The Role of Awareness and Cognitive Aptitudes in L2 Predictive Language Processing

Curcic, M.; Andringa, S.; Kuiken, F.

DOI
10.1111/lang.12321
Publication date
2019
Document Version
Final published version
Published in
Language Learning
License
CC BY-NC

Link to publication

Citation for published version (APA):
Curcic, M., Andringa, S., & Kuiken, F. (2019). The Role of Awareness and Cognitive Aptitudes in L2 Predictive Language Processing. *Language Learning*, *69*(S1), 42-71. https://doi.org/10.1111/lang.12321

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

The Role of Awareness and Cognitive Aptitudes in L2 Predictive Language Processing

Maja Curcic, Sible Andringa, and Folkert Kuiken

University of Amsterdam

This study investigated whether second language (L2) learners can develop predictive processing of determiners after a brief exposure to a novel language, and whether this depends on learners' awareness for the target structure and their cognitive aptitudes. One hundred L2 learners received auditory exposure to a miniature language based on Fijian that included a determiner–noun agreement pattern. Learners' processing of determiners was measured using a picture-matching task with eye tracking. We found that learners learned to anticipate the coming noun based on the determiner; they also gained a speed advantage. Learners' awareness played a crucial role in such anticipatory processing; only learners who were aware that determiners helped them during the test (i.e., prediction-aware learners) showed signs of anticipatory processing. The aptitude variables did not modulate learners' processing abilities, but there were links between aptitude and learners' abilities to develop different levels of awareness.

Keywords second language; processing; individual differences; awareness; cognitive aptitudes; eye tracking

---

The research presented in this article is part of the PhD research of Maja Curcic, supervised by Sible Andringa and Folkert Kuiken. We would like to express our deepest gratitude to Dirk Vet for his invaluable technical assistance, Eva van Lier for help in creating the target language of the study and recording instructions, and Ana Bosnic for recording the auditory materials. We also thank Ewa Dabrowska and three anonymous *Language Learning* reviewers for their extremely helpful comments and ideas for improving this manuscript.

Correspondence concerning this article should be addressed to Sible Andringa, Center for Language and Communication, University of Amsterdam, Spuistraat 134, 1012 VB, Amsterdam, Netherlands. E-mail: s.j.andringa@uva.nl

DOI: 10.1111/lang.12321

Introduction

Recent decades have witnessed a surge of studies investigating how language comprehension unfolds in real time. Such research strongly suggests that language processing is highly incremental in nature; sentences are processed in a word-by-word fashion, and listeners make rapid online predictions about the upcoming linguistic information rather than wait until the end of a sentence to interpret it (e.g., Altman & Mirković, 2009; Kuperberg & Jaeger, 2016). Methods that have been used to provide evidence for prediction in language processing range from event-related potentials (ERPs; e.g., Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) to eye tracking (e.g., Altmann & Kamide, 1999) and self-paced reading/listening (e.g., Brothers, Swaab, & Traxler, 2017). The established body of work provides evidence that language users may exploit a wide range of features in the discourse context to form expectations about upcoming information (e.g., Huettig, 2015; Kaan, 2014; Kuperberg & Jaeger, 2016), such as semantic (e.g., Altmann & Kamide, 1999), grammatical (e.g., Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Lew-Williams & Fernald, 2007), phonological (e.g., DeLong, Urbach, & Kutas, 2005), or prosodic features (e.g., Weber, Grice, & Crocker, 2006). This study investigated if second language (L2) learners can come to use determiners predictively in their L2. In grammar-based prediction, determiners have been among the most frequently investigated predictive cues, and determiner-based prediction has been found in several languages, such as Spanish (e.g., Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Lew-Williams & Fernald, 2007), French (e.g., Dahan et al., 2000; Foucart & Frenck-Mestre, 2011), and Dutch (e.g., Loerts, Wieling, & Schmid, 2013; Sabourin & Stowe, 2008).

Prediction as a feature of language processing is not uncontroversial; its precise nature and its role in language processing has been debated extensively (for excellent discussions, see Huettig, 2015; Kaan 2014; Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012). One of the issues concerns the mechanisms that generate predictions (e.g., Foucart, Ruiz-Tada, & Costa, 2015; Martin, Branzi, & Bar, 2018), which, according to Huettig (2015), may actually result from several different mechanisms that may include association and priming, the engagement of the production system in comprehension, and event-simulation processes. Another issue concerns the inevitability of prediction in information processing (Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012). Prediction has been argued to be a fundamental and inevitable property of language processing, allowing for conversation to run smoothly and comprehension to be accurate yet effortless (e.g., Federmeier, 2007; Kuperberg & Jaeger, 2016). Others, however, have questioned its significance (e.g., Huettig, 2015), pointing to studies that suggest that prediction is graded, that is, sentence contexts may vary in the degree to which they trigger certain predictions (see Kuperberg & Jaeger, 2016, for a discussion) and individuals may vary in the extent to which they generate predictions (e.g., Huettig, 2015; Kukona et al., 2016).

An issue that is particularly relevant to prediction in the L2 is the extent to which L2 learners can come to use their L2 to generate predictions in similar ways as native speakers (Kaan, 2014). Several studies have shown that L2 learners do not necessarily create predictions where native speakers typically do, even when they clearly possess knowledge of the target structure (e.g., Grüter, Lew-Williams, & Fernald, 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013). However, given that other studies have provided evidence of L2 prediction (further discussed below), it would seem important to investigate which factors potentially determine whether or not L2 learners form predictions (Kaan, 2014). The goal of the present study was to investigate the influence of awareness and particular aptitude components as causes of individual variation in L2 prediction. Therefore, the remainder of this review will specifically target potential sources of individual differences in acquiring predictive use of grammatical cues.

Identifying causes of individual differences is firmly rooted in second language acquisition (SLA) research in general and in L2-processing research more specifically. Studies that looked at L2 prediction showed that prediction may be graded and more or less likely to occur depending on a number of factors (Kaan, 2014). One such factor is the similarity between the first language (L1) and the L2 that may or may not allow for transfer of L1 processing routines (e.g., Foucart & Frenck-Mestre, 2011; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005; Tokowicz & Warren, 2010). Van Bergen and Flecken (2017) provided evidence for this. They studied L1 effects in predictive processing by looking at whether English, French, and German learners of Dutch as a L2 would be able to use Dutch placement verbs to predict object positions. They found that only German learners whose L1 had a similar way of encoding object positions could do this. Even if the L1 allows for a similar type of prediction, subtle differences between the languages may affect prediction, as shown by Dussias et al. (2013), who observed that Italian learners of Spanish generated predictions in Spanish based on the feminine determiner only. The authors speculated that the absence of prediction for masculine determiners may be because of the fact that Italian has two masculine determiners; this lack of congruence between the two languages may have prevented learners from using masculine determiners to predict. The same study also provides evidence that L2 learners can come to predict in nativelike ways even if their L1 does not possess similar features. The study included high- and low-proficient English learners of Spanish and found evidence of prediction in Spanish, despite the fact that English does not allow for determiner-based prediction. However, only learners who had a high proficiency of Spanish predicted, and this prediction was less pronounced than native-speaker prediction.

