Forcing prediction increases priming and adaptation in second language production

Theres Grüter¹, Yanxin Alice Zhu¹ & Carrie N. Jackson²

¹University of Hawai‘i at Mānoa
²Penn State University

Abstract

This paper presents a priming experiment with Korean learners of English designed to test the hypothesis that engagement in prediction and the consequent computation of prediction error will lead to increased priming and adaptation. Participants in the guessing-game condition, who had to predict a virtual partner's description of prime pictures, showed marginally greater immediate priming and significantly greater adaptation in terms of change from baseline to post-test than those in a standard repetition priming condition, consistent with error-based learning accounts of structural priming. Effects were largest for learners in the middle of the proficiency range. Findings from this study suggest that priming best facilitates L2 learning when learners are engaged in the proactive creation of expectations about upcoming information.

Introduction

Our success in daily life is critically dependent on our ability to adapt to our environment and to learn from our mistakes. This is evident in our successful use of language in everyday communication, as well as in our ability to learn a language from processing the linguistic and non-linguistic information in the input we experience. Traditionally, these two abilities—language production and comprehension on the one hand, and language learning and development on the other—were the focus of largely separate academic fields, with psycholinguists attending to language production and comprehension mechanisms in (mostly) native adult language users, and applied linguists and developmental psychologists attending to language learning. More recently, approaches from the broader field of cognitive science have sought to account for properties of language processing and learning via a set of more unified cognitive mechanisms. In particular, the view of language processing as driven, at least in substantial part, by prediction (Clark, 2013; Pickering & Gambi, 2018), and prediction error constituting a driving force in learning (Chang et al., 2006; Jaeger & Snider, 2013), has presented an exciting framework for unifying the study of language processing and language learning. The extension of this perspective to the field of Second Language Acquisition (SLA) is still in its infancy (but see Jackson &
Forcing prediction increases priming

Hopp, 2020; Kaan & Chun, 2018; for further discussion, Hopp, this volume), yet we believe holds great promise not only for increasing our understanding of the mechanisms that underlie language learning after early childhood, but also for a better integration of SLA in the wider field of cognitive science.

The study we present in this paper was motivated by the proposal that prediction error drives (second language) learning. Within this context, we define prediction broadly as expectations about information that has not yet been encountered. Within broader theoretical frameworks of predictive coding in cognition (e.g., Bar, 2007; Clark, 2013), prediction is assumed to operate largely outside of conscious awareness. However, the relation between prediction and conscious awareness, and in particular the extent to which attention and awareness guide the formation of expectations remains largely unknown (Meijs et al., 2018), and was listed among key questions for future research by Bar (2007). Against this backdrop, we deliberately adopt a broad definition of prediction that includes presumably implicit, contextually conditioned pre-activation of upcoming words in a sentence, as typically examined in psycholinguistic experiments on predictive sentence processing (e.g., Pickering & Gambi, 2018), as well as more explicit guessing about unknown information, as in the learning of new words (e.g., Potts et al., 2019). This broad definition is further motivated by our focus being not on the nature of prediction itself.
but on the downstream consequences of comparing predicted with actually encountered information, i.e., the computation of prediction error. Error-based learning has been a topic of relatively recent interest in the psycholinguistic literature on sentence processing, but has a long-standing history in learning theory (e.g., Rescorla & Wagner, 1972; see also Gambi, this volume). Drawing on insights from across these literatures, we hypothesized that if second language (L2) learners are put in a situation where they are forced to predict upcoming linguistic input, and forced to attend to potential discrepancies between their prediction and the actual input they then encounter, they will be more likely to use the encountered structures in their own subsequent production. To test this hypothesis, we conducted a written production priming experiment, focusing on double-object datives, with Korean L2 learners of English. The results show that participants who were forced to predict exhibited marginally greater immediate priming and significantly greater adaptation in terms of change from baseline to post-test, than participants who followed a standard repetition priming procedure.

**Structural priming and learning from prediction error**
Research on structural priming in language production and comprehension has shown that language users are more likely to use a particular syntactic structure when they have recently encountered that structure (Bock, 1986; Pickering & Ferreira, 2008). Such alignment with recent input was originally attributed to increased activation of the structure in question, making it temporarily more accessible to the production and comprehension systems (Pickering & Branigan, 1998). Residual activation accounts provide a straightforward explanation for immediate priming, that is, increased use of a primed structure in an immediately following target trial. Yet these accounts are less able to capture the observation that priming can persist over longer and variable time intervals (e.g., Bock & Griffin, 2000; Kaschak et al., 2014). Such long(er) term priming effects that persist beyond immediately adjacent trials gave rise to alternative accounts that attribute these effects to mechanisms of implicit learning (Chang et al., 2006; Dell & Chang, 2014). In this paper, we will refer to such longer term priming effects as adaptation; we use the term immediate priming for effects in immediately adjacent trials.

While the specifics of the proposed learning mechanisms remain a matter of on-going debate (Myslín & Levy, 2016), an assumption common to all implicit learning accounts is that it is error-driven (Dell & Chang, 2014; Jaeger & Snider, 2013). In other words, the adaptive behavior that we observe is assumed to be the result of the comprehender perceiving a
mismatch between a structure they encountered in the input and the
structure they would have expected in that context given their knowledge
and previous experience. For example, if upon observing lightning hit a
church, a speaker says *That church was just hit by lightning*, the use of a
passive will cause a comprehender to experience a certain amount of
surprisal given that it is generally more common to describe transitive
events in the active voice. Due to this mismatch between an observed
structure and one that would have been more likely given the
comprehender's previous experience, the comprehender's bias for
expecting and producing passives in the future will be adjusted in the
direction of increased likelihood of passives. The benefit of this
adjustment lies in the decrease of surprisal and disruption the
comprehender will experience the next time they encounter a passive to
describe a transitive event. The size of the adjustment will be a function of
the amount of surprisal experienced: the more unexpected the encountered
structure, the greater the adjustment. This is known as the "inverse
frequency effect," which is reflected in greater priming for infrequent
versus frequent constructions (Pickering & Ferreira, 2008).

