Title
Expectation Violation Leads to Generalization: The Effect of Prediction Error on the Acquisition of New Syntactic Structures

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Bovolenta, Giulia
Marsden, Emma

Publication Date
2021

Peer reviewed

Expectation Violation Leads to Generalization: The Effect of Prediction Error on the Acquisition of New Syntactic Structures

Giulia Bovolenta (giulia.bovolenta@york.ac.uk) and Emma Marsden (emma.marsden@york.ac.uk)
Department of Education, University of York, York, YO10 5DD, United Kingdom

Abstract

Prediction error is known to enhance priming effects for familiar syntactic structures; it also strengthens the formation of new declarative memories. Here, we investigate whether violating expectations may aid the acquisition of new abstract syntactic structures, too, by enhancing memory for individual instances which can then form the basis for abstraction. In a cross-situational artificial language learning paradigm, participants were exposed to novel syntactic structures in ways that either violated their expectations (Surprisal group) or that conformed to them (Control group). Results from a delayed post-test show that participants in the Surprisal group developed stronger representations of the structures' form-meaning mappings and were better able to generalize them to new instances, relative to the Control group.

Keywords: language acquisition; syntax; prediction error; cross-situational learning; artificial language learning

Introduction

Is it possible to 'surprise' a learner into acquiring a new structure in a foreign language? A growing body of literature suggests that unpredictable input favours language learning, in various ways. For instance, violated expectations about a structure's usage contribute to the acquisition of that structure in the long term (Robenalt & Goldberg, 2016; Goldberg, 2016). Additionally, *structural priming* – an increased likelihood to use or expect the syntactic structures we are exposed to, often called *adaptation* when it persists in the long term – is likely one of the mechanisms by which we tune into the patterns of our language (Peter & Rowland, 2019). There is evidence that prediction error drives adaptation to syntactic structure, both from computational modelling (Chang, Dell, & Bock, 2006) and empirical studies with both first language (L1) and second language (L2) speakers (Fazekas, Jessop, Pine, & Rowland, 2020; Montero-Melis & Jaeger, 2019). Fazekas et al. (2020) looked at the adaptation to two different syntactic structures in competition with each other: the direct object dative (DOD) and prepositional dative (PD). They used "prime surprisal", a method based on priming paradigms, relying on the fact that the two structures have different likelihoods of occurring with specific verbs. For instance, the verb *give* is more likely to occur with a DOD structure, while *bring* is more often used with a PD. Priming for either dative structure is generally greater when it is encountered with a non-typical verb (Peter, Chang, Pine, Blything, & Rowland, 2015; Jaeger & Snider, 2013). Fazekas et al. extended this paradigm by adding an immediate post-test to look for evidence of adaptation in production as a consequence of surprisal, in both children (5- to 6-year-old) and adults. In that post-test, both children and adults showed a greater shift towards producing the DOD if it had been presented with non-typical verbs in the priming procedure.

Structural priming and adaptation phenomena affect representations that have already been acquired; what changes as a consequence of exposure, and is further increased by prediction error, is the strength of existing structural representations. However, evidence shows that prediction error can also enhance the formation of new individual memories: events or associations which violate our expectations are remembered better than those that conform with them (*one-shot declarative learning*). Novel associations are better remembered if they violate an established pattern (Greve, Cooper, Tibon, & Henson, 2019; Brod, Hasselhorn, & Bunge, 2018). Stimuli that benefit from one-shot declarative learning include picture-word associations (Greve, Cooper, Kaula, Anderson, & Henson, 2017) and translation word pairs (De Loof, Ergo, Naert, Janssens, Talsma, Van Opstal, & Verguts, 2018). Surprising feedback, too, is better remembered. In Fazio & Marsh (2009), participants answered general knowledge questions (rating their confidence in their answers) and were then shown the correct answer, which was displayed in either red or green letters. When feedback was unexpected (either following a high-confidence incorrect answer, or a low-confidence correct one – in the latter case, it would unexpected because participants did not expect their answer to be the correct one) participants retained better memory of the feedback message. Specifically, they were better at remembering the font colour in which unexpected feedback was displayed, compared to feedback that was expected. This suggests that surprising feedback leads to a greater effort to encode it (known as the *surprise hypothesis)*, resulting in better 'source memory' (defined as memory for the conditions in which the feedback is encoded, including everything that gets encoded besides the content of the feedback itself). While the Fazio & Marsh (2019) study is not directly concerned with the acquisition of new linguistic structures, it shows that learners form stronger representations of material that is presented in a surprising fashion. This suggests the possibility that new linguistic structures, too, may be better remembered when they a presented in an unexpected way, a possibility which we will explore in the present study.

