Studies in Second Language Acquisition, **43** (2021), 133–154
doi:10.1017/S0272263120000443

Research Article

PREDICTIVE PROCESSING OF IMPLICIT CAUSALITY IN A SECOND LANGUAGE

A VISUAL-WORLD EYE-TRACKING STUDY

Hyunwoo Kim *

Yonsei University

Theres Grüter

University of Hawai'i at Mānoa

Abstract
Implicit causality (IC) is a well-known phenomenon whereby certain verbs appear to create biases to remention either their subject or object in a causal dependent clause. This study investigated to what extent Korean learners of English made use of IC information for predictive processing at a discourse level, and whether L2 proficiency played a modulating role in this process. Results from a visual-world eye-tracking experiment showed early use of IC information in both L1 and L2 listeners, yet the effect was weaker and emerged later in the L2 group. None of three independent and intercorrelated proficiency measures modulated L2 listeners' processing behavior. The findings suggest that L2 listeners are able to engage in prediction during real-time processing at a discourse level, although they did so to a more limited extent than native speakers in this study. We discuss these findings in light of similar evidence from other recent work.

INTRODUCTION

Discourse processing calls upon integration of multiple information sources, including explicit and implicit cues, requiring comprehenders to establish relations between a series of events being described (Graesser et al., 1994). Of the various types of cues employed to establish reference and coherence relations in discourse, verb-related referential biases are

This research was part of the first author's PhD dissertation research, funded by a Doctoral Dissertation Research Improvement grant from the National Science Foundation (BCS-1749240). We thank two anonymous reviewers and the editor of *SSLA* for their insightful feedback and suggestions. We are also grateful to Shinichiro Fukuda, Kamil Ud Deen, William O'Grady, and Bonnie D. Schwartz for their feedback, and Amber B. Camp and Fred Zenker for help with various aspects of this research. Supplementary materials are publicly available at https://www.iris-database.org/iris/app/home/detail?id=york:938107

* Correspondence concerning this article should be addressed to Hyunwoo Kim, Department of English Language and Literature, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea. E-mail: hyunwoo2@yonsei.ac.kr

known to play a crucial role (Kehler, 2002). For example, a substantial body of research shows that adult native speakers use information implicitly encoded by the verb to construct event representations associated with causally related clauses, such as those in (1).

(1) a. Jane bored Mary because she …
b. Jane feared Mary because she …

It is well established that certain interpersonal verbs trigger explanations that focus on one of the two arguments as the underlying cause of an event when followed by a *because*-clause. In (1a), for instance, there tends to be an overall preference for comprehenders to consider *Jane* as the underlying cause of the "bore" event, and hence to interpret *she* as coreferential with *Jane.* The verb *fear* in (1b), by contrast, is more likely to give rise to an interpretation in which *Mary*, the object of the main clause, is the original cause of the "fear" event, thus making *Mary* a more likely antecedent of *she.* The phenomenon that some interpersonal verbs like *bore* and *fear* induce biases to remention either their subject or object in a causal dependent clause when followed by an explanation is known as *Implicit Causality* (IC; Au, 1986; Brown & Fish, 1983; Garvey & Caramazza, 1974; Hartshorne, 2014). Verbs that induce a bias toward the subject as in (1a) are referred to as "NP1-biased verbs," and verbs with a bias toward the object as in (1b) are called "NP2-biased verbs." Numerous studies, discussed in more detail in the following text, have reported that native speakers make use of this information consistently and rapidly during the comprehension of sentences such as (1) (e.g., Cozijn et al., 2011; Ferstl et al., 2011; Hartshorne & Snedeker, 2013; Itzhak & Baum, 2015; Pyykkönen & Järvikivi, 2010).

While there is a general consensus that IC biases are associated with properties of verbs, researchers have long debated what aspects of a verb contribute to these biases. Some early accounts characterized IC as an arbitrary phenomenon where the strength and direction of a verb's bias is inherently encoded in verbs, just as grammatical gender is arbitrarily encoded in nouns in languages like Spanish (e.g., Garvey & Caramazza, 1974). Others claimed that language users' world knowledge about events typically described by IC verbs plays a key role in determining IC biases (Brown & Fish, 1983). More recent approaches have provided evidence that IC biases may be explained, at least in part, by the verb's semantic structure, in particular the thematic roles of its arguments (e.g., Bott & Solstad, 2014; Hartshorne & Snedeker, 2013). The origin of IC bias effects remains a matter of ongoing debate, which is relevant to but does not directly impact the goals of the present study (for further discussion, see Hartshorne, 2014).

Another issue of long-standing debate is the time course of the use of IC information during real-time processing. Two contrasting accounts can be identified: the integration account and the focusing account. The integration account proposes that IC information is only activated when disambiguating information becomes available, usually at the end of a causal dependent clause (Garnham et al., 1996; Stewart et al., 2000). In contrast, the focusing account claims an earlier activation of IC information, starting at or around the causal conjunction, and critically before any disambiguating information is encountered (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; Long & De Ley, 2000). Several recent studies using the visual-world eye-tracking paradigm have provided evidence consistent with the focusing account by demonstrating that native speakers

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

preferentially look at the causally implicated referent at or soon after they have processed the pronoun in a causal dependent clause (e.g., Cozijn et al., 2011; Itzhak & Baum, 2015; Järvikivi et al., 2017). The early activation of IC information suggests that speakers do not simply integrate incoming information in a passive, incremental manner, but they proactively generate expectations for upcoming information, providing support for predictive mechanisms operative at a discourse level in first language (L1) processing. We discuss these accounts and the evidence adduced to support them in the context of L1 processing in more detail in the next section.

Importantly, the predictive mechanisms assumed to be involved in discourse-level processing are probabilistic rather than deterministic in nature. In other words, an NP1-biased IC verb followed by a causal conjunction is assumed to lead the comprehender to expect that NP1 is *more likely* than NP2 to be rementioned in the dependent clause. It is this sense of "prediction" that is invoked here, following Van Berkum et al. (2007), who proposed comprehenders "use the implicit causality cue in something like 'David praised Linda because ...' *proactively*, and essentially *predict*, before the pronoun comes along, that the remainder of the sentence will tell us something about Linda" (p. 167; italics in the original). It is important to note that this usage of the term "prediction" differs from that in work on syntactic processing, where predictive effects are often deterministic, as for example in filler-gap constructions, where a filler deterministically predicts an upcoming gap (e.g., Omaki et al., 2015; for L2, see Kaan et al., 2010; for discussion of different usages of the term "prediction" in the pycholinguistic literature, see also DeLong et al., 2014 and Kuperberg & Jaeger, 2016).