Effects of L2 proficiency on prediction performance emphasize the graded nature of prediction, and they suggest that linguistic knowledge needs to be of a certain quality before learners can use it to generate predictions. Evidence for such links between prediction and representation quality can be found in both the L1 and the L2 literature. Kukona et al. (2016), for example, found that semantic prediction performance by L1 speakers in a visual-world eye-tracking paradigm was related to comprehension ability (a composite score of listening and reading comprehension tasks and vocabulary knowledge), decoding and fluency ability in reading, and rapid automatized naming (RAN). Borovsky, Elman, and Fernald (2012) similarly reported that prediction strength in the L1 in both children and adults was related to vocabulary size. Recent studies by Hopp (2013, 2016) showed the same for L2 learners. Hopp (2013) related gender-based prediction performance by near-native English learners of German to both gender-assignment abilities in a production task and to a measure of speed of lexical access and found that both were associated with prediction performance. In another study, Hopp (2016) used a pretest–posttest design to show that a group of English learners of German was able to use gender predictively after gender-assignment training. In this study, prediction performance also depended on learners' performance on the production task. Hopp also showed that native speakers' gender-based prediction was disrupted after presenting them with nontargetlike examples. Such studies provide evidence for the idea that L2 processing is not necessarily fundamentally different, but differences may stem from the fact that L2 knowledge is generally less well established and entrenched.

When exposed to linguistic input, L2 learners show substantial individual differences in whether or not they develop awareness of a particular target structure (e.g., Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler, 2015; Williams, 2005). Such awareness may be a crucial factor for predictions to arise, as it may enable more controlled processing of linguistic information. Huettig (2015) alludes to this when he suggests that prediction may result from an interplay between automatic processes and more controlled processes. Evidence that some degree of control modulates the extent of prediction in the L1 was provided by Brothers et al. (2017), who manipulated intentionality in a within-participants design and found that the effects of prediction were more pronounced in the task in which participants were explicitly invited to predict. The role of awareness in SLA has been intensively debated and studied (for reviews, see Andringa & Rebuschat, 2015; Leow & Donatelli, 2017; Rebuschat, 2015), for example, within the context of instructed L2 learning. Some studies have used online processing measures to determine what kind of processing learners can achieve after brief exposure to novel linguistic structures (e.g., Andringa & Curcic, 2015; Batterink & Neville, 2013; Davidson & Indefrey, 2009; Marsden, Williams, & Liu, 2013; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010). In some of these studies, awareness of the target was created through instruction. For instance, Davidson and Indefrey (2009) showed that learners can come to process adjective declension in their L2 in ways similar to native speakers after short explicit instruction. Andringa and Curcic (2015) investigated whether learners' awareness through instruction might be a shortcut to predictive processing. Learners in the study were briefly exposed to a miniature language based on Esperanto, and the target structure was differential object marking (DOM), according to which nouns denoting animate (but not inanimate) objects were preceded by the proposition "*a*" ("to''). Half of the learners were provided with explicit rule explanation. The study found no evidence of nativelike predictive processing of DOM. Another study by Andringa and Curcic (2016) focused on highly proficient learners of Spanish and investigated how learners' awareness that DOM could be useful during the test affected their predictive processing. The test included two phases, and, at the beginning of the second phase, learners were informed that DOM could help them during the test to arrive at the correct answer more quickly. Results provided clear evidence that learners, who hardly showed predictive processing in the first phase, clearly predicted in the second phase after they were made aware of the fact that DOM could be helpful to them.

Such studies suggest that awareness of L2 structures during exposure may not directly translate to nativelike predictive processing. However, raising learners' awareness that certain structures can be used during comprehension may help learners quickly develop predictive processing of those structures reminiscent of processing attested in L1 speakers, perhaps because this level of awareness encourages learners to make strategic predictions. More systematic research is needed that looks at how awareness may or may not influence L2 processing.

Finally, the extent to which L2 learners are able to generate predictions may depend on their cognitive resources (Kaan, 2014). However, determining which cognitive resources might facilitate prediction is not easy. The SLA field has a long tradition of attempting to establish links between success in L2 learning and learners' cognitive aptitudes that is not easily summarized (but see Wen, Biedroń, & Skehan, 2017). Three trends in the literature are worth noting. The first concerns the role of more general capacities, such as working memory and aspects of intelligence; while these were initially argued to be separate constructs from aptitude, they increasingly tend to be seen as partly overlapping or additional components of aptitude (DeKeyser & Koeth, 2011; Sasaki, 1996; Wen et al., 2017). A second observation concerns the recognition that the aptitude construct initially focused rather narrowly on learning in instructed settings and explicit types of learning abilities; the aptitude construct needed to be expanded toward different learning settings, most notably toward more implicit types of learning (e.g., Robinson, 2005, 2007). Several recent studies have included and found evidence for links between serial reaction-time tasks as a measure of implicit statistical learning and L2 learning success (e.g., Brooks & Kempe, 2013; Granena, 2013a; McDonough & Trofimovich, 2016). Finally, while aptitude has generally been seen as a multicomponential construct, recent models recognize more sharply that different components may be at play at different stages and for different processes in the L2 learning process (Skehan, 2016; Wen et al., 2017).

The picture that thus emerges from this is that a wide array of cognitive aptitudes may be involved in L2 learning and processing, at different stages and in different settings. However, studies linking such factors specifically to prediction are still scarce (Huettig, 2015; Kaan, 2014). So far, there have only been attempts to link working memory to prediction. For example, Huettig and Janse (2016) administered measures of working memory capacity and processing speed to older native speakers of Dutch and found that participants' composite working memory scores as well as their processing speed correlated with their ability to generate determiner-based predictions. Another study by Ito, Corley, and Pickering (2018) investigated semantic-based prediction by native and nonnative speakers of English under cognitive load conditions. The goal was to investigate if prediction is facilitated by cognitive resources, and they found that both groups were delayed in identifying predictable targets under cognitive load conditions. Kukona et al. (2016), however, included a complex reading-span task and a visuo-spatial memory task in their study and found no links with semantic prediction performance. Given the paucity of research in this area, much more research is needed that tries to link prediction to cognitive aptitudes.

The Present Study

Prediction has been claimed to be a core and crucial aspect of native-speaker language processing (e.g., Federmeier, 2007; Kuperberg & Jaeger, 2016) and may therefore be part and parcel of what L2 learners have to master in the course of acquiring a new language. However, some have questioned the centrality of prediction in language processing and argued that prediction is a graded phenomenon and may depend on the linguistic context and individual attributes (e.g., Huettig, 2015). Work on L2 prediction supports the notion of a graded phenomenon: While several studies have shown that L2 learners did not acquire a particular feature of the L2 to a degree that allowed them to use it as a predictive cue (e.g., Grüter, Lew-Williams, & Fernald, 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013), other studies have observed such prediction but found it depended on the similarity between the L1 and the L2 (Dussias et al., 2013; Van Bergen & Flecken, 2017) or level of proficiency or strength of representation (e.g., Hopp, 2013, 2016). Such findings suggest that nativelike prediction is ultimately attainable, but much more work is needed to investigate to what extent native speakers and L2 learners actually predict and which factors determine this (Huettig, 2015; Kaan, 2014). This study focused on potential sources of individual differences in learning to use the L2 predictively. While there are indications that prediction predicates on awareness for the target structure, very little is yet known about its precise role in learning to predict. Similarly, as yet, very few studies have attempted to link prediction to components of cognitive aptitude.

