Learning Through Prediction: A Case of Verb Bias Learning

Yi-Lun Weng[1], Amanda Owen Van Horne[2], and Zhenghan Qi[3,4]

[1] *Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, USA*

[2] *Department of Communication Sciences and Disorders, University of Delaware, Newark, DE, USA*

[3] *Department of Communication Sciences and Disorders, Northeastern University, Boston, MA, USA*

[4] *Department of Psychology, Northeastern University, Boston, MA, USA*

Author Note

Yi-Lun Weng https://orcid.org/0009-0002-6292-5028
Amanda Owen Van Horne https://orcid.org/0000-0002-3517-5885
Zhenghan Qi https://orcid.org/0000-0002-4812-8842

Correspondence concerning this article should be addressed to Zhenghan Qi, MD/PhD, Department of Communication Sciences and Disorders, Department of Psychology, Northeastern University, 360 Huntington Ave., Boston, MA 02115, United States. Email: z.qi@northeastern.edu.

Learning Through Prediction: A Case of Verb Bias Learning

Abstract

Linguistic prediction, which emerges from experience, is a pervasive process in language comprehension. However, how prediction develops as learning unfolds and how it drives the learning process remains unclear. This study examines three key questions: (1) whether learning is associated with growth in prediction, (2) whether stronger prediction errors are associated with greater learning, and (3) whether linguistic prediction skills are stable across tasks. Our results revealed that learners who successfully updated their verb biases showed greater growth in anticipatory looking patterns upon hearing the verb in ambiguous test sentences. Greater mismatch between learners' initial verb bias and the subsequent training condition predicted greater changes in learners' verb bias. Finally, individuals who tend to predict more in a separate comprehension task showed greater changes in anticipatory looking behavior and ambiguity resolution in the verb bias learning task. Taken together, these results provide empirical support for a prediction-based language learning framework.

Keywords: prediction, individual differences, error-based learning, eye movements, sentence processing

Word count: 9,931 words

Introduction

A remarkable ability of the human brain is its capacity to generate predictions based on incoming sensory input and test those predictions against prior expectations or beliefs (Clark, 2013; Friston & Kiebel, 2009). This predictive processing is not only observed in a specific domain but is found across various cognitive domains as well. It's what enables us to react quickly when a car swerves unexpectedly, foresee the trajectory of a bouncing ball, and finish someone's sentence before they do. In the realm of human communication, predictive processing also plays an important role in enhancing communication efficiency. As one person speaks, the other may start anticipating upcoming linguistic contents as the conversation progresses, thus allowing listener to minimize cognitive loads and reduce processing time, making interactions smoother and more engaging. Abundant evidence suggests that the brain constantly employs linguistic cues to anticipate upcoming words (Altmann & Kamide, 1999; Delong et al., 2005; Federmeier & Kutas,1999; Tanenhaus et al., 2000). For example, researchers have found that people exhibit predictive behaviors across various language tasks, including faster reactions for predictable words as compared to unpredictable ones across different paradigms (e.g., Lexical decision task: Arnon & Snider, 2010; Naming task: McClelland & O'Regan, 1981) and less time fixating on highly predictable words during sentence reading (Ehrlich & Rayner, 1981). Moreover, eye-tracking studies using the visual world paradigm have also demonstrated our brain's ability to generate predictions while comprehending sentences (Altmann & Kamide, 1999; Nation et al., 2003; Pickering & Gambi, 2018). As a sentence unfolds, people use linguistic cues, such as the semantics of verbs, to predict forthcoming content. This prediction process is manifested through anticipatory eye movements towards expected objects or scenes related to the predicted content. For example, Altmann and Kamide (1999) found that people looked at edible objects more than at inedible objects when hearing "*the man ate the*…" but did not do so when the verb "ate" was replaced with the verb "moved". In German main clauses, feminine sentence-initial noun phrase can serve either the agent or the patient role of the verb. Knoeferle et al. (2005) found that, in the presence of depicted events (e.g., a princess, who is being painted by a fencer, is washing a pirate), German speakers rapidly predicted upcoming noun phrase upon hearing the disambiguating verb. For example, they immediately looked at the pirate when hearing "*The princess washes …"*. These anticipatory eye movements are closely time-locked to the spoken utterances that enable the identification of items depicted by a concurrent visual scene (Tanenhaus et al., 1995) and have been regarded as a strong index of prediction in the language comprehension literature.

Given the considerable evidence indicating that prediction plays a crucial role in efficient language processing, an intriguing question arises: where does linguistic prediction stem from? In fact, to be able to make predictions during language processing, one must have learned and accumulated linguistic knowledge. Through learning, our brain builds internal language models of the world and stores information about linguistic patterns, semantic relationships, and syntactic structures. These internal representations serve as a foundation for generating predictions based on the likelihood of specific words or structures occurring in each context. In other words, prediction is a natural consequence of learning. One piece of evidence supporting this idea comes from statistical learning studies. Participants are passively exposed to sequences with embedded temporal regularities, where specific earlier stimuli serve as the antecedent of their corresponding later stimuli. Such repeated exposure often leads to faster response time in detecting the upcoming later stimuli in the input, implying that learners are increasingly certain about what the subsequent element is (Hunt & Aslin, 2001) and this is taken as evidence that they have learned the statistics pattern. For example, Misyak et al. (2010) showed that adults were able to use patterns acquired from an artificial grammar learning task to predict the end of an auditory sequence more quickly than the beginning of an auditory sequence. In sentence comprehension literature, listeners can adapt their syntactic prediction based on the distributional information in their experiences (e.g., Havron et al., 2019). For example, after a brief exposure that trained children to associate the ambiguous *la petite* with either a verb and a noun, children adapted syntactic prediction by making anticipatory gazes to the target before hearing the disambiguating word. To summarize, the knowledge of probabilistic structure forms the basis for generating expectations about upcoming units and the observed predictive responses are a natural outcome of learning which allows humans to process information efficiently.

In parallel to the account that prediction stems from learning, recent studies have suggested that prediction also plays a crucial role in motivating learning. For example, prediction-based learning accounts (Chang et al., 2000; Chang et al., 2006; Ramscar et al., 2013) propose that prediction motivates learning in two major ways: if learners experience a match between the existing linguistic knowledge and the actual input, the internal linguistic representation will be reinforced and learners will be able to process linguistic information more efficiently as the reducing cognitive demands helps with the speed and accuracy of processing (Christiansen & Chater, 2015). On the other hand, if learners experience a mismatch between the existing linguistic knowledge and the actual input during processing, the error signals will generate and invoke surprises, which are costly processes that draw attention and show advantage in memory (Corbetta & Shulman, 2002; Kishiyama et al., 2009), will then drive an updating of internal linguistic representations towards minimizing such surprisal in the future (Bock & Griffin, 2000; Chang et al., 2000; Dell & Chang, 2014; Peter & Rowland, 2019). Notably, these two scenarios represent how linguistic representations can be updated by either lower or higher prediction errors, computed through a comparison between the actual inputs and our prior knowledge about the probability with which certain linguistic representations occur in the context.

Most existing evidence supporting prediction-based learning theories comes from structural priming studies. The structural priming refers to the idea that when people encounter a particular syntactic construction, they are more likely to expect it again, or to reuse it in production, than they were before (Mahowald et al., 2016; Pickering & Branigan, 1998). Greater priming effect in less frequent syntactic structure, as known as inverse preference priming effect, has been regarded as supporting evidence for implicit error-based learning accounts (Bock et al., 2007; Bock & Griffin, 2000; Dell & Chang, 2014). Such inverse preference effect was also present in verb-specific syntactic adaptation. For example, the verb "show" can occur in both the double object dative structure (e.g., *Show the boy the book*) and the prepositional object dative (e.g., *Show the book to the boy*), but occurs more often in the double object dative structure. When "show" is followed by the less-common prepositional object dative, it will lead to prediction errors that then alter the connection weights of the experienced structure and update the individual's syntactic preferences (Arai & Mazuka, 2014; Bernolet & Hartsuiker, 2010; Peter et al., 2015; Fazekas et al., 2020). This error-based learning account is further supported by learning studies in children and adults. Verb-specific syntactic learning is stronger when the syntactic structures in the training conditions are less frequent and more surprising (Lin & Fisher, 2017). Prediction errors during learning, due to temporary syntactic or referential ambiguity, also led to greater subsequent changes in syntactic comprehension (e.g., Gambi & Messenger, 2023) and novel word semantic retention (Reuter et al., 2019; Gambi et al., 2021).

While evidence from these studies can address potential changes in behavior depending on the predictability of the input, they do not take into consideration the variations in an individual's prior linguistic knowledge. According to prediction-based learning accounts, the magnitude of prediction error is determined by the consistency between individuals' prior knowledge and the perceived input – stronger learning effects should be observed when the magnitude of prediction errors is larger (Chang et al., 2000; Chang et al., 2006). Notably, the quality and quantity of prior knowledge that individuals use for prediction is greatly influenced by individuals' learning experience. For example, some verbs such as "feel" allow two possible interpretations in the sentences such as "*Feel the frog with the feather*.", the prepositional phrase "with the feather" can be attached to the verb and given an instrument interpretation (i.e., using the feather to feel the frog), or attached to the noun and given a modifier interpretation (i.e., the frog that has the feather). Individuals' interpretation of the sentence is primarily determined by their prior language experience – some people may interpret the sentence with an instrument-biased approach, while others may interpret the sentence in a modifier-biased way (Snedeker & Trueswell, 2004). Lab-based learning studies demonstrated the malleability of such verb biases through brief verb-specific experiences (Qi et al., 2010; Ryskin et al, 2017). For example, Ryskin et al (2017) employed a computerized verb bias training paradigm where adult listeners follow auditory instructions containing unambiguous training and ambiguous testing sentences containing equi-biased verbs (e.g., *feel*). The visual display and the discourse information in training trials provided disambiguating cues, so that half of the verbs were trained to be paired with instrument *with* phrases and the other half of the verbs were trained to be paired with modifier *with* phrases (Figure 1). These brief experiences with specific verbs updated listeners' verb biases and led to more instrument- or modifier-biased processing outcomes.

However, the analytic methods that are commonly used to demonstrate learning effects across participants limit our ability to estimate the contribution of people's prior knowledge, which varies across individuals and verbs. Focusing on the performance at individual level may enable a better understanding of how prior linguistic knowledge interacts with the error signals generated by a mismatch in expectation and information received. By considering individual differences in linguistic knowledge and learning experiences, we may gain a better understanding of how prediction-based learning leads to linguistic knowledge acquisition. Therefore, to measure individual's prior knowledge, the present study adapted the verb bias learning paradigm from Ryskin et al. (2017) Experiment 2 to investigate whether adults generate new predictions after updating their verb biases and whether such predictions, especially the prediction errors (i.e., the mismatch between the prior verb bias knowledge and the training type), may potentially guide the learning process.