Despite the extensive research on the time course of IC processing in L1, relatively little is known about this issue in second language (L2) processing. Although there is some evidence that L2 learners are able to use verb-related bias information (Cheng & Almor, 2017, 2019; Kim & Grüter, 2019; Liu & Nicol, 2010), it remains largely unclear if, when, and to what extent L2 learners draw upon IC information to create proactive expectations about upcoming reference during real-time comprehension. This article attempts to address this question through a visual-world eye-tracking experiment conducted with both L1 and L2 speakers of English. Findings from this study will bear directly on the current debate on the role of prediction in L2 processing (Grüter et al., 2017; Kaan, 2014). In contrast to the well-attested evidence of prediction in L1 processing, there remains controversy over whether L2 learners proactively use linguistic information to the same extent as L1 speakers. While it appears that L2 learners are generally more restricted than L1 speakers in the predictive use of linguistic cues during processing (e.g., Grüter et al., 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013), the degree of anticipatory processing in L2 has been argued to be dependent upon several factors, including L2 proficiency, cross-linguistic competition, quality of L2 lexical representations, and cognitive abilities (Dijkgraaf et al., 2017; Hopp & Lemmerth, 2018; Kaan, 2014; Van Bergen & Flecken, 2017). The relative contribution of these and other potentially modulating factors, however, remains poorly understood (for review, see Kaan, 2014). Moreover, previous research on the role of prediction in L2 processing has mostly focused on lexical, morphosyntactic, and syntactic cues, whereas few studies have investigated predictive mechanisms in L2 discourse processing. This study aims to contribute to our understanding of this understudied domain by assessing to what extent L2 learners use information encoded in IC verbs to create proactive, probabilistic expectations about upcoming discourse referents in causal dependent clauses. To further investigate L2 proficiency as a key factor proposed to modulate prediction in L2 processing (Kaan, 2014), we strived to include sequential L2 learners (i.e., adult Korean-speaking learners who learned English as a foreign language in instructional settings), with a range in proficiency (intermediate to highly proficient), and assessed their English language skills through multiple independent measures. This allowed us to ask if and how L2 speakers make use of IC information for predictive processing during online spoken comprehension, and whether there are any modulating effects of L2 proficiency in this process.

USE OF IC INFORMATION IN L1 PROCESSING

Previous research has provided ample evidence that adult L1 speakers actively use IC bias in offline production and online comprehension. In offline sentence-completion tasks, speakers preferably provide continuations congruent with the verb's IC bias when they read sentences containing an IC verb in the main clause followed by *because* in the ensuing clause (e.g., Bott & Solstad, 2014; Featherstone & Sturt, 2010; Ferstl et al., 2011; Garvey & Caramazza, 1974; Hartshorne & Snedeker, 2013). In online comprehension tasks, comprehenders rapidly recruit IC information to generate expectations about who will be mentioned next and resolve the reference of an ambiguous pronoun in the second clause (e.g., Cozijn et al., 2011; Itzhak & Baum, 2015; Järvikivi et al., 2017; Pyykkönen & Järvikivi, 2010). These findings suggest that IC information serves as a robust cue for native speakers to construct event representations associated with causal coherence across clauses.

Unlike the convergent findings regarding the use of IC information in L1 processing, an issue that remains less conclusive is when this information becomes available during real-time language use. The integration account assumes that comprehenders defer their integration of IC information until disambiguating cues are provided as the discourse unfolds over time. Specifically, this account predicts that the IC effect should emerge at the point where the causality information fully disambiguates the embedded subject pronoun, that is at/near the end of the sentence or when disambiguating words are provided (Garnham et al., 1996; Stewart et al., 2000). In support of this account, Garnham et al. (1996) showed that effects of IC among English speakers emerge only after the presentation of disambiguating information. In a series of probe recognition tasks, participants read sentences containing an IC verb in the main clause, followed by either bias-consistent (e.g., *David approached Brian after school because he wanted some advice*) or bias-inconsistent continuations (e.g., *David approached Brian after school because he looked friendly*) on a word-by-word basis. During the task, the probe name indicating either the subject (*David*) or the object (*Brian*) of the main clause appeared at the top of the screen at one of three time points: before the pronoun (150 ms after the offset of *because*), after the pronoun (150 ms after the offset of *he*), and at the end of the sentence. Garnham et al. (1996) found that when participants were prompted to press a button as soon as they recognized the probe name, they had longer response times in the sentences with bias-inconsistent endings than in those with bias-consistent endings, but only when the probe name appeared at the end of the sentence; no differences between bias-consistent and -inconsistent sentences were observed when the probe name was presented before or after the pronoun. Although the researchers concluded that these results support the integration account, it should be noted that there exist some methodological limitations in their study. As Stewart et al. (2000) pointed out, the probe recognition task may have placed substantial demands on processing resources as participants were engaged in a dual task where they simultaneously read a target sentence and responded to a probe name, which could have prevented them from detecting the probe names in earlier regions. In addition, potential effects of IC bias were explored at only three regions—immediately preceding/following the pronoun and at the end of the sentence. It is thus possible that IC effects may have spilled over into the regions not specified in the task.

Several more recent studies of IC processing in L1 have used the visual-world eye-tracking paradigm, which involves lower task demands and allows for more ecologically valid and temporally sensitive investigations of real-time sentence processing. The linking hypothesis adopted in these studies is that listeners will be more likely to look at the image of a referent in a visual scene when they expect that referent to be rementioned in the upcoming discourse. (For discussion of this linking hypothesis and evidence from the visual-world paradigm in support of it, see Grüter et al., 2018.) These studies thus measure participants' relative likelihood to look at images of the NP1 and NP2 referents in a visual display while they listen to sentences that include NP1 and NP2 protagonists in the subject and object positions along with an IC verb in the main clause, followed by a *because*-clause that provides an explanation of the event described in the main clause. Collectively, findings from these studies have indicated that IC information becomes available much earlier than what would be predicted by the integration account (e.g., Cozijn et al., 2011; Itzhak & Baum, 2015; Järvikivi et al., 2017; Pyykkönen & Järvikivi, 2010). For example, in a study with native speakers of Dutch, Cozijn et al. (2011) found that speakers fixated on the bias-consistent referent significantly more often than the bias-inconsistent referent both for NP1- and NP2-biased verbs soon after the pronoun had been processed. Their findings are taken to support the focusing account because the IC effect emerged before participants encountered any disambiguating information in the subordinate *because*-clause. Similar results were reported by Itzhak and Baum (2015) from a study with native English speakers. While the main focus of their study was on the effect of prosody on speakers' use of IC information, they also reported a separate analysis of the condition without prosodic information (their "No-Accent" condition). In this condition, there were reliable effects of IC bias, indicated by significantly more fixations on NP1 referents following NP1-biased verbs, and more fixations on NP2 referents following NP2-biased verbs. These effects extended from pronoun onset to 400 ms after pronoun onset, consistent with the focusing account.[1]

In sum, while the debate on the timing issue of IC bias in L1 processing has not been resolved conclusively, the more recent evidence points to the general conclusion that native speakers can use IC information quite early during sentence processing, in particular before they encounter information that disambiguates the reference of the subject pronoun in the subordinate *because*-clause. For more extensive discussion of the debate on the time course of IC effects in L1 processing, the reader is referred to Koornneef et al. (2016).

USE OF IC INFORMATION IN L2 PROCESSING

Despite extensive evidence for native speakers' use of IC in reference interpretation and processing, only a few studies have examined this issue in L2 processing (Cheng & Almor, 2017, 2019; Contemori & Dussias, 2018, 2019; Liu & Nicol, 2010). These studies provide evidence that L2 learners are able to exploit IC information to resolve reference, although they may not always do so in the same way or to the same extent as L1 speakers. Moreover, the timing of L2 learners' use of this information remains largely unexplored. The only published evidence of the use of IC during real-time listening in a nonnative language comes from a visual-world experiment that included highly proficient Spanish–English bilinguals (Contemori & Dussias, 2018, 2019). In a first report of the findings from this study, Contemori and Dussias (2018) concluded that bilingual listeners can use IC information, but "show a significant delay in the online activation of the implicit causality bias" (p. 170) compared to monolingual English-speaking listeners. Notably, IC effects in this analysis were observed only with NP1 verbs; no significant effects were observed with NP2 verbs for either monolingual or bilingual listeners. Moreover, the conclusion of a delay in the bilingual group was based on the results of within-group analyses, in which the time point of divergence of looks to NP1 versus NP2 referents was determined separately in each group, and was found to occur approximately 400 ms earlier in the monolingual than in the bilingual group. Critically, in a combined analysis of data from both groups, no significant interactions with group were found. In a reanalysis of the data from this study, Contemori and Dussias (2019) came to the somewhat different conclusion that "bilinguals are not slower than monolinguals at processing the information associated with the IC of the verb" (p. 1). Thus while the study by Contemori and Dussias presents an important and notable first step toward the exploration of IC effects in L2 processing, the results from this study were not fully conclusive, indicating a need for further investigation. Moreover, it is important to note that the bilingual group in Contemori and Dussias (2018) was comprised of highly proficient speakers of English with mostly extensive length of residence in the United States and mean first exposure to English at age 6. It thus remains unclear to what extent findings from this study may generalize to other types of bilinguals, especially sequential learners and nondominant speakers. In an attempt to extend the investigation of the role of IC information in L2 processing, and to further explore the potentially modulating role of proficiency in this process, we conducted a visual-world eye-tracking experiment with nonimmersed Korean learners of English with varying L2 proficiency, assessed through multiple independent measures.

