, 2013).

However, what if knowledge pertinent to causal relations is not available because inferences are unfamiliar to readers? Crucially, such unfamiliar inferences are supposed to involve fundamentally different processes than the knowledge-based inferences discussed above; because what is inferred is novel, readers are required to deductively derive new information from the text rather than activate a corresponding idea in their existing knowledge base. To date, fewer studies have examined such inferences of unfamiliar ideas, and agreement has yet to be reached on whether they are made automatically during reading. Singer and colleagues (Singer et al., 1997; Singer & O’Connell, 2003) proposed that L1 readers make unfamiliar causal inferences during reading expository texts. They used *because-target sentences* that required causal inferences in two conditions: *explicit*, where inference ideas were presented before target sentences, and *implicit*, where those ideas were excluded (see Table 1). Participants read texts sentence by sentence on a PC screen and answered *inference questions*, in which they verified inference ideas with yes/no responses. The researchers hypothesized that if causal inferences are made during reading, reading times for targets sentences should be longer in the implicit than explicit condition due to additional processing required to infer the missing ideas. Conversely, correct response times for inference questions should not differ between the two conditions because the question probed the idea readers in the implicit condition should have inferred. Indeed, they obtained this pattern of results in a series of experiments with L1 readers when causal relations were signaled by the connective *because*. In contrast, Noordman et al. (1992) failed to evidence on-line generation of causal inferences about unfamiliar ideas, despite the fact that they used the same methodology as in Singer et al. (1997). Nevertheless, these studies only considered familiarity of inference ideas; in other words, they did not control familiarity of text content from which the inferences originated. Hence, possible effects of these two kinds of familiarity remain unresolved.

Meanwhile, some recent studies have suggested that inferring unfamiliar ideas can be influenced by reading proficiency. Millis, Magliano, and Todaro (2006) investigated L1 readers’ expository comprehension with a think-aloud method, where readers were asked to verbalize their thoughts while processing texts. The results showed that only proficient readers successfully understood unfamiliar causal relations via causal inferences. Researchers have reasoned that proficient readers can do so because they are better at *constructive activities* to make sense of texts by recruiting information relevant to the text domain from their knowledge base. Given this finding, it is possible that if readers have sufficient reading proficiency, and the text content is
familiar so that readers can use relevant knowledge, they will make unfamiliar causal inferences during reading. However, prior research on L2 reading has also shown that even proficient readers are strongly constrained by lower-level processing (e.g., word decoding, syntactic processing) when reading in L2. Horiba (2000) indicated that L2 readers tend to focus primarily on individual words and phrases, and they therefore have greater difficulty flexibly performing constructive processing, compared to L1 readers. Similarly, Yoshida (2012) demonstrated that EFL readers often fail to modulate reading processes in accordance with a task that requires integrative reading. Given these studies showing that L2 readers have difficulty with higher-level processing, such as inference generation, it is also possible that making unfamiliar inferences is difficult for EFL readers even if text content is familiar to them.

1.2 Text Familiarity and Reading Proficiency in Inference Generation

Familiarity of text content has widely been shown to have a large bearing on reading success. One compelling reading model that explains this is the memory-based view (Myers & O’Brien, 1998). The model illustrates reading processes by explaining the interplay between text information and readers’ relevant knowledge, assuming that reading is routinely guided by the passive activation of relevant knowledge and/or prior text information, called the resonance processes, in response to the incoming information. More specifically, when text information is familiar, readers’ knowledge strongly resonates with it, facilitating knowledge-driven fluent reading. Similarly, familiar text content may strongly induce the activation of specific inferences through resonance processes and consequently afford automatic inference generation.

However, research has also shown that readers’ knowledge cannot always be readily applied to such processes, and the success may depend on reading proficiency. McNamara and O’Reilly (2009) proposed that less proficient readers often fail to activate relevant knowledge because it is usually difficult for them to recognize when it is necessary to use such knowledge. This seems particularly likely for EFL readers, whose basic reading skills (e.g., word decoding) are less automated than L1 readers. In fact, Morishima (2013) showed that Japanese EFL university students do not exhibit passive activation of information, and hence are less susceptible to inconsistencies in texts, compared to L1 readers. Within the context of EFL, Yoshida (2003) compared learners with different proficiency levels and found that it is more difficult for low-proficiency learners to make knowledge-based inferences because their lower-level processing competes and draws cognitive resources from inference processing.