The overall goal of error-driven learning is to develop a system in
which expectations optimally match the input and minimize surprisal. The
hypothesized adjustments to the system due to prediction error can be
captured in terms of adjusted weights in a recurrent network, or a shift in
distributional expectations conditioned by surprisal within a Bayesian belief update model (for further discussion see Myslín & Levy, 2016). A critical assumption underlying all models of learning by prediction error is that language users generate predictions. This assumption seems uncontroversial in the context of broader claims that human brains are in essence "prediction machines," and that prediction is a pervasive property of human cognition and behavior far beyond language processing (Clark, 2013). Recent work has shown, however, that the contribution of top-down anticipatory processes relative to bottom-up integrative processes in various cognitive domains, including language comprehension, can vary substantially depending on multiple factors that are still not well understood. In the context of spoken language processing by adult native speakers, for example, participant-level factors such as age (e.g., Wlotko et al., 2012) and literacy (e.g., Mishra et al., 2012; for review, see Huettig & Pickering, 2019) have been shown to modulate language users' reliance on prediction. Similarly, task-related factors, such as the amount of preview time in visual-world experiments (Ferreira et al., 2013; Sorenson & Bailey, 2007) or the rate at which linguistic stimuli are presented (Ito et al., 2017), can modulate engagement in predictive processing. Such observations have led to on-going debate on the extent to which prediction is a driving, or even a necessary, force in language processing (Huettig, 2015; Ferreira & Chantavarin, 2018).
Forcing prediction increases priming

This view of prediction as a potentially less pervasive and more variable force in language processing has important consequences for accounts of language learning based on the computation of prediction error. If participant- and task-related factors modulate reliance on prediction, then variability in these factors should be predictive (no pun intended) of the amount of learning that occurs. In other words, reduced engagement in prediction, and thus reduced experience of prediction error, should lead to reduced learning. For implicit learning accounts of priming, this leads to the hypothesis that the size of priming effects will be modulated by participant- and task-related variability in reliance on prediction. As far as we know, no experimental study has directly tested this hypothesis in the context of syntactic priming. Studies on lexical priming, however, have shown that drawing participants' attention to the relatedness between primes and targets, either explicitly (Holcomb, 1988) or implicitly by manipulating the proportion of semantically associated prime–target pairs (Lau et al., 2013), can increase the degree to which the prime affects the processing of the target. Similarly, Brothers et al. (2017) demonstrated that effects of lexical prediction during sentence processing were enhanced when participants were explicitly instructed to try and predict the last word of a passage, as well as when the overall validity of predictive cues across the experiment was increased. These findings suggest that top-down goals and strategies, both implicit and explicit, can
modulate priming and prediction at the lexico-semantic level. Here we test whether explicit instructions to predict another speaker's description of a picture will increase the likelihood that the participant will subsequently produce the syntactic construction used by that speaker in their own descriptions. This study thus presents an initial attempt to extend the investigation of effects of top-down strategies on prediction to syntactic priming.

**Prediction in L2 processing-and structural priming**

Apart from the participant-level and task-related factors examined in research with native speakers, additional factors such as language proficiency and native-speaker status have been proposed to modulate engagement in prediction during language processing among non-native and bilingual speakers (Kaan et al., 2010; Kaan, 2014; Grüter et al., 2017; Peters et al., 2018). For example, Grüter et al. (2017) proposed that non-native speakers have *Reduced Ability to Generate Expectations* (RAGE) during sentence and discourse processing. The RAGE proposal was motivated by findings showing reduced effects of prediction among L2 versus native (L1) speakers in sentence and discourse processing studies using both online (Grüter et al., 2012; Martin et al., 2013) and offline
methodologies (Grüter et al., 2017). In the meantime, several studies have shown that L2 users are capable of engaging in predictive processing, sometimes to the same level as L1 users (e.g., Dijkgraaf et al., 2017). Others have reported smaller, delayed, or no effects of prediction among L2 users (e.g., Mitsugi & MacWhinney, 2016). Taken together, the current knowledge base suggests that L2 users can engage in prediction, but overall appear to do so to a lesser extent than native speakers. Why this is the case and what factors modulate L2 users' engagement in prediction, remain a matter of on-going investigation (for review, see Kaan, 2014).

What matters for present purposes is the descriptive generalization that, overall, L2 users appear to rely less on prediction during language processing than native speakers. In consequence, they should experience prediction errors less often, and thus encounter fewer opportunities to learn through the mechanisms proposed in accounts of error-based learning. Kaan and Chun (2018) explicitly appealed to this possibility in the context of a written structural priming study with Korean L2 learners of English targeting ditransitive constructions. Results showed cumulative adaptation effects in that the learners' likelihood to produce a construction increased the more often it had been encountered over the course of the experiment. Somewhat surprisingly, however, only weak and non-significant effects of immediate priming were observed. While acknowledging that such an explanation alone could not account for the
Forcing prediction increases priming

full range of observations in their experiment, the authors suggested that "the L2 learners may not have actively predicted the elements after the verb in the prime sentence," and therefore "[u]nder error-based views of priming, the learners will not have received an error signal when encountering a DO prime, and hence will not have adjusted their representations in response to the prime structure" (Kaan & Chun, 2018, p. 239). This is fully consistent with both error-based learning accounts of structural priming and with what we know about the role of prediction in L2 processing. As an explanation of Kaan and Chun's findings, however, it remains speculative, as their experimental paradigm did not directly measure or manipulate participants' engagement in predictive processing.