METHODS

PARTICIPANTS

This study included a total of 112 participants, consisting of 56 Korean learners of English (NNS) and 56 native English speakers (NS). Data from eight participants (four in the NNS and four in the NS group) were excluded due to low quality of the eye gaze record (see the following text for details), leaving 52 in the NNS (42 females, mean age = 25 years, $SD = 4.0$) and 52 in the NS group (31 females, mean age = 21.9 years, $SD = 2.8$) in the final analysis.

Participants in the NS group were recruited among the student population at University of Hawaii and self-identified as native speakers of English. They received partial course credit for their participation. The participants in the NNS group were recruited from Ewha Womans University and Sogang University in South Korea and received the Korean equivalent of $20 for their participation. According to their responses in the language background questionnaire, these participants had started learning English at the mean age of 8.8 years (SD = 2.8). Fifteen participants reported having stayed in an English-dominant country for more than 6 months (mean = 26 months, range: 6–60).

The NNS participants' English proficiency was measured through an English lexical decision task (LexTALE; Lemhöfer & Broersma, 2012), a written cloze test (Brown, 1980), as well as participants' self-rating of their speaking, listening, reading, and writing skills on a 1–10 scale. The construct of language proficiency is notoriously difficult to define and measure (Hulstijn, 2011), and there is currently no agreed-upon measurement of proficiency in Second Language Acquisition (SLA) research (Schoonen, 2011). For this reason, we opted to include three very different but all widely used measurement tools for overall English proficiency so as to be able to examine their concurrent validity before exploring the ability of these measures to explain relevant variance in predictive processing. The LexTALE is a 60-item lexical decision task in which learners indicate whether a string of letters constitutes an existing English word or not. In a validation study with Dutch- and Korean-speaking learners of English, LexTALE scores showed strong correlations (rs > .6) with other measures of proficiency, such as the Quick Placement Test (2001), a standardized test for general proficiency (Lemhöfer & Broersma, 2012). The Brown cloze test consists of a written passage from which 50 words were deleted, and learners are asked to fill in the gaps. It is one of the most widely used and researched cloze tests in the SLA literature, and has been demonstrated to have high reliability and validity across a broad range of learner groups (Brown & Grüter, in press). Finally, self-report measures of proficiency have been shown to have respectable reliability and validity in a number of contexts (Marian et al., 2007).

Results from these three tasks, summarized in Table 1, indicate a relatively broad range of proficiency within this NNS group. Proficiency measures were all at least moderately and significantly correlated with each other (all rs > .4, all ps < .01), with the exception of cloze test scores and self-ratings for listening (r = .263, p > .1).

MATERIALS

Linguistic stimuli for the eye-tracking experiment consisted of 24 items with NP1-biased verbs and 24 items with NP2-biased verbs. The IC verbs were carefully selected to satisfy the following criteria. First, verbs were selected from consistent VerbNet classes (Kipper et al., 2008), an extended version of Levin's (1993) taxonomy of verb argument structure, to maintain verbs' semantic structures as similar as possible within each bias type. The

TABLE 1. NNS participants' English language proficiency: descriptive statistics

Measures	Mean	Range
Self-ratings (1–10)		
Listening	7.5	4–10
Speaking	6.0	1–10
Reading	7.4	4–10
Writing	6.1	2–10
LexTALE (0–100)	70.6	50–96.25
Cloze test (0–50)	27.3	5–45

sentences in the NP1 condition included verbs in class 31.1 (Stimulus-Experiencer verbs, e.g., *bother*, *disturb*, *frighten*), and the sentences in the NP2 condition included verbs selected from class 31.2 (Experiencer-Stimulus verbs, e.g., *admire*, *fear*, *love*) and class 33 (Judgment verbs, e.g., *blame*, *criticize*, *thank*) in VerbNet. Second, we made sure that the selected verbs were listed in English textbooks used in Korean middle and high schools and in the vocabulary list for the Korean SAT test, to reduce potential difficulties associated with lexical processing for L2 learners. Third, the verbs were assessed for familiarity to L2 learners through an independent translation task with 20 Korean learners of English who did not participate in the main experiment. Items with fewer than 80% correct translations were discarded. Finally, the naturalness of verbs in constructions with human referents in subject and object positions was assessed through a fill-in-the-blank task (e.g., — *disturbed Owen*; *Nathan disturbed* —) with 23 native English speakers who did not participate in the main experiment. Only verbs for which human referents were inserted in both positions were retained.

Each item was comprised of three sentences: a context sentence, a critical sentence, and a follow-up question, as in (2). NP1 and NP2 referents were always of the same gender (half male, half female). To prevent participants from strategically associating the target pronoun in the *because* clause with a bias-consistent referent, half the critical sentences had an ending that was intended to be bias-consistent (2a), and half an ending intended to be bias-inconsistent ([2b]; Cozijn et al., 2011; Itzhak & Baum, 2015). Questions always asked about the referent of the ambiguous pronoun and were included only to keep participants attentive. Responses to these questions will be affected both by IC-bias of the verb as well as by the semantic content of the sentence ending and are thus not informative with regard to the research questions under investigation. For this reason, responses to the final question are not further analyzed, and no trials were excluded based on mouse-click responses.

(2) Examples of linguistic stimuli from the eye-tracking task (NP1-biased)

(a) Bias-consistent ending

(Context) Nathan and Owen used to study together at the library.

(Critical) Nathan disturbed Owen all the time because he needed help with his homework.

(Question) Who needed help with his homework?

(b) Bias-inconsistent ending

(Context) Patrick and Curtis were solving math problems in class.

(Critical) Patrick bothered Curtis every few minutes because he was the smartest kid in class.

(Question) Who was the smartest kid in class?

In addition to the 48 items with IC verbs, 48 items with predicates with no known IC biases were included as fillers, using connectors other than *because* (e.g., *before*, *although*, *while*, *but*, and *yet*; e.g., *Natalia tried to start a conversation but Lily didn't seem to notice*). A list of all items is provided in the supplementary materials.[2]

Linguistic stimuli were recorded by a female native speaker of English using broad-focus intonation (i.e., no emphasis was produced on any of the event participants). To assess whether the auditory stimuli were matched in terms of duration across conditions, independent samples *t*-tests were conducted to compare durations and onset times for

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

several regions of interest in the critical sentences between NP1 and NP2 conditions. Results showed that the critical sentences in the NP1 condition were not significantly different from those in the NP2 condition in terms of total duration (t(46) = 1.570, p = .123), duration of the NP2 region in the main clause (from NP2 onset to the onset of *because*) (t(46) = –1.387, p = .172), onset of *because* (t(46) = 0.865, p = .392), and onset of the pronoun in the *because*-clause (t(46) = 0.909, p = .368).