In the present study, the visual-world eye-tracking paradigm was used to investigate learners' abilities to develop determiner-based prediction within the context of a fully unknown miniature artificial language and to ascertain to what extent this is explained by awareness; awareness of different aspects of the structure was obtained through a careful postexperimental interview. In a more exploratory fashion, several aptitude measures were included. Because prediction minimally rests on sensitivity to the grammatical pattern, we included measures of phonetic encoding and language analysis on the basis of Skehan's Macro-SLA aptitude model (Skehan, 2016; Wen et al., 2017) that identifies these constructs as relevant for noticing and pattern recognition. The LLAMA D task (Meara, 2005) was used as a measure for phonetic encoding ability. In a validation study, Granena (2013b) found that this task measured a construct that is separate from working memory and analytic ability and suggests this task may be a measure of sequence-learning ability and more implicit types of aptitude. A nonverbal IQ task (Wechsler, 2008) was used as a measure of analytical ability. Finally, a measure of rote learning was included as this is associated with explicit learning (Doughty, 2018; Robinson, 2007) and may therefore be related to awareness or becoming aware. This measure was obtained from the word-learning phase of the present experiment. The use of a highly controlled miniature language learning setting allowed for control over the exposure and eliminated potential effects of instruction and learning context effects. We aimed to answer the following research questions:

1. Do phonetic encoding ability and analytical ability explain individual differences in learners' abilities to achieve determiner-based prediction?
2. Can awareness of the target structure explain individual differences in learners' ability to achieve nativelike predictive processing of determiners?

Method[1]

Participants

Participants in the study were 100 adult native speakers of Dutch between 19 and 35 years old ($M_{\text{age}} = 23.59$, $SD = 3.76$). They were students or highly educated adults, but they were never trained in linguistics. All participants reported having good hearing, normal or corrected-to-normal vision, no history of dyslexia, and no color-blindness. The participants reported speaking at least one foreign language. In a questionnaire, we asked them to list the languages they spoke and to assess their level of knowledge by means of can-do statements from the scales of the Common European Framework of Reference for Languages (Council of Europe, 2011). Learners knew between zero and five additional languages ($M = 2.8$, $SD = 1.1$). Some of them knew agreement languages, such as Spanish or Italian ($M = 0.97$, $SD = 0.8$, range = 0–3). We computed two "knowledge of foreign language" measures by awarding 0.5 points for each level of each language: (a) the FLK measure—reflecting the number and level of knowledge of all foreign languages and (b) the AgrFLK measure—reflecting the number and level of foreign languages that featured gender-based determiners and agreement between determiner and noun. On average, learners scored 8.71 ($SD = 3.56$) on the FLK measure and 2.15 ($SD = 2.31$) on the AgrFLK measure. At the end of the experiment, which lasted about 1 hour 45 minutes, learners were paid €15 for participation. The study was approved by the Ethics Committee.

The Target Language and the Target Structure

The target language used in this study was a miniature language constructed using lexical items from Fijian—an Austronesian language of the Malayo-Polynesian languages, spoken in Fiji. The language consisted of 20 nouns, 4 verbs, 6 adjectives, and 2 determiners (see Appendix S1 in the Supporting Information online). Where possible, both the original Fijian form and meaning were retained. However, whenever the original Fijian word had a complex form or a form similar to words present in participants' L1 or in languages frequently spoken among Dutch speakers (e.g., German, French, Spanish, and Italian), another Fijian word form was chosen and used to denote the original meaning. The grammar of the language was not based on Fijian. Verbs all appeared in third-person singular, marked by a *–t* added to the Fijian forms. The language had the same word order (i.e., Subject–Verb–Object) as Dutch, and it included a determiner–noun agreement pattern. The nouns were created by adding either *–is* or *–uk* to Fijian nouns, and they were preceded by a determiner that agreed with the noun in that masculine *lep* preceded nouns ending in *–uk* (e.g., *lep oseuk* ["the ball"]) whereas feminine *ris* preceded nouns ending in *–is* (e.g., *ris salis* ["the chair"]). The participants' L1, Dutch, has two definite determiners (*de*—common and *het*—neuter), and there is no feminine-versus-masculine distinction between them and no clear determiner–noun agreement (Booij, 2002). Dutch native speakers have been shown to use determiners predictively (Loerts et al., 2013; Sabourin & Stowe, 2008).

Language Exposure

Participants were told that they would be learning a new language by looking at images and listening to sentences describing the images. They were encouraged to learn the language as well as they could because their knowledge would be tested from time to time. Learners were not informed about the purpose of the experiment or the existence of any grammatical structure. The exposure started with a noun learning phase, followed by a noun test. After this, learners were exposed to phrases and sentences containing the target determiner–noun structure. We used E-prime software and a Tobii TX120 computer to present the materials. All instructions were presented in Dutch on a screen as well as auditorily. The experimental materials were always presented auditorily only and were recorded by a female native speaker of Serbian to give the language a foreign feel.

Noun Learning and Assessment

First, we wanted to make sure that all learners had identical and perfect knowledge of nouns before receiving exposure to the determiner–noun combinations. In the noun-learning phase, which lasted about 5 minutes, learners saw simple black-and-white images and heard bare nouns denoting the objects in the images. All 20 nouns were presented six times, and the order of trials was identical for all learners. Knowledge of the nouns was then tested by presenting four images on the screen. Participants heard a noun and were asked to click on the correct image, after which they were given feedback about the accuracy of their response so that they could continue learning. Each noun appeared as the target noun in two trials whereas the distractor nouns were randomly chosen by the program. The first 40 trials were identical for all learners. After this, learners were presented with only those nouns they had not yet learned. The test continued until learners reached 100% accuracy. We used learners' scores on the noun test (i.e., the number of items learners needed to reach 100% accuracy) as a measure of their rote memory ability.

Target Structure Exposure

In this part of the exposure, learners received more complex linguistic input. Apart from the 20 nouns taught in the previous phase, it included two determiners, three intransitive verbs, copula *be*, and six adjectives (see Appendix S2 in the Supporting Information online). In every trial, learners were exposed to the target structure; they saw an image representing a simple object or an intransitive activity and they heard a phrase or sentence describing the image. Learners' only task was to attend to the trials and learn as much as possible. They were not given any additional tasks. The items fell into one of the following four structural categories:

1. Determiner + noun
   Ris salis
   "The chair."
2. Determiner + adjective$_{color}$ + noun
   Ris matene salis
   "The blue chair."
3. Intransitive sentences: determiner + animate noun + verb$_{intransitive}$
   Ris burogis sisilit.
   "The dog is swimming."
4. Intransitive sentences: determiner + inanimate noun + is + adjective
   Ris salis na duka.
   "The chair is broken."

Learners heard 308 trials, and the overall length of the target structure exposure was about 25 minutes. The exposure was divided into two parts with a break in the middle. The first part consisted of 176 trials and lasted 15 minutes. The second part lasted 10 minutes, and learners saw 132 trials. Eighteen out of 20 nouns appeared 14 times, and two nouns were overrepresented to make the input more natural (Ellis & Ferreira-Junior, 2009), appearing 28 times each: the feminine noun *burogis* ("dog"), and the masculine noun *dawauk* ("cat"). The order of the trials was fixed and identical for all learners. Learners were equally exposed to feminine and masculine items, and the two determiner categories had the same number of overrepresented versus underrepresented nouns, animate versus inanimate nouns, nonbiological versus biological gender nouns (i.e., boy, girl, man, woman).