Nevertheless, the possibility raised by Kaan and Chun (2018) has potentially important implications for applied linguistics and SLA, where the view of structural priming as implicit learning has given rise to the development of pedagogical applications that seek to integrate priming into curricular activities, in the hope that such activities will facilitate learning of targeted structures (e.g., McDonough & Chaikitmongkol, 2010). Outcomes of these studies have been variable (for review, see Jackson, 2018). We hypothesize that one reason why structural priming activities with L2 learners have shown variable outcomes is that learners partook in these activities with no or only limited engagement in prediction. In the present study, we test this hypothesis by presenting one
group of L2 learners with a task that not only forced them to predict but also to assess whether their prediction was correct, while another group was presented with the same linguistic input in a task that required no prediction.

The role of L2 proficiency in predictive processing and structural priming

In a programmatic review of factors modulating L2 learners' engagement in prediction, Kaan (2014) included L2 proficiency as one such potential factor. Several studies have since investigated the role of proficiency in predictive processing directly. Somewhat surprisingly, despite a report of significant modulation of lexico-semantic prediction by proficiency in an early study by Chambers and Cooke (2009), more recent studies have reported no significant effects of proficiency on predictive processing at the level of lexico-semantics (Djikgraaf et al., 2017; Ito et al., 2018), morpho-syntax (Hopp, 2015; Mitsugi, 2018), and reference resolution in discourse (Grüter et al., 2017; Kim & Grüter, 2020). Notably, the studies that found significant effects of proficiency all focused on grammatical gender (Dussias et al., 2013; Hopp, 2013; see also Hopp, this volume), and where reported, knowledge of gender assignment correlated strongly with
Forcing prediction increases priming overall L2 proficiency (Hopp, 2013). This makes it difficult to tease apart whether the modulation observed in these studies was due to knowledge of the specific linguistic property under investigation or L2 proficiency more generally. Thus, while proficiency remains a perhaps intuitive candidate for modulating engagement in prediction during L2 processing, the empirical evidence so far has been less conclusive than one might expect.

In the context of structural priming, there appears to be general consensus that crosslinguistic structural priming becomes stronger with increasing L2 proficiency (Van Gompel & Arai, 2018). This has been attributed to increasingly shared syntactic representations across languages as L2 representations become more abstract with increasing experience and proficiency (Hartsuiker & Bernolet, 2017). Fewer studies have examined the role of proficiency in structural priming within the L2. Lower proficiency L2 learners appear to show greater priming when lexical material is shared between primes and targets ("lexical boost," Pickering & Branigan, 1998; Kim & McDonough, 2008), an effect that may be due to explicit imitation strategies in the absence of more abstract syntactic representations among less proficient speakers (Hartsuiker & Bernolet, 2017). Experiments without shared lexical material have shown conflicting results: While Schoonbaert et al. (2007, reanalyzed in Hartsuiker & Bernolet, 2017) found increasing proficiency associated with smaller priming effects in a study with Dutch-English bilinguals targeting
dative constructions in English, Bernolet et al. (2013) found greater priming effects for higher proficiency Dutch-English bilinguals on English genitive constructions. In an attempt to reconcile these findings, Hartsuiker and Bernolet (2017) suggested that in the absence of lexical overlap, we should generally see greater structural priming with increased proficiency as a result of the establishment of more abstract structural representations (analogous to the rationale for crosslinguistic structural priming), yet the explicit memory and imitation strategies that are assumed to account for the increased lexical boost effect in lower proficiency L2 speakers may also operate in the absence of lexical overlap depending on the nature of the construction involved.

In light of the limited and partially conflicting evidence from previous research on the role of proficiency in both predictive processing and structural priming in L2, we included an independent measure of proficiency in the present study with the primary goal of ensuring that proficiency did not present a confound in the critical comparison between the two experimental groups. In order to further explore potentially modulating effects of proficiency, we added proficiency scores as a continuous predictor to the model in a second, exploratory analysis step.

This study
The primary goal of this study is to test the hypothesis that arises from error-based accounts of learning, namely that increased engagement in prediction leads to increased learning. To this end, we conducted a written structural priming experiment with L1-Korean learners of English focusing on the dative alternation in English. Previous work has shown that native speakers of (American) English frequently use both double-object (DO, *The girl fed the squirrel some nuts*) and prepositional dative (PO, *The girl fed some nuts to the squirrel*) constructions, with a potential overall preference for the former (Bock & Griffin, 2000; Jaeger & Snider, 2013). L2 learners of English, on the other hand, tend to have a strong preference for using PDs. This has been shown for learners from a variety of L1 backgrounds (McDonough, 2006), and in particular for Korean learners of English (Kaan & Chun, 2018; Shin & Christianson, 2012). The fact that DOs constitute a dispreferred structure for Korean learners of English makes them an ideal target for the present study as priming should be facilitated by the inverse frequency effect. Conversely, priming would be difficult to show for POs under these circumstances since production of POs will likely be at or close to ceiling at baseline (see also Shin & Christianson, 2012). For this reason, only DOs are primed in this experiment (similar to Experiment 2 in McDonough, 2006). We will report effects of immediate priming, i.e., the likelihood of a DO produced
on a target trial immediately following a (DO) prime, as well as longer-term adaptation, i.e., an increase in the production of DOs in an immediate post-test as compared to production at baseline.

Critically, participants will be assigned to one of two groups: the guessing-game (GG) condition, in which they will be given a task that forces them to predict and attend to prediction error, or the control condition (CC), in which participants will encounter the same linguistic materials but in a task that does not require prediction. This allows us to address the following primary research question:

RQ: Does forced engagement in prediction lead to greater effects of (1) immediate priming, and (2) longer-term adaptation as measured by change from baseline to an immediate post-test?

