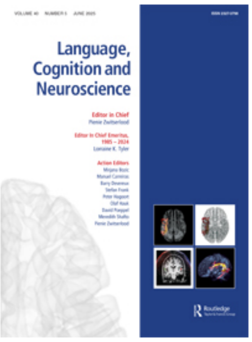

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: www.tandfonline.com/journals/plcp21

Indexing prediction error during syntactic priming via pupillometry

Shanthi Kumarage, Anton Malko & Evan Kidd

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2025.2506634

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Indexing prediction error during syntactic priming via pupillometry

Shanthi Kumarage [a,b], Anton Malko [c] and Evan Kidd [c]

[a]Language Development Department, Max Planck Institute of Psycholinguistics, Nijmegen, Netherlands; [b]School of Medicine and Psychology, The Australian National University, Canberra, Australia; [c]School of Literature, Languages and Linguistics, The Australian National University, Canberra, Australia

ABSTRACT

Prediction is argued to be a key feature of human cognition, including in syntactic processing. Prediction error has been linked to dynamic changes in syntactic representations in theoretical models of language processing. This mechanism is termed *error-based learning*. Evidence from syntactic priming research supports error-based learning accounts; however, measuring prediction error itself has not been a research focus. Here we present a study exploring the use of pupillometry as a measure of prediction error during syntactic priming. We found a larger pupil response to the more complex and less expected passive structure. In addition, the pupil response predicted priming while being weakly dependent on changes in expectations over the experiment. We conclude that the pupil response is not only sensitive to syntactic complexity in comprehension, but there is some evidence that its magnitude is related to the adjustment of dynamic mental representations for syntax that lead to syntactic priming.

ARTICLE HISTORY
Received 31 October 2024
Accepted 5 May 2025

KEYWORDS
Syntactic priming; prediction error; pupillometry; error-based learning; sentence processing

Introduction

One of the biggest challenges in psycholinguistic research is identifying the mechanisms by which humans acquire and process language. Early theoretical approaches were dominated by domain-specific proposals hypothesising language-specific mechanisms (following arguments made by Chomsky, 1959, 1965; and others, Fodor, 1983). However, the nature of language as a human skill instantiated in neural structures that evolved and were co-opted for language and communication implies that neurally plausible domain-general learning mechanisms at least play some determining role (Christiansen & Chater, 2008). In the current paper we consider one such mechanism: learning via *prediction error*. Predictive processing is argued to confer adaptive advantages by allowing the brain to develop, store, and use complex models of the environment rather than simple associations between sensory input and internal states (Clark, 2013). Learning via prediction error is plausibly instantiated in neural networks via recurrent connections and the backpropagation of error (Elman, 1990). That is, sensory input can be compared to previous predictive output and any mismatch converted to an error signal that adjusts the predictive model.

Prediction error plays a central role in neural network models of language, which typically learn via backpropagation of error (Rumelhart et al., 1986). Here we focus on one prominent computational model of language: Chang et al.'s (2006) proposed dual-path connectionist model of language production and syntax acquisition. The model comprises a meaning system, which encodes concepts and their roles within a message, and a separate sequencing system, which is a simple recurrent network that learns syntactic representations that allow it to correctly sequence words. The model makes next-word predictions during sentence comprehension and compares its predictions to the input. In the case of a mismatch, prediction error backpropagates through the system, and weights within the network are adjusted to reduce the likelihood of the error in future predictions, thus updating the model's syntactic representations. These same syntactic representations are used during sentence production, wherein the prediction mechanisms are used to incrementally output words to produce a grammatical sentence that accurately communicates the intended meaning. Therefore, the theory explains the acquisition of syntax through error-based learning and syntactic processing in the

CONTACT Shanthi Kumarage shanthi.kumarage@unimelb.edu.au School of Languages and Linguistics, Babel (Building 139), The University of Melbourne, Parkville 3010 VIC Australia

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23273798.2025.2506634.

services of production. Importantly, the theory connects language acquisition to adult language use: the error-based learning mechanism responsible for the learning of syntactic knowledge continues to operate in adulthood, albeit with a lower learning rate, suggesting that language use continues to modify syntactic representations across the lifespan via implicit learning (Dell & Chang, 2014).

Error-based learning in syntactic priming

Error-based learning has been invoked to explain *syntactic priming* – the tendency to persist in the use of a structure after previously hearing or using it (Bock, 1986). For example, hearing a passive prime, *the swimmer was eaten by the crocodile*, increases the likelihood of later describing a target using a passive, *the cyclist was swooped by the magpie*, rather than an active, *the magpie swooped the cyclist*. Priming independent of shared lexical content (both content words: Bock, 1986, and functional words: Bock, 1989; Ferreira, 2003; Hartsuiker et al., 1999), thematic structure (Bock et al., 1992), and prosody (Bock & Loebell, 1990) suggests that priming occurs at the level of representation of abstract syntactic structure (but see Ziegler et al., 2019). Evidence that priming reflects learning comes from observations of priming effects that endure beyond intervals that might otherwise be attributable to more transient activation mechanisms (Bock & Griffin, 2000; Kaschak et al., 2011).

In Chang et al.'s (2006) model, next-word prediction occurs during prime sentences. One layer of the model receives both the previously predicted word and perceived word, with the difference between them comprising the error signal. For example, hearing "*the cyclist was …*" may preference predictions such as ***riding*** *to work* or ***avoiding*** *the magpie* based on the sequence of words so far and the model's experience with English's canonical agent-first word order (Roland et al., 2007). Hearing a past rather than a present participle indicates that the sentence is the less frequent passive structure (*the cyclist was* ***swooped*** *by the magpie*). The network uses the error signal to adjust network weights, increasing the likelihood that a passive will be used in the future, thus leading to priming.

Therefore, a key prediction of error-based learning accounts, such as Chang et al. (2006), is that more surprising input produces greater prediction error and therefore greater adjustment of syntactic representations. Syntactic priming research provides evidence for this proposal. For example, priming effects are larger for less frequent structures than more frequent ones. Of the prepositional dative and double-object dative, whichever is less frequent is primed more strongly (the double object in Dutch: Bernolet & Hartsuiker, 2010; and the prepositional object in English: Jaeger & Snider, 2013; Kaschak et al., 2011). This *inverse-frequency effect* extends to the active-passive alternation (Bock, 1986), the mention or omission of the *that* complementiser (Ferreira, 2003), and relative clause attachment (Scheepers, 2003). Another way of manipulating the predictability of primes is by leveraging verb biases. Hearing *Bob threw Wendy a ball* should be more surprising than hearing *Bob gave Wendy a present* because *throw* occurs more frequently in the prepositional-object dative whilst *give* prefers the double object dative. Accordingly, researchers have observed *prime-surprisal effects*: stronger priming effects when the structure of a prime mismatches with the structure preferred by the verb in the prime sentence (Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2013; Peter et al., 2015; see also Fazekas et al., 2020).

Inverse frequency and prime-surprisal effects provide strong but indirect evidence for the role of prediction error in syntactic processing. They rely on the assumption that participants' language input closely matches the input sampled in corpora from which surprisal and frequency are calculated. A key challenge is getting measurable indices of cognitive processes like surprisal at the level of the individual (Kidd et al., 2018), which demonstrate that participants do experience the prediction error that experiments aim to induce. More direct evidence would consist of manipulating prime surprisal to demonstrate effects on an index of prediction error and linking that measure of prediction error to priming. To our knowledge, only two studies have attempted to do so. Arai and Chang (2024) investigated priming of the active/passive alternation in Japanese. They utilised the ambiguous case-marker *ni*, which marks both dative case in a sentence like *the boy talked* ***to his friend*** and the oblique argument in a passive: *the boy was hit* ***by his friend***. Since Japanese is a verb-final language, the ambiguity is not resolved until the verb. The words *boy-NOM* ***friend-ni*** could be interpreted as either *the boy* ***to his friend*** __ or *the boy was* ***by his friend*** __ until the verb is encountered. The authors manipulated participants' expectations of encountering a passive through the content of fillers, which either biased the interpretation of preverbal arguments towards a dative case-marked noun (*to his friend*) or not. Participants who were biased to a dative interpretation showed longer reading times on the verb in passive primes than those who were not, an indicator that they experienced greater prediction error. However, although priming was also larger when passives were less expected, longer reading time on passive prime verbs only numerically predicted participant's priming on the target and, surprisingly, longer reading time on active primes appeared to predict passive priming. In a re-analysis of a comprehension priming study of reduced relative clause structures (Tooley, 2020), Tooley and Brehm (2025) investigated whether reading time on the prime sentence predicted reading time on the target sentence. They found that shorter reading times on the prime sentence predicted shorter reading time on the target sentence. Assuming surprisal on the prime sentence results in a longer reading time, these results do not support an error-based learning account of priming.

One problem with using reading time as a measure of prediction error is that it is likely to involve many interacting processes that may obscure reliable measurement of prediction error at the level of the individual (Frinsel & Christiansen, 2024; Staub, 2021). Additionally, reading times typically speed up over the course of experiments (Prasad & Linzen, 2021), which could serve as a confound when using reading time as a predictor. In this study we explore a promising potential measure of prediction error that is more implicit and does not require deliberative action on behalf of the participant: pupil diameter.

Pupillometry

Pupil diameter is strongly associated with ambient light levels but also shows small but detectable changes due to cognitive processing (Sirois & Brisson, 2014). Specifically, pupil diameter can index emotional arousal, mental effort, top-down processes, and surprisal (Hepach & Westermann, 2016). Accordingly, pupillometry has been applied to a variety of questions concerning linguistic processing. Research on word recognition under different levels of noise, speech intelligibility, and speech rates has found that larger pupil size indexes cognitive effort under different listening conditions (Koch & Janse, 2016; Kramer et al., 2013; Kuchinsky et al., 2013; Zekveld et al., 2010). During sentence processing, cognitive effort induced by both ambiguity of reference (Vogelzang et al., 2016) and low semantic predictability (Winn, 2016) increases the pupil dilation response. Pupil dilation is also greater when the prosody of a sentence is incongruent with its information focus or syntactic structure (Engelhardt et al., 2010; Zellin et al., 2011; but see Aydın, 2023 for contradictory results) or when a word is semantically incongruent (Demberg & Sayeed, 2016; Häuser et al., 2018). In syntactic processing, pupil dilation has been observed in response to syntactic violations such as case marking (Aydın, 2023) and gender agreement (Demberg & Sayeed, 2016), but also to variations in syntactic complexity. Schluroff (1982) ranked a range of structures in English based on their Yngve depth (degree of left-branching; Yngve, 1960) and found that the magnitude of pupil dilation correlated with their syntactic complexity. Similarly, Stanners et al. (1972) found a complexity effect for structures with the same surface structure but which differed in thematic role assignment (*they are eager to please* vs. *they are easy to please*). Additionally, complexity effects have been found for several structures, including subject vs. object relative clauses in both written (Demberg & Sayeed, 2016; Just & Carpenter, 1993) and auditory comprehension (Demberg & Sayeed, 2016; Piquado et al., 2010), *wh*-phrases vs *whether* clauses (Just & Carpenter, 1993), affirmative and active sentences vs. negative and passive (Beatty, 1982), and SVO vs. OSV word order in Danish (Wendt et al., 2016).