In addition, a related question is whether an individual's linguistic prediction ability is consistent across language tasks. Previous studies have found a strong link between verb semantics and their bias towards certain argument structures, in addition to their frequency-based syntactic relationships (Argaman & Pearlmutter, 2002; Hare et al., 2004). For example, Argaman and Pearlmutter (2002) discovered that the probabilities of specific verb structures were closely connected to the likelihood of certain nouns being used as arguments, indicating that the selection of a verb's argument structure is influenced by its semantics. Therefore, if individuals rely on verb information to guide sentence processing, one might expect that people who are good at using extant verb semantic knowledge to predict upcoming events in sentence comprehension (e.g., *the man ate/moved the…*) would also be good at learning new relationships between verbs and syntactic structures because of repeated exposure to specific interpretations. Such a finding would support the hypothesis that the same underlying ability to deploy verb prediction is evident in both language processing and language learning.

The current study attempts to understand the dual role of prediction in learning from an individual difference perspective and to examine the stability of people's linguistic prediction skills. First, we ask whether learning is associated with growth in prediction. As the first step, we attempted to replicate the findings of verb bias learning effect in adults from Ryskin et al. (2017) study. Then we modified this learning task to enable observation of the gradual emergence of structural prediction over the course of learning, specifically hypothesizing that as learning progresses over three short training blocks, prediction will be evident in anticipatory looking patterns especially when learning is successful. Second, to address whether stronger prediction errors are associated with learning outcomes, we further modified the verb learning task by adding Pre-test trials. This allows us to quantify how divergent learners' initial anticipation about the structure was from the upcoming verb bias training as an index of prediction errors and then to evaluate the relationship between prediction errors and learning outcomes. We predict that greater magnitudes of prediction error should be linked to better learning outcomes. Finally, we examine whether individuals' linguistic prediction skills are stable across language comprehension and language learning tasks. Specifically, we carried out a conceptual replication of Nation et al. (2003). In this study, young listeners made faster anticipatory eye movements towards the direct object (e.g., *cake*), upon hearing an informative (e.g., *eat*) than a neutral verb (e.g., *choose*), which suggested that individuals make predictions during sentence comprehension based on semantic information in the verb. We extend this work by examining the correlation between the use of semantic information to predict during sentence comprehension and the structural prediction during verb bias learning.

Method

Transparency and Openness

We report how we determine our sample size and data inclusions and exclusions in the study below. This study's design and its analysis were preregistered on the Open Science Framework platform (https://osf.io/uhvd6). Data and codes used for analyses are available at Open Science Framework platform (https://osf.io/e3dh9/) in accordance with open science publication guidelines. In the results section we reported both pre-registered analyses that include replications of Ryskin et al. (2017) and Nation et al. (2003) and exploratory analyses that extend these results. All data were collected between April and July 2021 and all study procedures were carried out remotely. All procedures were reviewed and carried out in accordance with guidelines from the University of Delaware Institutional Review Board.

Participants

Informed consent was obtained for forty-one adults (34 female, mean age = 20.3 years, range = 18 – 35) from the University of Delaware community who participated in the experiment for course credit or payment. All were monolingual English-speakers and had normal or corrected-to-normal vision and hearing and no prior history of neurological issues.

During pre-registration, we conducted power analyses based on effect sizes from Ryskin et al. (2017) and estimated the need for 36 participants. The sample size was based on an *a priori* analysis (SIMR package in R; Green & MacLeod, 2016). The results of the power analysis suggested that with a sample size of N = 33, we would have a power of .90 to detect a standardized effect size of 2.17, assuming alpha = .05. Therefore, we registered recruitment of thirty-six participants to allow for attrition or examiner error. In addition, five participants were excluded from all analyses because of the low accuracy rate (below 75%) in training trials during the verb bias training session, demonstrating general inattention to the task. Thus, five additional participants were recruited, bringing our total consented participant numbers above the pre-registered number.

Apparatus and Stimuli

The task was programmed and hosted on Gorilla Experiment Builder (https://gorilla.sc/; Anwyl-Irvine et al., 2020) and the eye-tracking function was based on the WebGazer.js libraries (Papoutsaki, et al., 2016), a free and open-source library and is written entirely in JavaScript and can be integrated into webpages easily. The WebGazer.js uses a typical webcam that is present in laptops to infer the eye-gaze locations on a webpage in real time. It contains face and eye detection algorithms and incorporated and evaluated the data by using different facial feature detection libraries in JavaScript. The primary components of WebGazer.js are the tracker module and the regression module. The tracker controls how eyes in a facial image are detected, while the regression module governs how a regression model is learned and how gaze locations are predicted based on the eye patches extracted from the eye tracker. WebGazer.js has been shown to effectively detect eye fixations and successfully replicate findings of in-lab cognition and language studies with comparable accuracy (Semmelmann & Weigelt, 2018; Calabrich et al., 2021; Degen et al., 2021).

The experimental stimuli included 184 auditory sentences and corresponding visual stimuli that were developed to meet the criteria of six types of items divided amongst two tasks (verb bias learning task and verb semantics task). Specifically, we had three conditions (36 unambiguous instrument-trained trials, 36 unambiguous modifier-trained trials, and 40 ambiguous test trials) associated with the verb bias learning task modeled after Ryskin et al. (2017), two conditions (18 semantically predictable trials and 18 semantically unpredictable trials) associated with the verb semantics task modeled after Nation et al. (2003) and 36 filler trials.

Participants were presented with these experimental items as a single task and items were pseudo-randomly interleaved such that the two tasks and the filler items were not distinguishable. To that end, all the visual displays contained four items: a target animal, a target instrument/object, a distractor animal, and a distractor instrument/object in each of the four quadrants of the scene. Pictures (mostly photographs and some drawings) were selected to provide the clearest possible depiction of each object. Every animal was paired with a mini object. The mini object was either identical to the target instrument, identical to the distractor instrument, or distinct from both target and distractor instruments to either promote or minimize ambiguity.

Participants looked at a screen and listened to a pre-recorded speech at the same time. On each trial, participants first heard a sentence that described all the pictures on the screen (e.g., "*Here's a duck, a hat, another hat, a sponge, a bird, a sponge*."). After hearing the sentence, participants were asked to follow an instruction by clicking on the relevant objects and animals. After they clicked on the picture, they would see the feedback on the screen and hear a sound effect associated with the action. The goal of the feedback was to maintain participants' attention and motivation during the task. Following this shared structure, items systematically differed. We describe the stimuli associated with these two tasks in turn.

Verb Bias Learning Task

We adapted the verb bias learning task from Ryskin et al. (2017) Experiment 2 to allow observations of participants' gradual learning processes. In this task, participants were trained on eight equi-biased verbs (clean, cuddle, feel, hug, knock on, pinch, rub, squeeze); four verbs were instrument-trained, and four verbs were modifier-trained, with assignment of verbs counterbalanced across participants. These verbs were selected because, at the population level, they elicited comparable consideration of both the modifier interpretation and the instrument interpretation based on fixations to and mouse clicks on the target instrument in Ryskin et al. (2017) Experiment 1 results. Each verb was either trained in unambiguous instrument-biased condition or modifier-biased condition across 9 training trials with 5 ambiguous testing trials interspersed evenly: 2 pre-test trials before the first training trial, and then 1 post-test trial after every 3 training trials (Figure 1A).

During training, for an instrument-trained verb *rub*, immediately after the picture description (Figure 1B), the listener would hear a context question that should bias them toward an instrument interpretation, such as "*Hmm...what should you use to rub the bunny?*" followed by a prompt to respond "*I know! You should rub the bunny with the bottle*." Because both the auditory information and the visual display provided an unambiguous focus on the instrument-biased interpretation of the sentence "*Rub the bunny with the bottle*" should strengthen associations between the verb used in the sentence and the instrument-interpretation. Note that the instrument nouns were randomly chosen for the verb and may or may not be a plausible action in the real world. However, the action was always possible in the computerized task by dragging the instrument towards the animal. We made this design choice to minimize the impact of noun semantics on syntactic interpretation. Similarly, for a modifier-trained verb *rub*, the listener would hear a context question that should bias them toward a modifier-interpretation, like "*Hmm...which animal should you hug now?*" followed by a prompt to respond "*I know! You should hug* *the bunny with the sponge*". Because both the auditory information and the visual display provided an unambiguous focus on the modifier-biased interpretation of the sentence "*Hug the bunny with the sponge*" this should strengthen associations between the verb used in the sentence and the modifier-interpretation.

For both the pre-test and post-training testing trials, the visual display contained a target animal, a target instrument, a distractor animal, and a distractor instrument (Figure 1C). One mini object was identical to the target instrument and the other was identical to the distractor instrument to promote ambiguity. The test sentences, after the picture description, included a neutral context question "*Hmm…what should you do now?*" followed by a prompt to respond "*I know! You should hug the bird with the sponge*". Because both the visual display was ambiguous and the prompt was neutral, this allowed us to detect what the participants' preferred interpretation of the verb was in the moment.