Based on error-driven models of learning and structural priming, we expect consistently greater priming in the GG than in the CC group. As a second and more exploratory question, we also examine the role of L2 proficiency on the size of priming effects in each phase and group. Given the inconsistency of previous findings on the role of proficiency, we make no predictions regarding effects of proficiency.
Methods

Participants

Thirty-five native Korean-speaking L2 learners of English from the University of Hawai‘i student community, including students in short-term English language programs, participated in this study and were randomly assigned to one of two groups: guessing-game (GG, N = 18, 15 females) or control condition (CC, N = 17, 12 females). In addition to the priming experiment, all participants completed an English cloze test (Brown, 1980) as well as a language background questionnaire, which included self-ratings of their English language skills. Table 1 presents a summary of participants' background information and results of pairwise comparisons. Due to the small sizes, analyses were conducted using non-parametric statistics (Mann-Whitney U, Spearman's rho). Pairwise comparisons indicated no significant differences between the two groups for any of the

1 Data collection was cut short by the outbreak of the covid-19 pandemic, resulting in smaller than planned sample sizes. In view of previous production priming studies with similarly small Ns (e.g., Kaschak et al., 2011), we proceeded with analysis, but acknowledge limited statistical power.
Forcing prediction increases priming variables in Table 1. Cloze test scores and self-ratings were strongly correlated, \( r_S(32) = .76, p < .001. \)

<insert Table 1 about here>

Materials

The overall structure of the priming task is summarized in Table 2; a complete list of all linguistic materials is available at https://osf.io/c3af4/. Twenty-two ditransitive verbs were selected from the materials used by Jaeger and Snider (2013, Appendix A). Only verbs that appeared on vocabulary lists in English textbooks used in Korean middle and high schools and in the vocabulary list for the Korean SAT test were included. Six ditransitive verbs were used for target trials in both the baseline and post-test phase, with the same verbs used in both task phases but associated with different arguments (e.g., *give* was associated with *mother, girl, cake* in the baseline phase, and with *policeman, driver, ticket* in the post-test). Based on Jaeger and Snider's (2013) norming data, three ditransitive verbs in the baseline and post-test phase were biased towards a DO completion and three ditransitive verbs were biased towards a PO completion. The remaining 16 ditransitive verbs were used for prime and target trials in the priming phase, with 8 verbs used once each in primes
and 8 verbs used once each in targets. No verbs or other lexical items were
repeated between primes and targets. Based on Jaeger and Snider's (2013)
norming data, five ditransitive verbs used in the prime sentences and five
ditransitive verbs used in the target sentences were biased towards a DO
completion and the remaining three ditransitive prime verbs and three
ditransitive target verbs were biased towards a PO completion. Each
sentence was paired with a colored clip-art image illustrating the event,
with all arguments labelled and the verb printed below the image in its
infinitive form (see Figure 1). In addition, 48 similarly formatted sentence-
picture pairs were created using intransitive and simple transitive verbs.
These constituted prime trials in the baseline and post-test phase, as well
as prime and target trials in practice and filler items (see Table 2). Four
filler prime-target pairs each were included in the baseline and post-test
phase, and eight such pairs were included in the priming phase. Four
experimental lists were created in which the order of items within each
task phase (baseline, priming, post-test) was pseudorandomized. Thus, all
participants saw the same items in each test phase but in different orders.

<insert Table 2 about here>

<insert Figure 1 about here>
Procedure

All participants completed the language background questionnaire, followed by the priming experiment in either the GG or CC condition (described below). Immediately after the priming experiment, a semi-structured oral interview was conducted with the aim of evaluating the extent to which participants were (a) aware that the focus of the study was on the dative alternation, and (b) consciously trying to align their picture descriptions with those provided by the virtual partner. After the oral interview, participants completed the Brown cloze test, a 50-item written test that has shown high validity and reliability with comparable learner groups (Brown, 1980; Brown & Grüter, 2020. All testing was completed in a single session lasting approximately 60–90 minutes.

Guessing game (GG) condition

Participants in the GG condition were introduced to the experiment as follows: "In this activity, we would like to see how well you can GUESS how another person (Jessica) described pictures in English." They were then introduced to "Jessica," who was depicted and described to fit the stereotype of a native English-speaking American college student. Instructions then continued: "You will take turns guessing how Jessica described a picture, and describing pictures on your own. When the
picture has a GREEN frame, your job is to guess how Jessica described the picture. When the picture has a BLUE frame, your job is to describe the picture yourself. IMPORTANT: Please use ALL the words you see on the screen."

On prime trials in all three task phases, participants were presented with a labelled image, together with a picture of Jessica and the prompt "Write what you think Jessica wrote about this picture" (Figure 2, panel A). They then typed a sentence into a textbox. On the next screen (panel B), they were presented with Jessica's actual sentence—which always consisted of a DO construction in experimental items—together with the sentence participants had typed on the previous screen, and they were asked to indicate whether the two sentences were the same. This step was included to force participants to attend to potential differences between the predicted and the actual sentence, and thus to explicitly compute prediction error. This concluded the prime trial, and participants then proceeded to the next screen (target trial), where they were presented with an image as in Figure 1 and typed a sentence to describe it.

<insert Figures 2 & 3 about here>
Control condition (CC)

Participants in the CC condition were introduced to the experiment as follows: "In this activity, you will learn and use English words, and you will practice putting them together to make sentences to describe pictures. You will read and copy sentences that another person (Jessica) wrote, and you will write your own sentences." They were then introduced to Jessica in the same way as participants in the GG condition, and instructions continued: "You will take turns copying Jessica’s sentences, and describing pictures on your own. When the picture has a GREEN frame, your job is to read and copy Jessica’s sentence by typing it out again. When the picture has a BLUE frame, your job is to describe the picture yourself."