There is also more direct evidence that the effect of violation expectation on novel memory formation can aid language acquisition: Stahl & Feigenson (2017) showed that violation of expectations promotes vocabulary learning in young children. In their study, 3- to 6-year-old children were exposed to novel events which were either entirely possible or which violated core properties of the objects involved (e.g., a cup vanishing and reappearing in a different location). They were then taught the verb corresponding to the action (Experiment 1) or the noun denoting one of the objects (Experiment 2), and were tested immediately on its meaning. Children were significantly more accurate in their responses for verbs and nouns that they had learned in surprising events than for those they had learned in expected events (on which they performed at chance level). The effect was limited to nouns and actions involved in the surprising event: If children were taught the name for an object that was present during the event but did not participate in it, there was no learning effect (Experiment 4). This suggests that violated expectation did not aid learning simply by increasing attention[1] or arousal, but that it led children to revise their predictions about specific objects and events (Stahl & Feigenson, 2017).

We are now beginning to form a picture of the ways in which expectation violation can aid learning with regards to different aspects of language. If a learner already has the relevant abstract syntactic representation, encountering the structure in an unexpected context appears to strengthen that representation. Prediction error can also facilitate the acquisition of new declarative memories for lexical items, such as nouns or verbs, leading to stronger memory formation than non-surprising contexts. But what about the acquisition of *new, syntactic* representations among adult learners who have already established their L1 system? In this study, we address an unexplored gap in the literature, asking whether expectation violation could also aid the development of new abstract structural representations, including acquisition of their specific form-meaning mappings, rather than just strengthening existing ones. Following a usage-based approach to language acquisition, we assume that structural knowledge emerges through abstraction from individual learned exemplars (Bybee & Hopper, 2001; N. Ellis, 2002; N. Ellis, Römer, & O'Donnell, 2016). If expectation violation can aid memory for individual instances, then we hypothesise that it may also aid the acquisition of structural knowledge through abstraction from these individual instances.

The Present Study

To investigate whether expectation violation could aid the development of new syntactic representations, we carried out a controlled learning experiment using an artificial language (Yorwegian). We used an adapted version of the cross-situational learning paradigm used by Walker, Monaghan, Schoetensack, & Rebuschat (2020), in which participants simultaneously learn the vocabulary and grammar of a novel language by listening to new sentences and choosing between possible interpretations for them. Learners were first introduced to a default syntactic structure, the active construction, which they learned while they were also learning the vocabulary of the language. Then, once this structure had been learned and consolidated, participants were exposed on the second day to a more complex alternative, the passive construction. This simulates to some extent the real-life learning experience of many L2 English learners, who are likely to encounter the passive construction at a later stage in their learning due to its lower frequency and higher complexity, relative to the active construction. In this context, we manipulated the utterance containing the passive construction (in what we called a 'feedback' turn), so as to make it either unexpected (Surprisal group) or expected (Control group) relative to the pattern established during training. We hypothesised that participants in the Surprisal group would develop stronger representations for the passive sentences encountered in feedback, leading to improved learning of the passive syntactic structure itself.

[1] Surprisal does also lead to greater attention in infants, however, with evidence suggesting that there is an "optimal" level of surprisal beyond which attention decreases again (Kidd, Piantadosi., & Aslin, 2012).

Method

76 native English speakers (59 females, M_{AGE} = 31, *SD* = 7.62) were recruited from online research platform Prolific and completed the study over the course of three consecutive days, receiving a compensation of £12. Participants were all resident in the United Kingdom at the time of taking part in the study. Participants were randomly assigned to either the Surprisal (n = 39) or Control (n = 37) group on the first day of the study. The slight numerical imbalance between groups is a consequence of attrition, which is more difficult to control in an online study relative to a lab-based one (participants were evenly assigned to the two conditions on Day 1, but not all completed all three days: of 104 participants who began the experiment, only 76 finished it).