Thus far, research in language processing has typically interpreted the pupillary response as a measure of cognitive load or mental effort. However, many findings could also be interpreted as pupil size indexing prediction error or surprisal. Incongruencies between a sentences's syntax or semantics and its prosody (Engelhardt et al., 2010; Zellin et al., 2011) and in the semantic fit of a word within a sentence (Demberg & Sayeed, 2016; Häuser et al., 2018) represent violations of expectations. The same is true for syntactic violations. In Aydın (2023), Turkish speakers were presented with SVO transitive sentences (*the boy-NOM painted the desk*), in which the sentence-final object noun phrase occurred in either the grammatical accusative case (*desk-ACC*) or the unexpected and ungrammatical dative case (*desk-DAT*). Similarly, Demberg and Sayeed (2016) presented German speakers with sentences where the final noun was expected given the gender marking of the previous determiner and adjective (*Simone had a-MASC horrible-MASC dream-MASC*) or unexpected and mismatching (*Simone had a-FEM horrible-FEM dream-MASC*). In studies of complexity effects for syntactic structures, participants' accumulated syntactic knowledge would lead them to predict canonical or more frequent word orders over noncanonical ones. For instance, since Danish is an SVO language, speakers presumably expect this more frequent pattern in comparison to OSV (Wendt et al., 2016). For relative clauses, the well documented subject advantage means that English speakers expect a subject RC over an object RC (Just & Carpenter, 1993; MacDonald, 2013; Piquado et al., 2010). Relatedly, semantic knowledge would lead to high context sentences (*Bill stirred his coffee with a … spoon*) being more predictable than low context ones (*Jamie thought about a … spoon*; Winn, 2016). The suggestion is that pupil dilation may be an automated and implicit index of prediction error during language processing.

While this link has not been made explicitly in psycholinguistic research, outside of the field the suggestion that pupil dilation implicitly indexes prediction error is theoretically and empirically well-established. Theoretically, there is a neurobiological explanation for the link between pupil dilation and prediction error. Alongside the P3 event-related potential (ERP), the pupil dilation response is argued to reflect transient (phasic) noradrenergic activity in the locus coeruleus (LC; Gilzenrat et al., 2010; Murphy et al., 2011; Nieuwenhuis et al., 2005). Noradrenaline (NE) has neuromodulatory effects, increasing the responsivity of neurons to their inputs, and the LC projects to cortical processing areas (Aston-Jones & Cohen, 2005). Therefore, the prominent *adaptive gain theory* of LC-NE activity proposes that its function is to optimise behavioural responses to motivationally significant stimuli by increasing neuronal gain in relevant subsequent processing areas (Aston-Jones & Cohen, 2005). A similar and complementary argument applies to unexpected or infrequent stimuli: surprisal indicates the need for updating an internal model of the environment, with the phasic LC-NE response supporting this updating through enhanced neuronal gain in relevant processing areas (Dayan & Yu, 2006; Nieuwenhuis, 2011).

Empirically, an inverse association between the magnitude of the pupil response and the probability of stimuli has long been known (Friedman et al., 1973; Qiyuan et al., 1985). Further, the pupil response is contingent on participants' expectations: pupil dilation is greater when an expected reward or loss does not occur in a gambling task (Lavin et al., 2014; Preuschoff et al., 2011), when participants perform well on a difficult task or poorly on an easy one (Braem et al., 2015), and when an object moves counter to its expected trajectory or appears in an unexpected location (Harris et al., 2022; O'Reilly et al., 2013). In studies of sequence or cue learning, the probabilistic rules that govern participants' expectations are learnt during the task. Both less frequent transitions between stimuli (Alamia et al., 2019; Rutar et al., 2023) and omissions of expected stimuli following cues (Zhang et al., 2019) are associated with larger pupil responses. O'Reilly et al. (2013) note that prediction error and model-updating are two highly correlated, often experimentally confounded, but distinct cognitive events. Nassar et al. (2012) were able to measure each separately using a predictive inference task where participants predicted a series of numbers whose mean changed at random intervals throughout the experiment. They demonstrated that the magnitude of pupil size correlated with the difference between the predicted and presented number (i.e. prediction error), but also with the degree of change in participants' predictions for the next trial (i.e. model updating).

Cumulative effects in syntactic priming

Overall, the non-linguistic research explicitly links the pupil response to prediction error. In the present study, we combine pupillometry with syntactic priming. Under the assumption that priming is the outcome of the updating of syntactic representations driven by prediction error, we expect larger pupil dilation during prime trials to predict priming. However, this prediction must be qualified by the fact that such an effect is likely to be dynamic; that is, it may be moderated by cumulative effects across the course of the experiment. This is because, in syntactic priming studies, infrequent structures (e.g. the passive) are presented at a much higher probability than in natural language. If syntactic priming reflects implicit learning involving model updating, the effects of primes should accumulate, with additional encounters with a syntactic structure resulting in compounding updates in expectations and representations. That is, participants' expectations are not static, but change based on input.

Evidence supporting this assertion comes from comprehension and production priming studies. Several studies of syntactic priming in comprehension have observed *expectation adaptation*, where processing deficits associated with temporarily ambiguous structures (garden-path sentences) diminish as a function of the number of structures previously encountered (Farmer et al., 2014; Fine et al., 2013; Fine & Jaeger, 2016). Whilst not all studies have found that ambiguity effects change over the course of the experiment (Dempsey et al., 2020; Harrington Stack et al., 2018; Tooley et al., 2014), that may be a methodological limitation, as expectation adaptation studies have relied on self-paced reading. Prasad and Linzen (2021) suggest this method may be unsuitable for investigating adaptation effects because differences in reading times between structures become more difficult to detect as participants get faster at the task overall. In production priming, the "current run", or uninterrupted number of a structure (e.g. passive) in a row, including both primes encountered and targets produced, predicts priming (Jaeger & Snider, 2013; Kaschak et al., 2011). However, when using the total accumulated exposure to the structure across the experiment, the effect on priming depends on the structure investigated and whether the number of encounters or productions is used (Bernolet et al., 2016).

There are two points to consider in the observation of cumulative effects between comprehension and production. Firstly, implicit learning processes may affect production and comprehension measures differently. Prediction error will be highest at the beginning of the experiment, decreasing as more instances of the target structure are encountered. In comprehension, measures of processing difficulty due to reanalysis, such as reading time, or of prediction error, such as pupil dilation, should therefore decrease over time. However, although lower prediction error should result in smaller model updates at the end of an experiment, these are unlikely to correlate with decreased *production* of the structure. As updates to the baseline frequency of a structure are cumulative, the probability of producing the target structure should instead remain elevated or increase. Overall, this means that cumulative exposure to the target structure may weaken the association between comprehension measures and the probability of producing the target structure. Both previous studies investigating the influence of prime processing on target processing in syntactic priming may have been impacted by cumulative effects. They did not find that longer prime reading times were associated with facilitated target comprehension (Tooley & Brehm, 2025) or production (Arai & Chang, 2024). Arai and Chang (2024) point out that this could be because both shorter reading times and higher target production are expected later rather than earlier in the experiment given cumulative input. Unfortunately, their analyses involving trial were unable to converge to confirm this possibility.

A second issue is that comprehension and production priming studies have typically examined different types of structures: reduced relative clauses and main verb structures vs the dative and active/passive alternations (Arai et al., 2007; Tooley, 2023; Tooley & Bock, 2014). The present study addresses both of these considerations by including both a measure of production and a measure of prediction error associated with comprehension of the same active and passive structures. This allows the investigation of the time course of syntactic priming effects in both comprehension and production simultaneously.

The current study

The present study aimed to determine whether prediction error in syntactic processing can be directly measured by combining the syntactic priming and pupillometric methodologies. We test this using the active/passive alternation in English. The English passive is low in frequency compared to the active: there are 1.3–3.2 occurrences of the passive per 1000 words in spoken English, while actives are about 10 times more frequent (Roland et al., 2007; Xiao et al., 2006). Full passives including a by-phrase are even less frequent, making up only 2.5–5% of passives in spoken English (Xiao et al., 2006). The passive's patient-verb-agent thematic role order is non-canonical: in spoken English, patients occur before their verbs at most 5% of the time (Roland et al., 2007). Experimentally, this word order is even less likely for events with two animate arguments, as in our experimental materials, rather than events with an animacy contrast (Ferreira, 1994). All these features serve to make the passive, particularly as presented in our experiment, a highly unexpected structure compared to the active and an ideal test-case for our study.

Participants were presented with primes in the active or passive structure followed by a target picture of a transitive action to describe. We measured pupil diameter during prime sentences and recorded participants' productions in target sentences. We expected that passive primes would be associated with both a larger pupil response (in line with syntactic complexity effects, e.g. Just & Carpenter, 1993) and with greater production of passives (in line with syntactic priming effects, e.g. Bock, 1986). Crucially, we tested whether greater prediction error (i.e. the expected pupil dilation difference between actives and passives) increased participants' likelihood of being primed, as predicted by error-based learning accounts of syntactic processing (Chang et al., 2006). We also investigated cumulative effects in both the comprehension and production of passives. We tested whether there were changes in the magnitude of prediction error, syntactic priming, and the relation between them over the course of the experiment. We expected that the pupil response to passive primes would reduce over the course of the experiment as participants adapt their expectations to their higher frequency in the task (Fine et al., 2013). On the other hand, production of passives may increase over the task with greater cumulative exposure to passives (Kaschak et al., 2011). As such, passive production may become less tied to the surprisal induced by primes as participants increase their overall expected baseline frequency of passives.

Methods

Participants

Eighty individuals who spoke Australian English as a first language, were aged 18–35, with no history of developmental or acquired language disorder, and normal or corrected-to-normal vision were recruited from the

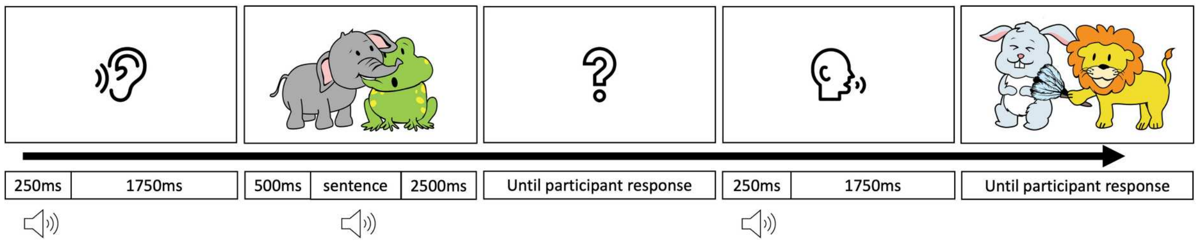

Figure 1. Structure of trials in the experiment.