For all task phases, prime trials in the CC condition consisted of a single screen in which participants were presented with a labelled image, together with a picture of Jessica, and Jessica's description of the picture (Figure 3). Participants then re-typed Jessica's sentence into a textbox. The copy–paste function was disabled so that participants were forced to re-type the sentence. They then proceeded to the next screen, consisting of target items in the same format as in the GG condition (Figure 1). The procedure in the control condition thus constitutes a typical production priming paradigm in the written mode. Production priming was chosen so that participants in both groups were engaged in the same basic activity,
typing a sentence, with the critical difference that participants in the CC condition simply repeated Jessica's sentence whereas those in the GG condition had to generate the sentence themselves in anticipation of what Jessica might have written.

In both conditions, participants completed two practice trials, which did not include ditransitives, before the beginning of the baseline phase, followed by the priming phase and the (immediate) post-test. As in other priming studies (e.g., Hartsuiker & Westenberg, 2000; Jackson & Ruf, 2018), transitions between these three task phases were not signaled to the participant, such that participants viewed the priming activity as one continuous task. The experiment was conducted in PsychoPy 2.0 (Peirce et al., 2019).

Data annotation and analysis

Sentences produced in target trials were annotated as "DO" when they contained a recipient NP followed by a theme NP and no preposition (e.g., *The mother gives the girl a cake*, 28.1% of all responses), as "PO_to" when it contained a theme NP, followed by *to* and a recipient NP (e.g., *Mom gives a cake to the girl*, 54.2%), and as "PO_otherPrep" when the recipient was preceded by a preposition other than *to* (e.g., *The mom gives the cake for girl*, 8.7%). These three sentence types together accounted for
91.0% of all target sentences produced (CC: 91.2%, GG: 90.8%). Sentences not fitting these criteria were excluded from further analysis (e.g., *The bride text message and send to groom*). "PO_to" and "PO_otherPrep" responses were collapsed for further analysis into a single category "PO". Non-target inflectional marking (tense, number) and article usage was disregarded for coding purposes. Sentences produced by participants in the GG condition on prime trials were annotated following the same criteria. DO and PO responses accounted for 85.4% of these data.

All analyses are based on data points consisting of DO and PO only. Mixed-effect logistic regression was employed to predict the likelihood of a DO sentence across task phase (baseline, priming, posttest) and group (GG, CC). Since the two groups did not differ by proficiency (see Table 1) and to avoid overfitting, proficiency was not included as a factor in the primary models. In order to explore the role of proficiency, we added cloze test scores to the model in a second step. Cloze test scores, rather than participants' self-ratings, were included as the measure of proficiency given the established reliability and validity of this cloze test with similar participant samples (Brown & Grüter, 2020). All analyses were conducted in R 3.6.0 (R Core Team, 2019) using the lmerTest package (Kuznetsova et al., 2017).
Results

Figure 4 presents the proportion of DOs produced by participants in both groups across the three task phases (baseline, priming, post-test). The production of DOs at baseline was low in both groups ($M_{CC} = .07, SD = .13$; $M_{GG} = .12, SD = .22$), with only 5 (of 17) and 6 (of 18) participants in the CC and GG groups, respectively, producing DOs at all. The proportion of DOs increased in both groups in the priming phase ($M_{CC} = .22, SD = .34$; $M_{GG} = .56, SD = .31$), with 7 CC participants and 16 GG participants producing DOs. These proportions remained similar in the post-test ($M_{CC} = .21, SD = .37$; $M_{GG} = .60, SD = .37$), with 7 CC and 16 GG participants producing DOs.

The likelihood of producing a DO was assessed in a mixed-effects logistic regression model with group (centered and contrast-coded, -.5 = CC, .5 = GG), task phase (dummy-coded, reference level: baseline) and their
Forcing prediction increases priming

interaction included as fixed effects, and participants and items as random effects.\(^2\) The output from this model is presented in Table 3.

<insert Table 3 about here>

Across both groups, significant increases in the likelihood of DOs were observed in the priming vs. baseline phase \((b = 2.59, p < .001)\), as well as in the post-test vs. baseline \((b = 2.60, p < .001)\). These effects were qualified by interactions with group, marginal for the increase in the priming phase, \(b = 1.32, p = .06\), and fully significant for the increase from baseline to post-test, \(b = 1.62, p = .03\). To further explore the interactions, separate models were fit to the data from each group. For the GG group, the production of DOs increased significantly compared to baseline in both the priming phase \((b = 3.14, p < .001)\) and the post-test \((b = 3.28, p < .001)\). The same was found for the CC group, albeit with smaller effect sizes (baseline-priming: \(b = 1.81, p < .001\); baseline-posttest: \(b = 1.65, p = \)

\(^2\) The maximal random effects structure justified by the design should include random participant slopes for task phase and random item slopes for group. The inclusion of these slopes was attempted, but led to boundary singularity errors. The largest model that fully converged included random intercepts only.
Forcing prediction increases priming.

In sum, these results show priming effects in both groups, for both immediate priming and longer term adaptation as measured by performance on the post-test. Critically, effects were larger in the GG than in the CC group, with the between-group difference approaching significance in the priming phase and reaching full significance in the post-test.

Production of predicted primes in the GG group

In order to examine the extent to which GG participants were successful in predicting that Jessica would produce DOs on ditransitive prime items during the priming phase, we analyzed their responses on prime items, i.e., the sentences they guessed that Jessica would produce to describe ditransitive events. Out of the 8 ditransitive primes, GG participants produced a mean of 3.0 DOs (SD = 2.0); no participant produced DOs on more than 6 out of 8 trials (range: 0-6). For comparison with the proportion of DOs produced on target items (Figure 4), we also calculated the mean proportion of DOs out of DO and PO responses only: $M = .44$

---

3 The model for the CC group data did not converge with random intercepts for both participants and items. Items were thus removed from the random effects structure of this model.
Forcing prediction increases priming

\((SD = .29)\). These results indicate that despite the explicit task of guessing what Jessica would say and Jessica's consistent production of DOs, participants were far from consistent in producing DOs when predicting what Jessica would say.