The present data set is based on data from two separate studies that were intended to examine the effects of exposure length and reliability. This means that half of the learners in the present data set received longer exposure: Their instruction included an additional set of 176 trials. These were the 176 trials from the first part of the exposure, but with a different ordering. Also, for half of the learners ($n = 50$), 4 out of 20 nouns—2 feminine and 2 masculine—were exceptions to the agreement pattern between determiners and nouns (feminine nouns received masculine determiner and vice versa), meaning their input was less reliable. Participants were always randomly assigned to the reliable or unreliable input groups. Before testing our research questions, we checked for interfering effects of both exposure length and reliability.

Picture Matching With Eye Tracking

In the picture-matching task, learners saw two images on the screen—one on the left and one on the right—and they heard a short phrase in the new language referring to one of the two images. The phrase consisted of a determiner, adjective, and noun (e.g., *lep ramase karetuk* ["the green bird"]). The learners' task was to choose the correct image by pressing either "z" (for the left image) or "m" (for the right image) on the keyboard. They were encouraged to be as fast and accurate as possible. All images were familiar to learners from the previous exposure phase.

The test materials consisted of 40 trials (see Appendix S3 in the Supporting Information online). In the construction of test trials, we systematically varied the gender of the target (correct) noun (masculine vs. feminine) and the gender of images in the visual scene (same gender vs. different gender) (see Appendix S3 in the Supporting Information online). Every trial fell into one of the following four experimental conditions: (a) feminine target, same-gender pictures; (b) feminine target, different-gender pictures; (c) masculine target, same-gender pictures; and (d) masculine target, different-gender pictures. There were 10 trials per condition. In every condition, there were eight trials featuring the regular nouns and two trials featuring the nouns that were exceptions for the unreliable input group; these were removed from the analysis for both learner groups.

Every noun appeared twice as the target noun—once in the same- and once in the different-gender conditions. Four previously introduced color adjectives were used to create a larger time window between determiner and noun, and they appeared equally often. Also, target images were presented on the left and on the right side of the screen equally often. While constructing the test materials, we took care not to bias participants' eye gazes and responses toward one of the two images in the visual scene by (a) never combining images of animate and inanimate objects in the visual scene, (b) never combining biological gender (man, woman, girl, and boy) with nonbiological gender nouns, and (c) using the same color adjective within a trial. The images used were the same as in the instruction (see Appendix S2 in the Supporting Information online), and they were presented side by side at 400 × 400 pixels on a 1280 × 1024 screen with 200 pixels in between. Looks on the image were counted as looks on target, which means that white areas around the object in the image were counted as looks on target. These areas varied depending on the object displayed.

After calibrating the eye tracker, participants did four practice items to make sure they understood the task. Trials would start after participants fixated for 100 milliseconds on a fixation cross in the middle of the screen. In each trial, images were presented for about 1,200 milliseconds before the sound was played and learners had a maximum of 5,000 milliseconds to choose the correct image. Learners had an average accuracy of 96.78% ($SD = 4.11$, range = 83.94–100) when choosing the target image, which means that they were able to recognize the target nouns very accurately.

Debriefing

After the language exposure and tests, we debriefed learners using a protocol that was designed to carefully and gradually probe learners' different levels of awareness: (a) gender awareness (i.e., awareness that *lep* and *ris* denote a gender distinction), (b) pattern awareness (awareness of the agreement pattern between determiners and nouns), and (c) prediction awareness (awareness that determiners were useful during the picture-matching test with eye tracking). We also determined when participants became gender and pattern aware. All debriefings were recorded. Appendix S4 in the Supporting Information online provides the protocol as well as detailed explanations about how responses were coded. All debriefings were coded by the first author. Twenty randomly chosen debriefing interviews were also coded by a linguist who was not involved in the study, and the protocol in Appendix S4 in the Supporting Information online was used. To determine interrater reliability, Cohen's Kappa was calculated for gender awareness (yes or no; Kappa = 0.773; $p < .001$), pattern awareness (yes or no; Kappa = 1; $p < .001$), prediction awareness (yes or no; Kappa = 1; $p < .001$), and for the moment (i.e., during exposure, eye tracking, grammaticality judgment tasks, or debriefing; or no awareness) at which participants became gender aware (Kappa = 0.735; $p < .001$) and pattern aware (Kappa = 0.925; $p < .001$).

Aptitude Measures

Three aptitude measures were administered to the entire sample: a measure of analytical ability (IQ test), phonetic encoding (LLAMA D), and rote memory (noun test). They are explained in the order in which they were administered.

The Noun Test

On average, learners needed 44.89 items ($SD = 5.91$, range = 40–71) to pass the test, and on this test, lower scores indicated better performance. It is important to note that 21 learners needed no more than 40 trials to achieve 100% accuracy, which means that the test did not measure the full range of rote memory ability present in the sample.

The Nonverbal IQ Test

The Matrix Reasoning subtest of the Wechsler Adult Intelligence Scale–4th edition (Wechsler, 2008) was used as a measure of learners' analytical ability. The test measures nonverbal abstract problem solving and inductive reasoning ability. It consisted of 26 items, and the learners' task was to choose the missing part of a visual pattern. There were five options from which to choose. The test was administered after the first part of the exposure to the target structure, and participants were given 20 minutes maximum to complete the test. On average, learners scored 22.9 ($SD = 1.86$, range = 17–26) on this test.

The LLAMA D Sound Recognition Task

The LLAMA D task was used to assess learners' phonetic encoding or discrimination ability.[2] The test is a part of the LLAMA language aptitude test battery (Meara, 2005). It measures the ability to recognize spoken language that one has been exposed to before, and it may tap the storage component of verbal working memory. Participants first listened to a sequence of 10 words in an unknown language, after which they did a 30-item test in which they listened to both previously heard words and novel words. Their task was to decide for each word whether they had heard it before. If they decided correctly, they gained points, and if they decided wrongly, they lost points. The maximum score was 75, and the task lasted about 5 minutes. It was administered after the debriefing session. On average, learners scored 27.9 ($SD = 13.75$, range = 0–65) on this test.

Data Analysis

During the picture-matching task, a Tobii TX120 eye tracker recorded the position of participants' eyes every 8.3 milliseconds, and for each recorded frame, we marked gaze accuracy as correct if a participant looked at the target image or incorrect if a participant looked at the distractor image. Frames for which eye data were missing or during which they did not look at either image were marked as missing and irrelevant, respectively. Altogether, 26.6% of the frames were excluded (17.84% missing, 8.76% irrelevant). Finally, trials were considered uninformative and potentially unreliable if more than 50% of the frames were missing or irrelevant, which led to the exclusion of 17.38% of the trials. Similarly, we excluded participants who lost more than two thirds of all test trials (i.e., 70%). In this way, four participants were excluded.

The goal of the statistical analyses was to see if learners could use determiners to anticipate the correct answer while hearing the determiner and adjective, but before hearing the disambiguating noun. We expected learners to show a preference for the target image only in different-gender trials, not in the same-gender trials. The determiner–adjective and noun phases were equally long in the same and different trial sets because each determiner–adjective combination and each noun occurred equally often in the two trial sets and because the same recordings were used in the two trial sets. The average length of the key determiner–adjective phase was 980 milliseconds ($SD = 79$), which was also the average boundary between the determiner–adjective phase and the noun phase. The average length of the noun phase was 776 milliseconds ($SD = 62$). We applied a correction of 200 milliseconds to account for the standardly observed delay in adult eye responses to linguistic materials (Dahan, Magnuson, Tanenhaus, & Hogan, 2001), meaning that zero milliseconds in the eye-tracking data presented below corresponds to 200 milliseconds after the onset of the determiner. This has been applied to all figures and time windows presented.