Materials

Participants were trained in an artificial language called Yorwegian, consisting of four nouns (*glim, blom, prag, meeb* – man, woman, boy, girl), eight verbs (*flug-, loom-, gram-, pod-, zal-, shen-, norg-, klig-* – call, chase, greet, interview, pay, photograph, scare, and threaten), one determiner (*lu* - the) and one preposition (*ka* - by). The specific word meaning pairs within the noun and verb categories were randomly assigned for every participant. All sentences were SVO, but there were two possible syntactic structures, differentiated by verbal inflection and use of the preposition *ka*. These were the Active structure (e.g., *Lu meeb flug**at** lu prag,* "The girl calls the boy") and the Passive (e.g., *Lu prag flug**es** **ka** lu meeb,* "The boy is called by the girl"). This type of passive construction is naturally found in Scandinavian languages. It was chosen so as to have a way of forming passive structures that would not be entirely familiar to L1 English speakers (as there is no equivalent of the BE auxiliary in Yorwegian), while still being ecologically valid.

We used a set of 208 black and white photographs depicting transitive actions, which we adapted from materials used by Segaert, Menenti, Weber, Petersson, & Hagoort (2012). The main set of training and testing pictures used on all three days (192 images) depicted the eight verbs: *call, chase, greet, interview, pay, photograph, scare,* and *threaten*. There were four characters which could fill the roles of Agent and Patient: *man, woman, girl* and *boy.* All possible combinations of different characters were included for each training verb, which yielded 12 possible Agent-Patient combinations (the Agent and Patient were always played by different characters). In the training set, the 12 Agent-Patient combinations were repeated for each of the eight verbs, yielding a total number of 96 possible scenes. Each scene was enacted twice, each with different actors, giving a total of 192 unique pictures. One set of 96 pictures was used in training blocks, on Day 1 and then again on Day 2, while the other set was reserved for testing blocks. Each picture could appear with one of two possible syntactic structures (Active and Passive constructions), for a total of 384 unique picture-sentence combination. Noun and verb meanings were randomly assigned for each participant. An additional generalisation set was also used (16 images). The pictures in this set depicted four additional transitive verbs (*dress, hug, pull,* and *push*) and were used in a generalisation structure test block on Day 3, to test participants' ability to process the syntactic structures they had been previously exposed to when used with novel verbs.

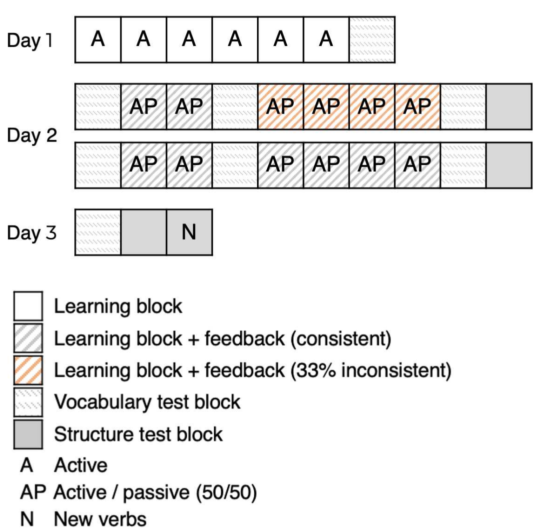

Figure 1: Cross-situational learning task schedule

Procedure

Participants did a cross-situational learning task online over the course of three consecutive days (Figure 1). The average total duration of the study was ~75min, with each of the three sessions taking approximately 25min. On each day, participants had to complete the session between 10am and 6pm. Subjects were randomly assigned to one of two groups, Surprisal or Control. On Day 1, the two groups followed the exact same protocol. On Day 2, participants followed the same procedure with the exception of four blocks (blocks 5-8), where we introduced the surprisal manipulation (described in the next section). On Day 3, both groups again followed the same protocol throughout. Participants were given a debriefing questionnaire at the end of the experiment to assess the development of any explicit rule knowledge.

Cross-situational Learning Task Participants received no explicit instruction on either the grammar rules or vocabulary of Yorwegian. They were taught using an adapted version of the cross-situational task used by Walker et al. (2020), which was also used for testing. Participants heard individual sentences in Yorwegian, while two pictures appeared on screen side by side. Their task was to select the picture that corresponded to the sentence they just heard (*target*) by pressing the left or right arrow on their keyboard; in normal learning and testing trials, they received no feedback on their answers. Trials were presented in blocks of 16 items each (Figure 1). There were four different types of trials: learning trials, vocabulary test trials, structure test trials, and learning trials with feedback (which included the critical between-group manipulation), as follows:

Learning trials. Distractor Agent, Patient and Verb were picked at random by the experimental software.