Proficiency

To explore whether L2 proficiency modulated the size of priming effects in this study, we added cloze test scores (centered) as an additional continuous fixed effect to the model reported in Table 3. This model, including all three predictors of interest—group, task phase, and proficiency—and their interactions, failed to converge. We therefore proceeded to conduct two separate models, one including only group, proficiency and their interaction (but not task phase, Model 1), the other including only task phase, proficiency and their interaction (but not group, Model 2).

Model 1 (DO ~ Group * scale(Proficiency) + (1 | Subject) + (1 | Item)) yielded significant main effects of group \((b = 2.02, SE = .67, z = 3.01, p = .003)\) and proficiency \((b = .87, SE = .34, z = 2.58, p = .01)\), but no interaction \((b = .32, SE = .66, z = .48, p = .63)\). The positive estimate for proficiency indicates that more proficient participants were more likely to produce DOs overall. The lack of an interaction indicates that this was
Forcing prediction increases priming

the case to the same extent for participants in both groups. The main effect of group shows that the two groups differed in their production of DOs across the experiment even when variability associated with proficiency was accounted for, thus providing further assurance that proficiency was not a confound in our primary analysis.

Model 2 (DO ~ TaskPhase * scale(Proficiency) + (1 | Subject) + (1 | Item)) showed the main effects of task phase observed in the primary model (baseline-priming: \( b = 4.12, SE = .78, z = 5.30, p < .001 \); baseline-posttest: \( b = 4.13, SE = .80, z = 5.19, p < .001 \)), as well as a main effect of proficiency (\( b = 2.80, SE = .67, z = 4.17, p < .001 \)), which was qualified by interactions with task phase (Proficiency*baseline-priming: \( b = -1.79, SE = .56, z = -3.19, p = .001 \); Proficiency*baseline-posttest: \( b = -2.09, SE = .58, z = -3.63, p < .001 \)). To better understand the interactions between proficiency and task phase, we visualized the marginal effects of these interaction terms using the *interactions* package in R (Long, 2019). As Figure 5 illustrates, increased production of DOs from baseline to priming and post-test phases is greatest among participants in the middle of the proficiency distribution. For those at the lower end, increases are small, as they produced few or no DOs throughout the experiment. At the higher end of the spectrum, participants produced DOs more consistently at baseline and DOs were also produced in later phases, but the difference between baseline and the subsequent two task phases was not as large as
Forcing prediction increases priming for participants in the middle of the proficiency range, who produced few or no DOs at baseline, and increased their production drastically over the course of the experiment.

Despite the absence of an interaction between proficiency and group in Model 1, we wanted to explore to what extent the interaction between proficiency and task phase observed in Model 2 obtained within each group. We therefore fit Model 2 to the data from each group separately. For both groups, the output mirrored the overall pattern, with significant main effects of task phase, proficiency, and negative interactions between the two (CC: Proficiency*baseline-priming: $b = -2.06$, $SE = .88$, $z = -2.33$, $p = .02$, Proficiency*baseline-posttest: $b = -2.48$, $SE = .93$, $z = -2.67$, $p = .008$; GG: Proficiency*baseline-priming: $b = -1.67$, $SE = .70$, $z = -2.37$, $p = .02$, Proficiency*baseline-posttest: $b = -1.94$, $SE = .73$, $z = -2.66$, $p = .008$). These results provide further indication that the role of proficiency was similar in the two groups.

*Exit interview*
The goal of the exit interview was to gauge whether participants may have employed explicit strategies, a possibility that seemed particularly relevant since the dative alternation is explicitly taught in EFL classrooms in Korea. To this end, we annotated and analyzed participants' responses to (1) the yes-no question "Do you think the way Jessica described the pictures influenced how YOU described them?," and (2) the follow-up question "In what way?" to a yes-response on the previous question. Responses were scored by two independent annotators, scoring (1) for 1=yes, 0=no, and (2) for 1=mention of dative alternation in some form, 0=mention of another linguistic property (e.g., article use), NA=no answer. Interrater agreement was high (34/35 and 33/35 respectively), and disagreements were resolved through discussion.

The majority of CC participants (13/17) and all GG participants (18/18) answered yes to the first question, indicating that most participants actively tried to align their linguistic choices with those of a virtual partner. Of the 13 CC participants who indicated that they actively sought to align with Jessica, 8 explicitly mentioned that they tried to do so by using more DOs. Other properties mentioned were usage of articles and verb tense. In the GG group, all but one participant (17/18) explicitly mentioned the dative alternation. These observations suggest that the majority of participants in both groups actively tried to align their responses with the linguistic models provided by a native speaker, but
those in the GG condition were overall more likely to focus on the dative alternation.

**Discussion**

Motivated by previous work on error-driven learning and implicit-learning accounts of structural priming, we set out to test whether forced engagement in prediction would lead to greater effects of (1) immediate priming, and (2) longer-term adaptation as measured by change from baseline to an immediate post-test. Korean L2 learners of English were presented with double-object primes from a virtual partner, Jessica. Those in the GG condition had to predict Jessica's descriptions and then compare their predictions to Jessica's actual descriptions; those in the CC condition merely had to retype Jessica's descriptions, following a standard production priming procedure. Both groups exhibited immediate priming, with significantly more DOs produced in the priming phase following DO primes than at baseline, as well as longer term adaptation as reflected by continually increased production of DOs at post-test. Critically, the increase in DO production at post-test was significantly greater in the GG than in the CC group. The persistence of priming beyond immediately adjacent trials is consistent with error-driven accounts of structural
Forcing prediction increases priming (Bock & Griffin, 2000; Dell & Chang, 2014) and suggests that the effects observed in both groups are unlikely to be driven by explicit memory or imitation strategies alone (Hartsuiker & Bernolet, 2017; Shin & Christianson, 2012). We note, however, that conclusions about longer-term adaptation and learning based on performance on an immediate post-test must be drawn with caution. While this was not logistically feasible in the present study, evidence from a delayed post-test would be desirable to better understand the permanence of what we interpret as learning effects and the observed difference between the forced-prediction (GG) and no-prediction (CC) treatments. Nevertheless, the observation that the increase in the use of DOs was substantially greater in the GG compared to the CC group, in both the priming phase and the immediate post-test, and both when measured in terms of mean proportion DOs produced and proportion of participants producing DOs, strongly suggests that being forced to predict the prime sentence facilitated L2 learners' use of DOs in their own subsequent productions.