Notes: This figure depicts the time course of a prime-target pair in the experiment. Participants viewed a 2000 ms fixation symbol accompanied by a 250 ms beep to indicate the type of trial. The prime picture appeared for 500 ms before the onset of the prime sentence and stayed on screen for 2500 ms after sentence offset. A question symbol appeared until participants indicated whether the picture matched the sentence audio with a keypress. This initiated the fixation symbol for a target sentence which was followed by a target picture that stayed on screen until the participant had described it and the experimenter progressed the task to the next trial.

Australian National University community. An additional three participants were tested but were not included because they did not meet the eligibility criteria. Participants received a 1-hour course credit or AUD$15 as compensation for their participation. The study was approved by the ANU Science and Medicine Delegated Ethical Review Committee (reference: 2022/710). The final sample ($N = 80$: $F = 53$, $M = 23$, NB = 3, undisclosed = 1) ranged in age from 18 to 33 years ($M = 22.55$, $SD = 4.29$). All participants used English for the majority of their interactions in an average week ($M = 98.1\%$, $SD = 5.6$), with 33 reporting knowing or using a language other than English. Sample size was estimated according to Mahowald et al.'s (2016) power analysis for syntactic priming effects. A sample of 80 participants and 32 items has sufficient power (>80%) to detect a large interaction effect (of the size of the lexical boost effect).[1]

Materials and design

Materials consisted of 32 prime pictures based on 8 transitive verbs (*bite, catch, carry, kick, pinch, push, kiss, lick*) and 32 target pictures based on a different set of 8 transitive verbs (*chase, drag, hit, prick, punch, shoot, tickle, feed*; see Figure 1 for example pictures). To control for animacy effects, both the agent and patient in each picture were drawn from a pool of 16 animate characters (*bear, cat, chicken, cow, dog, duck, elephant, frog, goat, horse, lion, monkey, mouse, pig, rabbit, turtle*), with each character occurring as agent and patient equally often. In addition, there were 128 filler pictures, depicting objects that could be described by a noun phrase (*a red apple*), more complex scenes that could be described using spatial prepositions (*the cups are on the chair*), single participant events that could be described by an intransitive construction (*the cow is sleeping*; *the dog is strong*), and three participant events that could be described by a dative construction (*the cow gives the chicken a present*). Most pictures were taken from Garcia et al. (2021, 2023), with some additional pictures drawn by the same artist to fulfil the requirements of this study. Experimental pictures were 800×450 pixels. All pictures were resized to 1920×1200 by padding them with white background. Experimental pictures were standardised to a mean luminance of 249.38 in the HSV colourspace (scale 0–255) using the *lumMatch* function from the *SHINE_color* MATLAB toolbox (Dal Ben, 2021; Willenbockel et al., 2010). We recorded audio descriptions of each prime and half the filler pictures by a female native speaker of Australian English. For experimental items both an active (*the pig is catching the cat*) and a passive description (*the cat is being caught by the pig*) were recorded. These descriptions were recorded in the present progressive form to avoid an adjectival interpretation. The recorded description did not match the picture for 19 filler items (e.g. *the fork is on the table* for a picture depicting *the scissors are in the box*).

The task consisted of 32 prime-target pairs, with prime (active or passive structure) manipulated within subjects. Prime and target contained no open class lexical overlap. No more than two primes of the same structure occurred in a row. Between 2 and 6 filler items intervened between each prime-target pair. Half the fillers were prime trials and half were target trials, with pseudorandom ordering of trials such that no more than three trials of the same type occurred in a row. The purpose of variable trial type and spacing of prime-target pairs was to mask the aims of the study from participants. We constructed 16 experimental lists counterbalancing the verb pairings in prime-target pairs, prime structure, and the direction of the action in prime and target pictures (LR or RL).

Apparatus

Stimuli were presented using Tobii Pro Lab (version 1.207.44884) software on the T60 eyetracker, which measured pupil diameter in mm at 60 Hz. Participants completed a 9-point calibration and validation procedure at the beginning of each of the three blocks of trials (mean validation accuracy 0.51 degrees). A Zoom H2n audio recording device was connected to the computer so the Tobii software automatically audio-recorded the task.

Procedure

Participants were seated in a dimly lit room (30 lux) and first completed a demographics questionnaire. They were then introduced to the three fixation symbols for listening, answering questions, and speaking (see Figure 1) and instructed to respond accordingly. On seeing the listening symbol, they were instructed to listen carefully to the upcoming sentence and pay attention to whether it matched the accompanying picture. When the question symbol appeared, participants responded with whether the picture matched (green button) or didn't match (red button) on a keypad. On seeing the speaking symbol, they were instructed to describe the upcoming picture. The instructions encouraged participants to respond with full sentences where possible and to remember the button locations rather than looking down to respond. There were six practice items, then three blocks of trials.

Figure 1 depicts the time course of a prime-target pair in the experiment. Both prime and target trials were preceded by a 2000 ms fixation symbol accompanied by a 250 ms beep, indicating to participants the start of a new trial and the type of trial. This also allowed extra time for the pupil to return to baseline following the previous trial (Mathôt & Vilotijević, 2022). The picture then appeared on screen.

In prime trials, we presented the picture for 500 ms before sentence onset to allow event apprehension to occur (Griffin & Bock, 2000). Therefore, participants could identify the event as transitive and form an expectation that the agent would occur first. We expected prediction error for passive primes to be induced in the very first part of the sentence, on hearing the first noun phrase. The peak pupil dilation typically occurs 1-1.5s from the point of difficulty (Just & Carpenter, 1993; Tromp et al., 2016). Recorded sentences varied from 1414 to 2147 ms in duration and were followed by 2500 ms of silence, which allowed sufficient time for the pupillary response to be observed. The picture remained on screen during this time so that the pupil response could be measured without changes in luminance caused by switching to a fixation symbol. Because the pupil response is typically task-evoked (see Zellin et al., 2011, p. 136), we introduced a picture-verification task on each prime trial. When a question symbol appeared after picture offset, participants responded by pressing a green button if the picture and audio description matched and a red button if they did not. The task automatically proceeded to the next trial after the participant responded. As participants needed to wait until the question symbol appeared to make a response, we minimised any cognitive processing associated with actually making the response (e.g. motor planning etc.) during the measurement of pupil diameter. About 20% of filler prime trials had mismatching audio descriptions to maintain attention to the task. On average, participants answered these attention check items correctly 97% of the time ($M = 0.97$, $SD = 0.06$, range: 0.68–1.00).

Target trials more closely resembled typical syntactic priming experiments. The picture appeared after the fixation symbol and remained on screen until the participant had described it and the experimenter progressed the task to the next trial.

Transcription and coding

Participants' responses were transcribed from the audio recording and scored as *active*, *passive* or *other*. Thirty trials were excluded due to recording equipment failure. If participants produced more than one sentence, only the first complete sentence was coded. If participants corrected themselves before producing a full sentence, the corrected form of the utterance was scored. Responses were scored as active if they contained an agent in the subject position, an appropriate transitive verb, and a patient in the object position and could be expressed in the alternate passive structure (e.g. *the elephant feeds the lion*). Passive responses needed to contain a patient in the subject position, an auxiliary verb (*was, got*) and appropriate transitive main verb, an agent in a by-phrase and be expressible in the alternate active structure (e.g. *the goat is being chased by the horse*). Transitive responses where the participant repeated the verb or a noun contained in the prime sentence were excluded. Other responses consisted of all other sentence forms, including datives (*the horse is feeding the duck some food*), intransitives (*the goat and the horse are walking*), and irreversible phrasal verbs (*a mouse running away from a chicken*). Overall, 90.2% of participants' sentences could be coded as Active or Passive.

Pupil data preprocessing

We performed data preprocessing and our analyses in R (version 4.2.1; R Core Team, 2022). The Tobii T60 eyetracker measures pupil size for both eyes. As recommended by Sirois and Brisson (2014), pupil size was regressed from each eye onto the other, thus imputing values missing from only one eye. Then, average pupil size across both eyes was calculated. Average pupil size was passed through an 11 sample Hanning filter using the *filter_data* function from the *PupillometryR* package (version 0.0.5; Forbes, 2020) to provide a smooth signal for blink detection. Blinks were detected using a velocity filter similar to the method described by Mathôt (2013). Due to the slower 60 Hz sampling frequency, our data typically showed a fast negative velocity before blinks but not the expected rapid increase in velocity following blinks. Therefore, we counted blinks as any velocity larger than − 0.02 mm change in the filtered data preceding a period of track loss. Blinks were extended to 50 ms on either side of the gap with no imputation of missing values (since generalised additive mixed models can handle missing data: van Rij et al., 2019). Filtered data was used for deciding which samples to remove during blink detection but we returned to the raw data for further preprocessing and analysis. We applied a velocity filter using a cut off of 0.15 mm change in pupil size between each sample, removing one sample on either side as well. Visual inspection of each trial for each participant showed that obvious outliers were removed without the exclusion of steep curves that made up the pupil response. Any samples where gaze position was outside the area of the screen were removed. After removing blinks, velocity outliers and gaze position outliers, we excluded trials with more than 25% of data removed in preprocessing and/or missing due to track loss. Baseline pupil size was calculated as the average pupil size in the 250 ms preceding sentence onset (i.e. 250–500 ms after picture onset). Trials where the baseline could not be calculated were excluded, as were trials where the baseline was more than two standard deviations away from the participant's mean baseline. The baseline was subtracted from pupil size to calculate baseline-corrected pupil diameter. Trials were cut to 3767 ms, the length of the shortest trial. In total 78.8% of trials ($N = 2016$) were included after excluding trials due to recording failure ($N = 9$), high percentage of missing or removed data ($N = 416$), and missing or improbable baseline values ($N = 119$).

Results

Syntactic complexity effect

Our main aim was to see whether prediction error, as indexed by pupil dilation, predicts priming. Therefore, we first analysed the time course of pupil dilation for active primes and passive primes. We hypothesised that passives should induce a greater pupil response than actives due to their low frequency, especially for the types of events depicted in our stimuli.