For the visual stimuli, 96 scenes, each comprised of two clipart images of human faces, were created. Each visual scene contained two areas of interest (AOIs), one corresponding to the main-clause subject (NP1) and the other to the main-clause object (NP2). Names were printed below each face, as illustrated in Figure 1. The position of bias-consistent and bias-inconsistent referents was counterbalanced between items.

Experimental and filler items were pseudorandomized so that no experimental items in the same bias condition occurred more than once in a row. Two lists were created in reversed orders, and participants were randomly assigned to one of the two lists.

In addition to the eye-tracking task, a translation task was conducted to assess L2 learners' knowledge of the target verbs. The items in the translation task consisted of the main clause portion of the 48 experimental items from the eye-tracking task (e.g., *Nathan disturbed Owen*).

PROCEDURE

All participants completed the eye-tracking task, after which the NNS group additionally completed the LexTALE, the cloze test, and the translation task. Prior to these tasks, all participants completed a language background questionnaire.

During the eye-tracking experiment, participants were seated at a comfortable distance from a laptop equipped with a remote eye-tracker below the screen. Prior to the experiment, they received written and oral instructions to listen to the sentences and answer a question. Eye movements were recorded from the right eye with an SMI REDn Scientific eye-tracker with a sampling rate of 60 Hz. The experiment began with calibration and three practice trials. Visual scenes were presented 1,000 ms before the onset of linguistic

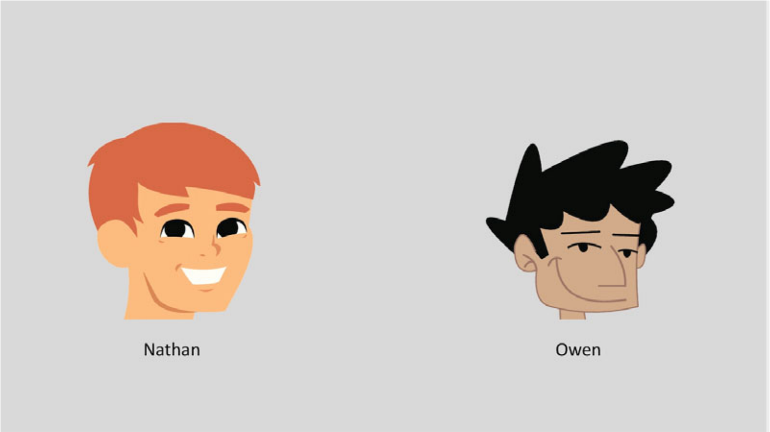

FIGURE 1. Sample of visual stimuli.

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

stimuli and remained on the screen for the duration of the context, critical, and question sentences. There was a 1,000 ms pause between the context and critical sentences and between the critical and question sentences. The question, which served primarily to keep participants engaged in the task, queried about the referent of the ambiguous pronoun in the critical sentence (experimental items) or one of the referents in the main clause (fillers). Participants responded to the question by clicking on one of the two images. No feedback was provided for their answers. After answering the question, participants proceeded to the next trial by pressing the spacebar. After finishing half the trials, participants took a break about 5–10 minutes. The next block started with recalibration.

Following the eye-tracking task, NNS additionally completed the LexTALE task, cloze test, and translation task. In the LexTALE, participants saw a series of letter strings (60 items total) on a computer screen and were asked to decide whether or not the string was an existing English word by clicking on the "yes" or "no" button on the screen. The cloze test consisted of three English paragraphs in which a total of 50 phrases had been replaced with blanks. Participants were asked to fill in the missing word for each blank. Responses were scored using acceptable-response scoring, as outlined in Brown (1980). During the translation task, the NNS participants were asked to provide a written Korean translation for each English sentence as accurately as they could. The entire test session took approximately 40 minutes for NS and 80–90 minutes for NNS participants.

DATA TREATMENT AND ANALYSIS

Following previous work using the visual world eye-tracking paradigm to investigate referential processing in real-time discourse comprehension (e.g., Arnold et al., 2000; Kaiser et al., 2009), the linking hypothesis underlying this study is that listeners will probabilistically look at visually represented referents that they expect to be mentioned in the upcoming discourse. Based on this assumption, our goal was to capture potential effects of IC information on the allocation of participants' eye gaze during the causal dependent clause. To this end, stimuli were aligned at the acoustic onset of *because* for further analysis of the eye-gaze data.

Participants' eye gaze during the critical sentence was classified automatically as fixations, saccades, and blinks using SMI Experiment Suite default settings. Fixation data were subsequently aggregated into 20 ms time bins. Trials with insufficient fixation data due to track loss were identified by calculating the proportion of sample points over the entire trial containing fixations. Trials with fixation data containing a number of sample points less than 2 standard deviations below the mean across all trials were excluded (1.5% in NS; 2.0% in NNS). Averaging over items, there was no item with a proportion of fixations below 2 standard deviations from the mean of items; thus no items were excluded. Six participants with an overall proportion of fixations below 2 standard deviations of the mean of all participants were identified and excluded from further analysis (3 NS; 3 NNS). Two additional participants were excluded, either due to calibration failure (1 NS) or because eye-tracking data were not recorded (1 NNS). This process left 52 participants each in the NS and NNS groups.

Time windows for analysis were determined following Itzhak and Baum (2015), but with some modifications. First, Itzhak and Baum (2015) included the NP2 region in the main clause as an analysis window, but they did not find an effect of IC bias in this early region. Thus, to minimize Type I error rates by conducting multiple analyses, we did not include the NP2 region in the analysis. Second, in Itzhak and Baum, the windows of analysis included a series of 200 ms windows following *because* (0–200 ms and 200–400 ms after the onset of *because*) and the pronoun (0–200, 200–400, 400–600, 600–800 ms after pronoun onset). An increased number of time windows, with each segment containing fewer sample points, is also likely to increase the chance for Type I error and may reduce statistical power. Therefore, we decided to include fewer and larger analysis windows in this study. In addition, given that L2 processing is often slower than L1 processing, we decided to include an additional, later time window as well. Based on these considerations, three successive temporal windows were determined as regions for analysis: from onset of *because* to pronoun offset (M = 520 ms, W1), from pronoun offset to 500 ms after pronoun offset (W2), and from 500 ms to 1,000 ms after pronoun offset (W3). Considering that it generally takes about 200 ms to plan and execute an eye movement (Matin et al., 1993), each analysis window was offset by 200 ms. That is, the analysis for the W1 segment included the time frame from 200 ms after the acoustic onset of *because* to 200 ms after pronoun offset. Likewise, the W2 and the W3 windows extended from 200 ms to 700 ms, and from 700 ms to 1,200 ms after pronoun offset, respectively.

For each of the three analysis windows, participants' preference for fixating on NP2 versus NP1 images was calculated for each trial by subtracting the number of 20 ms bins with looks to NP1 from the number of bins with looks to NP2 (see Grüter et al., 2018). For each time window, a linear mixed-effects model was constructed, testing for an effect of bias type (NP1-biased, NP2-biased). The analysis included *Group* (NS, NNS) and *Verb Bias* (NP1-biased, NP2-biased) as fixed effects (contrast-coded and centered around the mean), and participant and item as random effects. All models were constructed with the maximal random effects structure allowed by the design (Baayen, 2008; Barr et al., 2013; Jaeger, 2008), including a by-participant slope for *Verb Bias* and a by-item slope for *Group*. In case of convergence problems, the random effects structure was simplified by removing the by-item slope for *Group*. For each analysis, participants' proficiency scores (LexTALE scores, cloze test scores, self-ratings, combined scores) were added to the original model of the L2 data in a separate step, either as a continuous or categorical variable, to explore the potentially modulating role of L2 proficiency.

