languages

MDPI

Review

Prime surprisal as a tool for assessing error-based learning theories: a systematic review.

Authors

[1] affiliation

Abstract: Error-based learning theories of language acquisition are highly influential in language development research, yet the predictive learning mechanism they propose has proven difficult to test experimentally. Prime surprisal– the observation that structural priming is stronger following more surprising primes – has emerged as a promising methodology for resolving this issue, as it tests a key prediction of error-based learning theories: that surprising input leads to increased structure repetition as well as learning. However, as prime surprisal is a relatively new paradigm, it is worth evaluating how far this promise has been fulfilled. We have conducted a systemic review of PS studies to assess the strengths and limitations of existing approaches, with 13 contributions selected out of 66 search results. We found that alongside inconsistency in statistical power and how the level of surprisal is measured, the limited scope of current results cast doubt on whether PS can be used as a general tool to assess error-based learning. We suggest two key directions for future research. Firstly, targeting the scope of the prime surprisal effect itself with reliable statistical power and appropriate surprisal measurements, across a greater variety of languages and grammatical structures. Secondly, using the prime surprisal method as a tool to assess the scope of an error-based learning mechanism utilizing conditions in which prime surprisal has been reliably established.

Keywords: prime surprisal; error-based learning; linguistic prediction; syntactic priming; systemic review; language acquisition

1. Introduction

Since the first investigations of prime surprisal (PS), this paradigm has held the promise of shedding light on prediction's role in language processing and, crucially, language development (e.g., Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2008). The PS effect shows that surprising structures are repeated more often than predictable ones and it offers a distinctive tool **for exploring prediction's role in language processing and learning**. Linguistic predictions are notoriously difficult to target experimentally. Rather than assessing reactions to the critical linguistic stimulus, they require the measurement of cognitive processes that start before the stimulus appears. Targeting the role of linguistic predictions is however crucial, as they are theorised to play a fundamental role in language learning (see e.g., Chang, Dell & Bock, 2006; Rabagliati, Gambi & Pickering, 2016). As such PS-based studies have already led to significant advances in our understanding of prediction's role in language acquisition (e.g., [Omitted for peer-review]; Jaeger & Snider, 2013; Peter, Chang, Pine, Blything & Rowland, 2015). Yet at the same time, much remains unknown or unclear, and important issues regarding our understanding and use of PS in prediction research are still to be resolved.

Prediction – the ability to anticipate upcoming events – is widely studied in relation to many aspects of human life. It underpins our ability to do various activities from playing music to participating in a volleyball match (see e.g., Novembre and Keller,2011;

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Urgesi, Savonitto, Fabbro and Aglioti,2012). Predictive processes are therefore a widely studied phenomenon across many branches of cognitive science, and the language sciences are no exception (see e.g. Ryskin and Nieuwland, 2023). Pre-activating upcoming language might aid communication in several ways. For instance, by anticipating what our conversational partner will say next, we can start planning our own response ahead of time. This in turn leads to swift turn-taking and smooth dialogues (e.g., Magyari, Bastiansen, de Ruiter & Levinson, 2014).

Of all the roles that prediction might play in communication, perhaps the most interesting possibility is that it is key to acquiring language in the first place. Error-based theories (e.g., Chang, Dell & Bock, 2006; Ramscar, Dye, & McCauley, 2013) suggest that language learners continuously predict the next word while listening to others talk. They then compare their own predictions to upcoming words and update their linguistic knowledge based on any potential differences. Error-based learning (EBL) theories have gained wide support due to their large explanatory power and the clear, computationally replicable account they provide (Chang, Dell & Bock, 2006; Peter & Rowland, 2019).

Given EBL theories' strong focus on how previous linguistic input affects subsequent language use, the structural priming paradigm was a logical starting point to target these accounts. Priming studies typically include linguistic structures that can alternate between different forms that have similar meanings. English structural priming studies (the language most often used in this line of research) commonly feature dative or transitive structures (see Mahowald, James, Futrell and Gibson (2016) for a meta-analysis of syntactic priming in language production). The goal of such studies is to assess whether participants are more likely to repeat previously-processed linguistic structures as opposed to using the alternative form (e.g. Bock, 1986; Messenger, Branigan, McLean & Sorace, 2012). For instance, in a dative priming study, a structural priming effect would show that participants were more likely to reuse the double object dative (DOD) after processing a DOD than after a prepositional dative (PD). The structural priming effect is now well established (Mahowald et al., 2016), and is, in itself, a source of support for EBL accounts. Priming studies demonstrate that previous linguistic input shapes subsequent language use, a premise which is central to error-based theories. However, structural priming by itself does not target the role of surprisal, which limits the extent to which the paradigm is suited to further assessing EBL theories.