Following recent recommendations, we fit a generalised additive mixed model to our pupillometric data (GAMM; van Rij et al., 2019; Wieling, 2018; Wood, 2017). This analysis approach confers several benefits in analysing time series data with high variability between trials (van Rij et al., 2019). Firstly, GAMMs allow the modelling of non-linear relationships, such as that between time and pupil size. In contrast, extracting features of the pupil response, such as peak or latency, simplifies the analysis of the non-linear pupil response but results in the exclusion of trials where there is no clear peak. Secondly, GAMMs make it possible to control for autocorrelation between samples in time series data, therefore decreasing the likelihood of Type 1 error due to autocorrelated errors. Finally, the inclusion of random effects for items and subjects accounts for variability within and between subjects. We ran our analyses using the *mgcv* (version 1.9.1; Wood, 2011) and *itsadug* (version 2.4.1; van Rij et al., 2022) R packages and closely followed the procedures outlined by van Rij et al. (2019).

GAMMs fit non-linear regression lines (smooths) over a continuous predictor. When including a factor variable, the model estimates as many smooths as levels of the factor (e.g. one for each condition in the experiment). For a detailed comparison of approaches to coding factor variables in GAMMs see Wieling (2018). Briefly, factor variables can be coded such that a smooth is estimated for each condition (factor coding) or such that a reference smooth and difference smooth are estimated. The advantage of the latter approach is that because model estimates indicate whether smooths are significantly different from 0, the difference smooth indicates whether the difference between conditions is significant. Under this approach, the smooths can be non-centred (binary difference coding), therefore including the intercept difference between conditions in the difference smooth, or centred (ordered factor coding), which allows the difference to be attributed to either an intercept difference or non-linear difference, or both. We chose ordered factor coding as it is the most informative given our design.

Our GAMM modelled pupil size over time, estimating the effect of prime condition (*active*: 0 or *passive*: 1) as an ordered factor. The effect of condition was modelled with two parametric terms: an intercept term, and an intercept difference term (*Condition*); and two non-linear terms: a reference smooth (for the *active*

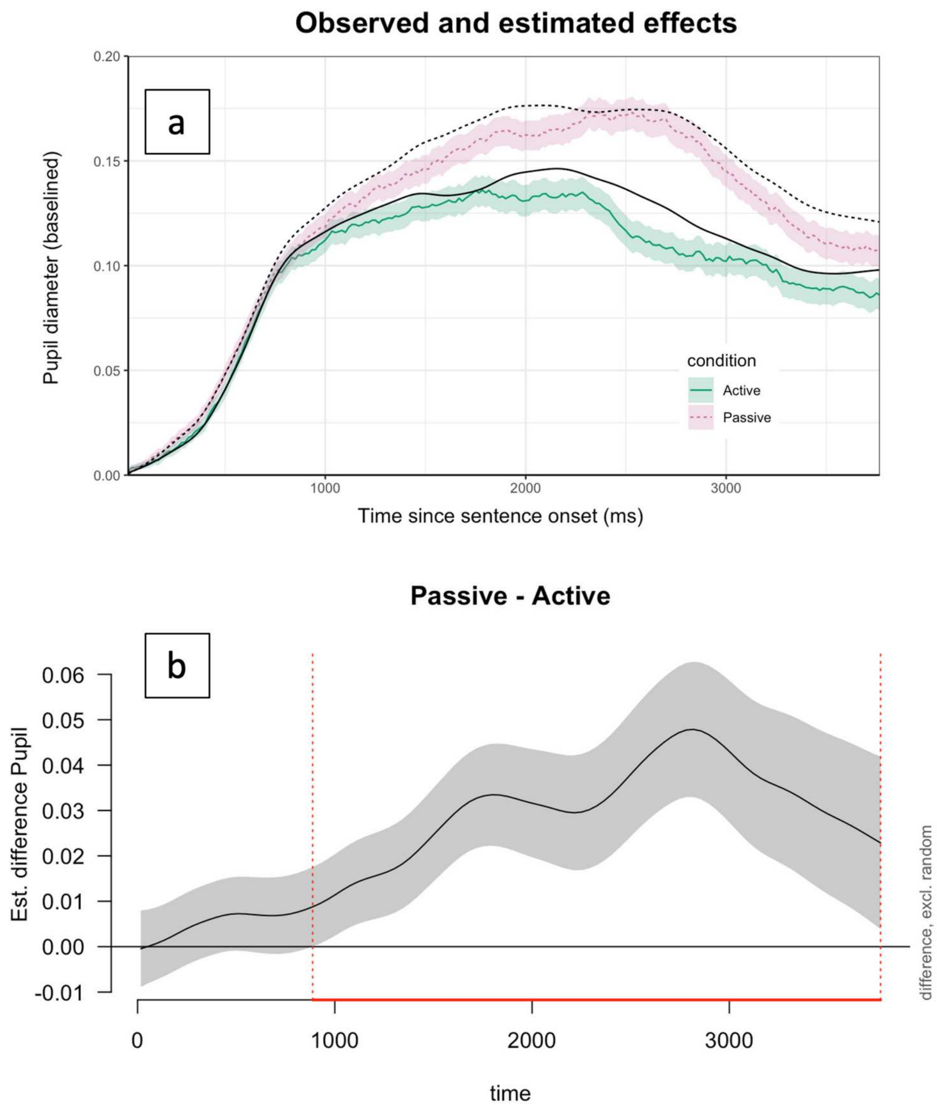

Figure 2. Model-fitted effects on pupillary time course including (a) fitted effects for active and passive primes and (b) fitted difference smooth between conditions.

Notes: Coloured lines and shading representing standard error in (a) reflect observed pupil diameter and black lines reflect model fitted pupil diameter. Shading in (b) represents 95% confidence interval around the model-estimated difference in pupil diameter between conditions. Model-fitted effects are estimated at the median of other predictors in the model (i.e. gaze position) and excluding the contribution of random effects.

condition; *Time*) and a difference smooth (for the *passive* compared to *active* condition; *Time:Condition*). The angle of the eye at different gaze positions can systematically distort the measurement of pupil size (pupil foreshortening error: Brisson et al., 2013). Therefore, we controlled for gaze position by including a non-linear interaction between the X and Y coordinates of gaze position (*GazeX.GazeY*).

For a detailed discussion of random effects structures in GAMMs, we refer readers to van Rij et al. (2019). A full random effects structure includes a random smooth (non-linear regression line) for each "event" (i.e. each item for each participant; we use the term *event* instead of *trial* to avoid later confusion with trial number). We could not include random smooths for each event ($N = 2016$) due to the high computational demands of doing so. We instead included random slopes and intercepts for each event, and random smooths by participant and item nested in verb (as recommended by van Rij et al., 2019). We performed model diagnostics using the *gam.check* function and an autocorrelation function (ACF) plot. Based on these, we increased the number of basis dimensions for smooths (*k*) as necessary, fit the model with a scaled-*t* distribution to meet the assumption of normally distributed residuals, and added an AR-1 model to account for autocorrelation of residuals using the lag 1 correlation value of $\rho = .939$ (van Rij et al., 2019).

The model showed that the pupil response was significantly larger following passive prime than active

Table 1. Summary of GAMM modelling pupillary response.

	β	SE	*t*	*p*
Parametric coefficients				
Intercept	0.10	0.01	7.88	<.001
Condition	0.03	0.01	4.44	<.001
	edf	Ref.df	F	*p*
Smooth terms (fixed effects)				
Time	27.17	32.68	8.60	<.001
Time:Condition	15.14	19.79	7.78	<.001
GazeX,GazeY	170.62	192.57	8.70	<.001
Smooth terms (random effects)				
Time, Participant	660.23	719.00	1022.86	<.001
Time, Item	248.60	287.00	387.32	<.001
Event	1405.17	2014.00	88.07	<.001
Time, Event	1724.57	2014.00	126.01	<.001

Table 2. Model summary of GAMM modelling impact of trial on pupillary response.

	β	SE	*t*	*p*
Parametric coefficients				
Intercept	0.10	0.01	7.98	<.001
Condition	0.03	0.01	4.45	<.001
	edf	Ref.df	F	*p*
Smooth terms (fixed effects)				
Time	26.93	32.45	8.66	<.001
Time:Condition	15.80	20.59	7.75	<.001
Trial	1.70	1.75	12.57	<.001
Trial:Condition	1.56	1.59	0.19	0.73
Time:Trial	11.52	13.12	12.93	<.001
Time:Trial:Condition	11.73	13.61	2.59	<.001
GazeX,GazeY	170.62	192.57	8.70	<.001
Smooth terms (random effects)				
Time, Participant	623.24	719.00	84.01	<.001
Time, Item	247.93	287.00	270.37	<.001
Trial, Participant	98.89	700.00	8.99	<.01
Event	1358.75	2012.00	63.31	<.001
Time, Event	1717.48	2012.00	93.22	<.001

prime sentences. Interpreting significance in GAMMs requires a combination of visualisation and interpreting model summary statistics. Figure 2(a) plots the observed and model fitted baselined pupil diameter for active and passive prime sentences. The pupil response appears larger following passive than active primes, a finding that is statistically confirmed in the model summary (Table 1) and model-fitted difference curve (Figure 2(b)). The intercept difference (parametric effect for Condition) shows that pupil size is significantly higher in passive than active sentences, $\beta = 0.03$, $t = 4.44$, $p < .001$. For smooths, the estimated degrees of freedom (edf) indicates how wiggly the regression line is, with lower edfs indicating a smoother line. The reference degrees of freedom (Ref.df) indicates the degrees of freedom associated with the *F*-test of significance. A significant *p*-value indicates that the regression line is significantly different from zero. Because we explicitly modelled the difference between conditions (Time:Condition), we can conclude that the time course of the pupil response is significantly different between active and passive sentences, $F = 7.78$, Ref.df = 19.79, $p < .001$. Although both terms indicate a significant difference between conditions, they do not indicate the time period when this difference occurs. Figure 2(b) plots the model-fitted difference between passive and active primes, combining the intercept and non-linear differences between conditions. The pupillary response to passive sentences is significantly larger than the response to active sentences from about 900 ms after sentence-onset onwards, as indicated by the confidence interval excluding 0.

Cumulative effects on pupil response

Our first analysis established that, as expected, the pupil response was larger when participants heard passive rather than active prime sentences. This finding is in line with our hypothesis that pupil size can index prediction error for infrequent and surprising structures. Our second aim was to investigate cumulative effects in both the comprehension and production of passives during syntactic priming. Here we investigate cumulative effects on the comprehension of passives as measured by the pupil response.