For the coding of the translation task, two coders, the first author and a highly proficient English-speaking learner of Korean blind to the purpose of the study, annotated participants' translations for accuracy. Participants' translations were coded as correct when they appeared as entries in the NAVER English–Korean dictionary (https://dict.naver.com/). Trials from the eye-gaze data were removed if the participant's response to that item on the translation task was incorrect (9%) or gave rise to intercoder disagreement (1%).

RESULTS

ANALYSES OF EYE GAZE

Figure 2 illustrates the average time-course of fixations over the course of the critical sentence. During the main clause, we see in both groups and conditions the expected sequence of looks to NP1 followed by looks to NP2, in line with the order in which the two

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

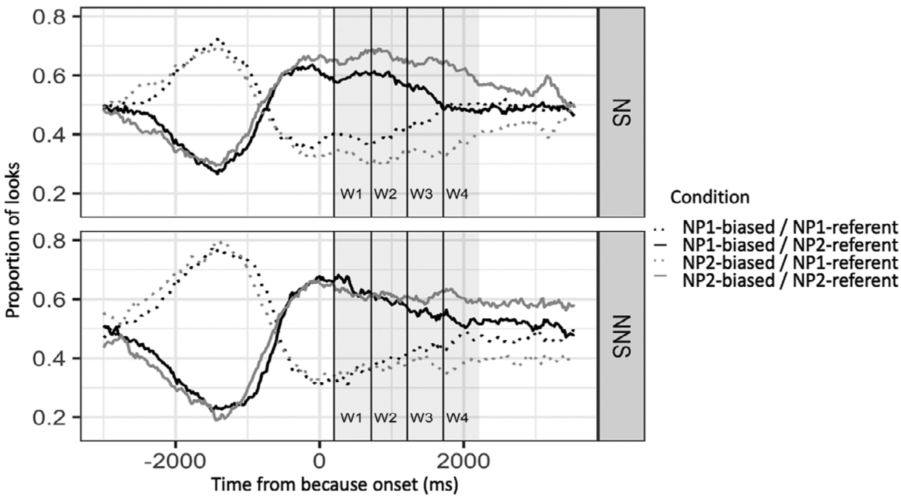

FIGURE 2. Overview of the time course of fixations across critical sentences by verb bias (NP1-biased, NP2-biased) and AOI (NP1-referent, NP2-referent) starting from *because* onset (0 ms) for NS (upper panel) and NNS (lower panel) groups. Proportions of looks to each AOI are calculated out of all fixations to NP1- or NP2-referents. Means represent means over trials. W1 = because + pronoun; W2 = 0–500 ms after pronoun offset; W3 = 500–1,000 ms after pronoun offset; W4 = 1,000–1,500 ms after pronoun offset.

referents are mentioned. By the onset of *because* (W1 region), listeners were still predominantly fixating on the most recently mentioned referent, that is, NP2. Visual inspection of these graphs suggests that despite this overall preference for persevering on the most recently mentioned referent, the NS group was more likely to look at NP1 referents in sentences with NP1-biased verbs (black dotted line) than in sentences with NP2-biased verbs (gray dotted line), and similarly more likely to fixate on NP2 referents with NP2-biased (gray solid line) than with NP1-biased verbs (black solid line). Differences between the two bias-type conditions appear to emerge even before the onset of *because.* In contrast, differences between conditions in the NNS group do not appear to emerge until substantially later. In the W3 region, a pattern similar to that in the NS group is beginning to emerge and remains stable throughout the remainder of the sentence.

To assess when effects of verb bias emerge in each group, linear mixed-effects regression (lmer) was conducted on participants' preference for fixating on NP2 versus NP1 referents for each window. Table 2 presents a summary of the output from these models. In the first window (W1), there were no main effects of *Group* or *Verb Bias*, but an interaction emerged between *Group* and *Verb Bias* ($b = -3.394$, $p = .020$). Separate analyses examining this interaction were performed for each group, with alpha adjusted to .025, with the model including *Verb Bias* as a fixed effect. The results showed a main effect of *Verb Bias* in the NS ($b = 2.897$, $p = .008$), but not in the NNS group ($b = -0.446$, $p = .710$), indicating that only the NS group showed sensitivity to the IC bias in this early window.

In the W2 region, there was an effect of *Verb Bias* ($b = 1.865$, $p = .045$), with a greater preference for fixating on NP2- versus NP1-referents in the NP2-bias than in the NP1-bias

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

TABLE 2. Results of the mixed-effects logistic regression

Window	Predictor	*b*	*SE*	*p*
W1	(Intercept)	5.113	0.711	<.001
(because + pronoun)	Group	−0.605	1.334	.652
	Verb bias	1.108	0.863	.206
	Group × Verb bias	**−3.394**	**1.397**	**.020**
W2	(Intercept)	4.407	0.687	<.001
(0–500 ms after pronoun offset)	Group	−1.686	1.225	.173
	Verb bias	**1.865**	**0.903**	**.045**
	Group × Verb bias	−2.192	1.288	.095
W3	(Intercept)	3.411	0.696	<.001
(500–1,000 ms after pronoun offset)	Group	−1.104	1.230	.372
	Verb bias	**3.171**	**0.975**	**.002**
	Group × Verb bias	−2.025	1.434	.166

Note: Formula for each model: lmer(type ~ verb.bias*group + (1 + verb.bias|participant) + (1 + group|item)).

condition. There was no effect of *Group*, but a trend toward an interaction ($b = -2.192$, $p = .095$). For comparison with the analysis of the data in W1 and a full exploration of potential differences between the NS and the NNS groups, we conducted analogous exploratory follow-up analyses within each group in all subsequent windows regardless of the significance of the interaction. These analyses revealed an effect of *Verb Bias* in the NS ($b = 2.989$, $p = .010$), but not in the NNS group ($b = 0.835$, $p = .477$) in the W2 region, suggesting that, consistent with the results from the previous window, the NSs' looking patterns, but not the NNSs', were influenced by IC bias in this region.

In the third window (W3), a main effect of *Verb Bias* ($b = 3.171$, $p = .002$) emerged without a significant effect of *Group* or an interaction, suggesting that *Verb Bias* influenced both groups. Follow-up analyses conducted within each group demonstrated a robust effect of *Verb Bias* for the NS ($b = 4.241$, $p < .001$) and a trend toward a weaker effect in the same direction for the NNS group ($b = 2.216$, $p = .089$). These results suggest an emerging role of IC bias in the NNS group in this region.

To further examine this late emerging effect in the NNS group, we decided to conduct exploratory analyses in the following 500 ms region: from 1,000 ms to 1,500 ms after pronoun offset (W4). A main effect of *Verb Bias*, not modulated by *Group*, emerged in this window ($b = 4.330$, $p < .001$). Separate analyses for each group showed a continually robust effect of *Verb Bias* in the NS group ($b = 5.430$, $p = .002$) and a weaker but increased (compared to the previous window) effect in the NNS group ($b = 3.331$, $p = .012$).