Exploratory analyses including L2 proficiency showed no difference between the two groups, but revealed an overall pattern that may help reconcile contradictory findings from previous within-L2 priming studies. Our analyses showed consistent interactions between proficiency and the size of priming effects, both from baseline to the priming phase and from baseline to post-test. The visualization of these
interactions revealed a non-linear relation, with less priming among the lowest and highest proficient participants, and greater priming among those in the middle of the distribution. The absence of priming among low-proficiency speakers is consistent with the proposal that priming requires some form of abstract representation of the structure in question (Hartsuiker & Bernolet, 2017; McDonough & Fulga, 2015). These low-proficiency participants may not yet have had a sufficiently stable representation of DOs to benefit from error-driven learning. Reduced priming among our highest proficiency learners, on the other hand, is consistent with surprisal-based accounts, in that Figure 5 indicated that this reduced effect was driven by greater likelihood to produce DOs at baseline. This suggests that DOs caused less surprisal among these participants, and thus less adaptation as per the inverse frequency effect. The observation that priming was most effective among learners in the middle range of the proficiency spectrum suggests that there might be a "sweet spot" in development when the benefits of error-driven learning are greatest. We hypothesize that this is when an abstract representation has begun to be established but is not yet stable enough to support production in the absence of activation through priming.

We included an exit interview in this study to gauge whether participants engaged in explicit processes during the priming experiment. It is generally assumed that participants in priming experiments are
unaware of their adaptation to primed structures and the nature of the structures under investigation (Bock & Griffin, 2000). There is strong evidence that priming occurs under conditions where explicit strategies are not available (Ferreira et al., 2008), yet few priming studies with healthy L1 or L2 adults report to what extent participants were aware of the purpose of the experiment (for exceptions, see Jackson & Ruf, 2018; Myslín & Levy, 2015). Results from the exit interview show that many participants in both groups were (a) aware of the construction under investigation, and (b) consciously tried to align their productions with those encountered on primes. This observation suggests that the assumption that priming is fully unconscious may not always be warranted, especially in studies with L2 learners. Thus, conclusions about implicit learning should be drawn with caution, and future studies addressing these issues—both with L1 and L2 users—may benefit from including a measure of participants' explicit awareness.

Responses on the exit interview indicated that while the majority of participants in both groups consciously tried to align with the virtual partner, awareness of that partners' use of DO constructions was reported by more participants in the GG (17/18) than in the CC (8/17) group. Given the small sample sizes in the present study, we refrain from further statistical comparison of subgroups of participants who did and did not report awareness of DOs. We note, however, that such comparisons would
Forcing prediction increases priming

be valuable in a larger study to better understand to what extent the increased adaptation effects observed in the GG group derive from the forced-prediction treatment leading to increased attention and awareness more generally, rather than the computation of prediction error as we hypothesized here. Yet to the extent that awareness is related to prediction violation (Lupyan & Clark, 2015), dissociating the two may turn out to be less than straightforward.

More generally, it is clear that additional work is needed to better understand which aspects of the guessing-game manipulation led to the increased priming effects observed in this study. With the goal of involving participants in the control condition in as similar a task as possible, we had decided on a production priming paradigm in which CC participants had to retype the prime sentence. Thus, all participants produced a ditransitive sentence on prime trials; however, only GG participants actively generated one and explicitly compared their own utterance against a model. We therefore cannot exclude the possibility that the observed between-group difference reflects better learning due to other mechanisms involved in generating vs repeating a sentence, rather than due to the computation of prediction error. Recent work by Potts et al. (2019) on novel vocabulary learning has shown better learning success when participants had to guess the meaning of a novel word first than when they were presented with form-meaning pairs from the start.
Importantly, this advantage emerged only in contexts where there was an information gap that participants had a desire to fill. This led the authors to propose that "the act of generating a response to an unfamiliar cue, where the correct response is not yet known, stimulates a desire to close an information gap, leading to enhanced motivation to encode corrective feedback" (Potts et al., 2019, p. 1039; see also Gambi, this volume). It therefore remains possible that the increased priming effects observed in the GG group in this study were due in part to increased motivation to close an information gap present only in the GG condition. Future work is needed to tease apart what specific aspects of the GG treatment beside the computation of prediction error contributed to the greater effects of priming and adaptation that were observed in this study. This will enhance not only our understanding of the role of prediction in L2 processing and learning, but ultimately our ability to harness these benefits to support L2 learning.

In conclusion, the small-scale study reported here presents a first step towards the investigation of how explicit manipulation of L2 learners' engagement in prediction may affect learning in the form of structural priming and adaptation. Findings lend some support to the hypothesis that forced prediction can increase both immediate priming and longer term adaptation, suggesting that the explicit computation of prediction error may benefit L2 learning in the sense of increased likelihood to produce a
previously dispreferred construction. However, the present findings do not allow us to fully tease apart to what extent the differences observed between the two groups in this study are attributable to the computation of prediction error by participants in the guessing game condition, and to what extent other factors related to the guessing-game manipulation, such as increased awareness and motivation to close an information gap, contributed to these different outcomes. Separating these explanations at both empirical and conceptual levels is a challenge that we must leave for future work to address.