In contrast, some studies have reported that text familiarity can compensate for the lack of proficiency in less skilled readers. Miller and Keenan (2011) showed that the comprehension deficiency experienced by less proficient L2 readers is largely mitigated by familiar expository texts, compared to a condition where text familiarity and L2 reading proficiency are both low. Based on these findings, they concluded that high text familiarity can promote higher-level
processing that is mostly driven by readers’ prior knowledge, which in turn alleviates difficulties in lower-level processing caused by insufficient reading skills.

In sum, the literature has posed two different views regarding the relationship between text familiarity and reading proficiency: (a) a certain level of reading proficiency is required before text familiarity can have facilitative effects, and (b) text familiarity serves as a compensatory mechanism for the lack of reading proficiency. In this context, the present study explored if EFL readers can activate relevant knowledge to generate familiar causal inferences from familiar texts. By addressing this, the present study aimed to show which of the above views best applies to Japanese EFL readers.

2. The Present Study

The purpose of the present study was to explore EFL readers’ on-line causal inference generation from expository texts considering two kinds of familiarity (inference ideas and text content), and L2 reading proficiency. To this end, experimental texts for three familiarity conditions were selected in a pilot study. Participants were given no specific task in the experiment so that strategic processing would not trigger unexpected inferences.

To assess on-line inference generation, the present study followed Singer et al. (1997) who employed two measures of inference processing: reading times for target sentences requiring causal inferences, and correct response times for inference questions querying the inferences. This methodology was adopted because it has been shown that using these two measures provides complementary ways of examining the on-line status of inference generation (e.g., Noordman et al., 1992; Singer et al., 1997; Singer and O’Connell, 2003). Specifically, according to the first index, reading times are significantly longer for target sentences when inferences are implicit (i.e., implicit condition) than when they are explicitly stated (i.e., explicit condition) because it takes additional time to infer implicit ideas. However, reading times alone are not informative about the content of the inferences. Therefore, correct response times for inference questions are also considered; the second index is based on comparable inference response times between the implicit and explicit conditions. This is because if readers made inferences during reading, the corresponding ideas should be active in memory, such that implicit- and explicit-inference questions can be verified at a similar speed. With this methodology, the present study addressed the following research question (RQ).

RQ: When two kinds of familiarity (inference ideas and text content) and L2 reading proficiency are considered, under what conditions do Japanese EFL readers make causal inferences during the reading of expository texts?
3. Method

3.1 Participants

Fifty-five Japanese university students majoring in humanities, literature, or social studies participated in this study. Data from five participants were removed due to trouble with experimental tools. The remaining 50 participants were grouped into two reading proficiency groups (25 per group) based on their L2 reading proficiency test scores; the high group scored significantly better \((M = 17.52, SD = 3.38, \text{Min/Max} = 14/26)\) than the low group \((M = 9.48, SD = 2.86, \text{Min/Max} = 4/13)\), \(t(48) = -9.82, p < .001, r = .80\).

3.2 Materials

**L2 reading proficiency test.** A 26-item reading test was constructed from the pre-first and second grades of the STEP test (Obunsha, 2005a, 2005b). It is important to note that the present study interpreted reading proficiency as an ability to appropriately and accurately understand what the text communicates, rather than specific grammatical or vocabulary knowledge. This is based on the assumption that the reader has to first interpret what is explicitly stated in the text in order to infer implicit ideas. The reading section of the STEP test was widely used in experimental as well as educational settings to assess such a discourse comprehension ability. The reliability of the test was acceptable, Cronbach’s \(\alpha = .82\).

**Reading materials.** Expository texts for three text and inference familiarity conditions, *Familiar* (FA), *Partially Unfamiliar* (PU), and *Unfamiliar* (UF), were used.