We added several terms to our original analysis of the pupil response in order to include the effect of trial number (coded 1–32). These were: a reference smooth for trial (for the *active* condition; *Trial*), a difference smooth for trial (for the *passive* compared to *active* condition; *Trial:Condition*), a reference tensor product interaction between time and trial (for the *active* condition; *Time:Trial*), and a difference tensor product interaction (for the *passive* compared to *active* condition; *Time: Trial:Condition*). This implements separate main and interaction effects within a GAMM as well as directly modelling whether the interaction effect differed between active and passive primes (i.e. the three-way interaction effect). We also added a random smooth for trial by participant but not by item because each item only occurred at four trial positions across the lists. It was again necessary to increase the basis dimensions for smooths (*k*) above the default, fit the model with a scaled-*t* distribution, and include an AR-1 model. Table 2 reports the summary of the final model and Figure 3 visualises the pupil response over the course of the experiment. The plots on the left are contour plots mapping the difference in pupil size between passive and active primes along the scales of time after sentence onset (*x*-axis) by trial number (*y*-axis). The dotted white line indicates the trial number at which model-predicted active and passive pupil responses are plotted on the right-hand side. The reference and difference smooths for time remain

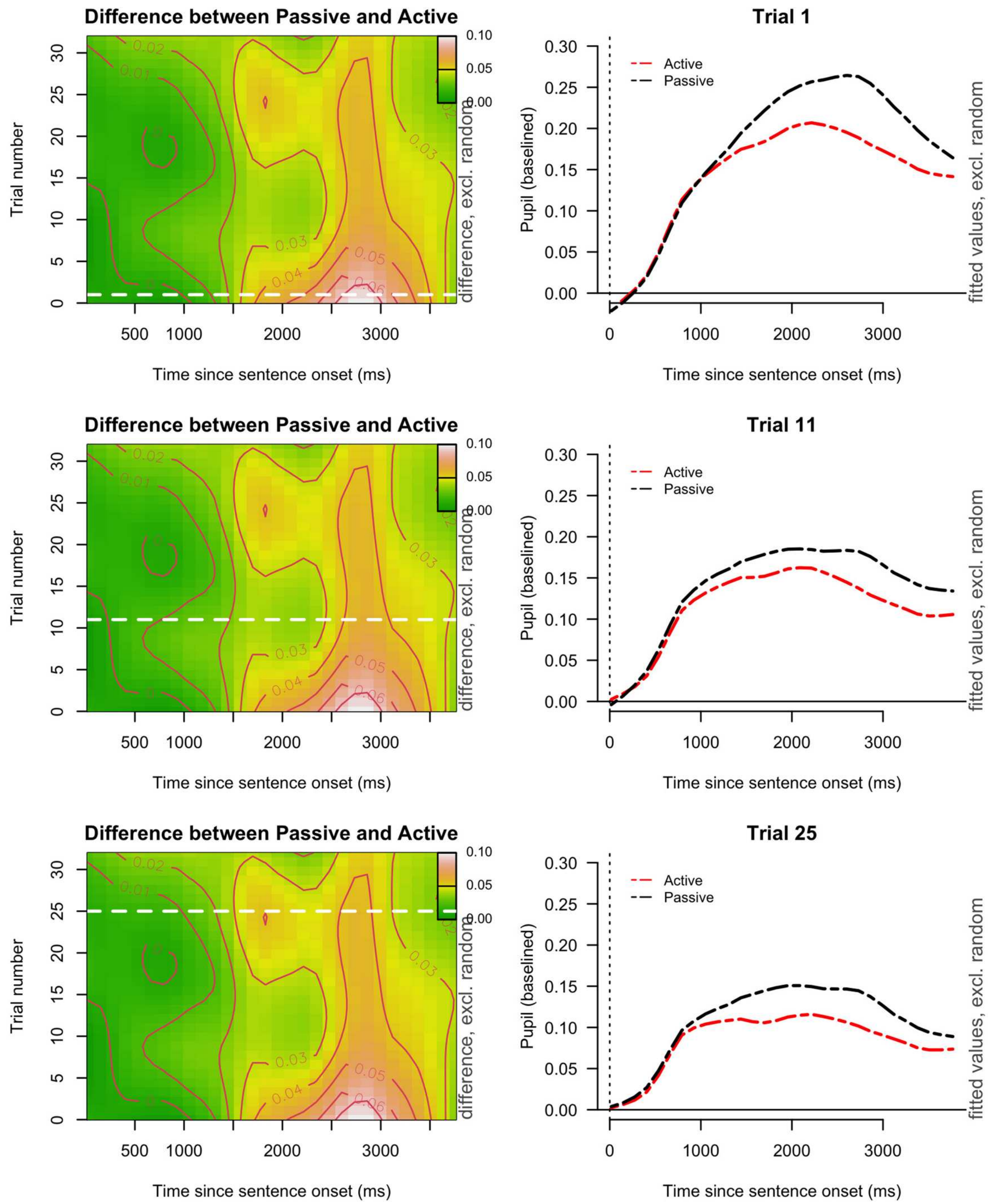

Figure 3. Pupillary response over time and trials.

Notes: Model-fitted effects are estimated at the median of other predictors in the model (i.e. gaze position) and excluding the contribution of random effects.

significant, again indicating a significant pupil response that differs between active and passive primes. The contour plots illustrate that the largest difference between conditions occurs between 2500 and 3000 ms. The reference and difference smooths for trial indicate a near linear main effect of trial (edf close to 1), which does not differ between active and passive primes. Since pupil size is baselined, this reflects an overall decrease in the size of the pupil *response* (not size) over the course of the experiment, as shown in the right-hand side plots in Figure 3. Finally, there is a significant interaction between time and trial, which significantly differs between conditions. The difference between passive and active primes is strongest at the very earliest trials, as indicated by the areas of white (large difference) at the bottom of the contour plots.

The difference between conditions is evident across the pupil response at earlier and later trials but focused between 2500-3000 ms in middle trials.

Predicting syntactic priming using pupillometry

The GAMM analyses confirmed that there was a larger pupil response for passive than active primes, a difference which attenuates as more passives are encountered over the course of the experiment. Assuming our linking hypothesis that the pupil response indexes prediction error, this pattern of results suggests that the infrequent passive elicits greater prediction error that attenuates with cumulative exposure across the experiment. We now turn from participants' comprehension of passives during syntactic priming to their production. Our key hypothesis was that prediction error in processing the prime, as indexed by pupillometry, would predict participants' production of passives after passive primes. A secondary hypothesis was that changed expectations due to cumulative input may moderate the production and priming of passives and the influence of prediction error. In this analysis we tested both these hypotheses.

Operationalising prediction error from the pupil dilation response

The pupil response has typically been investigated as a dependent variable rather than as a predictor (but see Contier et al., 2024). We originally operationalised prediction error as the average model-predicted pupil size across the whole prime sentence, using model predictions from our first GAMM analysis. We chose to use model-predicted rather than measured pupil size because the model controls for gaze position. At the request of an anonymous reviewer, we investigated alternative operationalisations. Contier et al. (2024) used cluster-based permutation to determine an approximate period during which pupil size differs between conditions and averaged pupil size only during that window. We did the same, however we used our first GAMM to estimate the period of significance. GAMMs can estimate the onset and offset of effects whilst cluster-based permutation should not be used to do so (Ito & Knoeferle, 2022; Sassenhagen & Draschkow, 2019). Similarly to Contier et al. (2024), this time window was fairly long: from 888 ms onwards (see Figure 2(b)). Contier et al. (2024) conducted an exploratory analysis further restricting the time window for averaging pupil size to 500 ms around the peak of the pupil dilation response. Similarly, we calculated the peak of the difference between conditions (2817 ms, see Figure 2(b)) and calculated another measure of prediction error averaging pupil size in the 500 ms window around this. In addition, we calculated the average pupil size for these three windows – whole sentence, period of significance only, 500 ms window – for the preprocessed measured pupil size as well as for the model-predicted pupil size. Plots comparing the model estimates for all 6 measures of prediction error are available in the Appendix and the full analyses are available in our online materials. Unlike Contier et al. (2024), who found that the 500 ms window around the peak better predicted their effect of interest, we found negligible differences between model estimates for the three time windows. The pattern of results was also the same regardless of whether model-predicted or measured (but preprocessed) pupil size was used. We therefore retained our original model using average model-predicted pupil size across the whole prime sentence as our operationalisation of prediction error.

Final model

Table 3 summarises the proportion of active, passive and other responses that participants produced in each experimental condition. *Other* responses were excluded from our analyses. We ran a Bayesian mixed effects logistic model to predict participants' production of passives using the *brms* R package (version 2.20.4; Bürkner, 2017). Bayesian models are more likely to converge with complex random effects structures than frequentist ones (Eager & Roy, 2017). The model included prime (effect coded: active: – 0.5, passive: 0.5), the z-scored measure of prediction error, trial (coded 0-31, i.e. shifted so that trial 1 is the reference point), and their interactions as predictors. We included the full random effects structure, with an intercept and slopes for prime, prediction error, trial and their interactions by participants, and an intercept and slopes for prime, prediction error, and their interaction by items nested in verbs. The model was run with 3000 iterations, 1000 of them warm-up, and 10 chains using weakly informative priors (for all fixed coefficients on the logit scale we chose a normal distribution with a mean of 0 and standard deviation of 2). The value of *adapt_delta*, which decreases the step sizes taken by the model, was increased to 0.95 to prevent divergent transitions.

Table 3. Number and proportion of active, passive and other responses in each condition.

Prime	Active		Passive		Other	
	N	%	N	%	N	%
Active	1119	88.5	41	3.2	105	8.3
Passive	1009	79.8	141	11.1	115	9.1

Note: Trials that were excluded due to recording failure are not included in this table (15 per condition).

Table 4. Model summary of model predicting syntactic priming with prediction error and trial number.

Effect	Estimate	95% credible interval	Posterior probability
Intercept	−3.43	−4.13 | − 2.77	>.999
Prime	1.59	0.61 | 2.59	.995
Prediction error	−0.09	−0.56 | 0.37	.623
Trial	−0.03	−0.06 | 0.01	.889
Prime*Prediction error	1.00	0.14 | 1.88	.971
Prime*Trial	−0.01	−0.04 | 0.06	.400
Prediction error*Trial	−0.00	−0.03 | 0.02	.442
Prime*Prediction error*Trial	−0.04	−0.09 | 0.00	.934

Note: The posterior probability that an effect is smaller or larger than zero is calculated in a directional (one-tailed) hypothesis test.

Convergence diagnostics indicated reliable convergence and estimates of posteriors: the maximum Rhat was 1.01, the minimum bulk effective sample size (BESS) was 2247 and the minimum tail effective sample size (TESS) was 5003. A posterior predictive check showed good model fit.

Table 4 reports the model summary statistics. The credible interval indicates the range of values within which the effect has a 95% chance of falling given the data. The posterior probability indicates the chance that an effect falls above or below zero given the data. We interpret a posterior probability of >.95 as strong evidence for an effect given the data, of >.85 as weak evidence for an effect, and of close to .5 as no evidence for an effect (see Engelmann et al., 2019). When interactions are included in the model, effects must be interpreted at the reference level (0) of other variables. We made trial 1 the reference point so that non-cumulative effects establish whether an effect is present at the beginning of the experiment (i.e. trial coded 0–31).