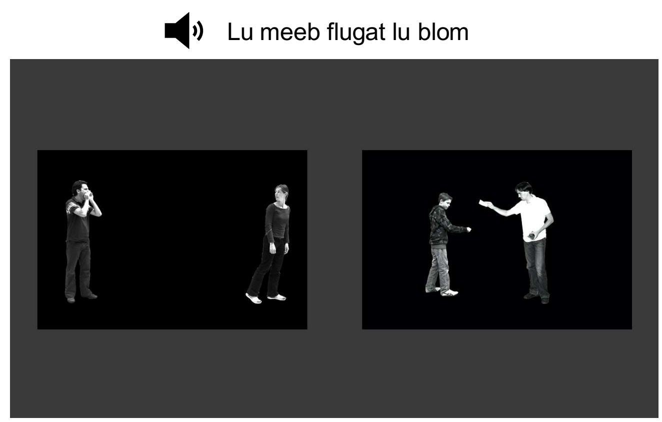

Figure 2: Example of a learning trial

Learning trials with feedback. On Day 2, all learning blocks (Blocks 2-3 and 5-8), contained a proportion of learning trials with feedback. 12 out of 16 learning trials in each of these blocks were followed by feedback on the answer just given: after making their choice (in a learning trial), participants were shown the correct picture which they should have picked, regardless of whether they had picked it or not (in a feedback screen). The saw the correct picture displayed on its own, in the centre of the screen, and they also heard the sentence which they had responded to once again. More precisely, they heard a sentence with the same agent, patient and verb as the one they had responded to, but the syntactic structure used to describe the scene could either be the same (congruent feedback) or different (incongruent). In Blocks 2 – 3, all feedback was congruent and evenly spread across structures: both groups received feedback on 6 passive and 6 active learning trials per block, and the sentence they heard during feedback matched the one they had responded to in both content and structure.

In Blocks 5 – 8, we introduced the between-group 'surprisal' manipulation. Feedback was still given on 12 out of 16 trials, and both groups still received congruent feedback on 8 of these trials (4 active and 4 passive). The remaining 4 learning trials with feedback were manipulated so that the feedback they were followed by was congruent for the Control group, but incongruent for the Surprisal group. In the Control group, these four critical trials required participants to respond to a passive sentence, while in the Surprisal group participants would respond to an active one. This was done to ensure that the feedback itself – the sentence learners were exposed after giving their answer, as they saw the correct picture again – would be in the passive for both groups. In this way, both groups received feedback on their answers 12 times over the course of each block, and of these 12 times, it was accompanied by a passive structure 8 times, and by an active one 4 times[2].

Over the course of the whole experiment, participants saw 16 critical learning trials with feedback (with incongruent feedback for the Surprisal group, but congruent for Control). Each of these critical trials was followed by a structure test trial, which is described below.

Structure test trials. All parameters in the pictures were kept constant apart from Agent and Patient roles, which were reversed from target to distractor (e.g., if the target picture was *The girl interviews the man*, the distractor would be *The man interviews the girl*). The following parameters were always randomly chosen: the position of target and distractor picture on screen (left / right), and the position of Agent and Patient characters inside the pictures (left / right). Structure test trials were included in structure test blocks and also immediately following critical feedback trials.

Noun test trials. All parameters in the pictures were kept constant apart from the Patient noun (e.g., if the target picture was *The girl interviews the man*, the distractor could be either *The girl interviews the woman* or *The girl interviews the boy*). Noun test trials were included in vocabulary test blocks only.

Verb test trials. All parameters in the pictures were kept constant apart from Verb. Verb test trials were included in vocabulary test blocks only.

Debriefing Questionnaire At the end of Day 3, participants were administered a debriefing questionnaire. The first part of the questionnaire included questions on the participants' educational and language background. The second part included specific questions on the experiment itself, aimed at probing participants' awareness of the structures and of the functional distinction between them ("Did you notice that a new type of sentence was introduced on Day 2 (yesterday's session)?", and if Yes, "What were the two types of sentence you learned, and what do you think the difference was between them?").