In FA texts, both texts and inferences were familiar. Participants were most expected to make causal inferences during reading these texts because familiar contexts may facilitate knowledge-driven processes, which promote knowledge-based inference generation. In PU texts, text content was familiar, but inference ideas were unfamiliar. These texts, in which participants were required to infer unfamiliar ideas to understand familiar context, were the primary focus of this study. Finally, in UF texts, both texts and inferences were unfamiliar. It was likely that participants would have difficulty making inferences from UF texts because of limited available relevant knowledge. Note that texts where text content was unfamiliar and inference ideas were familiar were not created for two reasons. First, because the primary concern of the study is whether new information (i.e., unfamiliar ideas) can be derived through inference generation, the reverse condition (i.e., familiar inferences from unfamiliar texts) is not considered so relevant. Second, four familiarity conditions in a within-participants design would require participants to process a huge number of texts, which was expected to increase their burden too much.

**Construction of reading materials.** Experimental texts were selected by creating 45 candidate texts by following the procedure specified in Singer et al. (1997). First, I constructed target sentences that could be expressed with *because* (e.g., *Miso is recommended for keeping the body warm because it contains a rich amount of salt*) by referring to entries in a science
encyclopedia. Then, the surrounding contexts were created from entry descriptions. Next, three Japanese graduate students majoring in English education and I identified candidate inference ideas (e.g., Salt keeps the body warm) necessary for linking causal relations in the target sentences. After that, discussion was held to determine ideas that were commonly identified by three or more of us. Integrating the identified ideas yielded inferences that were used in a pilot study. Finally, low-frequency words (JACET8000 level 5 or more; JACET, 2003) were rephrased into simpler words, and a native English speaker checked the materials.

Pilot study. To determine suitable texts for the experiment, a pilot study was conducted with 65 expository materials (45 were newly constructed via the above procedure, and 20 were from Hosoda, 2014). The purpose of the pilot study was twofold: (a) to identify texts satisfying the three familiarity conditions (FA, PU, and UF), and (b) to confirm the viability of inference generation. Forty Japanese undergraduates read the 65 texts, excluding the inference ideas, sentence by sentence on a personal computer (PC) screen, and rated the familiarity of each text on a 7-point scale (1 = totally unfamiliar, 7 = very familiar). After reading all the texts, a booklet was provided for a recognition task where participants were asked to judge if each of the 65 inference ideas had been stated in the texts they had just read. Finally, they rated familiarity of the 65 inferences on the same scale as in the text familiarity ratings.

Mean familiarity ratings of the text and inferences were calculated. For both of inference and text familiarity, those with mean ratings of 5 or more were deemed familiar, and those with 3 or less were unfamiliar. As a result, 14, 12, and 15 texts met the criteria for FA (i.e., familiar texts and inferences), PU (i.e., familiar texts and unfamiliar inferences), and UF (i.e., unfamiliar texts and inferences) texts, respectively. To confirm the viability of inference generation, I selected the texts where most participants falsely judged that the inference ideas had been stated; participants falsely recognized the missing inference ideas because they inferred and encoded them. Finally, 12 texts were selected for each familiarity condition. Twelve filler texts were added for a total of 48 texts in the experiment.

Table 1 shows an example of PU texts. Each text had a target sentence with the connective because, which requires causal inferences. There was explicit or implicit condition in a given text. In the explicit condition, texts included explicit sentences that directly stated inference ideas just before the target sentences. In the implicit condition, the explicit sentences were excluded, such that participants had to generate causal inferences to understand the target sentences. Additionally, the experimental texts were accompanied by two verification questions: an inference question probing the inference idea, and a detail question targeting specific details to encourage careful reading. The correct response was “yes” for half of the detail questions, and “no” for the other half. Filler and control texts (see below) were followed by two detail questions and no inference question. The questions were presented in Japanese so that L2 decoding ability, which was not the focus of this study, would not influence response times.
Table 1  An Example Partially Unfamiliar Text and Verification Questions