Finally, to further explore whether the native and nonnative speakers' use of IC bias increased over the course of the experiment, potentially reflecting learning or adaptation effects (e.g., Kaan et al., 2019), we conducted additional analyses for each time window including *Verb Bias* (NP1-biased, NP2-biased) and *Trial Number* as fixed effects for each group. Results from the NS group showed that the main effect of *Verb Bias* emerging in the three windows did not interact with *Trial Number.* Likewise, results from the NNS group showed that *Trial Number* did not interact with the weak effect of *Verb Bias* in W3 or the main effect of *Verb Bias* in the exploratory window. These results indicate that effects of IC bias are consistent over the course of this experiment and thus unlikely to reflect learning or adaptation within the experiment.

In sum, the NS group showed sustained evidence of using IC information, starting from (at least) the onset of *because* and stretching up to 1,500 ms after pronoun offset, whereas similar effects did not start to emerge until substantially later in the NNS group, with a weak effect of verb bias emerging in the W3 region, and becoming somewhat stronger in the 1,000–1,500 ms window. These findings indicate that IC information affects L2 learners' referential processing during online comprehension, yet to a smaller extent and at a later point compared to native speakers.

ROLE OF PROFICIENCY

To probe for effects of proficiency, scores from the two independent proficiency measures (LexTALE, cloze), an average self-rating score (averaged across scores in the four domains), and a combined score averaged over the z-scores of these three measures were individually added to the model of the NNS data. In separate exploratory models, each proficiency measure was added either as a continuous (transformed to z-score) or a categorical variable. For the operationalization of proficiency as a categorical variable, L2 participants were divided into two proficiency groups based on a median split: NNS-H group with scores at or above the median, and NNS-L group with scores below the median. Table 3 provides an overview of groupings by median score for each proficiency measure.

Figure 3 illustrates the mean time-course of fixations during the critical sentences as a function of verb bias type for each proficiency group (NNS-H, NNS-L), created based on the mean split of the combined z-scores. Visual comparison of looking patterns in the NNS-H versus NNS-L groups suggests that the late effect of verb bias may be somewhat greater for higher-proficiency learners than lower-proficiency learners.

Modulating effects of proficiency on L2 use of IC information were modeled in each of the three original time windows (W1, W2, W3), as well as in the additional exploratory window between 1,000–1,500 ms after pronoun offset, where the effect of verb bias appeared strongest in the NNS group. Each model included *Verb Bias* (NP1, NP2) and proficiency measure (either continuous or categorical) as fixed effects along with the maximal random effects structure allowed by the design. Results of these analyses showed that none of the proficiency measures significantly interacted with *Verb Bias* in any of the time windows, regardless of whether the measure was added to the model as a continuous or categorical variable. In model comparison using the anova() function in R, adding an interaction between *Verb Bias* and *Proficiency* to the model containing the two

TABLE 3. Results of median-split for each proficiency measure

	Higher proficiency group (NNS-H)		Lower proficiency group (NNS-L)	
Proficiency measure	Sample size	Mean score (*SD*)	Sample size	Mean score (*SD*)
LexTALE (median = 68.75)	28	80.4 (8.3)	24	59.2 (5.1)
Cloze test (median = 26)	28	34.3 (5.6)	24	19.1 (5.7)
Self-ratings (median = 6.8)	26	7.9 (0.7)	26	5.6 (0.9)
Combined z-score (median = −0.11)	26	0.7 (0.5)	26	−0.7 (0.4)

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

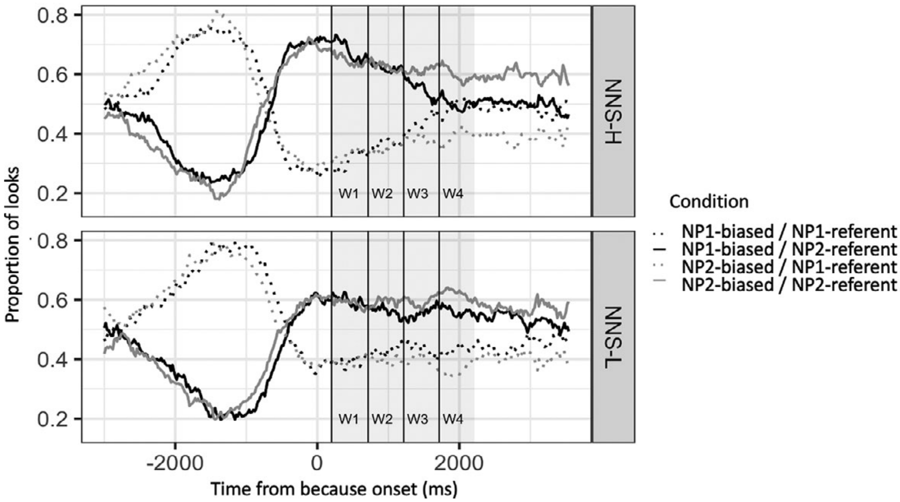

FIGURE 3. Overview of the time course of fixations across the critical sentence by verb bias (NP1-biased, NP2-biased) and AOI (NP1-referent, NP2-referent) starting from *because* onset (0 ms) for NNS-H (upper panel) and NNS-L (lower panel) groups.

fixed effects did not improve overall model fit, no matter what proficiency measure was used. These results indicate that proficiency did not play a measurable role in L2 learners' use of IC information in this experiment.

To further probe what in Figure 3 appears to be a trend toward a stronger effect of *Verb Bias* in the higher proficiency group later in the sentence, we inspected the effect of *Verb Bias* in NNS-H and NNS-L groups created by median-split of each of the proficiency measures summarized in Table 3 for each time window. The results of the separate analyses for each proficiency group showed no effect of *Verb Bias* for any of the proficiency groups in W1, W2, and W3. In the window between 1,000 ms and 1,500 ms after pronoun offset, when the L2 learners were divided by their self-ratings and combined scores, the effect of *Verb Bias* emerged somewhat more strongly for the NNS-H group (self-ratings: $b = 3.578$, $p = .043$; combined scores: $b = 3.918$, $p = .021$), than for the NNS-L group (self-ratings: $b = 2.868$, $p = .070$; combined scores: $b = 2.418$, $p = .119$). Yet overall, these findings indicate that the effect of predicate type was not robustly modulated by proficiency in this experiment.

DISCUSSION

The goal of this study was to investigate the role of implicit causality on real-time referential processing in a second language, and to explore the contribution of proficiency as a potentially modulating factor. To this end, we used the visual world paradigm to measure native and nonnative speakers' fixations on bias-consistent versus -inconsistent referents in successive temporal windows starting at the onset of *because* in a causal dependent clause following a matrix clause containing either an NP1- or an NP2-biased verb.

https://doi.org/10.1017/S0272263120000443 Published online by Cambridge University Press

Results indicated effects of IC in both groups, yet these effects differed both in size and timing. In the NS group, we observed an early and robust effect of verb bias, which persisted from the onset of *because* up to 1,500 ms after pronoun offset. The native speakers' sustained preference for fixating the bias-consistent referent from as early as *because* onset suggests they access and utilize IC information rapidly during real-time comprehension. This pattern is consistent with findings from previous visual-world studies with native speakers of Dutch (Cozijn et al., 2011), Finnish (Järvikivi et al., 2017; Pyykkönen & Järvikivi, 2010), and English (Itzhak & Baum, 2015), and provides additional evidence that IC information is readily available to native speakers well before they encounter disambiguating information. Collectively, this evidence aligns with the focusing account (Greene & McKoon, 1995; Koornneef & Van Berkum, 2006; Long & De Ley, 2000).