Confirming that we replicated the syntactic priming effect, the effect of prime had strong evidence at the beginning of the experiment for an average level of surprisal (average pupil size z-score = 0). There was also strong evidence for the interaction between prime and prediction error, indicating support for our key hypothesis at trial 1. An examination of conditional effects showed that larger prediction error increased the likelihood of producing a passive after a passive prime (β = 0.41 [−0.08, 0.90], posterior prob. = .917) but decreased the likelihood of producing a passive after an active prime (β = −0.59 [−1.36, 0.16], posterior prob. = .902). In sum, the more surprising participants found the prime, the more likely they were to reuse its structure (i.e. be primed).

We now turn to the model terms involving trial, that is, cumulative effects over the course of the experiment. There was weak evidence for the main effect of trial, indicating that on average passive production is estimated to slightly decrease over the course of the experiment, if we hold prediction error at its average value (pupil size z-score = 0). There was no evidence for the prime by trial interaction effect, suggesting that the priming effect was stable over the experiment, again at the average value of prediction error. Both these effects should be interpreted in light of the three-way interaction effect, which received weak evidence. Figure 4 plots the model-predicted production of passives at various values of prediction error following active and passive primes at trials 1, 11, and 25 (the same trials plotted in Figure 3). The first panel (trial 1) reflects the prime by prediction error interaction term in our model. At the beginning of the experiment, greater pupil dilation induced by the prime corresponded to a stronger priming effect. Larger prediction errors following an active prime were associated with higher production of actives (i.e. in Figure 4, greater values of

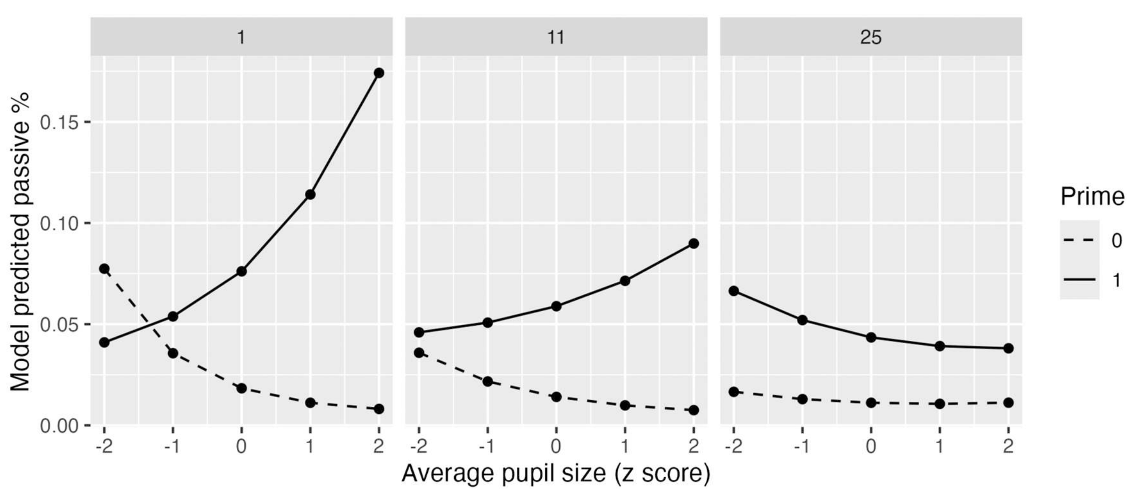

Figure 4. Model predicted interaction between prime and measure of prediction error over the course of the experiment.
Notes: Model-fitted effects are estimated at the median of other predictors in the model (i.e. gaze position) and excluding the contribution of random effects.

pupil size lead to a decreased likelihood of passive production). Conversely, greater prediction error after a passive prime was linked to a higher passive production. However, the effect decreased as the experiment progressed. This was evident when we used a different rescaled trial predictor in the model, estimating other effects at the mid-point of the experiment – trial 16. The interaction effect between prime and prediction error went from strong to weak, $\beta = 0.38$ [−0.19, 0.93], posterior prob. = .871 (see online materials for full model).

Discussion

In this paper we aimed to index prediction error using pupillometry to predict syntactic priming, and in doing so directly tested the suggestion that priming is driven by error-based learning (Chang et al., 2006). We found that the pupillary response was larger during passive than active primes, replicating syntactic complexity effects (Beatty, 1982; Demberg & Sayeed, 2016; Just & Carpenter, 1993; Piquado et al., 2010). We also replicated the syntactic priming effect (Mahowald et al., 2016), with participants producing more passive responses after passive than active primes. Using mean model-predicted pupil size during prime presentation as a measure, we found evidence that prediction error predicted syntactic priming, which varied across the experiment. Investigating cumulative effects showed that the difference in pupil dilation between active and passive primes was largest at the beginning of the experiment, and it was here where pupil dilation was strongly linked to priming. As the experiment progressed, the difference in pupil dilation across conditions reduced, and there was weak evidence that the relationship between it and priming weakened, as participants continued to produce passives at a similar rate.

Our first hypothesis was that syntactic priming is driven by prediction error which can be indexed by pupillometry. Chang et al. (2006) propose that an error-based learning mechanism underlies syntactic priming. During prime comprehension, the model makes next-word predictions and compares them to the actual input. The active-passive alternation provides a test case for this mechanism given contrasts between the structures in terms of frequency and canonical word order (Roland et al., 2007; Xiao et al., 2006). In an active sentence (*the elephant is biting the frog*), the sequence of word classes and syntactic categories is associated with minimal prediction error, whereas a large error signal is produced for passives when participants hear a second auxiliary and past participle (*the frog is* ***being bitten*** *by the elephant*), since a present participle (corresponding to an active) is more expected. This error backpropagates through the system, which adjusts the network weights that comprise syntactic representations and increases the likelihood of a passive structure (i.e. priming). The greater the prediction error, the greater adjustment in network weights and therefore increase in passive production. To test this proposition, we operationalised prediction error as pupil dilation, drawing upon arguments outside of psycholinguistics (e.g. Preuschoff et al., 2011). If this is a reasonable assumption, the error-based learning account predicts that a larger pupil response to a prime sentence will predict syntactic priming. Since we replicated both the syntactic complexity effect and the syntactic priming effect, we could test this key hypothesis.

Overall, we found evidence in support of this hypothesis. At the beginning of the experiment, there was strong evidence for a relationship between average pupil size during prime presentation and participants' likelihood of being primed. The more prediction error participants experienced, the more likely they were to repeat the structure used in the prime sentence. The effect received strong evidence in all but one variation of prediction error measures (these reflected possible choices of time window and the use of measured vs. model-predicted pupil size). The strength of this effect was potentially influenced by cumulative effects, which we discuss below. Thus, we take our result as promising evidence that pupil size indexes trial-by-trial prediction error. This result aligns with findings from other domains where researchers have directly manipulated participants' expectations and found that the pupil response is proportional to the magnitude of prediction error (e.g. O'Reilly et al., 2013; Preuschoff et al., 2011). In these studies, participants' expectations can be manipulated directly by creating stimuli based on known distributions (e.g. probability of a random number between 1 and 10 being higher or lower than a previous number; Preuschoff et al., 2011) or defined probabilistic rules (e.g. transitional probabilities between stimuli; Alamia et al., 2019). However, in the linguistic domain we draw upon participants' lifelong experience with linguistic structures, which vary along parameters other than frequency (e.g. structural complexity). Our study suggests that pupil dilation as a measure of prediction error can be extended to accumulated syntactic knowledge.

The second major aim of our study was to investigate cumulative effects in comprehension and production during syntactic priming. The evidence for these effects was less conclusive but still suggestive of the effects of adapting to cumulative input. Under Chang et al.'s (2006) error-based learning account, representations and expectations of syntactic structures are dynamic and affected by each encounter with a syntactic structure. The account predicts that as a structure becomes more expected, it induces less prediction error and is produced at a higher rate. To investigate cumulative effects in our study, we incorporated trial number as a predictor in our analyses. The difference between prediction error for active and passive primes as measured by pupil dilation was greatest in the first five trials (Figure 3). After that point, the difference was smaller. In terms of production measures, there was weak evidence for a slight reduction in passive production over the course of the experiment but no evidence for a change in the syntactic priming effect. The slight reduction in passive production over the experiment can be explained by a particularly high production of passives in the first few trials when participants found passives particularly surprising.

This pattern of findings is consistent with participants rapidly adapting to the context of the experimental task where passives are far more frequent than in everyday speech (Roland et al., 2007; Xiao et al., 2006). Instead of the influence of passive primes accumulating over the course of the experiment and demonstrating greater priming in the second half, participants appeared to quickly change their expectations and productions of passives. After the initial phase of surprisal and boost in passive production, they settled into an expectation for a higher baseline of passives than in everyday speech. This pattern of results aligns with proposals that for predictive processing to be efficient, rapid adaptation to variation in syntactic preferences by speakers and contexts must be possible (Fine et al., 2013). Since syntactic priming in one context does not always transfer to other contexts (Heyselaar & Segaert, 2022 vs Kaschak et al., 2014), an outstanding question for error-based learning accounts is how to reconcile longer term incremental changes in syntactic representations with dynamic context-specific adaptations.

The ability to observe cumulative effects somewhat varied between comprehension (i.e. pupil size during prime trails) and production (passive production during target trials). If pupil dilation indexes prediction error, then production of passives during target trials represents the outcome of model updating. This can make cumulative effects in production measures difficult to observe: if baseline frequency is updated rapidly, later productions of a structure will reflect this higher baseline, with minimal influence from surprisal and model updating following the immediately previous prime. Effectively, in the context of our experiment, a passive prime may become less surprising because of adaptation, which could then maintain higher rates of passive use.

This explanation accounts for an effect we found suggestive evidence for: the strong association between prediction error and syntactic priming early in our experiment, which decoupled later on. There was weak evidence for this effect in our model, supported by the interaction between prime and prediction error going from strong to weak evidence when estimated at the mid-point vs beginning of the experiment. There have been some efforts to distinguish prediction error and model-updating experimentally. O'Reilly et al. (2013) presented coloured dots that occurred in runs of a similar location, with a change in location signalled by change in dot colour (model update trial), interspersed with grey dots that could occur anywhere but did not provide predictive value (surprise alone trials). They were able to distinguish the involvement of different brain areas in the two processes. Nassar et al. (2012) showed that at change points in the mean of a series of presented numbers, surprisal as measured by pupil dilation was larger and predicted the magnitude of change in participants' predictions of the next number (model updating). In syntactic priming, we cannot measure the magnitude of model updating on a particular trial, only whether or not the primed structure was produced. Instead, it may be possible to compare cumulative effects for structures with different baseline frequencies, which are likely to produce different magnitudes of model updating. For example, passives are so infrequent (1–3 in every 1000 words; Roland et al., 2007; Xiao et al., 2006) that encountering several in a short period may result in rapid adaptation, whereas the distribution between double object and prepositional object datives is more balanced and could result in more incremental changes. Bernolet et al.'s (2016) finding that the number of primes encountered marginally predicted dative production but not passive production provides some preliminary support for this assertion.