<table>
<thead>
<tr>
<th></th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Miso is a traditional Japanese seasoning, made from soybeans, sea salt, and rice.</td>
</tr>
<tr>
<td>b</td>
<td>Among them, salt has the effect of raising body temperature. (explicit).</td>
</tr>
<tr>
<td>c</td>
<td>Miso is recommended for keeping the body warm because it contains a rich amount of salt. (target).</td>
</tr>
<tr>
<td>d</td>
<td>Nowadays, food companies produce miso in large quantities and the home-made style has become rare.</td>
</tr>
<tr>
<td>e</td>
<td>塩分は体を温かく保つ (inference)</td>
</tr>
<tr>
<td>f</td>
<td>手作りのミソは今でもよくみられる (detail)</td>
</tr>
</tbody>
</table>

*Note.* Reading and response times for the underlined parts were recorded.

Three material sets were created with the 48 texts. Within each set, the 36 experimental texts were evenly assigned to the three text conditions; explicit, implicit, and control. Control texts were used for the memory task, which is not reported in this paper, and did not include explicit sentences. Filler texts included explicit sentences to equate the number of texts with and without explicit sentences. The assignment of each text to the three text conditions was counterbalanced across sets.

*Familiarity ratings.* To confirm text and inference familiarity for the experimental texts, text and inference familiarity ratings were collected using the same 7-point scale as in the pilot study (1 = *totally unfamiliar*, 7 = *very familiar*).

### 3.3 Procedure

Participants were individually tested in a session that lasted about 80 minutes. They took the L2 reading proficiency test for 30 minutes, and then proceeded to a reading task.

In the reading task, a “ready?” signal was displayed at the center of the PC screen. After participants pressed a “yes” button on the Response-Pad RB-730, the first sentence appeared. Participants were instructed to read each sentence for understanding in a self-paced manner. Reading times for target sentences were recorded by Super Lab 4.5. After one text was processed, the screen went blank for 2,500 milliseconds (ms) and a 500-ms fixation (****) followed. Then, the first verification question appeared. Participants were asked to verify the question as quickly and accurately as possible by pressing the “yes” or “no” button. Note that participants were asked to respond “no” even if the question was true based on common sense; they were instructed to base their judgment only on the text they had read. This was because as inference ideas for FA texts were familiar, inference questions for those texts could potentially be verified without inferring from the texts, that is, only with reference to common sense (participants practiced before the experiment, and possible effects of these instructions will be discussed later). Response times and accuracy were recorded. Following another blank (2,500 ms) and a fixation (500 ms), the second verification question appeared. After answering this second question, another fixation (####) was
presented for 500 ms, and then participants were asked to rate the familiarity of the text on a 7-point scale (see above). After a 1000-ms blank, another “ready?” signal was presented. This sequence was repeated for 48 texts that were presented in a random order. Participants took a five-minute break halfway through the reading task.

After the reading task, a booklet was provided containing the inferences for the 48 passages, and participants were asked to rate inference familiarity on a 7-point scale.

3.4 Data Analysis

Data from those cells were removed that received mean familiarity ratings of 5 or more when they were originally intended as unfamiliar, or 3 or less when they were originally intended as familiar. This led to the exclusion of 4.25 % of the original data.

Target reading and inference response times were divided by the number of syllables, as the number of words in target sentences and inference questions varied. To investigate on-line causal inference generation, 2 (L2 reading proficiency: High, Low) × 3 (Familiarity: FA, PU, UF) × 2 (Explicitness: Implicit, Explicit) three-way mixed analyses of variance (ANOВAs) were conducted on mean target reading and inference question response times. L2 reading proficiency was a between-participants variable and the other two variables were within-participants variables. A significance criterion of \( \alpha = .05 \) was used for all analyses. Marginally significant effects were considered non-significant. In addition to the statistical analyses, descriptive analysis of error rates for inference questions are reported to examine inference accuracy\(^1\).