Acknowledgments

We are grateful to Hyunwoo Kim for assistance with the creation of experimental materials for this study. We thank Holger Hopp, Edith Kaan and two anonymous reviewers for their very helpful feedback on an earlier version of this paper.

References


Brown, J. D., & Grüter, T. (2020). The same cloze for all occasions?


*Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1634), Article 20120394. https://doi.org/10.1098/rstb.2012.0394
Forcing prediction increases priming


Forcing prediction increases priming

https://doi.org/10.1177/0267658312437990

https://doi.org/10.1075/lab.15011.gru


https://doi.org/10.1016/S0010-0277(99)00080-3


Forcing prediction increases priming

*Language Acquisition, 40*(3), 677-692.

https://doi.org/10.1017/S0272263117000365


https://doi.org/10.1075/lab.4.2.05kaa


https://doi.org/10.1080/01690965.2011.641387

Forcing prediction increases priming


Forcing prediction increases priming

https://doi.org/10.16910/jemr.5.1.3


https://doi.org/10.1017/S1366728914000881


Forcing prediction increases priming


Table 1

*Participant demographics (means, standard deviations and ranges) and between-group comparisons*

<table>
<thead>
<tr>
<th></th>
<th>Guessing Game (GG) group</th>
<th>Control Condition (CC) group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N=18)</td>
<td>(N=17)</td>
</tr>
<tr>
<td>Age</td>
<td>23.8 (SD = 5.1)</td>
<td>22.9 (SD = 3.8)</td>
</tr>
<tr>
<td></td>
<td>(18-42)</td>
<td>(19-34)</td>
</tr>
<tr>
<td>Age of first exposure to English</td>
<td>9.4 (SD = 4.0)</td>
<td>9.1 (SD = 3.6)</td>
</tr>
<tr>
<td></td>
<td>(5-21)</td>
<td>(3-18)</td>
</tr>
<tr>
<td>Length of stay in English speaking environment (in months)</td>
<td>16 (SD = 28)</td>
<td>22 (SD = 36)</td>
</tr>
<tr>
<td></td>
<td>(0-108)</td>
<td>(0-109)</td>
</tr>
<tr>
<td>Cloze test score (/50, acceptable-answer scoring)*</td>
<td>30.2 (SD = 10.0)</td>
<td>26.8 (SD = 10.7)</td>
</tr>
<tr>
<td></td>
<td>(17-47)</td>
<td>(9-45)</td>
</tr>
<tr>
<td>Self-rating of overall English language ability (0-10)</td>
<td>6.1 (SD = 1.7)</td>
<td>5.5 (SD = 2.4)</td>
</tr>
<tr>
<td></td>
<td>(4-9)</td>
<td>(2-10)</td>
</tr>
</tbody>
</table>

*One participant in the GG group did not complete the cloze test.*
Table 2

*Priming experiment: structure and materials*

<table>
<thead>
<tr>
<th>Phase</th>
<th>Experimental items</th>
<th>Fillers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number and structure of prime-target pairs</td>
<td>number and structure of prime-target pairs</td>
</tr>
<tr>
<td>Practice</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>prime: (in)transitive</td>
<td>target: (in)transitive</td>
</tr>
<tr>
<td>Baseline</td>
<td>6 prime: (in)transitive</td>
<td>4 prime: (in)transitive</td>
</tr>
<tr>
<td></td>
<td>target: ditransitive</td>
<td>target: (in)transitive</td>
</tr>
<tr>
<td>Priming</td>
<td>8 prime: ditransitive: DO</td>
<td>8 prime: (in)transitive</td>
</tr>
<tr>
<td></td>
<td>target: ditransitive</td>
<td>target: (in)transitive</td>
</tr>
<tr>
<td>Posttest</td>
<td>6 prime: (in)transitive</td>
<td>4 prime: (in)transitive</td>
</tr>
<tr>
<td></td>
<td>target: ditransitive</td>
<td>target: (in)transitive</td>
</tr>
</tbody>
</table>
Forcing prediction increases priming

Table 3

Model statement and summary of fixed effects in mixed effects logistic regression

Formula: \( DO \sim \text{Group} \times \text{TaskPhase} + (1 \mid \text{Subject}) + (1 \mid \text{Item}) \)

<table>
<thead>
<tr>
<th>Fixed effects:</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.47</td>
<td>0.53</td>
<td>-6.52</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Group</td>
<td>1.06</td>
<td>0.88</td>
<td>1.21</td>
<td>0.23</td>
</tr>
<tr>
<td>TaskPhase (baseline-priming)</td>
<td>2.59</td>
<td>0.49</td>
<td>5.32</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>TaskPhase (baseline-posttest)</td>
<td>2.60</td>
<td>0.51</td>
<td>5.09</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Group : TaskPhase (baseline-priming)</td>
<td>1.32</td>
<td>0.69</td>
<td>1.90</td>
<td>0.06</td>
</tr>
<tr>
<td>Group : TaskPhase (baseline-posttest)</td>
<td>1.62</td>
<td>0.72</td>
<td>2.24</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Figure 1

Example of visual stimuli
Forcing prediction increases priming

Figure 2

Prime trial in the guessing-game (GG) condition

A

Please judge whether your sentence and Jessica's sentence are exactly the same. Press the green key if the two sentences are exactly the same, press the red key if they are different.

Jessica: The official awards the athlete a trophy.
You: The official awards the trophy to the athlete.

B

Write what you think Jessica wrote about this picture.
The official awards the trophy to the athlete.
Figure 3

*Prime trial in the control condition (CC)*

Jessica: The official awards the athlete a trophy.

Please type Jessica's sentence...
The official awards the athlete a trophy.
Forcing prediction increases priming

Figure 4

Proportion of DOs out of all DO and PO utterances by group and task phase. Error bars indicate 95\% confidence intervals on means by participants.
Figure 5

Visualization of marginal interaction terms using the interact_plot function in R; data from both groups combined. Error bands show 95% CIs.