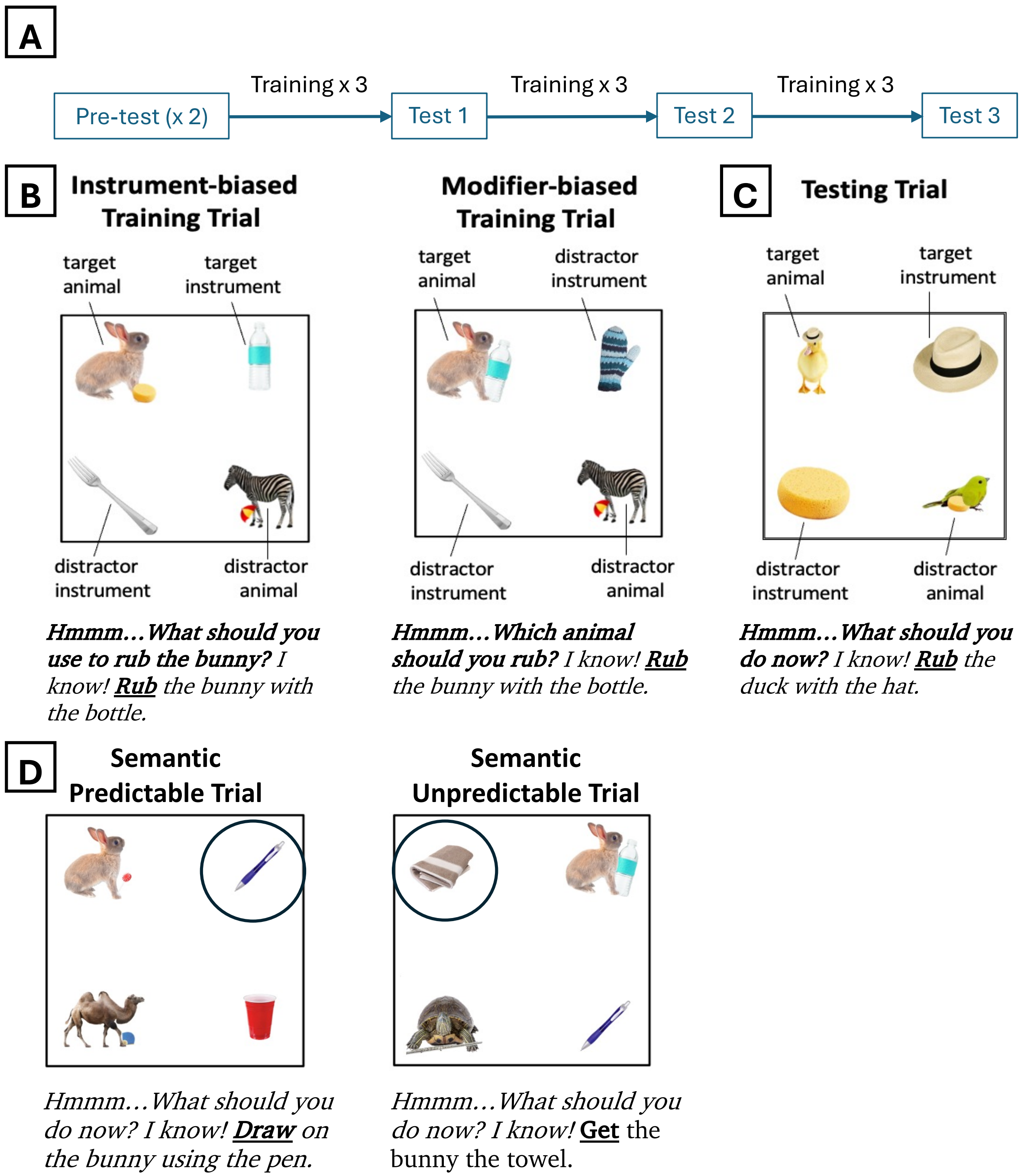


Figure 1 Procedure and stimuli for the verb bias learning and the semantic prediction tasks. (A) The training and test schedule for each verb. (B) Unambiguous verb bias training trials. (C) Ambiguous verb bias test trials. (D) Unambiguous trials for predictable and unpredictable conditions (target objects are circled).

Verb Semantic Prediction Task

This task was inspired by, but not a direct replication of, Nation et al. (2003) study as we wanted to ensure that the displays were like the verb bias learning task just described. For semantically unpredictable and predictable items, the critical sentence contained either a neutral verb (e.g., *get* or *make*) or a supportive verb (e.g., *draw* or *bake*) that served to restrict the following noun. These items were unambiguous and thus used a different sentence frame from the verb-bias training sentences. The verbs selected did not overlap with the eight equi-biased verbs in the verb bias learning task.

The visual display (Figure 1D, for the sentence "Draw on the bunny using the pen") included a semantically relevant/target object (e.g., pen), a semantically irrelevant/distractor object (e.g., towel), a target animal (e.g., bunny) and a distractor animal (e.g., turtle). Each animal was paired with a mini object that was either identical to the target instrument or the distractor instrument to promote predictability. For the eighteen semantically predictable trials, the four pictures were first introduced in the same way as in the verb-bias learning task, and then the participant heard a semantically predictable instruction, such as "*Draw on the bunny using the pen*". A participant sensitive to the constraints of the verb "draw" should engage in anticipatory looking to the relevant object on the screen that can be used for drawing, the pen, and thus should be quick to respond to the correct item. For the eighteen semantically unpredictable trials, the four pictures on the display were semantically unrelated to the verb and the verbs intentionally carried limited semantic information. Following picture introduction, the listener heard the semantically unpredictable instruction, such as "*Get the camel the cup*." Because a verb like "get" contains limited cues to upcoming sentence information, the participants should not be able to anticipate which picture to look at until the sentence is concluded.

Filler Trials

Filler trials served to control the spacing of other trials relative to each other and ensure that the task had sufficient variability in sentence frames. On 36 filler trials, neither a modifier, nor an instrument interpretation was plausible and there was no specific semantic association amongst the picture displays. The visual display included two animals and two instruments, as well as mini-instruments associated with each animal. Verbs were selected to be neutral regarding semantic information as compared to the display. The sentence structures were also distinct from the other item types (e.g., "*Wiggle the panda*.").

No with-prepositional phrase was contained in the sentences. The verb selected was not one of the trained verbs nor was it one of the verbs used in the semantically predictable or unpredictable trials.

Feedback

For feedback for the unambiguous training trials, regardless of their responses, participants saw the correct action that would be associated with the trained interpretation. That is, for instrument-biased training, participants saw a picture depicting an instrument interpretation and heard a relevant sound effect. For modifier-biased training, participants saw a picture depicting the event as a modifier interpretation and heard a relevant sound effect. For the ambiguous test trials, the participant would see a response that matched their interpretation of the task and hear the related sound effect. For example, if they clicked on an instrument, they would see an instrument depiction of the event; if they clicked on the modifier interpretation, they would see a modifier depiction of the event. If participants picked one of the distractor choices, they would be shown - at random - either an instrument or a modifier depiction of the sentence on the grounds that this may reorient them to the task at hand. For feedback for the filler trials, regardless of response, participants saw a picture depicting the event that matched the instruction of the trial.

Procedure

Before the task began, participants were asked to close all browser tabs that could produce pop-ups or alerts that would interfere with the study on their computers. Then experimenters provided detailed instructions about how to position themselves following the recommendations for best practices on online eye-tracking studies (Semmelmann & Weigelt, 2018). For example, participants were instructed to sit directly in front of the monitor to ensure the visibility of their face, to avoid moving their head around, to keep lights in front of them rather than behind them so that the webcam could clearly detect their faces. After participants positioned themselves appropriately using the live feed from the webcam on their computer, their screen entered full-screen mode automatically. The entire experiment session was monitored by the experimenters using screen sharing and recording function with Zoom.

The experiment started with a calibration/validation session: participants saw a sequence of five green dots appear on their screen, each lasting for three seconds. Participants were told to fixate directly at each green dot until it disappeared. The validation procedure was essentially identical to the calibration procedure, except that the dots were red. If participants failed to calibrate at least one of the points (i.e., the estimated location of the calibration point is different from the validation point), they needed to repeat the calibration/validation procedure. Once participants passed the initial calibration/validation procedure, the task would start. The same calibration/validation procedure was scheduled for every 36 experimental trials and there was a total of four calibration/validations throughout the whole experiment.

The trial order was pseudo-randomized such that for a given verb, two Pre-test trials, serving as baseline, were arranged before the first training trial, and each of the remaining three test trials was preceded by three training trials. Trials with the same verb were never adjacent to each other. Also, no more than three trials of the same type (Modifier-trained trial, Instrument-trained trial, Test trial, Semantically predictable trial, Semantically unpredictable trial, or Filler trial) were arranged in a row. The lag between a training trial for a given verb and the next test trial for that verb was, on average, 3.45 trials (Median = 3, SD = 1.9). Two lists were constructed to counterbalance the verb-training pairings (e.g., for half of the subjects, the verb "pet" was instrument-trained and for the other half it was modifier-trained). Participants were randomly assigned to one of the two lists.

Validation of webcam eye-tracking measures:

Previous research has documented multiple methods to validate webcam-based eye-tracking methods (see Bott et al., 2017; Semmelmann & Weigelt, 2018). Rigorous replications of classic psycholinguistic effects, including verb bias effects during online sentence processing (Ryskin et al., 2017, Experiment 1) have been reported using online data collection with automated fixation coding by webgazer (James et al., 2025). To validate our own webcam eye-tracking measures, we compared time courses of fixation patterns during sentence comprehension between the sample in the current study and a separate sample in a lab-based study with a Tobii X-120 eye-tracker. We also followed Bott et al. (2017) and employed two approaches to validate the use of webgazer-coded fixation data in the verb semantic

prediction task. The test trials in the verb bias learning task were not used for this analysis due to the ambiguity feature of the equi-biased verbs and the learning nature of the task. In our supplementary materials, we reported the following evidence supporting the validity of our eye-tracking measures.
1) Similar fixation time courses for sentences containing equi-biased verbs measured with Tobii eye-tracker and webcam (Supplementary Figure 1).

2) An acceptable split-half reliability of individual subjects' fixation measures in the verb semantic prediction task (Supplementary Method Page 2)

3) Similar predictability effect revealed by automated and manual coding of a subset of eye-movement data in the verb semantic prediction task (Supplementary Method Page 2).

Results

Preregistered Analyses 1: Whether Learning is Associated with Growth in Prediction

Prior to the main individual difference analyses, we first attempted to replicate the action and eye movement findings from Ryskin et al. (2017) which showed that adults' representations of verb bias are malleable over a short time scale of learning. Throughout the analyses for this task, the independent variables were Training Types, which were entered as deviation coded contrasts: Instrument-trained (−0.50) vs. Modifier-trained (0.50), and Testing Time Points (Pre-test, Testing Time Point 1, Testing Time Point 2, and Testing Time Point 3). In our pre-registration, we proposed a dummy-coding scheme so that every testing time point was compared to the Pre-test. However, as all three test time points were spaced apart with a consistent number of training trials, it is more appropriate to treat time points as a numeric continuous variable. Therefore, in all the following analyses testing verb bias learning effect, we treated the testing time points as a mean-centered numeric variable (-1.5, -0.5, 0.5, 1.5). Both random intercepts for verbs and participants as well as random by-verb and by-subject slopes for Training Types and Testing Time Points are included in the initial model. The optimal random effect structure was determined using the rePCA function from the RePsychLing Package in R (Bates et al., 2015; Matuschek et al., 2017). We only built the statistical model to test clicks on the target instrument because of the mirroring pattern for participants' clicks on the target animal at the end of the sentences.

Moreover, adults' eye movements were also analyzed in the same three time windows as in Ryskin et al. (2017). The verb time window started at the onset of the verb (e.g., rub) and ended at the onset of the first noun (e.g., duck), with an average duration of 572 milliseconds. The N1 time window began at the onset of the first noun and ended at the onset of the second noun (e.g., hat), with an average duration of 1040 milliseconds. The N2 time window started at the onset of the second noun and ended 1000 milliseconds later (the maximum duration of the instrument). All time windows were offset by 200 ms to account for the time needed to program and launch an eye movement (Hallett, 1986). For each time window, we analyzed the training effect on the logit transformed proportion of fixations to either the target animal or the target instrument using mixed-effects linear regression.

Lastly, to further expand our understanding and determine if learning is associated with growth in prediction, we analyzed adults' anticipatory looks, which indicate the listener's attempt to parse the sentence before it is fully presented and can signal the deployment of newly acquired information for predictive purposes. We identified individuals' first fixation (lasting > 200 ms) upon hearing the verb in each trial (starting 200 ms after the onset of the verb allowing time to launch an eye-movement). The relationship between learning and anticipatory looking behavior was examined in two ways. First, we used a mixed-effects logistic regression model to examine the effects of training type on participants' first fixation patterns. The dependent variable was whether participants' first fixation was directed towards the instruments on the display. We considered learning-induced predictions to have occurred if we observed a significant interaction between Training Types and Testing Time Points. As the data pattern for listeners' first fixation on either animal was a mirror image of that for either instrument, we refrained from duplicating the analyses. Second, we examined whether individual participants who showed greater changes in their verb-biases over the course of learning (Test 3 – Test 0), as measured by their offline mouse clicks on the target and online proportion of fixations to the target during the N2 window, also showed greater changes in their anticipatory looking behavior (first fixation). Our primary analyses were accomplished by computing each participant's overall change scores collapsing across two verb-bias training types to avoid inflation in the degrees of freedom.