To analyze the gaze data, we used the cluster-based permutation analysis (described in Dink & Ferguson, 2015; used in Maris & Oostenveld, 2007) from the EyetrackingR package (Dink & Ferguson, 2015). This analysis allowed us to establish the onset and duration of predictive processing. The advantage of this procedure is that it provides a good correction for false alarms while retaining data sensitivity. Cluster-based permutation analysis involves two steps. In the first, *t* tests are run on time bins of a particular size, which may yield clusters of adjacent bins for which a significant difference is observed between same and different trials. In the second step, similar analyses are run on randomly shuffled data, which is reiterated hundreds of times. The outcomes of the second step are then compared against the observed significant clusters in the first step, which yields probabilities that express the likelihood that the results are chance findings. In our analyses, we used 50-milliseond time bins and the procedure of reshuffling and testing was repeated 2500 times. To investigate if and how the awareness and the cognitive aptitude variables were related to prediction, we created a prediction variable that was used as dependent variable by subtracting the proportion of looks in the same-gender trials from the proportion of looks in the different-gender trials in each 50-millisecond time bin for each participant. We then used linear models to assess for each time bin whether the difference between same- and different-gender trials were modulated by awareness and aptitude.

The goal of the reaction-time (RT) analysis was to find out whether learners gained a speed advantage in choosing the correct image when determiners were predictive of the correct response and whether different levels of awareness and aptitudes modulated the speed gains. The RTs were measured from the onset of the noun until the button press. The analysis was conducted on correct responses only. Before analyzing the results, we excluded all exception test trials from both reliable input and unreliable input learner groups, which means that participants in this study were compared on exactly the same item sets. We used the cluster-based permutation analysis to check if prediction was related to input reliability or exposure length. We found no such effects given that no significant time clusters were detected in the first step of the cluster-based permutation analyses. This informed our decision to merge the two data sets and to ignore these variables in further analyses.

Results

Awareness

Most learners reported that they focused on learning the noun, verb, and adjective meanings during the exposure and paid less attention to other aspects of the language. However, as Figure 1 shows, more than half of the learners—62 out of 100—reported that they had become aware of the gender distinction during the exposure. Twenty-four learners were found to be pattern aware: They noticed the agreement pattern between the determiner and the noun during the exposure and were able to reproduce it. As can be seen from Figure 1, pattern-aware learners were always also gender aware, except for one learner who noticed the pattern but was unaware of the gender distinction. Some learners also reported

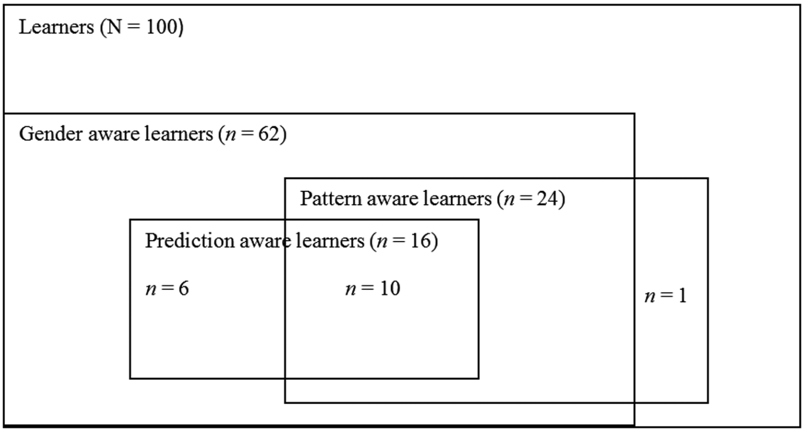

Figure 1 Overview of learners' different levels of awareness as reported during debriefing.

being prediction aware: They noticed that determiners sometimes helped them during the test to arrive at the correct answer faster ($n = 16$). These learners were always also gender aware and were in some cases ($n = 10$), but not always ($n = 6$), also aware of the pattern. The debriefing data imply that gender awareness was a prerequisite for developing other, deeper levels of awareness in our study.

Eye Movements

The eye-movement data were analyzed in four steps. First, we analyzed the data to see if learners as a group showed predictive processing of determiners. The dependent variable was the proportion of looks toward the target image, and the within-subject factor was trial type (same- vs. different-gender images). In step one of the cluster-based permutation analysis, we found a time cluster indicative of prediction between 550 and 1,250 milliseconds (Figure 2, shaded area). In step two of the analysis, this effect was confirmed as significant ($t = 31.72$, $p = .003$). This means that learners as a group showed predictive processing of determiners about 400 milliseconds before they could launch noun-based looks (the onset of the noun corresponds to 980 milliseconds, the vertical line in Figure 2). They looked more at the target than the nontarget images in different-gender trials as compared to same-gender trials, and this effect spilled over into the noun phase.

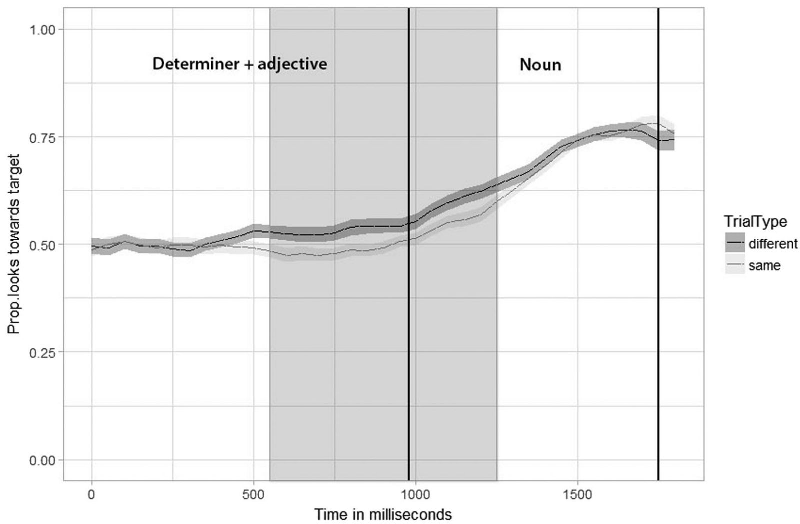

Figure 2 Proportion of learners' looks towards the target image over time in the same-gender and different-gender trials and the predictive processing effect observed between 550 and 1,250 milliseconds (shaded area).