Connections with the adaptive gain theory

Under the adaptive gain theory, phasic LC-NE activity was initially linked to the P3 ERP (Nieuwenhuis et al., 2005) and then to task-evoked pupil dilation (Murphy et al., 2011). The release of NE is thought to enhance neuronal gain and therefore optimise behavioural responses to motivationally significant stimuli (Aston-Jones & Cohen, 2005) and to support learning following novel or unexpected stimuli (Nieuwenhuis, 2011). We found that the pupil response was larger when listening to a less predictable passive sentence than active sentence, and predicted the likelihood of being primed in line with error-based learning accounts of syntactic priming. Our study therefore contributes evidence that pupil dilation can index prediction error even from implicit predictive processing of stimuli that are not task-relevant (i.e. motivationally significant). For example, Alamia et al. (2019) found that stimuli needed to be attended to in order to observe a pupil response to infrequent stimuli transitions, but that these statistical regularities did not need to be task-relevant or reach conscious awareness. Similarly, Damsma and Van Rijn (2017) found that Dutch university students' pupils dilated in response to salient omissions in standard Western drum beats even when they were instructed to ignore the audio to focus on an alternative task. In our study, participants were required to indicate whether the picture matched the sentence audio, which acted as a cover task and a way to maintain attention on the stimuli. However, tracking the frequency of active and passive stimuli (within filler stimuli of different structures) was not task-relevant.

An implication stemming from the adaptive gain theory is the integration of electroencephalography (EEG) and pupillometry effects. In language processing, the P600 is argued to be equivalent to the P3 (Sassenhagen et al., 2014; Sassenhagen & Fiebach, 2019). It follows that similar effects should be observed between the P600 and pupil dilation. Indeed, like the pupil response, the P600 is observed in response to syntactic violations (e.g. Hagoort et al., 1993). However, a recent study measuring both responses following syntactic violations did not find conclusive evidence that the two were correlated (Contier et al., 2024). Further, another language-related ERP, the N400, is also argued to reflect prediction error (Bornkessel-Schlesewsky & Schlesewsky, 2019; Rabovsky et al., 2018) and has been found to correlate with pupil dilation (Kuipers & Thierry, 2011). Some researchers characterise the N400 as a prediction error signal and the P600 as reflecting the behavioural consequences of predictive processing (Bornkessel-Schlesewsky & Schlesewsky, 2019), whilst others argue they both reflect error signals but at different levels of linguistic representation (Fitz & Chang, 2019). Syntactic priming offers a context where the link between the P600, N400, and pupil dilation effects could be further investigated. Studies investigating syntactic priming in comprehension using ERPs have generally found that both the N400 and P600 are reduced following a prime of the same structure (Chen et al., 2013; Ledoux et al., 2007; Tooley et al., 2014). However, this attenuation has only been found when prime and target share a verb (Chen et al., 2013; Tooley et al., 2009), with only one structure, reduced relative clauses, investigated. Rather than observing their attenuation from comprehended prime to comprehended target, the magnitude of these ERPs following a prime could be linked to the likelihood of production of the primed structure as we did here with pupil size (Tooley, 2023). Links between the responses could also be investigated if pupil size is monitored alongside EEG, as in Contier et al. (2024).

Future directions

This study found that pupil dilation is a promising measure of prediction error during syntactic priming. Pupillometry offers several advantages for syntactic priming research. Firstly, it is an alternative receptive measure of language processing. Comprehension priming studies have often relied on self-paced reading, a method with known disadvantages (Frinsel & Christiansen, 2024; Prasad & Linzen, 2021). Secondly, it is a non-invasive measure that can easily be integrated into production priming tasks. Tooley has recently pointed out that investigating how prime processing influences target processing in syntactic priming offers opportunities for a deeper mechanistic understanding of language processing (Tooley, 2023; Tooley & Brehm, 2025). Finally, we suggest that pupil dilation is an automated and implicit index of prediction error rooted in a neurobiological response. Corpus-based estimates of surprisal rely on the language input sampled being representative of the varied experience of individual participants. Pending further research, pupillometry could offer an approach to measuring individualised responses to input. We outline some specific potential applications below.

We chose a structural alternation which reliably produces large priming effects as a test case, but a variety of manipulations of prime content that are intended to increase prediction error increase syntactic priming and could be adopted by future studies using pupillometry (e.g. manipulating expectations: Arai & Chang, 2024; verb-bias effects: Bernolet & Hartsuiker, 2010; inverse frequency effects: Jaeger & Snider, 2013). For example, the dative alternation enables cross-linguistic investigations where the frequency of different dative forms are in complementary distribution, in addition to the investigation of verb-biases. The double object dative is less frequent in Dutch, whilst the prepositional object dative is less frequent in English (cf., Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2013). If the pupillary response to primes of the same structure varies according to the structure's distribution within the language, this could more definitively tie pupil size to prediction error. The same reasoning applies if the pupil response to primes of the same dative structure differs for verbs biased towards or against the structure. Chang et al.'s (2006) error-based learning account is a theory of acquisition as well as processing. Another prediction of the account is that syntactic priming is larger for those who are subject to greater prediction error because their syntactic representations are still developing (Kumarage et al., 2022; Rowland et al., 2012).

We also found some evidence for dynamic changes to expectations and syntactic representations during the experiment. This issue could be further investigated by manipulating the distribution of input over the course of an experiment. Pupil dilation increases at change points in non-linguistic input (Nassar et al., 2012) and could be compared at points in the experiment where prediction error is expected to be larger and smaller and correlated with syntactic priming at these times. For example, Jaeger and Snider (2013) compared comprehension priming when participants encountered primes in block order (e.g. all double object datives then all prepositional object datives) and alternating order. In the blocked condition, prediction error should show attenuation over the first block, then a sharp increase and attenuation again over the second block.

Finally, future research may also address the limitations of this study. One limitation was the uncertainty in deciding on how to operationalise prediction error from the pupil response. While GAMMs allow sophisticated modelling of the time course of the pupil response, it is less clear how to extract an equally sophisticated trial-by-trial measure of prediction error from that model. Average pupil size during the prime sentence, or a restricted period of that sentence, is a reasonable but coarse operationalisation. We did not have the computational power to include random smooths for each trial for each participant. Extracting these random effects from a model that did include them is one option for a more precise measure.

Conclusion

In conclusion, we found a larger pupil response elicited by passive when compared to active sentences, in line with previous syntactic complexity effects (e.g. Just & Carpenter, 1993), and replicated the widely reported syntactic priming effect (Mahowald et al., 2016). Overall, pupil size strongly predicted syntactic priming at the beginning of the experiment. However, investigating cumulative effects showed weak evidence that participants rapidly adjusted their baseline expectations and productions of the passive, in line with rapid adaptation to linguistic context (Fine et al., 2013). These results are consistent with Chang et al.'s (2006) error-based learning account of syntactic processing and acquisition. In light of the difficulty of measuring and operationalising cognitive processes, our findings open a range of possible applications within research that investigates error-based learning and the links between expectations, frequencies, and prediction error in studies of syntactic processing.

Note

1. We note that we used Bayesian rather than frequentist statistics as in Mahowald et al.'s (2016) power simulations. Bayesian and frequentist models do typically produce very similar results unless using informative priors. We used weakly informative priors but our results did not change when using uninformative priors. Sample sizes that result in low power in frequentist model will result in wide credible intervals in Bayesian models, which can be interpreted similarly.

Acknowledgements

We thank Jet Bustamante and Maggie Otto for their work producing our experimental materials and Dr Rowena Garcia for sharing her materials with us.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

This research was supported by the Max Planck Society (Max Planck Instituut voor Psycholinguïstiek), the Australian Research Council [grant number DP210102836], and the Australian National University.

Data availability statement

Our processed data sheets and analysis scripts are available on the Open Science Framework (https://osf.io/7gpy4/).

ORCID

Shanthi Kumarage http://orcid.org/0009-0006-3578-2737
Anton Malko http://orcid.org/0000-0002-0089-1072
Evan Kidd http://orcid.org/0000-0003-4301-2290

References

Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., & Zenon, A. (2019). Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience*, *39*(27), 5369–5376. https://doi.org/10.1523/JNEUROSCI.3010-18.2019

Arai, M., & Chang, F. (2024). Influence of word-based prediction errors on syntactic priming. *Seijo University Economic Papers*, *243*, 97–120.

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Aydın, Ö., & Uzun, İ. P. (2023). Pupil dilation response to prosody and syntax during auditory sentence processing. *Journal of Psycholinguistic Research*, *52*(1), 153–177. https://doi.org/10.1007/s10936-021-09830-y

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Bernolet, S., Collina, S., & Hartsuiker, R. J. (2016). The persistence of syntactic priming revisited. *Journal of Memory and Language*, *91*, 99–116. https://doi.org/10.1016/j.jml.2016.01.002

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. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, *31*(2), 163–186. https://doi.org/10.1016/0010-0277(89)90022-X

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192. https://doi.org/10.1037//0096-3455.129.2.177

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*(1), 1–39. https://doi.org/10.1016/0010-0277(90)90035-I

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*, 150–171. https://doi.org/10.1037/0033-295X.99.1.150

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*, Article 298. https://doi.org/10.3389/fpsyg.2019.00298

Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E., & Notebaert, W. (2015). Open your eyes for prediction errors. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 374–380. https://doi.org/10.3758/s13415-014-0333-4

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, *45*(4), 1322–1331. https://doi.org/10.3758/s13428-013-0327-0

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chen, Q., Xu, X., Tan, D., Zhang, J., & Zhong, Y. (2013). Syntactic priming in Chinese sentence comprehension: Evidence from event-related potentials. *Brain and Cognition*, *83*(1), 142–152. https://doi.org/10.1016/j.bandc.2013.07.005

Chomsky, N. (1959). Review of Skinner's verbal behavior. *Language*, *35*, 26–58.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509. https://doi.org/10.1017/S0140525X08004998

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

Contier, F., Wartenburger, I., Weymar, M., & Rabovsky, M. (2024). Are the P600 and P3 ERP components linked to the task-evoked pupillary response as a correlate of norepinephrine activity? *Psychophysiology*, *61*(7), Article e14565. https://doi.org/10.1111/psyp.14565

Dal Ben, R. (2021). *SHINE_color: Controlling low-level properties of colorful images*. PsyArXiv. https://doi.org/10.31234/osf.io/fec6x

Damsma, A., & Van Rijn, H. (2017). Pupillary response indexes the metrical hierarchy of unattended rhythmic violations. *Brain and Cognition*, *111*, 95–103. https://doi.org/10.1016/j.bandc.2016.10.004

Dayan, P., & Yu, A. (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network (Bristol, England)*, *17*, 335–350. https://doi.org/10.1080/09548980601004024

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 B: Biological Sciences*, *369*(1634), Article 20120394. https://doi.org/10.1098/rstb.2012.0394

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS One*, *11*(1), e0146194. https://doi.org/10.1371/journal.pone.0146194

Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(10), 1906–1921. https://doi.org/10.1037/xlm0000881

Eager, C., & Roy, J. (2017). *Mixed effects models are sometimes terrible*. arXiv. https://doi.org/10.48550/arXiv.1701.04585

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*(4), 639–645. https://doi.org/10.1080/17470210903469864

Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., Theakston, A., & Lieven, E. (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology*, *110*, 30–69. https://doi.org/10.1016/j.cogpsych.2019.02.001

Farmer, T. A., Fine, A. B., Yan, S., Cheimariou, S., & Jaeger, T. F. (2014). Error-driven adaptation of higher-level expectations during reading. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*, 2181–2186.