Results For Preregistered Analyses 1

The results of participants' mouse clicking behavior in the test trials suggest a similar pattern as Ryskin et al. (2017). Over the course of learning, participants' choices of the target instrument increased for the Instrument-trained condition and decreased for the Modifier-trained condition. Similarly, participants' choices of the target animal increased for the Modifier-trained verbs and decreased for the Instrument-trained verbs (Figure 2A). The mixed-effects logistic regression analysis on target instrument clicks suggested a marginal main effect of Training Types ($\beta = -0.45$, $SE = 0.23$, $z = -1.93$, $p = .054$, $OR = 0.64$ [95% CI 0.40 to 1.01]), that is, participants were more likely to choose the target instrument in the Instrument-trained than in the Modifier-trained condition. There was also a marginal main effect of Testing Time Points ($\beta = -0.19$, $SE = 0.11$, $z = -1.65$, $p = .098$, $OR = 0.83$ [95% CI 0.66 to 1.04]). Importantly, the interaction between the Training Type and the Testing Time Point was significant ($\beta = -0.38$, $SE = 0.13$, $z = -2.95$, $p = .003$, $OR = 0.68$ [95% CI 0.53 to 0.88]), suggesting the condition difference grew over the course of training, providing a conceptual replication of the training effect reported in Ryskin et al. (2017).

The eye movement behaviors, however, show a large individual variability. Validation of the webcam-based eye-tracking measures and coding is confirmed and reported in the Supplementary Methods. To investigate how real-time sentence processing changes over the course of training, we examined the proportion of target animal and target instrument fixations during verb, noun-1, and noun-2 time windows. The results of mixed-effect regression analyses showed no significant main effect of Training Types or interactions between Training Types and Testing Time Points in any time window (Supplementary Table 1). To investigate whether participants generate new predictions for verb-specific knowledge over the course of learning, we identified participants' first fixation in each trial as a measure for prediction and analyzed the training effect. The mean onset of the first fixation was 877 ms ($\pm$ 104.4 ms), which coincides with the noun 1 onset time 881 ms ($\pm$ 25.5 ms). The results of mixed-effects logistic regression model revealed no main effect of Training Types or interactions between Training Types and Testing Time Points (Supplementary Table 2).

In subsequent analyses, we address our first key research question, that is, whether variabilities in learning outcomes, measured by both offline behavior and online eye movement, are associated with individuals' growth in anticipatory looking behavior. We utilized Pearson correlations to assess the relationship between the change in first fixation and the change in corresponding clicking and proportion of fixation at the end of the sentence. We hypothesized that learners whose final interpretation of the test sentences showed greater changes during verb-bias training would exhibit more pronounced changes in anticipatory looks towards the trained target. Conversely, individuals who were less successful in learning the new verb biases would generate fewer changes in anticipatory looks towards the trained target. Targets were defined as target animal picture for modifier-trained trials and target instrument picture for instrument-trained trials. As shown in Figure 2B, greater changes in anticipatory looks over the course of training were marginally related to greater changes in offline ambiguity resolution ($R = 0.32$, $p = 0.06$) and significantly related to greater changes in online eye-movement ($R = 0.62$, $p < 0.001$). Similar findings were found using a more stringent definition of anticipatory looking, the proportion of animal or instrument fixation within the verb window (before the Noun 1 onset). Greater increase in anticipatory looking in the verb window was associated with greater increase in target fixation in the Noun 2 window ($R = 0.72$, $p < 0.001$). These results suggest that despite a lack of overall training effect across participants in our sample, the variabilities in listeners' anticipatory looking behavior systematically track the degrees of verb-bias learning: people who showed greater changes in their final interpretation based on the updated verb biases also showed greater changes in their anticipatory looks towards potential upcoming referents upon hearing the verbs.

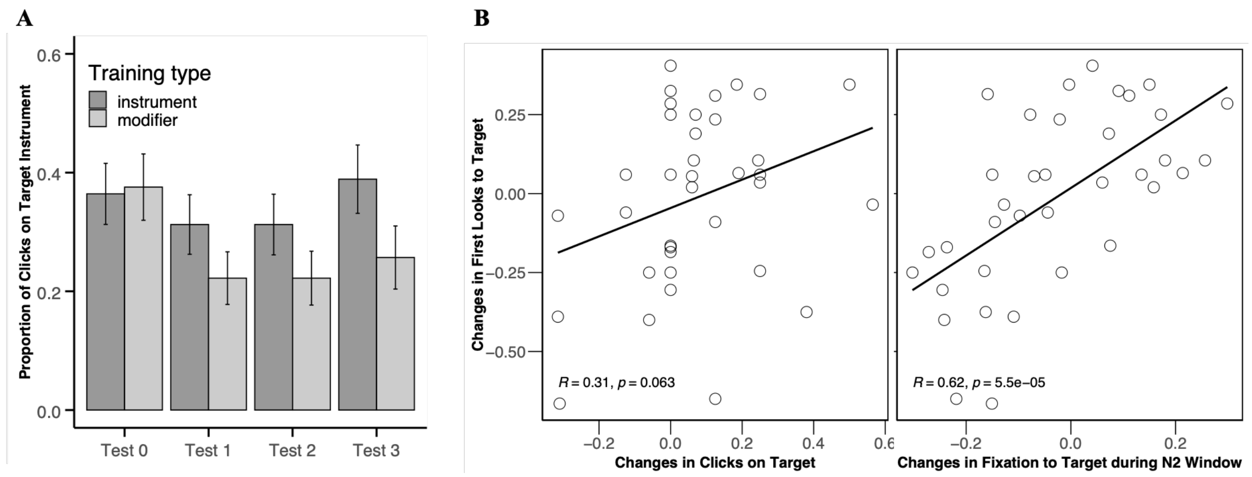


Figure 2. Verb bias learning and anticipatory looking behavior: (A) Proportion of offline mouse clicks on target instrument over the course of learning. The x-axis represents the four testing timepoints and the y axis represents the proportion of clicks on the target instrument in both Instrument-trained and Modifier-trained conditions. Error bars represent within-subject standard error of the mean. (B) Changes in anticipatory looking patterns (y-axis) are associated with changes in verb-bias guided disambiguation (x-axis, mouse clicks on target (left) and proportion of target fixation in the N2 time window (right)). All change scores were computed by the differences between Test Time 3 and Test Time 0.

Exploratory Analyses: Whether Stronger Prediction Errors are Associated with Greater Learning

To fully comprehend the role of prediction in learning, we carried out an exploratory analysis to test the error-based learning account. This account consists of learners continuously predicting upcoming input based on their prior knowledge, receiving feedback from the actual input, and adjusting these predictions as the sentence unfolds. The most significant change in learning should occur when the input deviates the most from their prior knowledge, as the error signal is strongest. As the input aligns more closely with learners' expectations, the learning effect diminishes, leading to more efficient processing. However, as this is an exploratory analysis, selecting a specific metric (e.g., clicking or eye gaze behaviors) to operationalize prediction errors and connect them to learning outcomes was uncertain. Hence, we adopted a multiverse approach to this analysis (Steegen et al., 2016), which is particularly beneficial when various metrics could be used without a clear hypothesis-driven rationale. By transparently presenting multiple ways of selecting data metrics, we aim to avoid type I and type II errors and enhance reproducibility. A robust result would exhibit consistent outcomes regardless of the method chosen for reporting the data.

Thus, we conducted an examination of metrics derived from the Pre-test trials, aiming to identify indices of prior verb bias knowledge before training. First, for each verb for each participant, we extracted the eye-movement metrics that represent the real-time interpretation *opposite from* subsequent training. (e.g., *target-animal* fixation for a verb to be trained in an *instrument-biased* condition, or *target-instrument* fixation for a verb to be trained in a *modifier-biased* condition). Then we tested whether the degree of such divergence between pre-test processing and subsequent training modulates learning by assessing the interaction between pre-test fixation and testing time points in a mixed-effects linear regression model for participant's target fixation (e.g. target-instrument fixation for instrument-trained verbs). We did not perform these analyses for first fixation or click behavior, because it is difficult to estimate verb-specific biases from only two pre-test trials for these binary metrics.

Second, we extracted participant-level metrics across trials to characterize each individual's levels of divergence between pre-test processing and subsequent training. These metrics include (1) mouse clicking behaviors, (2) proportion of fixations in the N2 time window, and (3) first fixation. We then correlated these Pre-test divergent metrics with various metrics representing the magnitude of the learning outcome (i.e., difference scores between Time 3 and Time 0). Our focus was to ascertain whether prediction influences learning, and we considered evidence supporting this idea if individuals exhibited more substantial changes in verb-bias knowledge after training when their pre-test verb bias knowledge diverged more from the upcoming training condition.

Results For Exploratory Analyses

Across all three eye-movement analysis windows, we found greater pre-test divergence from subsequent training significantly predicted more growth in fixations to the target over the course of learning (Pre-test Divergent Fixation * Testing Time Points: Verb Region: $\beta = 0.08$, $SE = 0.04$, $t = 2.20$, $p = 0.03$, OR = 1.08 [95% CI 1.01 to 1.16]; Noun 1 Region: $\beta = 0.14$, $SE = 0.03$, $t = 4.29$, $p < 0.001$, OR = 1.15 [95% CI 1.08 to 1.23]; Noun 2 Region: $\beta = 0.07$, $SE = 0.03$, $t = 2.01$, $p = 0.04$, OR = 1.07 [95% CI 1.00 to 1.15]). Figure 3A illustrates these interactions: verbs and participants that started with higher levels of divergence from subsequent training showed greater learning than those who started with lower levels of divergence. Supplementary Table 3 presents the full model outputs of these analyses.

The multiverse correlation analyses performed at each individual participant level across verbs suggested a similar pattern (Figure 3B): participants who showed greater level of divergence from subsequent training exhibited greater changes in corresponding behavior. After Bonferroni correction for multiple tests, participants' changes in their mouse click behavior ($R = 0.46$, $p < 0.001$) and first fixation ($R = 0.57$, $p < 0.001$) remained significantly associated with the levels of their pre-test divergence from subsequent training.

To summarize, the more divergent participants' initial verb bias knowledge was from subsequent training types (e.g., initial modifier-biased interpretation but received the instrument-biased training), the greater the learning effects, linking prediction errors to learning outcomes.