Results

A total of 70 participants were included in the analysis. Four participants were excluded for failing to listen to the items before giving their responses (the criterion response time for this exclusion decision was under 1s on at least six trials per block, in any given block). One participant was excluded due to suspect unfair means (such as taking notes, based on response times over 10s and 100% accuracy from Block 1 of the cross-situational learning task on Day 1). One participant was excluded for failing to finish the Day 2 task in one sitting. We report data from the three structure test blocks (one at the end of Day 2, and two on Day 3) and from the debriefing questionnaire. Data from structure trials during the learning phase was analysed but is not reported for space reasons (no significant differences were observed between groups in these trials).

Structure Test Blocks

We analysed accuracy data as a binary outcome (correct / incorrect) at the trial level. We used generalized linear mixed-effect models (GLMER) for binomial data, which we implemented in R version 4.0.3 (R Core Team, 2020) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). Following Barr, Levy, Scheepers, & Tily (2013) we used the maximal random structure supported by the model. For each model, we first created a formula containing the maximal fixed effect structure and the maximal random effect structure (random intercepts by subject and item as well as random slopes for subjects and items by each of the fixed effect predictors, and their interactions). We used the package *buildmer* (Voeten, 2020) to automatically identify the maximal random structure that would allow the model to converge. We then used *buildmer* again on the resulting formula to do stepwise backwards model selection using likelihood-ratio tests, eliminating fixed-effect predictors one by one (starting from higher-level interactions) and only retaining them if they significantly improved model fit. We analysed data from the three structure test blocks in individual *glmer* models, entering Group, Structure (Active vs. Passive) and their interaction as predictors in the initial model for each. We report the coefficients of the mixed-effect models converted to odds ratios (*OR*) to provide a measure of effect size, together with the statistical significance of the effects (*p* values).

[2] The numerical imbalance between active and passive feedback was intentional, given that the passive was the 'minority' structure that we intended to boost through increased exposure. This is also why only 12 of 16 learning trials per block were followed by feedback – giving feedback on all 16 would have caused participants to see an equal amount of active and passive feedback.

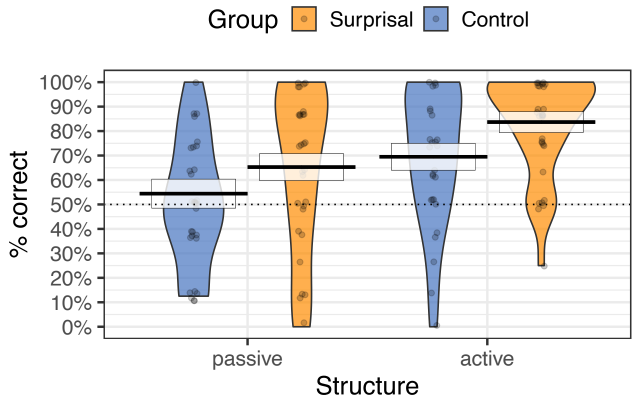

Figure 3: Day 3 structure test block (new verbs). Dotted line represents chance (50%) level accuracy. Shaded rectangles mark 95% CIs.

Day 2 (trained verbs) In the structure test block on Day 2, there was a numerical trend towards higher accuracy in the Surprisal group and for Active sentences (Surprisal group: Active *M* = 74.7% (*SD* = 43.6%), Passive *M* = 66.7% (*SD* = 47.2%); Control group: Active *M* = 65.4% (*SD* = 47.6%), Passive *M* = 57.4% (*SD* = 49.5%), but there were no statistically significant effects of either Group or Structure.

Day 3 (trained verbs) In this test block, we found a significant effect of Structure (*OR* = 2.80, 95% CI [1.50 – 5.23], *p* = .001), due higher accuracy for Active relative to Passive, in both groups (Surprisal group: Active *M* = 78.4% (*SD* = 41.2%), Passive *M* = 62.8% (*SD* = 48.4%); Control group: Active *M* = 75.4% (*SD* = 43.2%), Passive *M* = 58.5% (*SD* = 49.4%). There was no significant effect of Group.[3]

Day 3 (new verbs) In the generalisation structure test block (Figure 2), we found significant main effects of Group (*OR* = 2.42, 95% CI [1.28 – 4.61], *p* = .007) and Condition (*OR* = 3.05, 95% CI [1.56 – 5.95], *p* = .001), but no interaction between the two. Subjects in the Surprisal group were more accurate than the Control group, and both groups had higher accuracy for active structures (Surprisal group: Active *M* = 83.7% (*SD* = 37%), Passive *M* = 65.3% (*SD* = 47.7%); Control group: Active *M* = 69.5% (*SD* = 46.1%), Passive *M* = 54.4% (*SD* = 49.9%).