Turning to the results of the L2 group, our analyses indicated an emerging, nonsignificant effect of verb bias in the time window between 500 ms and 1,000 ms after pronoun offset. An additional exploratory analysis suggested that the size of this effect increased in a subsequent time window (1,000–1,500 ms after pronoun offset). When we analyzed the NS and NNS data within the same model, *Verb Bias* significantly interacted with *Group* only in the first time window (W1), suggesting that despite the absence of significant effects of verb bias within the L2 group alone in any time window, processing patterns became quite similar between the two groups after this early window. Taken together, we interpret these results as indicating that L2 listeners are able to access and make use of IC information during real-time listening, albeit at a somewhat slower rate, and potentially to a lesser extent, than what is seen in native language processing.

Importantly, as in native language processing, the (delayed and weaker) effects of IC in the L2 group nevertheless occurred at a point well before potentially disambiguating information could reasonably have been processed. On average, the late, exploratory time window spanning 1,000–1,500 ms after pronoun offset, where the strongest IC effect was observed in the L2 group, corresponds roughly to the one or two words after the verb in the *because*-clause in the linguistic signal (e.g., *Nathan disturbed Owen all the time because he needed* ***help with*** *his homework*). Yet, as half the experimental items were constructed to have a bias-consistent, and the other half a bias-inconsistent continuation in the *because*-clause, this emerging late effect of IC in the L2 group is unlikely to be attributable to information encountered later in the signal. As such, the pattern observed in the L2 group here also aligns with the focusing account, rather than the integration account proposed for IC processing among native speakers.

Consistent with a number of other recent studies investigating prediction in various other linguistic domains in L2 processing (e.g., Dijkgraaf et al., 2017; Foucart et al., 2016; Hopp, 2013; Ito et al., 2018), we found that the Korean learners of English in this study engaged in predictive behavior. At the same time, the effect differed in timing and size when compared to observations from L1 processing, an observation that has also been reported in other recent work on L2 prediction (e.g., Kaan et al., 2016). This pattern is consistent with the RAGE hypothesis, which states that L2 users have *r*educed *a*bility to *g*enerate *e*xpectations during processing (Grüter et al., 2017). Whether the reduced and/or delayed effects observed here and elsewhere are due to a reduced *ability* to make use of information predictively, or reduced engagement in predictive processes due to the potentially reduced utility of prediction given the nature of L2 knowledge, must remain for future work to disentangle.

One possible explanation for L2 learners' reduced engagement in predictive behavior may lie in learners' difficulties with accessing and retrieving lexical representations for the verbs and integrating the information to incrementally update their discourse models during real-time processing. Several studies have provided empirical evidence that L2 learners are more restricted compared to native speakers in terms of their ability to consistently access and retrieve lexical information due to lower-quality lexical representations (Lexical Bottleneck Hypothesis; Hopp, 2014, 2018), and to integrate multiple sources of information (Hopp, 2010; Roberts et al., 2008; Sorace & Filiaci, 2006) during online processing. In the experiment presented here, successful recruitment of IC information for discourse processing requires participants to access and retrieve properties of IC verbs, integrate this information with the connector *because* to establish the coherence relation between the clauses, and create mental models of the events to generate expectations about who or what will be mentioned in the following clause. While these processes appear to be accomplished relatively easily by native speakers, as shown in the early effect of IC observed in the NS group in this study, they may be more taxing for L2 learners, which could account for the delayed and weaker effect of verb bias in the L2 group in this study.

On the assumption that lexical access and retrieval become less taxing with increasing proficiency, this explanation would predict that engagement in predictive processing should be modulated by L2 proficiency. However, we found no evidence for L2 proficiency as a modulating factor, despite the inclusion of three separate and intercorrelated proficiency measures, and a learner sample with substantial range in proficiency. This null result may seem surprising, especially in light of Kaan's (2014, p. 268) observation that "various studies have shown that predictive behavior increases with increasing proficiency" (Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013). It appears worth noting, however, that our findings align with several more recent studies that explicitly aimed to assess the role of proficiency in L2 predictive processing. In particular, looking at prediction at a lexical-semantic level, neither Dijkgraaf et al. (2017) nor Ito et al. (2018) observed significant effects of proficiency. Furthermore, two studies (Hopp, 2015; Mitsugi, 2018) found significant effects of proficiency on later processes of information integration, while proficiency did not modulate earlier predictive effects in the same experiments. These latter findings indicate that the proficiency measures included in these studies were sensitive enough to capture relevant variance in language skill among the respective L2 samples; yet this variance was not associated with variance in predictive processing.

Collectively, the findings from this more recent body of work, including the results from the present study, raise questions about the generalizability of Kaan's (2014) original observation that proficiency is a modulating factor in predictive behavior in L2 processing. It may be relevant to note that two of the three studies originally cited by Kaan in support of this claim investigated prediction in the context of grammatical gender. Importantly, measures of general L2 proficiency typically correlate with L2 learners' performance on offline measures of gender assignment. This potentially confounding factor was explicitly reported in Hopp (2013). The same holds for a more recent study on the processing of grammatical gender by Russian learners of German, which found greater effects of prediction in the more highly proficient learner group (Hopp & Lemmerth, 2018); yet just as in Hopp (2013), accuracy on a gender assignment task correlated strongly with overall proficiency in this study, making it impossible to disentangle whether the observed differences were due to proficiency in general, or conditioned more directly by participants' knowledge of the linguistic property under investigation in the experiment.

Plausible limitations in study design, including limited variance in proficiency among the sample investigated; lack of sensitivity of the specific proficiency measures employed; and reduced power due to limited sample size must always remain under consideration when interpreting null effects. It is possible, for example, that results from learners with more extensive immersive L2 experience would be different (but see Kim & Grüter, 2018, for findings showing that immersed and nonimmersed Korean learners of English did not show different IC biases in an offline sentence completion task). We believe that collective evidence across multiple studies, including the present one, with various L2 groups, different proficiency measures, and targeting a variety of linguistic constructions, has now accumulated to suggest that overall L2 proficiency does not play a critical role in L2 speakers' engagement in predictive processing. If future work can confirm a differential role of proficiency in predictive versus integrative processing—as suggested by Hopp (2015) and Mitsugi (2018)—this would constitute an important new insight into the nature of L2 processing.

Given the present evidence, we can offer only a speculative note as to why the role of proficiency may be limited in L2 predictive processing. Kuperberg and Jaeger (2016, p. 32) posited that the degree to which comprehenders engage in prediction is "a function of its expected utility, which, in turn, may depend on comprehenders' goals and their estimates of the relative reliability of their prior knowledge and the bottom-up input." Proficiency measures aim to assess the extent of learners' prior knowledge—*not* learners' estimates of the reliability of that knowledge relative to that of the bottom-up input. While we may expect the two to be related, it remains unclear how strong that relation really is. We may well imagine a learner with extensive L2 knowledge who nevertheless lacks confidence in the utility of that knowledge, and vice versa. Preliminary evidence that bilinguals' self-perception may relate to their engagement in predictive processing was presented by Peters et al. (2018), who showed differences in predictive processing behavior between bilinguals who self-identified as native speakers of the language, versus those who did not. However, self-identified native-speaker status correlated with vocabulary size in their sample, thus making it impossible to clearly separate this factor from L2 proficiency. Future work will be needed to define and operationalize individual differences between L2 learners that capture Kuperberg and Jaeger's (2016) notion of speakers' estimates of the relative reliability of different types of information.