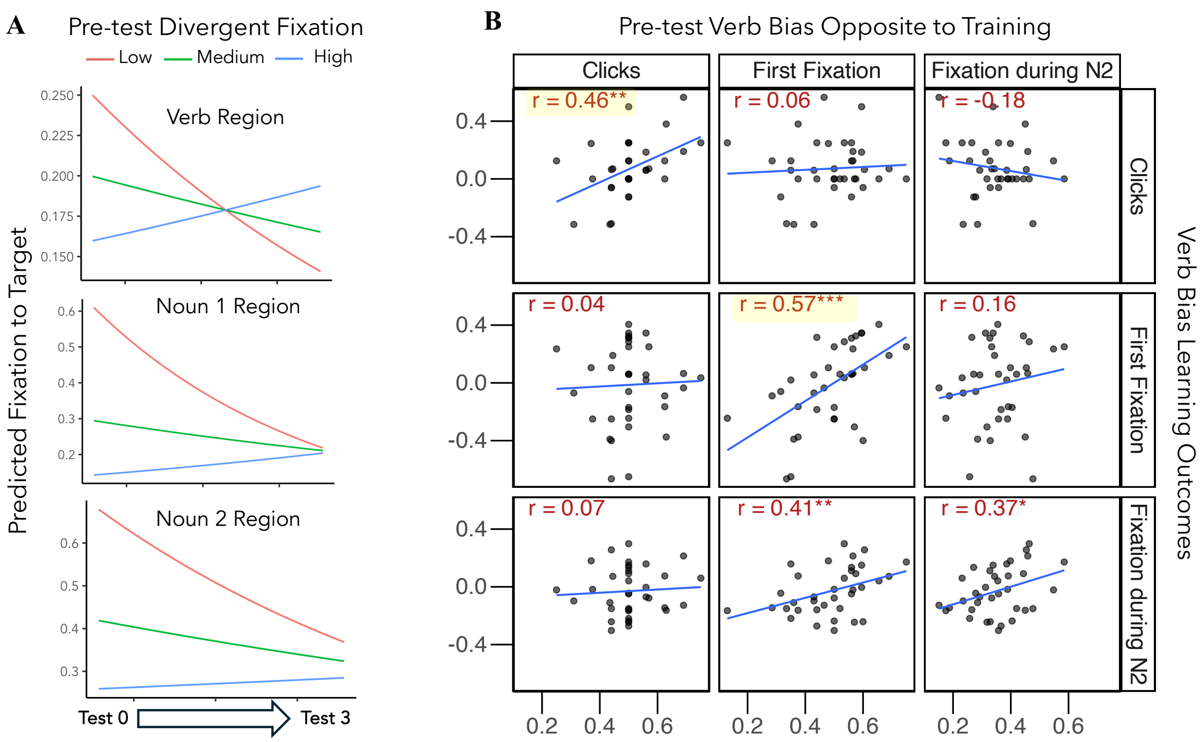


Figure 3. Relationship between pre-test verb bias knowledge and verb bias learning outcomes. (A). Significant modulation effects of pre-test divergent fixation (e.g., % of looks to target instrument for verbs subsequently trained in the modifier-biased condition) on the growth of target fixation over the course of training. (B) Correlations between pre-test processing that was opposite to subsequent training and verb bias learning outcomes (change scores in first fixations to targets, proportion of looks to targets during N2 window, and clicks on targets). * uncorrected p < 0.05; ** uncorrected p < 0.01; ***

uncorrected $p < 0.001$. Highlighted correlations remain significant after Bonferroni correction for multiple comparisons.

Preregistered Analyses 2: The Stability of Linguistic Prediction Skills Across Tasks

Having explored the dual role of prediction in learning, we now seek to understand the consistency of individuals' linguistic prediction skills across different tasks. To achieve this, we investigated their linguistic prediction abilities using a language comprehension task adapted from a study by Nation et al. (2003). In their research, they found that skilled comprehenders engaged in anticipatory looking based on verb semantics during sentence comprehension. In our study, we also wanted to test whether skilled language users, adults, used their knowledge of verb semantics to predict upcoming information. To do this, we compared the proportion of fixations on the semantically relevant object within a predictive time window between predictable and unpredictable conditions. The predictive time window extended from the onset of the first noun to the onset of the second noun. We used the participants' proportion of fixations on the target object ("pen") as the dependent variable in our analysis. To analyze the data, we employed mixed-effects models, incorporating random intercepts for participants and items, along with random by-participant slopes for semantic predictability conditions. We represented predictability conditions as a deviation coded contrast: semantically predictable (-0.50) versus semantically unpredictable (0.50). Our main criterion for considering successful replication of the task was the observation of a significant anticipatory looking effect on the target in the predictive time window.

We also examined adults' first fixation patterns using the same approach as we performed on the verb bias learning task. We first identified individuals' first fixations upon hearing the verb and analyzed patterns related to verb informativeness using mixed-effects logistic regression. The dependent variable was whether participants' first fixation was on the target item. Similarly, we included random intercepts for participants and items in the model. We aimed to determine if participants demonstrated a higher proportion of first fixations on the target item in the semantically predictable condition compared to the semantically unpredictable condition, indicating their utilization of verb semantic information to predict forthcoming linguistic content.

Lastly, we related an individual's ability to make use of semantic information for prediction with how quickly that individual adjusts their verb bias knowledge during verb bias training. We predicted that adults who are good at semantic prediction during sentence comprehension should also show greater evidence of newly learned verb bias knowledge. We operationalize the semantic prediction measure as the condition difference in 1) proportion of target fixation during the anticipatory time window and 2) the proportion of first fixations to the target. We performed two sets of analyses. First, we estimated the extent to which individual's semantic prediction, as a fixed effect, modulates learning of verb biases measured by the growth of participants' anticipatory looking behavior in mixed-level linear regression models. We look for a three-way interaction between semantic prediction, testing order (test 3 vs. test 0) and training type (Instrument vs. Modifier) as an indicator of reliable modulation across verbs. Second, considering a potential asymmetry between instrument and modifier-biased training, we performed a total of twelve correlations between two verb semantic prediction effect measures and three verb-bias learning measures (difference scores of first fixation, proportion of fixations during N2 time window, and proportion of clicks on the trained target between Time 3 and Time 0) for each type of training. The significant findings were reported after Bonferroni corrections for multiple comparisons.

Results For Preregistered Analyses 2

The proportion of fixations to the semantically relevant object were plotted by predictability conditions in Figure 4. As mentioned previously, the predictive time window was defined as the period when the semantic information from the verb is potentially available, but the target object has not yet been named. Thus, fixations in this time window index an anticipatory process for the upcoming linguistic content. As expected, there was a significant difference between the predictable and unpredictable conditions for the fixations to the target in the predictive time window ($\beta = -0.05$, $SE = 0.02$, $t = -2.22$, $p = 0.03$, $\eta^2 = 0.12$ [95% CI: 0.01 to 1.00]). The analyses on the first fixation yielded similar but a weaker condition effect ($\beta = -0.27$, $SE = 0.18$, $z = -1.50$, $p = 0.13$, $OR = 0.76$ [95% CI: 0.54 to 1.09]). Restricting analyses within the subset of trials where participants made the first fixation during the predictive time window (mean onset of the first fixation is 858 ms ± 19.6 ms. 92% of the trials) showed a similar marginal effect (β = -0.32, *SE* = 0.18, z = -1.73, *p* = 0.08, *OR* = 0.72 [95% CI: 0.51 to 1.04]).

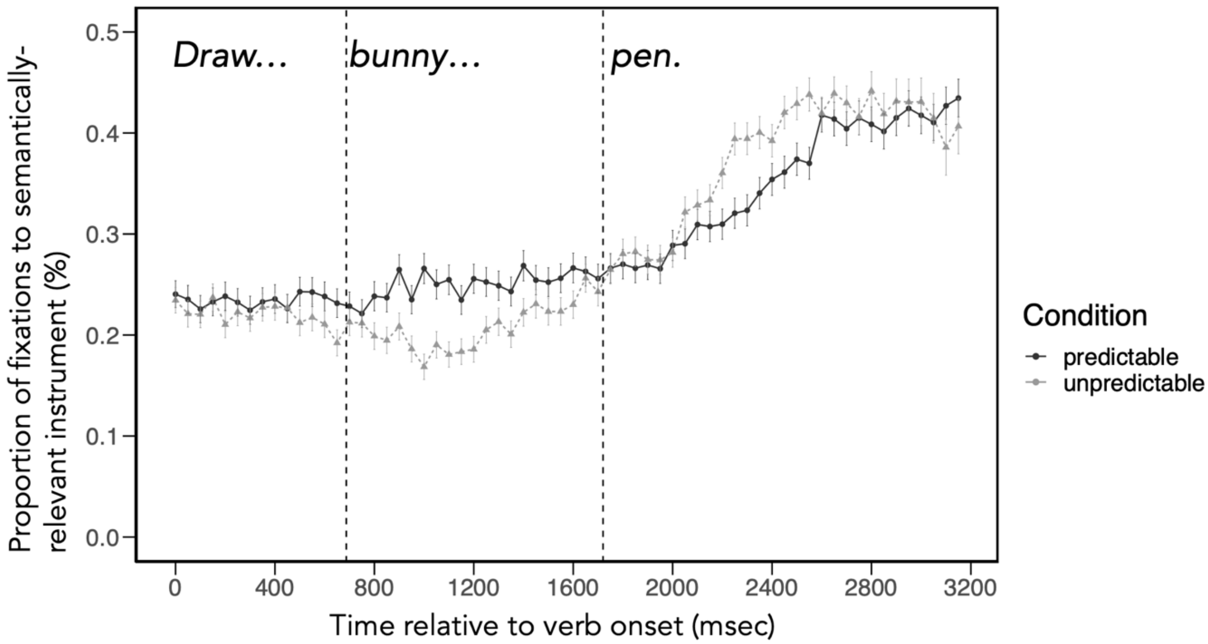


Figure 4. Eye Movement Results for Verb Semantic Prediction Task. Proportion of the fixations to the semantically relevant instrument (e.g., pen) in both predictable and unpredictable conditions over time in milliseconds. 0 ms corresponds to the onset of the verb. The data are aligned at verb onset; dashed vertical lines indicate approximate onsets of N1 (e.g., bunny) and N2 (e.g., pen), offset by 200 ms.

The key question for this pre-registered analysis is whether linguistic prediction is stable between language comprehension and language learning. We found individuals' semantic prediction marginally modulated the training effect on listeners' anticipatory looking patterns such that those with greater semantic prediction effect also shows greater change in anticipatory looking patterns over the course of learning (three-way interaction: semantic prediction effect * training type * testing order: β = 1.29, *SE* = 0.73, z = 1.76, *p* = 0.079, *OR* = 3.65 [95% CI: 0.86, 15.4], see Figure 5A for an illustration of the interaction using median split of the semantic prediction effect). We replicated these findings using the proportion of fixation during the verb region as the earliest metric of anticipatory looking behavior: greater semantic predication is significantly associated with greater training effect on listeners' fixation to target animal upon hearing the verb (three-way interaction: semantic prediction effect * training type * testing order: $\beta = 0.84$, $SE = 0.37$, $t = 2.27$, $p = 0.024$, $OR = 2.32$ [95% CI: 1.12, 4.83]). Semantic prediction as measured by the proportion of looks to target during the anticipatory window did not yield any significant three-way interactions.