By comparing priming with more- or less-frequent primes, other studies found additional support for EBL theories in the form of enhanced priming after more surprising structures. This inverse frequency effect has been observed in various modalities, such as morphological (Moder, 1992) as well as semantic priming (see e.g. Goldinger et al., 1989) and inhibitory phonological priming (e.g., Luce et al., 2000 or see Kapatsinski, 2006, 2007, for a review). Crucially for this work, the inverse frequency effect has also been observed in syntactic priming, showing that overall less-frequent syntactic structures, such as passive sentences, tend to prime more strongly than more-frequent active structures (Bock, 1986, Ferreira, 2003). However, while the inverse frequency effect is in itself supportive of EBL theories, it still does not consider a key element for these theories: the role of the immediate linguistic context (Jaeger & Snider, 2008, 2013).

The PS paradigm takes the evaluation of EBL theories a step further by assessing how the prime structure's surrounding context influences the size of the priming effect. These studies typically rely on the observation that each verb is more likely to appear with one of the possible sentence structures than the alternative one (e.g., Bernolet & Hartsuiker, 2010; Peter et al., 2015). For instance, when considering the dative structure overall, the DOD variant is more frequent in adult language use (British National Corpus, BNC Consortium, 2007). However, when we observe verb-specific patterns, we find that some

verbs favour the PD structure. For instance, the verb ***pass*** occurs more often in a PD structure than in a DOD structure, but the verb ***give*** prefers the DOD structure (BNC Consortium, 2007). As a result, a sentence including the verb ***pass*** is more surprising in a DOD structure (e.g., The nurse passed the patient an apple) than in a PD structure (e.g., The nurse passed an apple to the patient). Conversely, a sentence including the verb ***give*** is more surprising in a PD structure (e.g., The postman gave the package to the recipient) than in a DOD structure (e.g., The postman gave the recipient the package).

The PS effect is particularly suited for assessing the Dual-Path model (Chang et al., 2006; Dell & Chang, 2014), an EBL account that has been applied to syntactic acquisition. This model operates via next-word prediction and predicts stronger priming effects after surprising as opposed to predictable sentences. The Dual-Path model suggests that after hearing a verb, listeners predict the following words according to the syntactic structure that most often follows the given verb. For instance, in a dative priming study, if participants hear the DOD-biased verb *give*, they are most likely to predict that the sentence will be a recipient and the sentence will feature a DOD structure. if the sentence unfolds in such a way that it is not compatible with a DOD structure, for example because the second post-verbal noun is preceded by the preposition to indicating a PD structure, the learning mechanism produces an error signal. This error signal both increases the probability of immediate structure repetition (PS) and results in enhanced learning effects. However, if they hear a recipient indicating a DOD structure after a DOD-biased verb, participants are more likely to successfully predict the upcoming structure, and thus no error signal will be produced. In this case, the probability of immediate structure repetition is not increased and no enhancement in learning occurs.

Some priming studies have indeed found enhanced priming effects after surprising primes when structures appeared with mismatching as opposed to matching verbs. Such PS effects have been found in both production (e.g. Jaeger & Snider, 2013, Peter et al., 2015) and comprehension (e.g., Fernandes, 2015; Fine & Jaeger, 2013), as well as in languages other than English (Dutch: Bernolet & Hartsuiker, 2010). Importantly, from the perspective of language acquisition research, the priming effect's sensitivity to surprisal has also been demonstrated in studies with child participants ([Omitted for peer-review]). In addition, more recent studies (e.g., [Omitted for peer-review]) have confirmed that surprisal contributes to cumulative, delayed (as well as immediate) structure repetition, providing experimental evidence for surprisal's connection to learning.