CONCLUSION

This study employed the visual-world eye-tracking paradigm to investigate native and nonnative listeners' use of IC information for predictive processing and probed the potentially modulating role of L2 proficiency in this process. Results indicated early use of IC information in both groups, consistent with the focusing account of IC processing. Yet the effect was weaker and emerged later in the L2 compared to the L1 group. These findings suggest that L2 learners are able to draw upon IC information to generate expectations about discourse continuations during real-time listening, but they engage in predictive processing to a more limited extent than native speakers. Despite the range of proficiency in the L2 group, none of the three independent and intercorrelated proficiency measures modulated learners' predictive use of IC information in this study, adding to accumulating evidence suggesting that proficiency does not play a decisive role in L2 predictive processing. Taken together, the patterns observed in this study support the claim that native and nonnative speakers engage in qualitatively similar predictive processing behavior (Kaan, 2014), yet to a different extent and with different timing. Proficiency did not emerge as an explanatory factor for these differences, thus calling upon future work to further explore other causal factors that may underlie them.

NOTES

[1]As rightly noted by a reviewer, because these effects were not observed prior to the pronoun (except by Pyykkönen & Järvikivi, 2010, but see Cozijn et al., 2011, for discussion of a potential confound), a retroactive integration account, not involving prediction, cannot in principle be excluded (see also Grüter et al., 2018, for discussion). We agree. Yet we also note that even if the comprehender were to wait for the occurrence of the pronoun to engage in referential processing, this processing would have to remain probabilistic and critically reliant on expectations about discourse continuations based on information encoded by the preceding predicate and causal conjunction. We thus follow the previous literature in the assumption that IC effects at or shortly after the pronoun but before disambiguating information in the dependent clause involve expectation-based proactive processing, i.e., prediction in the sense discussed earlier.

[2]A reviewer inquired about IC bias of the Korean translation equivalents of the English IC verbs used in the experiment. We did not assess IC bias of Korean translation equivalents in this study, yet previous studies that investigated IC biases cross-linguistically found overwhelming uniformity for IC biases of translation equivalents across a number of languages (Bott & Solstad, 2014; Hartshorne et al., 2013), including English and Korean (Kim & Grüter, 2019).

REFERENCES

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*, B13–B26.

Au, T. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*, 104–122.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bott, O., & Solstad, T. (2014). From verb to discourse: a novel account of implicit causality. In B. Hemforth, B. Schmiedtová, & C. Fabricius-Hansen (Eds.), *Psycholinguistic approaches to meaning and understanding across languages* (pp. 213–251). International Publishing.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, *64*, 311–317.

Brown, J. D., & Grüter, T. (in press). The same cloze for all occasions? Using the Brown 1980 cloze test for measuring proficiency in SLA research. *International Review of Applied Linguistics in Language Teaching*.

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*, 237–273.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1029–1040.

Cheng, W., & Almor, A. (2017). The effect of implicit causality and consequentiality on nonnative pronoun resolution. *Applied Psycholinguistics*, *38*, 1–26.

Cheng, W., & Almor, A. (2019). A Bayesian approach to establishing coreference in second language discourse: Evidence from implicit causality and consequentiality verbs. *Bilingualism: Language and Cognition*, *22*, 456–475.

Contemori, C., & Dussias, E. P. (2018). Prediction at the discourse level in L2 English speakers: An eye-tracking study. In A. B. Bertolini, & M. J. Kaplan (Eds.), *Proceedings of the 42nd Annual Boston University Conference on language development* (pp. 159–171). Cascadilla.

Contemori, C., & Dussias, P. E. (2019). Prediction at the discourse level in Spanish–English bilinguals: An eye-tracking study. *Frontiers in Psychology*, *10*, 1–9.

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, *64*, 381–403.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*, 631–645.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*, 917–930.

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*, 353–387.

Featherstone, C. R., & Sturt, P. (2010). Because there was a cause for concern: An investigation into a word-specific prediction account of the implicit-causality effect. *The Quarterly Journal of Experimental Psychology*, *63*, 3–15.

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, *43*, 124–135.

Foucart, A., Romero-Rivas, C., Gort, B. L., & Costa, A. (2016). Discourse comprehension in L2: Making sense of what is not explicitly said. *Brain and Language*, *163*, 32–41.

Garnham, A., Traxler, M., Oakhill, J., & Gernsabacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, *35*, 517–543.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, *5*, 459–464.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Greene, S. B., & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, *6*, 262–270.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*, 191–215.

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, *7*, 199–229.

Grüter, T., Takeda, A., Rohde, H., & Schafer, A. J. (2018). Intersentential coreference expectations reflect mental models of events. *Cognition*, *177*, 172–176.

Hartshorne, J. K. (2014). What is implicit causality? *Language, Cognition and Neuroscience*, *29*, 804–824.

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*, 1474–1508.

Hartshorne, J. K., Sudo, Y., & Uruwashi, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, *60*, 179–196.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between nonnative and native speakers. *Lingua*, *120*, 901–931.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*, 33–56.

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*, 250–278.

Hopp, H. (2015). Semantics and morphosyntax in L2 predictive sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*, 277–306.

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27.

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*, 171–199.

Hulstijn, J. H. (2011). Language proficiency in native and non-native speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, *8*, 229–249.

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism, Language and Cognition*, *21*, 251–264.

Itzhak, I., & Baum, S. R. (2015). Misleading bias-driven expectations in referential processing and the facilitative role of contrastive accent. *Journal of Psycholinguistic Research*, *44*, 623–650.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Järvikivi, J., Van Gompel, R. P., & Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, *46*, 525–550.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*, 257–282.

Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In C. J.-W. Zwart & M. D. Vries (Eds.), *Structure preserved* (pp. 207–214). John Benjamins Publishing Company.

Kaan, E., Futch, C., Fuertes, R. F., Mujcinovic, S., & de la Fuente, E. Á. (2019). Adaptation to syntactic structures in native and nonnative sentence comprehension. *Applied Psycholinguistics*, *40*, 3–27.

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, *19*, 1–18.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*, 55–80.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.

Kim, H., & Grüter, T. (2018). *Crosslinguistic activation of referential bias: Effects replicate and are unaffected by L2 proficiency*. In *Poster presented at 8th Generative Approaches to Language Acquisition (GALANA)*. Indiana: University.

Kim, H., & Grüter, T. (2019). Cross-linguistic activation of implicit causality biases in Korean learners of English. *Bilingualism: Language and Cognition*, *22*, 441–455.

Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2008). A large-scale classification of English verbs. *Language Resources and Evaluation Journal*, *42*, 21–40.

Koornneef, A., Dotlačil, J., van den Broek, P., & Sanders, T. (2016). The influence of linguistic and cognitive factors on the time course of verb-based implicit causality. *The Quarterly Journal of Experimental Psychology*, *69*, 455–481.

Koornneef, A., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*, 445–465.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*, 447–464.

Liu, R. & Nicol, J. (2010). Online processing of anaphora by advanced English learners. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Proceedings of the 2008 second language research forum* (pp. 150–165). Cascadilla Proceedings Project.

Long, D. L., & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, *42*, 545–570.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*, 574–588.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.

Mitsugi, S. (2018). Generating predictions based on semantic categories in a second language: A case of numeral classifiers in Japanese. *International Review of Applied Linguistics in Language Teaching* (published online 2018-07-03). https://doi.org/10.1515/iral-2017-0118.

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *6*, 1–18.

Peters, R., Grüter, T., & Borovsky, A. (2018). Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied Psycholinguistics*, *39*, 1439–1469.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence or implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*, 5–16.

Quick Placement Test. (2001). Oxford University Press.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, *30*, 333–357.

Schoonen, R. (2011). How language ability is assessed. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. *II*, pp. 701–716). Routledge.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*, 339–368.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*, 423–443.

Van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulate the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26–42.

Van Berkum, J. J. A., Koornneef, A., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158–171.