To further understand the relationship between semantic prediction and verb bias learning holistically (across trials), we performed correlation analyses. After correcting for multiple tests, our analyses suggested that stronger semantic prediction effect, measured by proportion of first fixation to the target, was strongly related to greater learning for modifier-trained verbs, measured by the changes in the target-animal fixation at the end of the sentence (Figure 5B; $R = 0.50$, $p = 0.002$) and marginally related to changes in anticipatory looking to animals ($R = 0.33$, uncorrected $p = 0.052$). Mediation analysis (*mediation* R package*:* Tingley et al., 2014) suggested that 30% (95% CI: 0.07 to 0.60) of the effect of sematic prediction on the changes of final interpretation during verb bias learning was carried through its effect on the changes of anticipatory fixation (Figure 5C; indirect effect: $\beta = 0.24$, 95% CI: 0.04 to 0.46, $p = 0.02$; direct effect: $\beta = 0.57$, 95% CI: 0.20 to 0.90, $p = 0.004$). Restricting analyses within the subset of trials where participants made the first fixation during the predictive time window yielded similar results in both the correlation analyses ($R = 0.54$, $p < 0.001$) and mediation analyses (indirect effect: $\beta = 0.24$, 95% CI: 0.06 to 0.47, $p = 0.006$; direct effect: $\beta = 0.57$, 95% CI: 0.21 to 0.88, $p = 0.004$). These results suggest that stronger semantic prediction is related to greater learning-induced changes in predictive processing and the outcome of ambiguity resolution.

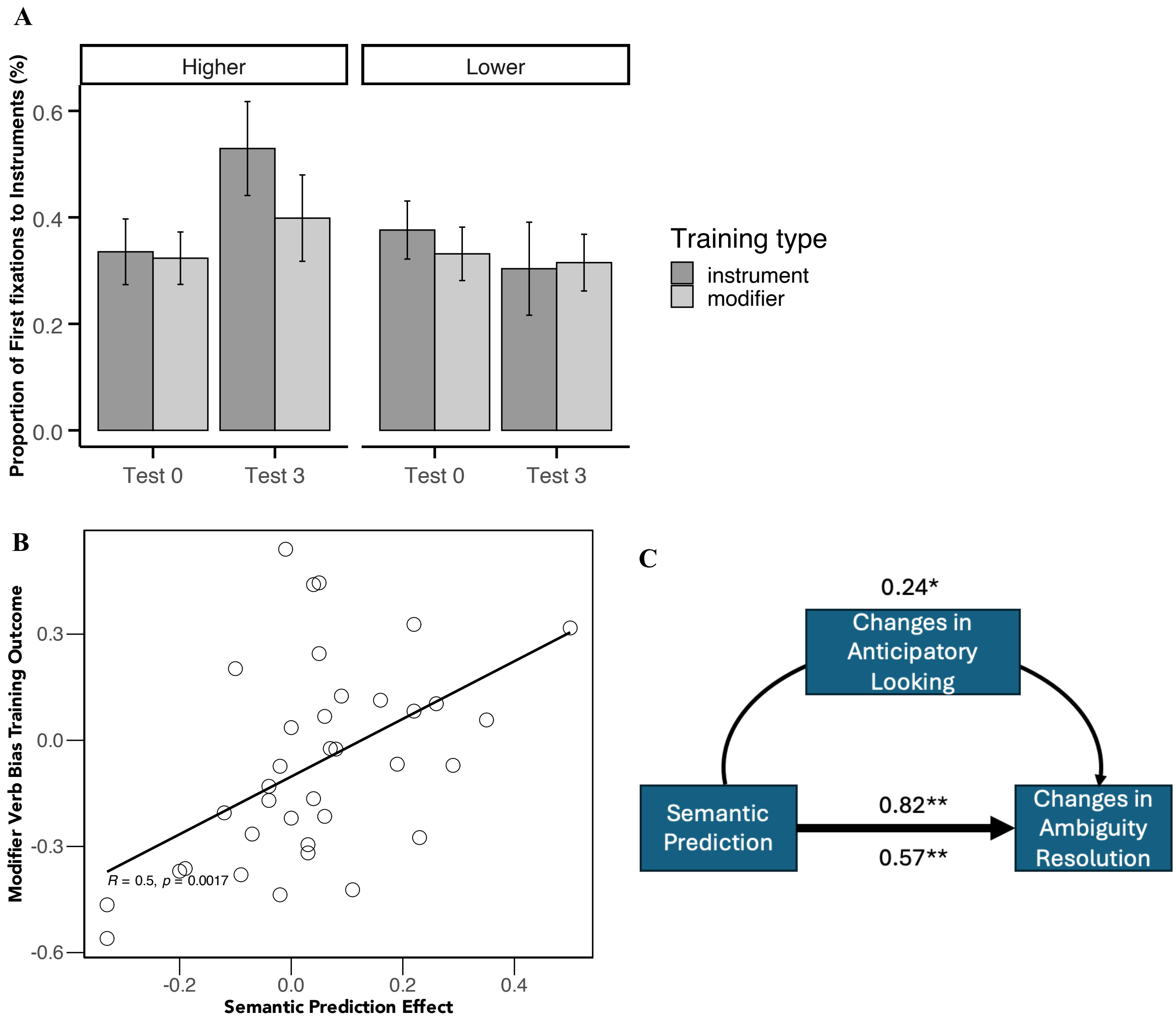


Figure 5. **(A)** An illustration of how individual's semantic prediction effect (median-split, Higher vs. Lower) modulates changes in anticipatory looking behavior during verb bias learning. Semantic prediction effect was measured by the difference in the proportion of first fixation to the target between the predictable and the unpredictable conditions and was statistically analyzed as a continuous variable. **(B)** Correlations between semantic prediction (same as above) and verb bias learning outcomes (the changes in proportion of fixations to the target animal during the Noun 2 region, Test 3 – Test 0); Bonferroni-corrected *p's* < 0.05. (B) The path model of the mediation analysis with the coefficient estimates (indirect, direct, and total effects): the effect of sematic prediction on the learning-induced changes in ambiguity resolution (target animal fixation during the Noun 2 window) was carried through its effect on the changes of anticipatory fixation (proportion of first fixations to either animal). *, $p < 0.05$; **, $p < 0.01$.

Discussion

The goals of the present study are to understand the dual role of prediction in learning and examine the stability of individuals' linguistic prediction across language tasks. To begin with, we examined whether learning is associated with growth in prediction by asking whether individuals generate predictions for newly learned verb-specific knowledge after a short period of learning using the verb bias learning task adapted from Ryskin et al. (2017). Our findings replicated the malleability of adults' verb bias in the clicking results and revealed the impact of individual variations in the eye fixation data. Crucially, by employing individuals' first fixations as a measure of prediction, we found that adults' first fixations to the trained target were significantly associated with their learning outcomes. Learners who displayed a greater proportion of fixations to the trained target during the N2 time window or a higher proportion of clicks on the trained target following the training were more prone to directing their initial eye gaze towards the trained target after hearing the verb. These results suggest that successful learners could effectively utilize the newly learned verb bias knowledge to generate predictions shortly after being exposed to the training. Second, we asked whether there is available evidence that stronger prediction errors may also be associated with greater learning. By employing diverse metrics to assess individuals' initial verb bias knowledge, we found that individuals and verbs with more divergent a priori biases from the upcoming training condition exhibited a greater shift in verb bias knowledge. These findings provide supportive evidence for the error-based learning framework which posits the discrepancies between individuals' prediction and actual input generate a strong error signal, leading to a higher rate and magnitude of change compared to situations where predictions align more closely with input. Finally, we investigated the stability of individuals' linguistic prediction skills across various language tasks. The results indicated that, in line with existing literature (e.g., Nation et al., 2003), adults demonstrated the ability to anticipate upcoming linguistic items based on the semantic information of verbs (e.g., "bake"), relying on the unfolding relationship between the scene and the sentence (e.g., "the cake"). Crucially, our findings suggest that those who excelled in using verb semantics to predict sentences during the comprehension task also exhibited stronger anticipatory looking patterns and better performances on the verb bias learning task. Taken together, these findings offer compelling evidence supporting the bi-directional relationship between prediction and learning and suggest that adults' linguistic prediction skills tend to be consistent across different language tasks.

In the verb bias learning task, we successfully replicated the action findings from Ryskin et al. (2017), which showed that adults' syntactic parsing choices are well aligned with their verb-bias training experiences. This suggests that adults can modify their biases surrounding the interpretation of individual verbs based on their accumulated experience with specific verb-structure co-occurrences. However, we did not observe a significant interaction of testing time points and training conditions on fixation patterns in the N2 time window. While we are confident that the inter-subject variabilities in training effects present meaningful opportunities for us to discover relationships between prediction and learning, it is still possible that the magnitude of the overall training effect was affected by the webcam technology, as suggested by previous studies (e.g., Semmelman & Weigelt, 2018). Therefore, it is important for future studies to consider adjustments for the expected effect size in webcam-based eye-tracking research when estimating sufficient statistical power.

However, the findings of our pre-registered analyses went beyond just replication. We found a significant correlation between adults' first looks at the trained target and their learning outcomes. This correlation indicates that the ability to generate predictions is closely linked to successful learning. Participants demonstrated their capacity to utilize newly acquired linguistic information to direct their anticipatory eye movements, effectively guiding online sentence processing. This finding aligns well with the findings in the statistical learning literature, where participants' reaction times become faster in predictable conditions after implicit learning (e.g., Hunt & Aslin, 2001). However, our study differs from traditional statistical learning research as we operationalized a classic learning task with explicit feedback in each test trial and a direct measure of anticipation. Through this approach, we demonstrated how adults updated their existing verb bias knowledge and developed new anticipatory looks for correct items as learning unfolds. Our findings establish a connection between the predictive learning mechanism and the learning process itself.

Another key discovery in our study is that the degree to which initial verb bias mismatches subsequent inputs predicts verb-specific malleability. These findings align well with prediction-based models of learning (e.g., Chang et al., 2006; Dell & Chang, 2014; Elman, 1990). In particular, the results suggest that generating a preliminary prediction and subsequently adjusting linguistic representations in response to a prediction error, rather than a confirmed prediction, could be an effective and efficient way to enhance successful learning in subsequent moments. Within error-based learning frameworks, error signals arise when there is a mismatch between the listener's predictions and the perceived input. Therefore, learners who made strongly divergent predictions exhibited a faster learning rate and experienced more significant outcome changes. The notable correlation between individuals' initial verb bias experience in the Pre-test trials and the subsequent changes in verb bias because of learning in our study thus indicates that learning might be sensitive to the magnitude of prediction error. This insight highlights the intricate relationship between prediction and learning processes, shedding light on the mechanisms that drive language development and facilitate language comprehension.