Table 1 Results of the cluster-based permutation analysis—effects of prediction awareness

Time cluster	Sum statistic	Time range in ms	Probability
1	2	250–300	0.53
2	24.86	350–850	0.007
3	2.03	900–950	0.52
4	2.06	1,150–1,200	0.49

In the second step, we looked for links between predictive behavior and (a) gender awareness ($n = 62$), (b) pattern awareness ($n = 24$), and (c) prediction awareness ($n = 16$). In each analysis, one of the three levels of awareness was entered as a fixed effect. The dependent variable was always the prediction score that expressed the difference in proportion of looks to same- and different-gender trials. We found no effects of gender awareness and pattern awareness. However, we did find a significant effect of prediction awareness, such that learners who had reported this awareness ($n = 16$) showed higher levels of predictive processing than those who had not ($n = 80$). In step one of the analysis, four potential time clusters were detected (see Table 1), but

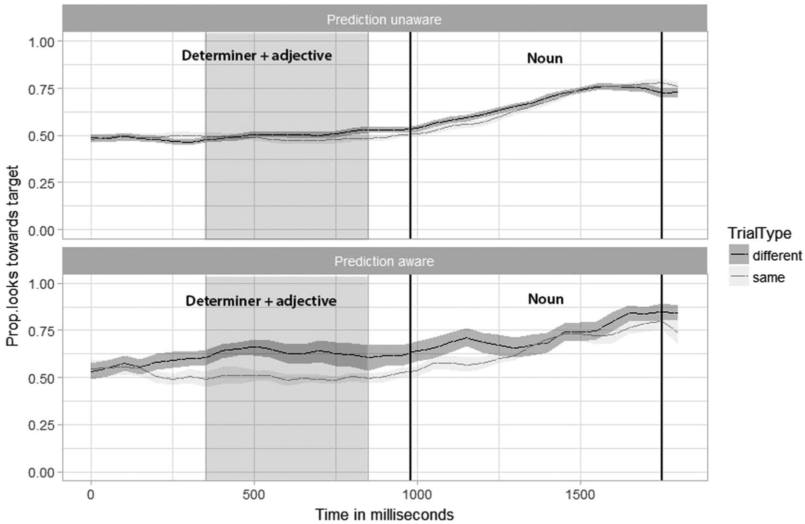

Figure 3 Proportion of prediction aware versus unaware learners' looks toward the target image over time in the same-gender and different-gender trials and the time cluster during which the effect of prediction awareness was observed.

Table 2 Results of the cluster-based permutation analysis—online processing effects of prediction unaware learners

Time cluster	Sum statistic	Time range in ms	Probability
1	2.13	850–900	0.46
2	2.01	1,050–1,100	0.53

only the largest one turned out to be significant after step two (see Figure 3, shaded area: 350–850 milliseconds, $b = 24.86$, $p = .007$). The other three time clusters detected in step one lasted only 50 milliseconds and did not reach significance. The results imply that learners who were prediction aware showed earlier and stronger predictive processing compared to learners without this level of awareness.

To gauge to what extent the overall prediction effect observed was driven by the prediction-aware learners, we ran a separate analysis on learners without prediction awareness. In step one, two potentially significant time clusters were detected between 850 and 900 milliseconds and between 1,050 and 1,100 milliseconds (see Table 2). However, neither cluster survived step two, meaning that these should be treated as chance findings. There was no evidence of predictive behavior by learners without prediction awareness.

Table 3 Results of the cluster-based permutation analysis—effects of aptitude measures on predictive processing

Time cluster	Sum statistic	Time range in ms	Probability
LLAMA D			
1	2.14	550–600	0.44
2	2.1	1,700–1,750	0.46
Noun test scores			
1	4.08	1,250–1,350	0.33
IQ—analytical ability			
1	4.39	600–700	0.31

As a third step, we investigated to what extent predictive processing was related to the aptitude measures. The analyses revealed no effects of phonetic encoding, rote memory, and analytical ability. While in step one some potentially significant but brief time clusters were identified, these clusters turned out to be chance findings in step two (see Table 3).

As a final step, we used generalized linear models to investigate if aptitude and awareness were related. In these analyses, we ran several models with the awareness variables as the dependent variable and different aptitudes as fixed effects. We found that learners' gender awareness was predicted by rote memory as measured by the noun test ($b = -0.1$, $SE = 0.04$, $p = .02$, 95% confidence interval [CI] [–0.18, –0.02], $OR = 0.91$) and by analytical ability as measured by the IQ test ($b = 0.24$, $SE = 0.12$, $p = .04$, 95% CI [0.003, 0.47], $OR = 1.27$). Learners' pattern awareness was predicted by the LLAMA D measure of phonetic encoding ($b = 0.06$, $SE = 0.02$, $p = .002$, 95% CI [0.02, 0.1], $OR = 1.06$), and there was a tendency that LLAMA D also predicted prediction awareness ($b = 0.04$, $SE = 0.02$, $p = .066$, 95% CI [–0.003, 0.08], $OR = 1.04$).

Response Times

To assess if learners would show a determiner-based speed advantage in choosing the correct target, we used linear mixed models. Following the recommendation of Barr, Levy, Scheepers, and Tily (2013), we always applied maximal random-effects structures. The dependent variable was RT in milliseconds. The model tried always included trial type (same or different gender) as a fixed effect as well as the random intercepts for participants and items and the random by-participant slope for trial type. Awareness and aptitude variables were added to this baseline model separately to assess whether they modulated a potential speed advantage. For awareness, we ran models with (a)

Table 4 Overview of learners' mean reaction times (RTs) and their standard deviations (*SD*) split by trial type and prediction awareness

Prediction awareness	Different gender trials RTs (*SD*)	Same gender trials RTs (*SD*)
Aware (n = 16)	1,869 (498)	2,152 (216)
Unaware (n = 84)	2,294 (237)	2,303 (259)

gender awareness, (b) pattern awareness, and (c) prediction awareness, each time included as fixed effect and as random by-item slope. For aptitude, we ran models with (a) phonetic encoding, (b) rote memory, and (c) analytical ability, each time included as a fixed effect and as random by-item slope.

For the baseline model, the results showed that there was no main effect of trial type—there was no overall speed advantage when determiners were predictive of the correct answer ($b = –53.74$, $SE = 48.14$, $t = –1.12$, $p = .26$, 95% CI [–148.61, 41.1]). For the different levels of awareness, we found that gender awareness did not lead to a speed advantage. However, both pattern awareness and prediction awareness led to a speed advantage: There were significant interactions between trial type (same vs. different gender) and pattern awareness ($b = –132.25$, $SE = 62.16$, $t = –2.13$, $p = .03$, 95% CI [–256.58, –7.92]) and between trial type and prediction awareness ($b = –274.75$, $SE = 66.3$, $t = –4.14$, $p < 0.001$, 95% CI [–407.35, –142.14]), such that awareness was associated with faster responses on different gender trials. Given that pattern awareness and prediction awareness overlapped substantially, we also checked whether pattern-aware/prediction-unaware learners ($n = 14$) gained a speed advantage. We found no speed advantage in this subgroup—that is, no main effect of trial type ($b = –1.02$, $SE = 62.98$, $t = –0.02$, $p = .99$, 95% CI [–126.99, 124.95]), which might mean that only prediction awareness really contributed to learners' speed gains. Table 4 shows that prediction-aware learners were able to choose the correct image almost 300 milliseconds faster in the different-gender trials compared to the same-gender trials, while no such advantage was present for the prediction-unaware learners. For the effects of the aptitudes on speed advantage gains, we found no evidence that any of the aptitude measures modulated learners' speed gains.

Discussion

Predictive language processing has been observed frequently in both L1 and L2 users and has been argued to be a core feature of language processing, vital to our ability to arrive at a proper understanding of incoming speech quickly and effortlessly. It seems that L2 learners can learn to use features of their L2 predictively but that this is more or less likely depending on a number of factors, such as the characteristics of the L1 (e.g., Dussias et al., 2013; Van Bergen & Flecken, 2017), and the quality of the L2 knowledge (e.g., Hopp, 2013, 2016). Little is yet known about other potentially mediating factors, such as learners' cognitive resources and awareness, while there are good reasons to assume that these may affect L2 prediction (e.g., Brooks & Kempe, 2013; Huettig & Janse, 2016; Kukona et al., 2016). The present study used a miniature language learning paradigm to expose learners to an entirely new language with gender marked determiners to establish whether awareness and cognitive aptitudes might be related to the ability to learn to use determiners as a cue to predict upcoming information.