4. Results and Discussion

4.1 Familiarity Ratings

First, it is necessary to confirm that text and inference familiarity ratings were appropriate for each familiarity condition. Overall, mean text/inference familiarity ratings for FA, PU, and UF texts were 5.27/6.26, 5.21/2.24, and 2.57/2.21, respectively (see Table 2). To examine rating differences between the two proficiency groups and between the three familiarity conditions, 2 (Proficiency: High, Low) × 3 (Familiarity: FA, PU, UF) two-way ANOВAs were performed on text and inference familiarity ratings. Results showed significant main effects of Familiarity on both text, \( F(2, 33) = 556.63, p < .001, \eta^2_p = .97 \), and inference, \( F(2, 33) = 3250.83, p < .001, \eta^2_p = .99 \), familiarity ratings. Neither the Proficiency main effect nor the interactions were significant (all ps > .050). Bonferroni-adjusted post hoc tests on the Familiarity effects showed that text familiarity ratings for FA and PU texts were similar (\( p = .511 \)) but significantly higher than UF texts (both ps < .001), and inference familiarity ratings were significantly higher for FA than PU and UF texts (ps < .001), which were comparable (\( p = .568 \)). Accordingly, the analyses confirm that (a) the two types of familiarity were rated appropriately for FA, PU, and UF texts, and (b) this was the case for both proficiency groups (i.e., no effects involving Proficiency were significant).
Table 2 Mean Text and Inference Familiarity Ratings

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Familiar</th>
<th>Partially Unfamiliar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Text</td>
<td>Inference</td>
<td>Text</td>
</tr>
<tr>
<td>High (n = 25)</td>
<td>5.24 (.19)</td>
<td>6.24 (.11)</td>
<td>5.22 (.24)</td>
</tr>
<tr>
<td>Low (n = 25)</td>
<td>5.30 (.29)</td>
<td>6.28 (.28)</td>
<td>5.20 (.27)</td>
</tr>
<tr>
<td>Total (N = 50)</td>
<td>5.27 (.24)</td>
<td>6.26 (.21)</td>
<td>5.21 (.25)</td>
</tr>
</tbody>
</table>

Note. Familiarity ratings ranged from 1 (totally unfamiliar) to 7 (very familiar). Standard deviations are in parentheses.

4.2 Target Reading Times

Table 3 shows mean reading times for target sentences. The three-way ANOVA found significant main effects of Proficiency, $F(1, 48) = 9.78, p = .003, \eta_p^2 = .17$, Familiarity, $F(2, 96) = 16.04, p < .001, \eta_p^2 = .25$, and Explicitness, $F(1, 48) = 30.11, p < .001, \eta_p^2 = .39$. Additionally, the Familiarity × Explicitness interaction was significant, $F(2, 96) = 4.28, p = .017, \eta_p^2 = .08$. No other effects were significant (all $ps > .100$). Follow-up analyses on the Familiarity × Explicitness interaction showed that target reading times were significantly longer in the implicit than explicit condition for FA ($p = .003$) and PU ($p < .001$), but not for UF texts ($p = 110$).

Thus, the target reading times support on-line causal inferences for FA and PU texts (implicit > explicit) but not UF texts (implicit ≈ explicit). In addition, these results did not differ between the two proficiency groups, as revealed by the fact that the three-way Proficiency × Familiarity × Explicitness interaction was not significant, $F(2, 96) = 0.58, p < .562, \eta_p^2 = .01$.

Table 3 Mean Reading Times for Target Sentences (in milliseconds)

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Familiar</th>
<th>Partially Unfamiliar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Implicit</td>
<td>Explicit</td>
<td>Implicit</td>
</tr>
<tr>
<td>High (n = 25)</td>
<td>351.23</td>
<td>312.10</td>
<td>416.63</td>
</tr>
<tr>
<td></td>
<td>(96.06)</td>
<td>(80.20)</td>
<td>(117.87)</td>
</tr>
<tr>
<td>Low (n = 25)</td>
<td>413.06</td>
<td>349.53</td>
<td>495.94</td>
</tr>
<tr>
<td></td>
<td>(133.16)</td>
<td>(96.34)</td>
<td>(146.35)</td>
</tr>
<tr>
<td>Total (N = 50)</td>
<td>382.14</td>
<td>330.81</td>
<td>456.29</td>
</tr>
<tr>
<td></td>
<td>(119.08)</td>
<td>(89.74)</td>
<td>(137.48)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses.