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), Article 180877. https://doi.org/10.1098/rsos.180877

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33*(6), 715–736. https://doi.org/10.1006/jmla.1994.1034

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language*, *48*(2), 379–398. https://doi.org/10.1016/S0749-596X(02)00523-5

Fine, A. B., & Jaeger, T. F. (2016). The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1362–1376. https://doi.org/10.1037/xlm0000236

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, *8*(10), Article e77661. https://doi.org/10.1371/journal.pone.0077661

Fitz, H., & Chang, F.. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Forbes, S. H. (2020). Pupillometryr: An R package for preparing and analysing pupillometry data. *Journal of Open Source Software*, *5*(50), Article 2285. https://doi.org/10.21105/joss.02285

Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology*, *34*(5), 475–484. https://doi.org/10.1016/0013-4694(73)90065-5

Frinsel, F. F., & Christiansen, M. H. (2024). Capturing individual differences in sentence processing: How reliable is the self-paced Reading task? *Behavior Research Methods*, *56*, 6248–6257. https://doi.org/10.3758/s13428-024-02355-x

Garcia, R., Garrido Rodriguez, G., & Kidd, E. (2021). Developmental effects in the online use of morphosyntactic cues in sentence processing: Evidence from Tagalog. *Cognition*, *216*, Article 104859. https://doi.org/10.1016/j.cognition.2021.104859

Garcia, R., Roeser, J., & Kidd, E. (2023). Finding your voice: Voice-specific effects in Tagalog reveal the limits of word order priming. *Cognition*, *236*, Article 105424. https://doi.org/10.1016/j.cognition.2023.105424

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483. https://doi.org/10.1080/01690969308407585

Harrington Stack, C. M., James, A. N., & Watson, D.G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*, *46*(6), 864–877. https://doi.org/10.3758/s13421-018-0808-6

Harris, D. J., Arthur, T., Vine, S. J., Liu, J., Abd Rahman, H. R., Han, F., & Wilson, M. R. (2022). Task-evoked pupillary responses track precision-weighted prediction errors and learning rate during interceptive visuomotor actions. *Scientific Reports*, *12*(1), Article 22098. https://doi.org/10.1038/s41598-022-26544-w

Hartsuiker, R. J., Kolk, H. H. J., & Huiskamp, P. (1999). Priming word order in sentence production. *The Quarterly Journal of Experimental Psychology Section A*, *52*(1), 129–147. https://doi.org/10.1080/713755798

Häuser, K. I., Demberg, V., & Kray, J. (2018). Surprisal modulates dual-task performance in older adults: Pupillometry shows age-related trade-offs in task performance and time-course of language processing. *Psychology and Aging*, *33*(8), 1168–1180. https://doi.org/10.1037/pag0000316

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development*, *17*(3), 359–377. https://doi.org/10.1080/15248372.2015.1135801

Heyselaar, E., & Segaert, K. (2022). Structural priming persists for (at least) one month in young adults, but not in healthy older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(8), 1219–1234. https://doi.org/10.1037/xlm0001123

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83. https://doi.org/10.1016/j.cognition.2012.10.013

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *47*(2), 310–339. https://doi.org/10.1037/h0078820

Kaschak, M. P., Kutta, T. J., & Coyle, J. M. (2014). Long and short term cumulative structural priming effects. *Language, Cognition and Neuroscience*, *29*(6), 728–743. Scopus. https://doi.org/10.1080/01690965.2011.641387

Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, *18*(6), 1133–1139. https://doi.org/10.3758/s13423-011-0157-y

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*(3), 381–388. https://doi.org/10.3758/s13421-010-0042-3

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, *139*(4), 1618–1636. https://doi.org/10.1121/1.4944032

Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2013). Processing load during listening: The influence of task characteristics on the pupil response. *Language and Cognitive Processes*, *28*(4), 426–442. https://doi.org/10.1080/01690965.2011.642267

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23–34. https://doi.org/10.1111/j.1469-8986.2012.01477.x

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*, Article 61. https://doi.org/10.3389/fnhum.2011.00061

Kumarage, S., Donnelly, S., & Kidd, E. (2022). Implicit learning of structure across time: A longitudinal investigation of syntactic priming in young English-acquiring children. *Journal of Memory and Language*, *127*, Article e104374. https://doi.org/10.1016/j.jml.2022.104374

Lavin, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, *7*, Article 218. https://doi.org/10.3389/fnbeh.2013.00218

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: Evidence from event-related potentials. *Psychological Science*, *18*(2), 135–143. Scopus. https://doi.org/10.1111/j.1467-9280.2007.01863.x

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, Article 226. https://doi.org/10.3389/fpsyg.2013.00226

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

Mathôt, S. (2013). *A simple way to reconstruct pupil size during eye blinks*. Figshare. https://doi.org/10.6084/m9.figshare.688001

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077. https://doi.org/10.3758/s13428-022-01957-7

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil–linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nieuwenhuis, S. (2011). Learning, the P3, and the locus coeruleus-norepinephrine system. In R. Mars, J. Sallet, M. Rushworth, & N. Yeung (Eds.), *Neural basis of motivational and cognitive control* (pp. 209–222). MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38), E3660–E3669. https://doi.org/10.1073/pnas.1305373110

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. https://doi.org/10.1016/j.jml.2014.12.002

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569. https://doi.org/10.1111/j.1469-8986.2009.00947.x

Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *47*(7), 1156–1172. https://doi.org/10.1037/xlm0001046

Preuschoff, K., 't Hart, M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*, Article 115. https://doi.org/10.3389/fnins.2011.00115

Qiyuan, J., Richer, F., Wagoner, B. L., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology*, *22*(5), 530–534.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, *125*(1), 49–63. https://doi.org/10.1016/j.cognition.2012.06.008

Rumelhart, D. E., Hintont, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. https://doi.org/10.1038/323533a0

Rutar, D., Colizoli, O., Selen, L., Spieß, L., Kwisthout, J., & Hunnius, S. (2023). Differentiating between Bayesian parameter learning and structure learning based on behavioural and pupil measures. *PLoS One*, *18*(2), e0270619. https://doi.org/10.1371/journal.pone.0270619

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), Article e13335. https://doi.org/10.1111/psyp.13335

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, *200*, 425–436. https://doi.org/10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, *89*(3), 179–205. https://doi.org/10.1016/S0010-0277(03)00119-7

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, *17*(1), 133–145. https://doi.org/10.1016/0093-934X(82)90010-4

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Stanners, R. F., Headley, D. B., & Clark, W. R. (1972). The pupillary response to sentences: Influences of listening set and deep structure. *Journal of Verbal Learning and Verbal Behavior*, *11*(2), 257–263. https://doi.org/10.1016/S0022-5371(72)80086-0

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, *116*, Article e104190. https://doi.org/10.1016/j.jml.2020.104190

Tooley, K. M. (2020). Contrasting mechanistic accounts of the lexical boost. *Memory & Cognition*, *48*, 815–838. https://doi.org/10.3758/s13421-020-01019-3

Tooley, K. M. (2023). Structural priming during comprehension: A pattern from many pieces. *Psychonomic Bulletin & Review*, *30*(3), 882–896. https://doi.org/10.3758/s13423-022-02209-7

Tooley, K. M., & Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition*, *132*(2), 101–136. https://doi.org/10.1016/j.cognition.2014.04.002

Tooley, K. M., & Brehm, L. (2025). Putting the prime in priming: Using prime processing behavior to predict target structural processing. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-025-02643-3.

Tooley, K. M., Swaab, T. Y., Boudewyn, M. A., Zirnstein, M., & Traxler, M. J. (2014). Evidence for priming across intervening sentences during on-line sentence comprehension. *Language and Cognitive Processes*, *29*(3), 289–311. https://doi.org/10.1080/01690965.2013.770892

Tooley, K. M., Traxler, M. J., & Swaab, T. Y. (2009). Electrophysiological and behavioral evidence of syntactic priming in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 19–45. https://doi.org/10.1037/a0013984

Tromp, J., Hagoort, P., & Meyer, A. S. (2016). Pupillometry reveals increased pupil size during indirect request comprehension. *Quarterly Journal of Experimental Psychology*, *69*(6), 1093–1108. https://doi.org/10.1080/17470218.2015.1065282

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 1–22. https://doi.org/10.1177/2331216519832483

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2022). itsadug: Interpreting time series and autocorrelated data using GAMMs. R Package, Version 2.4.

Vogelzang, M., Hendriks, P., & Van Rijn, H. (2016). Pupillary responses reflect ambiguity resolution in pronoun processing. *Language, Cognition and Neuroscience*, *31*(7), 876–885. https://doi.org/10.1080/23273798.2016.1155718

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology*, *7*, Article 345. https://doi.org/10.3389/fpsyg.2016.00345

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*, 1–17. https://doi.org/10.1177/2331216516669723

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3–36.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast*, *6*(1), 109–149. https://doi.org/10.1075/lic.6.1.05xia

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, *104*(5), 444–466.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear & Hearing*, *31*(4), 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zellin, M., Pannekamp, A., Toepel, U., & Van Der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*(3), 133–141. https://doi.org/10.1016/j.ijpsycho.2011.05.009

Zhang, F., Jaffe-Dax, S., Wilson, R. C., & Emberson, L. L. (2019). Prediction in infants and adults: A pupillometry study. *Developmental Science*, *22*(4), Article e12780. https://doi.org/10.1111/desc.12780

Ziegler, J., Bencini, G., Goldberg, A., & Snedeker, J. (2019). How abstract is syntax? Evidence from structural priming. *Cognition*, *193*, Article 104045. https://doi.org/10.1016/j.cognition.2019.104045