An interesting outcome of the debriefing procedure was that learners seemed to go through several stages of awareness in acquiring determiner-based prediction. We observed progressively deeper levels of awareness for the target structure. One-third of the participants simply did not work out the determiner system, but a majority did develop awareness for the gender distinction. Some of these gender-aware learners also noted the agreement between the determiner and the noun ending (pattern awareness), and some learners even became aware of their own predictive processing (prediction awareness). Both in terms of eye movements and RTs, our data show that only prediction-aware learners developed determiner-based prediction, which suggests that learning was contingent upon developing awareness for the target. Because gender awareness seems to have been a prerequisite for predictive awareness to develop (all prediction-aware learners were also gender aware), our data strongly suggest that prediction awareness was the endpoint of the learning process (attained only by some) that ultimately led to predictive behavior. This pattern of development is in line with a heavily debated claim in SLA research that structures have to be noticed before they can be learned (e.g., Schmidt, 2001; see Andringa & Rebuschat, 2015, for a discussion) and reminiscent of Bialystok's (1994) analysis and control framework, in which she proposed that language learning is essentially a process of developing awareness for the structure of language and learning to use that structure effortlessly. The results are also in line with Brooks and Kempe (2013), who also clearly observed that self-developed awareness was the best predictor of L2 learning in their artificial language learning study.

The present findings might also be construed as being in line with studies that suggest that L2 knowledge needs to be of a certain quality before learners can use it to generate predictions (e.g., Hopp, 2013, 2016). Perhaps some degree of explicit awareness is a requirement for L2 prediction, or even prediction more generally. However, Andringa and Curcic (2015) did not find that awareness as generated through instruction led to predictive behavior. In this study, the exposure was brief and the L2 knowledge probably insufficiently consolidated. In the larger L2 literature, abundant naturalistic exposure has similarly been found to be more beneficial for online processing than explicit classroom instruction (e.g., Frenck-Mestre, 2002; Pliatsikas & Marinis, 2013).

We did not observe clear links between predictive behavior and our aptitude measures. Very few studies have been done so far; they have looked at working memory, all using different measures and study designs, and the evidence they provide about links between working memory and prediction is mixed (e.g., Huettig & Janse, 2016; Ito, Corley, & Pickering, 2018; Kukona et al., 2016). This study did not observe links between prediction and three aptitude components that have been hypothesized to be linked to noticing and pattern recognition (Skehan, 2016). However, we did observe a tendency that learners' phonetic encoding ability predicted whether learners would develop prediction awareness, which was in turn associated with predictive processing. This might imply that phonetic encoding is not related to predictive behavior per se but facilitates the analysis process that leads to predictive behavior. The observed links between rote learning and IQ with the development of gender awareness perhaps similarly suggests that these cognitive aptitudes may help learners to analyze the input they are exposed to and notice the patterns present in it. High rote learning aptitude, for example, may help establish well-entrenched associations between nouns and their determiners as a prerequisite for further analysis. While we did not investigate to what extent awareness mediates potential links between cognitive abilities and learning, our findings are similar to those of Brooks and Kempe (2013), who observed that nonverbal intelligence and sequence-learning ability were predictive of how likely learners were to discover rules in the input. Such results emphasize the multidimensional nature of aptitude for learning and its complex interactions with factors such as the nature of the input and learning stages (e.g., DeKeyser, 2012; Dörnyei, 2006; Robinson, 2005; Skehan, 2002). Much more research is needed, as it is difficult to formulate well-informed hypotheses on the basis of current empirical findings about the aptitudes that may underlie the learning of particular structures, in particular contexts, at particular stages of learning.

An intriguing question is whether the predictive processing that we observed in this study was completely strategic and intentional. One might argue that the prediction observed here was the result of task-awareness and strategic behavior. While we cannot rule out this possibility, we have some reasons to argue that prediction was to a considerable extent based on intuitions developed during the exposure, which in turn led to awareness of predictive processing. First, learners who reported prediction awareness during debriefing described it as a feeling they had during the test, rather than a strategy that they intentionally adopted. This observation, although not systematically investigated in our studies, implies that prediction awareness may have developed as a result of learners becoming aware of their own predictive processing. Second, because of the nature of the target structure used in our study, we argue that knowledge deeper than prediction awareness was needed to show predictive processing. A simple comparison with Andringa and Curcic (2016) can illustrate this point. This study demonstrated that provision of explicit instruction about differential object marking led to higher rates of predictive processing, which was indicative of strategic prediction. Once learners were told that the preposition announces an animate object in a sentence, it was easy for them to predict without any deeper knowledge of the structure. However, such strategic performance would not be possible with determiner-based prediction, where even if learners were told that they can predict based on the determiner, they would not be able to do it without having well-entrenched knowledge of all different determiner–noun combinations. Therefore, even if one would assume that all processing found in our study was deliberate and strategic from the very start, learners needed substantial knowledge of the target structure to show prediction, and prediction awareness in itself would not have been enough.

Our data have important methodological implications for both L1 and L2 online processing research. The potential role that awareness may play in learners' or native speakers' online processing is very important to investigate given that most L1 and L2 research implicitly assumes that early processing observed with time-sensitive measures (e.g., ERPs, eye movements, RTs) may guarantee that the observed processing is implicit. However, using a particular measure of processing cannot guarantee that the processing observed is explicit or implicit (Godfroid & Winke, 2015; Morgan-Short, Faretta-Stutenberg, & Bartlet-Hsu, 2015). Our data underscore this and point to the importance of thorough debriefing (Rebuschat et al., 2015). Both L1- and L2-processing studies frequently report debriefing at the level of asking learners about the purpose of the experiment. This study shows that this will not allow researchers to detect more subtle levels of awareness. For example, in our study, when asked, none of the 100 learners were aware of the general focus of the experiment. Yet, when probed, 62 learners were gender aware, and 16 learners reported task-related prediction awareness in the form of a feeling that determiners were sometimes helpful in choosing the correct answer more quickly during the test. The importance of thorough debriefing extends far beyond the visual-world eye-tracking paradigm and L2 learners, but it is equally if not even more important in studies that rely on native speakers and methods based on gauging learners' sensitivity to grammaticality or plausibility violations.

While it may seem a contradiction in terms to use a language that does not have native speakers to study issues of attainment in L2 learning, we hope this study has shown that there are benefits to be gained from such an approach. It shows that determiner-based prediction was an achievable goal within this limited language after relatively brief exposure and that this approach offers the advantage of studying learning outcomes in highly similar and controlled conditions. The data show that even brief exposure may already lead to large differences in learning outcomes. It may well be that studying language learning in such an isolated context is crucial for making progress in identifying the cognitive characteristics that are predictive of L2 learning and linking particular aptitudes to particular stages of learning and particular types of input. More research in this area will be needed that focuses on different aspects of these factors (e.g., different aptitudes, different kinds of awareness) and relates them to processing of different language structures. Finally, it would be necessary to conduct studies that also investigate to what extent native speakers differ in their ability to show predictive processing and what lies behind such processing, that is, whether L1 processing is purely implicit or potentially accompanied by some level of awareness, as shown in this study. Studies on L1 processing almost exclusively focus on group results and normally find evidence of predictive processing at group levels, which perhaps falsely creates the impression that every single L1 speaker in the group demonstrates the processing under investigation. Recent research has provided substantial evidence that native speakers vary greatly in their L1 knowledge and proficiency levels, especially when it comes to the acquisition of more complex structures (e.g., Dabrowska, 2012). More extensive research into the range of variation in L1 processing would enable setting fair expectations when evaluating processing of L2 learners in terms of whether or not they can reach nativelike processing levels.