To examine the stability of individuals' prediction abilities across different language tasks, we conducted an evaluation of adults' linguistic prediction using a modified version of the language comprehension task from Nation et al. (2003). We hope to replicate the effects of anticipatory looks observed in the semantic prediction literature for sentence comprehension (e.g., Altmann & Kamide, 1999; Altmann & Kamide, 2007; Kamide et al., 2003; Kamide et al., 2003) and extend this work by establishing a connection between participants' first fixation patterns in the comprehension task and their prediction measures and learning performances in the verb bias learning task. Consistent with the existing literature, our study confirmed that adults displayed more fixations on the semantically related target before it was explicitly referenced. This finding supports the notion that adults can pre-activate the representation of the forthcoming noun during sentence processing based on the semantic representation of the verb. Moreover, similar to the first fixation results in the verb bias learning task, some but not all individuals exhibited anticipatory looking behavior at a very early stage in the process. They directed their first fixation to the trained target immediately after hearing the verb, which served as the first word of the sentence, and attempted to predict the last word of the spoken utterance. These anticipatory effects observed in the first fixation align with the findings from scene-viewing literature, which suggest that eye fixation sites are closely linked to meaningful scenes, providing cues about the likely location of specific objects (Henderson et al., 1999; Oliva et al., 2003). Through our investigation into the connections between individuals' linguistic prediction skills across diverse language tasks, our study enhances our comprehension of how humans effectively generalize the learning mechanism across various linguistic contexts. This, in turn, deepens our understanding of how we learn and update knowledge about the world throughout our lives.

The replicated and novel findings in our study contribute valuable insights to existing knowledge about the dual role of prediction in learning and the stability of linguistic prediction skills across different tasks. However, it is also essential to acknowledge several limitations in our current approach, which require further investigation in future studies. First, while the current study kept our stimuli as close as possible to Ryskin et al (2017) to facilitate comparisons across studies, our stimuli did not manipulate event plausibility, which is another important cue for sentence processing and have been shown to interact with listeners' usage of verb biases (Garnsey et al., 1997; Kidd et al., 2011; Yacovone et al., 2021). Future research should systematically investigate how plausibility changes learners' reliance on linguistic distributional information for verb bias learning. The second limitation is that the current work does not establish a definitive causal relationship between prediction and learning. Future research should aim to explore causality more explicitly to better understand the dynamics between prediction and learning. The third potential concern is that the current results came from a limited number of test trials, with only one test trial per testing time point for the purpose of replicating the original task. This limited sample size may reduce the statistical power of the study and potentially limit the generalizability of the results. To strengthen the findings, future studies should consider increasing the number of test trials and assessing participants' verb bias learning performance at multiple time points, which can yield more robust outcomes. The fourth limitation is related to the web-based eye-tracking method. Despite the evidence of qualitatively similar time courses of our fixation data with the in-lab eye-tracking data, web-based eye-tracking method is susceptible to variables such as reduced task engagement and inconsistent software (e.g., browser) and hardware (e.g., screen sizes) (Slim & Hartsuike, 2023; Slim et al, 2024). Our prediction and learning measures are difference scores between conditions or time points, which control for these confounding factors. However, these findings require further replications in a much larger sample. Addressing these limitations will facilitate a deeper understanding of the interplay between prediction and learning, paving the way for more comprehensive and insightful research in this area.

Taken together, the present study is one of the first attempts to thoroughly examine the close relationship between prediction and learning. We not only successfully replicated two crucial literature findings suggesting that adults are able to use verb cues to anticipate upcoming information during sentence comprehension (Nation et al., 2003) and their interpretations of the ambiguous syntactic structure were guided by the newly learned verb biases (Ryskin et al., 2017), but also extends beyond the existing literature in several ways. First, we modified the verb learning task with test trials throughout the training period (before, during, and after training). This allowed us to discern the evolving effects of training and predictive behaviors over time. Second, we tested the hypotheses by implementing a novel webcam-based eye-tracking technique, a popular tool in computer science domain (e.g., Smartphone eye-tracking: Valliappan et al., 2020; TurkerGaze: Xu et al., 2015). Lastly, different from previous studies that calculates an average over an interval of time, we validated that individuals' first fixation can potentially serve as a measure to reflect individual differences in linguistic prediction ability, thus suggesting that predictive behaviors can occur very early during the language processing stage. Overall, the novel approach and significant findings in the current study contribute to a deeper understanding of the intricate interplay between prediction and learning across various language domains. As a result, the current study opens new avenues for future research and offers valuable insights into the mechanisms of lifelong learning.

Constraints on Generality

Our sample consisted of monolingual English-native speakers aged 18 – 35 from the Newark, Delaware area, recruited online and with typical language development. Most of participants were students from the University of Delaware. The findings are most relevant to populations with similar ages, education backgrounds, and English language proficiency. There is a need for further research in populations such as young children, teenagers, senior citizens, non-native English speakers, and clinical populations.

Author's notes

All data, code, analyses, stimuli, and pre-registrations are available from our project's Open Science Framework repository at this link: https://osf.io/e3dh9/. Hypotheses, sample sizes, and analysis plans were pre-registered. We have no conflicts of interest to disclose. The authors would like to express their gratitude to Dr. Rachel Ryskin for sharing the verb bias learning task materials and to the researchers who offered feedback at the 2022 Human Sentence Processing conference.

References

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, 28(4), 525-560. https://doi.org/10.1080/01690965.2012.658072

Arai, M., & Mazuka, R. (2014). The development of Japanese passive syntax as indexed by structural priming in comprehension. *Quarterly Journal of Experimental Psychology*, 67(1), 60-78. https://doi.org/10.1080/17470218.2013.790454

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. https://doi.org/10.3758/s13428-019-01237

Argaman, V., Pearlmutter, N. J., & Garnsey, S. M. (2002). Lexical semantics as a basis for argument structure frequency biases. *Sentence processing and the lexicon: Formal, Computational and Experimental Perspectives*. Amsterdam: John Benjamins. https://doi.org/10.1075/nlp.4.16arg

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, 62(1), 67-82. https://doi.org/10.1016/j.jml.2009.09.005

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition*, 129(2), 205-211. https://doi.org/10.1016/j.cognition.2013.07.003

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed- effects models using lme4. *Journal of Statistical Software*, 67(1). https://doi.org/10.18637/jss.v067.i01

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, 114(3), 455-461. https://doi.org/10.1016/j.cognition.2009.11.005

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning?. *Journal of Experimental Psychology: General*, 129(2), 177. https://doi.org/10.1037//0096-3445.129.2.177

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3), 437-458. https://doi.org/10.1016/j.cognition.2006.07.003

Bott, N. T., Lange, A., Rentz, D., Buffalo, E., Clopton, P., & Zola, S. (2017). Web camera based eye tracking to assess visual memory on a visual paired comparison task. *Frontiers in Neuroscience*, 11, 370. https://doi.org/10.3389/fnins.2017.00370

Bubic, A., Yves von Cramon, D., Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontier in Human Neuroscience,* 4, 1–15.

Calabrich, S. L., Oppenheim, G. M., & Jones, M. W. (2021). Episodic Memory Cues in Acquisition of Novel Visual-Phonological Associations: a Webcam-Based Eye-Tracking Study. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43. Retrieved from https://escholarship.org/uc/item/76b3c54t

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29, 217-230. https://doi.org/10.1023/A:1005101313330

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–72. https://doi.org/10.1037/0033-295X.113.2.234

Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology*, 6, 1182.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356-371. https://doi.org/10.1016/j.cognition.2009.10.009

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43. Retrieved from https://escholarship.org/uc/item/6182t9jb

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641-655. https://doi.org/10.1016/S0022-5371(81)90220-6

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E

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.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469-495. https://doi.org/10.1006/jmla.1999.2660

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not in children. *Cognition*, *211*, 104650.

Gambi, C., & Messenger, K. (2023). The Role of Prediction Error in 4-Year-Olds' Learning of English Direct Object Datives. *Languages*, *8*(4), 276. https://doi.org/10.3390/languages8040276.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Hare, M., McRae, K., & Elman, J. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, 19(2), 181-224.

Havron, N., de Carvalho, A., Fiévet, A. C., & Christophe, A. (2019). Three-to four-year-old children rapidly adapt their predictions and use them to learn novel word meanings. *Child development*, *90*(1), 82-90.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5), 438-443. https://doi.org/10.1111/1467-9280.00183

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130(4), 658. https://doi.org/10.1037/0096-3445.130.4.658

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Linguistics and Language Compass*, *2*(4), 647–670. https://doi.org/10.1111/j.1749- 818X.2008.00072.x

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

Kishiyama, M. M., Yonelinas, A. P., and Knight, R. T. (2009). Novelty enhancements in memory are dependent on lateral prefrontal cortex. *J. Neurosci.* 29, 8114–8118.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1), 95-127. https://doi.org/10.1016/j.cognition.2004.03.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1– 26. https://doi.org/10.18637/jss.v082.i13

Lin, Y., & Fisher, C. (2017). Error-based learning: A mechanism for linking verbs to syntax. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 39):754–759

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, 91, 5-27. https://doi.org/10.1016/j.jml.2016.03.009

McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 7(3), 634. https://doi.org/10.1037/0096-1523.7.3.634

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1, 31. https://doi.org/10.3389/fpsyg.2010.00031

Nation, K., Marshall, C. M., & Altmann, G.T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314– 329. https://doi.org/10.1016/j.jecp.2003.09.001

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. In *Proceedings 2003 International Conference on Image Processing* (Vol. 1, pp. I-253). IEEE. https://doi.org/10.1167/3.9.3

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: scalable webcam eye tracking using user interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, 3839–3845

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.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4), 633-651. https://doi.org/10.1006/jmla.1998.2592

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Qi, Z., Yuan, S., Fisher, C.L. (2011). Where does verb bias come from: Experience with particular verbs affects online sentence processing, In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of 35th Boston University Conference on Language Development*. 500-512. Somerville, MA: Cascadilla Press.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language*, *89*(4), 760– 793. https://doi.org/DOI: https://doi.org/10.1353/lan.2013.0068

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental psychology*, *55*(8), 1656-1665.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 781. https://doi.org/10.1037/xlm0000341

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50, 451-465. https://doi.org/10.3758/s13428-017-0913-7

Slim, M. S., Kandel, M., Yacovone, A., & Snedeker, J. (2024). Webcams as windows to the mind? A direct comparison between in-lab and web-based eye-tracking methods. *Open Mind*, *8*, 1369-1424.