Debriefing Questionnaire

21 out of 36 subjects in the Surprisal group and 14 out of 34 subjects in the Control group developed sufficient explicit knowledge of the structures to be able to verbalise their respective functions. To assess whether the experimental manipulation had made participants in the Surprisal group more likely to develop explicit knowledge of the Active / Passive distinction, we constructed a simple logistic regression with explicit knowledge as a binary outcome and Group as predictor. While the Surprisal group had a numerically higher rate of explicit knowledge, the effect was not significant (*OR* = 0.50, 95% CI [0.19 – 1.28], *p* = .15). However, a post-hoc power analysis carried out using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) showed the debriefing questionnaire to be underpowered (0.24 power), meaning that we cannot draw definitive conclusions form the lack of a significant effect.

Discussion

Our research question concerned the effect of prediction error on the acquisition of new structural knowledge. We hypothesised that violating expectations at the item level would lead to stronger abstract structural knowledge in the Surprisal group: Our results broadly support this hypothesis. We found that participants in the Surprisal group performed significantly better than those in the Control group in a structure comprehension test using novel verbs, which shows that the Surprisal group had developed stronger abstract knowledge than the Control group, and were able to use that knowledge to generalize that structure to a new lexicon.[4] The effects we observed, however, were not limited to the passive construction as we had hypothesised, given that the manipulation was on passive items only. In the comprehension test, the advantage for the Surprisal group was found across both structures. We consider these findings below, offering possible interpretations for the observed pattern of results and discussing their implications.

In the structure test on Day 3 (new verbs), we found a main effect of Structure and one of Group: Both groups were better at selecting the correct interpretation of active sentences than they were for passive ones, and the Surprisal group was overall more accurate than the Control group. The effect of structure is compatible with our experimental design: Given that participants had received more and earlier exposure to this structure than to the passive, it is not surprising that they developed higher accuracy on it. We also expected the Surprisal group to perform better than the Control group in the structure test, which was confirmed. However, the effect was found for both Active and Passive structures (and was numerically greater for active ones), whereas we had expected to find an advantage specifically

[3] We can only draw limited conclusions from the results of the Day 3 (trained verbs) test block, however, as this block was affected by a counterbalancing error which meant that half of the participants (equally spread among groups) saw the exact same items as in the Day 2 structure test (while the other half saw the same pictures but described using the opposite structure, which was the intended design). This does not affect the following test block (Day 3, new verbs), which used entirely novel Agent – Verb – Patient combinations, with verbs not encountered during training.

[4] While we claim that participants developed abstract structural knowledge, we do not make any specific claims with regards to the relationship between the novel Yorwegian structures and those in the participants' L1 (English). It is possible that participants simply learned an extension of the English active / passive distinction, to which they added the novel morphemes. But it may also be the case that they acquired the new structures as separate representations, which would them become linked to their L1 representations for the active / passive structure.

for passive sentences, given that they were the target of our experimental manipulation. There is, however, a plausible way in which better knowledge of the passive construction could also lead to higher accuracy on active trials, by providing negative evidence that could help participants rule out the incorrect alternative. In our structure test, the competitor (incorrect) picture always depicted the same action happening with agent and patient roles reversed, meaning the two constructions were effectively put in competition against each other. If the sentence was in the active form, e.g., *Lu meeb flug**at** lu prag* ("The girl calls the boy"), then the target picture would depict a girl calling a boy, while the competitor would depict a girl being called by a boy. This means that a sentence with the same nouns in the same positions as the target sentence could be used to describe the competitor picture, but only if it had different morphosyntax, that is, *Lu meeb flug**es ka** lu prag,* ("The girl is called by the boy"). Being sensitive to this distinction would help participants make the correct choice by ruling out the competitor picture, that is, by providing negative evidence of what the active sentence could *not* describe. Crucially, however, this requires specific sensitivity to the morphosyntactic distinction, which would in turn depend on accurate knowledge of the passive construction, as well as the active. Relying only on vocabulary would not be of help in this context, as both pictures could be described by sentences containing the same verb and nouns in the same order.