Final revised version accepted 2 August 2018

Notes

1 The data presented in this study were collected for the purposes of two other studies. This implies that the data were collected at two moments in time. The two studies and the two corresponding data sets differed in the duration of exposure that learners received. For the purposes of the present study, we decided to merge these two data sets to obtain more statistical power. This was possible because exposure length did not affect determiner-based prediction. In these studies, participants were also given grammaticality judgment tasks to probe their knowledge, but these were always presented after the picture-matching task with eye tracking.

2 More information on the LLAMA D test and the whole LLAMA aptitude battery can be found at: http://www.lognostics.co.uk/tools/llama/.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Andringa, S. J., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition. *Studies in Second Language Acquisition*, *37*, 237–268. https://doi.org/10.1017/S0272263115000017

Andringa, S., & Curcic, M. (2016, June). *A validation study: Is visual world eye-tracking suitable for studying implicit learning?* Poster presented at the Fifth Implicit Learning Seminar, Lancaster, England.

Andringa, S. J., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning. *Studies in Second Language Acquisition*, *37*, 185–196. https://doi.org/10.1017/S027226311500008X

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. https://doi.org/10.1016/j.jml.2012.11.001

Batterink, L., & Neville, H. J. (2013). Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience*, *25*, 936–951. https://doi.org/10.1162/jocn_a_00354

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*, 157–168. https://doi.org/10.1017/S0272263100012857

Booij, G. (2002). *The morphology of Dutch*. Oxford, England: Oxford University Press.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*, 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition*, *41*, 281–296. https://doi.org/10.3758/s13421-012-0262-9

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Council of Europe (2011). *Common European framework of reference for languages: Learning, teaching, assessment*. Council of Europe.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*, 219–253. https://doi.org/10.1075/lab.2.3.01dab

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534. https://doi.org/10.1080/01690960143000074

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*, 465–480. https://doi.org/10.1006/jmla.1999.2688

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience*, *21*, 433–446. https://doi.org/10.1162/jocn.2008.21031

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*(Suppl. 2), 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. *2*, pp. 395–406). London, England: Routledge.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. https://doi.org/10.1038/nn1504

Dink, J. W., & Ferguson, B. (2015). *EyetrackingR: An R library for eye-tracking data analysis*. Retrieved from http://www.eyetrackingr.com

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*, 42–68. https://doi.org/10.1075/aila.19.05dor

Doughty, C. J. (2018). Cognitive language aptitude. *Language Learning*, *68*, in press. https://doi.org/10.1111/lang.12322

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*, 1–35. https://doi.org/10.1017/S0272263112000915

Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, *93*, 370–385. https://doi.org/10.1111/j.1540-4781.2009.00896.x

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*, 379–399. https://doi.org/10.1017/S136672891000012X

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say “book”? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, *30*, 768–780. https://doi.org/10.1080/23273798.2015.1016047

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In J. Altarriba & R. Herridia (Eds.), *Bilingual sentence processing* (pp. 218–236). Amsterdam, Netherlands: Elsevier.

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners’ eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*. Amsterdam: John Benjamins.

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*, *35*, 105–129. https://doi.org/10.1075/lllt.35.04gra

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 665–703. https://doi.org/10.1111/lang.12018

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. https://doi.org/10.1177/0267658312437990

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*, 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*, 277–307. https://doi.org/10.1177/0267658315624960

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*, 80–93. https://doi.org/10.1080/23273798.2015.1047459

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. https://doi.org/10.1017/S1366728917000050

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*, 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kukona, A., Braze, D., Johns, C. L., Mencl, E., Van Dyke, J. A., Magnuson, J. S., . . . Tabor, W. (2016). The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica*, *171*, 72–84. https://doi.org/10.1016/j.actpsy.2016.09.009

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59. https://doi.org/10.1080/23273798.2015.1102299

Leow, R. P., & Donatelli, L. (2017). The role of (un)awareness in SLA. *Language Teaching*, *50*, 189–211. https://doi.org/10.1017/S0261444817000039

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

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. https://doi.org/10.1016/j.jml.2010.07.003

Loerts, H., Wieling, M., & Schmid, M. (2013). Neuter is not common in Dutch: eye movements reveal asymmetrical gender processing. *Journal of Psycholinguistic Research*, *42*, 551–570. https://doi.org/10.1007/s10936-012-9234-2

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marsden, E., Williams, J., & Liu, X. (2013). Learning novel morphology: The role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, *35*, 619–654. https://doi.org/10.1017/S0272263113000296

Martin, C., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1079). https://doi.org/10.1038/s41598-018-19499-4

Martin, C., 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. https://doi.org/10.1016/j.jml.2013.08.001

McDonough, K., & Trofimovich, P. (2016). The role of statistical learning and working memory in L2 speakers' pattern learning. *The Modern Language Journal*, *100*, 428–445. https://doi.org/10.1111/modl.12331

Meara, P. (2005). *Llama language aptitude tests*. Swansea, England: Lognostics.

Morgan-Short, K., Faretta-Stutenberg, M., & Bartlet-Hsu, L. (2015). Contributions of event-related potential research to issues in explicit and implicit second language acquisition. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 349–384). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/sibil.48.15mor

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*, 154–193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*, 167–182. https://doi.org/10.1017/S136672891200017X

Rebuschat, P. (Ed.). (2015). *Implicit and Explicit Learning of Languages*. Amsterdam: Benjamins.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, *37*(2), 299–334. https://doi.org/10.1017/S0272263115000145

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. https://doi.org/10.1017/S0267190505000036

Robinson, P. (2007). Aptitude, abilities, contexts and practice. In R. DeKeyser (Ed.), *Practice in second language learning: Perspectives from linguistics and psychology* (pp. 256–286). Cambridge, England: Cambridge University Press.

Sabourin, L., & Stowe, L. (2008). Second language processing: When are first and second language processed similarly? *Second Language Research*, *24*, 397–430. https://doi.org/10.1191/0267658306sr259oa

Sasaki, M. (1996). Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses. New York, NY: Peter Lan.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–94). Amsterdam, Netherlands: John Benjamins.

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17–40). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/bpa.3.02ske

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*, 173–204. https://doi.org/10.1017/S0272263105050102

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, *22*, 1097–1106. https://doi.org/10.1080/09541440903325178
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. https://doi.org/10.1016/j.jml.2016.05.003
Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 443–467. https://doi.org/10.1037/0278-7393.31.3.443
Van Petten, C., & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176–180. https://doi.org/10.1016/j.ijpsycho.2011.09.015
Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*, B63–B72. https://doi.org/10.1016/j.cognition.2005.07.001
Wechsler D. (2008). *Wechsler adult intelligence scale* (4th ed.). San Antonio, TX: Pearson.
Wen, Z. E., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*, 1–31. https://doi.org/10.1017/S0261444816000276
Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269–304. https://doi.org/10.1017/S0272263105050138

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Target Language Words Used in the Exposure.
Appendix S2. Target Structure Exposure Items.
Appendix S3. Items of the Test with Eye Tracking.
Appendix S4. Debriefing Protocol.