4.3 Inference Response Times

Table 4 shows mean correct response times for inference questions. The three-way ANOVA yielded significant main effects of Proficiency, $F(1, 48) = 6.77, p = .012, \eta_p^2 = .12$, Familiarity, $F(2, 96) = 26.19, p < .001, \eta_p^2 = .35$, and Explicitness, $F(1, 48) = 16.32, p < .001, \eta_p^2 = .25$. More important, the three-way Proficiency × Familiarity × Explicitness interaction was
significant, $F(2, 96) = 3.34, p < .040, \eta^2_p = .07$. To interpret this, follow-up analyses were conducted focusing on an aspect of findings related to the research question, that is, the simple Familiarity × Explicitness interaction within each proficiency group.

For the low group, the simple Familiarity × Explicitness interaction was not significant, $F(2, 48) = 1.68, p = .198, \eta^2_p = .07$ (see Figure 1). In contrast, the simple main effect of Explicitness was significant, $F(1, 24) = 15.09, p < .001, \eta^2_p = .39$, indicating that the low group took significantly longer to verify inference questions in the implicit than explicit condition regardless of Familiarity condition (i.e., for FA, PU, or UF texts). This is incompatible with the on-line inference criterion, which suggests that response times should be comparable across the Explicitness conditions. Hence, these results suggest that low-proficiency readers failed to make causal inferences during reading expository texts.

Next, for the high group, the simple Familiarity × Explicitness interaction was significant, $F(2, 48) = 5.27, p = .009, \eta^2_p = .18$ (see Figure 2). Follow-up analyses revealed that response times did not significantly differ between the Explicitness conditions for FA ($p = .450$) and PU texts ($p = .322$) but were significantly longer in the implicit condition for UF texts ($p = .016$). Thus, inference response times in the high group support on-line causal inference generation for FA and PU texts (i.e., implicit $\approx$ explicit), but not for UF texts (i.e., implicit $>$ explicit).

Table 4 Mean Correct Response Times for Inference Questions (in milliseconds)

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Familiar</th>
<th>Partially Unfamiliar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Implicit</td>
<td>Explicit</td>
<td>Implicit</td>
</tr>
<tr>
<td>High (n = 25)</td>
<td>100.97</td>
<td>105.11</td>
<td>111.61</td>
</tr>
<tr>
<td>Low (n = 25)</td>
<td>115.15</td>
<td>109.18</td>
<td>133.62</td>
</tr>
<tr>
<td>Total (N = 50)</td>
<td>108.06</td>
<td>107.17</td>
<td>122.62</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses.

Figure 1. Inference response times for the low group.  
Figure 2. Inference response times for the high group.

215
4.4 Inference Error Rates

Finally, error rates for inference questions were computed. As shown in Table 5, accuracy was comparable across implicit and explicit conditions for FA texts for both proficiency groups. However, for PU texts, error rates were 12% larger in the implicit than explicit condition for the low group, whereas for the high group, implicit-error rates were only 2% larger than explicit ones. This is consistent with the above finding that the high group made on-line causal inferences for PU texts, but the low group did not. In addition, for UF texts, both proficiency groups’ error rates increased to the larger extent than the other texts in the implicit versus explicit condition, suggesting that it is difficult to generate inferences from unfamiliar texts.

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Familiar</th>
<th>Partially Unfamiliar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Implicit</td>
<td>Explicit</td>
<td>Implicit</td>
</tr>
<tr>
<td>High (n = 25)</td>
<td>.03 (.08)</td>
<td>.02 (.07)</td>
<td>.08 (.12)</td>
</tr>
<tr>
<td>Low (n = 25)</td>
<td>.04 (.09)</td>
<td>.03 (.08)</td>
<td>.20 (.25)</td>
</tr>
<tr>
<td>Total (N = 50)</td>
<td>.04 (.09)</td>
<td>.03 (.08)</td>
<td>.13 (.20)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses.