Another potential explanation for our findings is that surprising feedback did lead to better structural learning, but not in the way we had hypothesised. In this study, we opted to generate surprisal by violating expectations about experimental setup, rather than expectations about a specific structure being used in a specific context, as done in verb surprisal studies (e.g., Fazekas et al., 2020). The reason we opted to do this was to avoid conflating the effect of surprisal with that of variety: To violate expectations, participants in the Surprisal group would have to be exposed to the passive structure in novel contexts, besides those to which both groups were exposed, which would result in the Surprisal group encountering passive sentences in a greater variety of contexts compared to Control. Evidence from artificial language studies shows that the acquisition of novel structures is positively influenced by context variability (e.g., Gómez, 2002), so we sought to avoid potential confounds by violating expectations in a way which would not result in greater context variability. However, in doing so, we unwittingly ran into another potential confound. It is possible that what drove the effect of the surprisal feedback trials was actually the juxtaposition of an active and passive sentence used in sequence to describe the same event, rather than the passive feedback sentence being better encoded due to it being unexpected. This would have showed learners that the two structures could be used to describe the same event, potentially prompting them to pay more attention to the specific form-meaning mappings in the two structures. If learners follow a "uniqueness principle" and assume that any given meaning can only be encoded by one grammatical form (Pinker, 2009), then the presence of two superficially equivalent forms may trigger a search for functional distinctions that may justify the existence of both forms in the grammar. If that were the case, we may expect the Surprisal group to have greater awareness of the functional distinction between the structures.

Unfortunately, we are not able to assess the possibility that juxtaposition of structures is what was driving learning. Firstly, while the debriefing questionnaire did not show any significant differences in awareness between groups, limitations with the tool we used mean we cannot draw any definitive conclusions. The questionnaire had low statistical power, and retrospective verbal report is generally not a very sensitive measure of awareness (Rebuschat, 2013). Secondly, the link we make between the juxtaposition explanation and the emergence of awareness is speculative; as it stands, we have no way to confirm or rule out this explanation given the currently available data. One future development of this research will address the issue by including a measure of item memory, testing for specific memory of the feedback sentences received in the critical feedback trials. If participants do show better memory for passive feedback sentences encountered in the surprising condition, this will lend support to our original hypothesis, that the surprisal manipulation improved memory for individual items, which in turn lead to better generalisation. However, this would not entirely rule out a role for the second potential mechanism just described (i.e., juxtaposition of two structures leading to more accurate representations of structure-meaning mappings). In order to fully investigate this point, further research could include a different way to generate surprisal, that does not result in juxtaposition of an active with a passive sentence describing the same picture. If the same effects are observed, it would suggest that the effect of our experimental manipulation was not primarily driven by the nature of our experimental design.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013-4). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31.

Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

De Loof, E., Ergo, K., Naert, L., Janssens, C., Talsma, D., Van Opstal, F., & Verguts, T. (2018). Signed reward prediction errors drive declarative learning. *PloS One*, *13*(1), e0189212.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*.

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*, *7*(11), 180877.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*(1), 88–92.

Goldberg, A. E. (2016). Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. *Language and Cognition*, *8*(3), 369–390.

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, *13*(5), 431–436.

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165.

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology. General*, *148*(2), 325–341.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, *7*(5), e36399.

Montero-Melis, G., & Jaeger, T. F. (2019). Changing expectations mediate adaptation in L2 production. *Bilingualism: Language and Cognition*, 1–16.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1–15.

Peter, M. S., & Rowland, C. F. (2019). Aligning Developmental and Processing Accounts of Implicit and Statistical Learning. *Topics in Cognitive Science*, *11*(3), 555–572.

Pinker, S. (2009). *Language Learnability and Language Development*. Harvard University Press.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rebuschat, P. (2013). Measuring Implicit and Explicit Knowledge in Second Language Research: Measuring Implicit and Explicit Knowledge. *Language Learning*, *63*(3), 595–626.

Robenalt, C., & Goldberg, A. E. (2016). Nonnative Speakers Do Not Take Competing Alternative Expressions Into Account the Way Native Speakers Do: L2 Learners and Competition Dynamics. *Language Learning*, *66*(1), 60–93.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension--an FMRI study. *Cerebral Cortex*, *22*(7), 1662–1670.

Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition*, *163*, 1–14.

Voeten, C. C. (2020). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. https://cran.r-project.org/package=buildmer

Walker, N., Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the acquisition of vocabulary and grammar: An individual differences approach. *Language Learning*, *70*(52).