Snedeker, J., Trueswell, J.C. (2004). The developing constraints on parsing decisions: the role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology,* 49(3), 238-299.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712. https://doi.org/10.1177/1745691616658637

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634. https://doi.org/10.1126/science.7777863

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29, 557-580. https://doi.org/10.1023/A:1026464108329

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). *mediation: R package for causal mediation analysis*. *Journal of Statistical Software, 59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., ... & Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications*, 11(1), 4553. https://doi.org/10.1038/s41467-020-18360-5

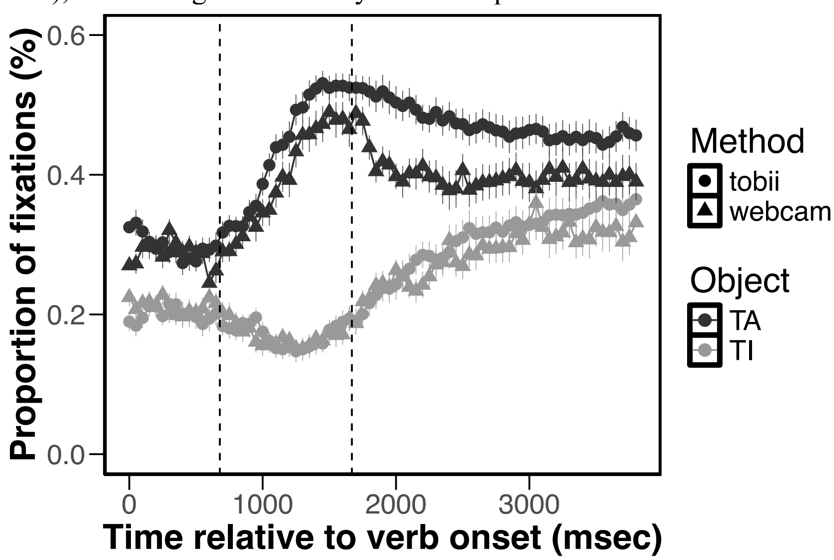
Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015). Turkergaze: Crowdsourcing saliency with webcam based eye tracking. arXiv preprint arXiv:1504.06755.

Supplementary Methods

Eye Movement Data Validation

The webcam-based eye-tracking technique and WebGazer.js provided a total of 263,514 frame estimates in the verb bias test trials. We excluded approximately 10.66% of the estimates (N=28,102) with face convergence value lower than 0.5 based on suggestions by Anwyl-Irvine et al. (2018) to ensure accurate gaze estimation. Also, none of the participants was excluded based on the data exclusion criteria on the pre-registered analysis (i.e., participants who have usable data from fewer than 13 out of 40 possible trials would be removed from the analysis).

First, we compared the time courses of fixation patterns to target between the auto-coded eye-tracking data from WebGazer.js and the in-lab eye-tracking data from Tobii X-120. The in-lab sample is a group of adult participants (N = 25) from a different study. They listened to the same test sentences as the current verb bias learning task, but without any preceding training trials. The webcam eye-tracker showed similar trajectories as the in-lab eye-tracking data (Supplementary Figure 1). We performed a deviation analysis comparing the time courses of the two methods. For the fixations to the target animal, the difference between the two methods seems to emerge after the onset of the second noun (e.g., hat). However, such difference did not withstand FDR corrections for multiple comparisons (all p's > 0.14). The time courses of fixations to the target instrument did not differ between the two methods (all p's > 0.50), confirming the similarity in the temporal resolution of the online and in-lab eye-tracking methods.



Supplementary Figure 1. Similar time courses of online sentence processing (equi-biased verbs) between online (N = 36) and in-lab samples (N = 25). The two vertical dotted lines represent the onset of the first (e.g., duck) and the second (e.g., hat) nouns, respectively, in the sentence (e.g., Rub the duck with the hat). TA: target animal; TI: target instrument.

Then we followed Bott et al. (2017) and employed two approaches to validate the quality of our data using the eye gaze data in the verb semantic prediction task. Because our main task is a verb bias learning task, we expect changes of fixation patterns across trials. The semantic prediction task, i.e., informative vs. neutral verbs, offers a better opportunity to measure stable processing within and across individuals.

We examined whether the proportion of looks during sentence processing achieved adequate internal consistency in a time window where we expect the least ambiguity. We chose the N2 time window of the eighteen semantically predictable trials to assess the split-half reliability of the proportion of fixation to the target object (e.g., "crayon" in "*Draw on the lion using the crayon*"). We split the trials into odd and even trials and then averaged the proportion of fixations to the target in the N2 time window across all the odd or even trials of the semantically predictable trials for each participant.

We also tested the difference between the manually frame-by-frame coded data collected from zoom recorded videos and the eye gaze data collected from WebGazer.js across 36 semantically predictable and unpredictable sentences in four randomly selected participants. If the two datasets are similar, we expected a lack of difference between the two coding methods, and they should generate positively correlated measures across all trials. Using Datavyu (Datavyu Team, 2014), two trained coders watched the playback of zoom video recordings frame-by-frame and marked the location of listeners' fixation (top left, top right, bottom left, bottom right, and away/blinks) from the onset of the sentence until 2.8s later (70 frames, the shortest sentence duration so that the coding windows are consistent across trials). The frame-by-frame inter-coder agreement is 90.2% on average. Proportion of fixation to the semantically relevant instrument during the predictable Noun 1 window (e.g., "bunny" in "*Draw on the bunny using the pen*") was then computed for each trial for each coding method, logit-transformed, and entered as the dependent variable in a mixed-effects linear regression model. We tested the fixed effects of conditions (predictable vs. unpredictable), coding methods (manual vs. auto), and the interaction between the two factors using the lmer function in the lme4 R package (Kuznetsova et al., 2017). Random intercept of participants and trials were included. Bayes factors (BFs) were computed to quantify the degree by which the probability shifts away from or towards the null hypothesis after observing the data. BFs below 3 are considered weak evidence for the alternative hypothesis (Lee & Wagenmakers, 2013). To further assess the similarity between the data, we performed Spearman correlation between the two coding methods across all trials.

Both approaches confirmed the validity of the auto-coded eye-tracking data from WebGazer.js. The split-half reliability analyses suggest an acceptable internal consistency (R = 0.65, adjusted Spearman-Brown prophecy formula = 0.79). The comparison between the auto-coded and manually coded data suggests a similar pattern between the two datasets. The results of mixed-effects regression model revealed no significant difference between the two coding methods ($\beta = -0.28$, $SE = 0.31$, $t = -0.92$, $p = 0.36$, BF = 0.58) nor was there an interaction between coding methods and condition ($\beta = -0.91$, $SE = 0.62$, $t = -1.47$, $p = 0.14$, BF = 0.13). Across all 209 sentences, the auto-coded and the manually frame-by-frame coded data are modestly associated ($R = 0.47$, $p < 0.001$).

Supplementary Results

To investigate how real-time sentence processing changes over the course of training, we examined the proportion of target animal and target instrument fixations during verb, noun-1, and noun-2 time windows. The results of mixed-effect regression analyses showed no significant main effect of Training Types or interactions between Training Types and Testing Time Points in any time window (Supplementary Table 1).

Supplementary Table 1: Effects of verb bias learning on the proportion of fixations to target instrument and target animal during online sentence processing.

	Fixations to Target Instrument				**Fixations to Target Animal**			
Fixed effects	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Verb Time Window								
(Intercept)	-1.78	0.09	-18.75	0.00***	-1.51	0.07	-20.36	0.00***
Training Types	0.02	0.13	0.16	0.87	-0.15	0.14	-1.10	0.27
Testing time points	0.02	0.06	0.21	0.84	-0.13	0.06	-2.17	0.03*

Types x Time points	-0.02	0.12	-0.21	0.84	-0.04	0.12	-0.32	0.75
N1 Time Window								
(Intercept)	-1.92	0.06	-29.54	0.00***	-0.86	0.11	-7.57	0.00***
Training Types	0.02	0.11	0.20	0.85	0.13	0.12	1.07	0.29
Testing time points	0.03	0.05	0.66	0.51	-0.22	0.05	-4.07	0.00***
Types x Time points	0.05	0.10	0.48	0.63	-0.04	0.11	-0.35	0.73
N2 Time Window								
(Intercept)	-1.48	0.11	-14.08	0.00***	-0.69	0.16	-4.39	0.00***
Training Types	-0.08	0.12	-0.67	0.50*	0.09	0.14	0.69	0.49
Testing time points	0.14	0.05	-2.54	0.01*	-0.07	0.06	-1.09	0.28
Types x Time points	-0.11	0.11	-1.02	0.31	0.07	0.12	0.63	0.53

$*p < .05$; $** p < .01$; $***p < .001$

To investigate whether participants generate new predictions for verb-specific knowledge over the course of learning, we identified participants' first fixation in each trial as a measure for prediction and analyzed the training effect. The results of mixed-effects logistic regression model revealed that there was a significant main effect of Testing Time Points but there was no main effect of Training Types or interactions between Training Types and Testing Time Points (Supplementary Table 2).

Supplementary Table 2: Effects of verb bias learning on the proportion of trials with first fixation to either instrument.

Fixed effects	***β***	***SE***	**z**	***p***
(Intercept)	-0.38	0.06	-5.85	0.00***
Training Types	-0.09	0.12	-0.69	0.49
Testing time points	0.14	0.05	2.70	0.007**
Types x Time points	-0.10	0.11	-0.95	0.34

$*p < .05$; $** p < .01$; $***p < .001$

To test the role of prediction in learning, we carried out an exploratory analysis to test the error-based learning account. For each verb for each participant, we extracted the eye-movement metrics that represent the real-time interpretation *opposite from* subsequent training. (e.g., *target-animal* fixation for a verb to be trained in an *instrument-biased* condition, or *target-instrument* fixation for a verb to be trained in a *modifier-biased* condition). Then we tested whether the degree of such divergence between pre-test processing and subsequent training modulates learning by assessing the interaction between pre-test fixation and testing time points in a mixed-effects linear regression model for participant's target fixation (e.g. target-instrument fixation for instrument-trained verbs). Across all three eye-movement analysis windows, we found greater pre-test divergence from subsequent training significantly predicted more growth in fixations to the target over the course of learning (Supplementary Table 3).

Supplementary Table 3: Effects of pre-test processing that was opposite to subsequent learning on target fixation over the course of verb bias learning.

Fixed effects	*β*	*SE*	*t*	*p*
Verb Time Window				
(Intercept)	-1.74	0.12	-14.44	0.00***
Divergent Fixation	-0.02	0.04	-0.62	0.53
Testing time points	0.05	0.08	0.67	0.50
Divergent Fixation x Time points	**0.08**	**0.04**	**2.20**	**0.03***
N1 Time Window				
(Intercept)	-1.67	0.11	-15.86	0.00***
Divergent Fixation	-0.24	0.04	-6.52	0.00***
Testing time points	0.06	0.07	0.86	0.39
Divergent Fixation x Time points	**0.14**	**0.03**	**4.29**	**0.00*****
N2 Time Window				
(Intercept)	-1.22	0.14	-8.81	0.00***
Divergent Fixation	-0.19	0.04	-4.61	0.00***
Testing time points	0.00	0.06	0.00	0.99
Divergent Fixation x Time points	**0.07**	**0.04**	**2.01**	**0.04***

$^*p < .05$; $^{**}p < .01$; $^{***}p < .001$