4.5 Discussion of Findings Regarding Low-proficiency Readers

For the low group, on-line causal inference generation was not supported regardless of Familiarity condition, as correct response times for inference questions were significantly longer in the implicit condition for all texts. In particular, the failure of low-proficiency readers to make the inferences during the reading of FA texts supports the view that one needs to have a certain level of reading proficiency to utilize relevant knowledge (e.g., McNamara & O'Reilly, 2009). In other words, there may be a certain threshold of L2 reading proficiency that must be reached before EFL learners can benefit from text familiarity in inference generation.

At the same time, however, this observation runs counter to the assumption of the validation model that understanding causal relations inherently entails activating relevant knowledge underlying the causality (See Section 1.1). Regarding this, an inspection of inference error rates might provide an explanation. That is, the low group’s inference error rates for FA texts were comparable between the implicit (M = .04, SD = .09) and explicit conditions (M = .03, SD = .08). This raises the possibility they made the inferences when answering the question, that is, when asked to do so by the question. Specifically, longer response times for verifying implicit inference questions might reflect attempts to consciously infer implicit ideas at that time. Then, they successfully made the inferences for FA texts, yielding relatively similar error rates across the Explicitness conditions. It is necessary to note that in this study participants were asked to answer “no” when the question was true based on common sense. This was intended to ensure that they could not provide correct answers unless they inferred it from the texts (see the Procedure section).
With this taken into account, it is possible that the low group intentionally or strategically generated causal inferences when deciding if queried ideas were inferable from the text. This view in turn suggests that low-proficiency learners’ difficulties in inference generation seem not to lie in generation itself, but in the automaticity of generation during reading.

In contrast, results for PU texts indicated that it was not easy for low-proficiency readers to infer unfamiliar ideas, even when the inference question explicitly required it. This is reflected in inference error rates, which were larger in the implicit versus explicit condition (Implicit: $M = .20$, $SD = .25$ vs. Explicit: $M = .08$, $SD = .16$). Past research states that besides processing literal-level information, making unfamiliar inferences requires readers’ engagement in constructive activities to make sense of texts (Millis et al., 2006). Thus, unfamiliar inferences might be particularly difficult for low-proficiency readers, whose cognitive resources are typically limited to lower-level processing. To address this, an independent samples $t$ test was additionally conducted comparing error rates for detail questions between the two proficiency groups. This revealed significantly lower accuracy for the low group ($M = .20$, $SD = .08$) than the high group ($M = .13$, $SD = .05$), $t(48) = 3.30$, $p = .002$, $r = .43$. Thus, low-proficiency readers seemed to struggle more with literal-level understanding compared to the high group, which in turn makes it more challenging to generate unfamiliar inferences. In sum, these results suggest that (a) low-proficiency readers can make familiar inferences when the task requires them to do so, but (b) making unfamiliar inferences is difficult for them due to limited cognitive resources.

4.6 Discussion of Findings Regarding High-proficiency Readers

For high-proficiency readers, the results suggested that they made causal inferences during reading FA and PU texts. Target reading times for FA and PU texts were significantly longer in the implicit condition, whereas correct inference response times were relatively similar across the Explicitness conditions. Research to date has only considered one type of familiarity (text or inference) at a time (e.g., Noordman et al., 1992; Singer et al., 1997), and no agreement has been reached regarding conditions under which EFL learners make causal inferences during expository reading. In contrast, the present study tested two types of familiarity simultaneously and extended previous findings by showing that EFL readers can make on-line causal inferences when both L2 reading proficiency and text familiarity are high.

In particular, the evidence that the high group generated causal inferences for PU texts suggests that EFL readers can infer what they did not already know if their L2 reading proficiency and familiarity of text content are sufficiently high. In cognitive terms, generating unfamiliar inferences can be interpreted as deriving new information, necessitated by a coherent understanding of the text. At this point, it might be easier for high-proficiency readers, who can process literal-level information relatively efficiently, to engage in such constructive activities if texts are short, plain, and familiar, like the ones used in this study. Indeed, two pieces of evidence support this view. First, the high group showed significantly more accurate understanding of text...
details than the low group, as reflected by error rates for detail questions (see the previous section). Second, when looking at PU texts, inference error rates for the high group were comparable between the implicit ($M = .08, SD = .12$) and explicit ($M = .06, SD = .11$) conditions. These converge to suggest that the high group might be able to devote more cognitive resources to inference processing than the low group by efficiently processing literal-level information. Of course, claims about processes that occurred during reading must be interpreted with caution, as this study only included quantitative analyses. Nevertheless, given high-proficiency readers' relatively accurate understanding of explicit and implicit information, it is possible that they made the inferences to causally understand expository texts whose content is familiar.

In contrast, UF text results indicated that even proficient readers had difficulty making the inferences from unfamiliar texts. This meshes with previous findings showing that knowledge-driven reading is impaired for unfamiliar texts, increasing the cognitive burden on lower-level processing (e.g., Horiba, 2000). Thus, readers might be overloaded trying to achieve literal-level understanding during reading unfamiliar texts, and thus inference processing suffered.

5. Conclusion

The present study examined EFL readers' causal inference generation during reading expository texts, considering two kinds of familiarity (inference ideas and text content) and L2 reading proficiency. The answer to the RQ is that EFL readers make on-line causal inferences when both text familiarity and L2 proficiency are sufficiently high; proficient readers made causal inferences during reading familiar texts. On the other hand, less proficient readers did not so even when text content is familiar. However, inference error rates suggest that they made familiar inferences when required by the task. Of primary significance is the fact that proficient readers made unfamiliar inferences when text content was familiar. Making unfamiliar inferences can be construed as a process of learning, given that what is inferred did not exist in readers' prior knowledge (Singer & O'Connell, 2003). From this perspective, the present study is particularly informative in that it demonstrates that inference generation plays an important role in the acquisition of implicit ideas from texts, as well as in the coherent understanding of explicit information.

Furthermore, the present findings may have pedagogical implications for reading instruction tailored to students with different reading skills. First, students with high reading proficiency are likely to make inferences from expository texts if they have sufficient domain knowledge. Therefore, after providing background information on texts, teachers should promote inference generation to enhance students' ability to derive implicit meaning. This includes having students explain causal mechanisms in the texts (i.e., self-explanation) or providing "why" questions that require meaningful integration of multiple text ideas during reading. In addition, students should be given inference questions after reading. This will allow them to confirm the
content of inferences and stabilize the inferred ideas. At the same time, it is important to pay attention to students with low reading proficiency, because they may have difficulty making inferences even from familiar materials. For such students, teachers should first try to reduce the burden with lower-level processing to make inference processing more likely. For example, it may be helpful to increase discourse markers that signal relations between ideas or insert keywords to increase argument and conceptual overlap. Furthermore, the fact that the low group displayed evidence of task-induced inference generation suggests that providing inference questions after reading may be effective for leading low-proficiency students to inference-level understanding.

Finally, it is important to mention limitations of this study that merit attention in future research. First, details of reading processes or strategies employed during comprehension are not fully clarified. Making unfamiliar inferences should involve constructive activities to search for meaning. However, it is not prudent to derive definitive conclusions from the present quantitative findings alone; studies are needed to more thoroughly investigate how readers process texts, with measures such as the think-aloud. Second, it is necessary to assess learning outcomes from inference processing. Productive tasks asking readers to apply information to new situations will meet this purpose, including providing explanations for how events unfold or problem-solving procedures. Third, we still do not know the exact component of L2 reading proficiency that most contributes to inference generation. Although the present study considered L2 reading proficiency as an ability to appropriately and accurately understand what the text communicates, it is likely that an ability to access prior text or relevant knowledge matters with longer or more complicated texts. Future work addressing these will give clues about how to develop students who learn new knowledge from texts, as well as a better understanding of discourse processes in EFL learners.

Note

1. Descriptive interpretation was deemed appropriate because floor effects were observed according to the criteria in Oshio (2004).

Acknowledgements

I deeply thank Prof. Yuji USHIRO and three anonymous reviewers for their valuable comments.

References