However, despite the promise of PS as experimental support for EBL theories and the accumulating evidence it provides, there are reasons to be cautious in our interpretation of existing results. PS is still a relatively new method, with the first study using the paradigm published 15 years ago (Jaeger & Snider, 2008). To date, most PS studies look at a very limited set of languages, chiefly English. Importantly – given PS's proposed relevance to language acquisition research – studies targeting children are also limited in number. In addition, there are several ways in which surprisal can be defined and measured (such as binary versus continuous measures of surprisal, e.g., Jaeger & Snider, 2013 versus Peter et al., 2015), making results difficult to compare. This in turn poses a challenge for any effort to generalise existing results as support for EBL theories.

It is also important to consider the relevance of PS as a tool for assessing error-based language development in future studies. According to the Dual-path model, immediate PS effects are the result of the same error-based learning mechanism that is behind long-term language acquisition. This makes PS studies an excellent instrument for assessing error-based learning experimentally. They have the potential to target a long-term learning mechanism in one relatively short experiment as opposed to more resource-intensive longitudinal studies. Consequently, more and more studies are using PS to assess

whether specific groups of people, for instance children of various ages (e.g. Peter et al., 2015) or L2 learners (e.g., Kaan & Chun, 2018) can make linguistic predictions.

However, if PS is to be used as a tool for widely assessing predictive abilities in diverse participant groups, including learners of English as well as other languages, it is crucial to establish under what circumstances this effect reliably occurs. This is especially key for participant groups that consistently showed predictive abilities in other experimental settings, such as adult monolingual speakers (see Ryskin and Nieuwland (2023) for a review). Otherwise, the lack of PS is difficult to interpret in studies featuring new participant groups. For instance, if a PS study is conducted with second-language learning adolescents and no PS effect is found, it would be difficult to establish whether this group is not using linguistic prediction or another factor (for example, low statistical power or a problematic measure of surprisal) has led to this result. In other words, PS can only be a useful instrument for targeting linguistic predictions if the circumstances in which we would expect it to appear reliably are firmly established.

To achieve more clarity regarding these issues, we conducted a systematic review of verb-based PS studies to date. We will discuss the cases in which PS was and was not found, and the role of diverging experimental choices in shaping these results. We will then discuss the limitations of the PS paradigm as a tool for assessing error-based learning, including the extent to which existing results can be taken as evidence for or against error-based theories. Based on our assessment of the current literature, we argue that PS only reliably appears under a specific set of conditions, and that specifying these conditions is essential for further developing this paradigm. We will also set out future research directions focusing on how PS can be best utilized to assess predictive language processing.

2. Materials and Methods

2.1. Defining prime surprisal

As studies vary greatly in how they define and measure PS (see section 3.2), we will begin our review by discussing the PS definition we used in our search criteria. We aimed to keep our sample relatively homogeneous. Thus, we used a narrow definition of PS. However, in order to include the maximum number of relevant studies, we have incorporated dissertations and conference proceedings as well as published, peer-reviewed articles. As our approach aimed to encompass the maximum number of studies that targeted error-based learning by using the PS paradigm, we decided to concentrate on studies measuring the surprisal values conditional on the verb, as this was the most common approach in our sample. As this choice might mirror a bias in the literature itself, further studies should consider measuring other surprisal-based effects that support EBL theories, such as the inverse frequency effect in priming mentioned above, or surprisal induced by non-verb based frequencies, discussed below.

For our search, we have defined PS studies as studies that contain experimental or corpus data from human participants and compare the size of syntactic priming following more versus less predictable prime sentences. We included studies where the level of surprisal was determined based on verb-structure cooccurrence frequencies, the prime sentences were grammatical (did not feature non-alternating verbs in the dispreferred structure) and both the prime and target sentence were from the same human language.

While adapting a narrow PS definition was necessary to allow for comparison between studies, it unfortunately means that some potentially relevant work is beyond the scope of this review. For instance, Bovolenta and Mardsen (2021, 2023) found prime-surprisal-

based learning effects in artificial language learning studies, which demonstrated how surprisal can affect even newly-learned structures (see also Janciauskas, 2017). Buckle and colleagues (Buckle, Lieven & Theakston, 2017) showed that PS effects can be driven not only by verb-structure frequencies, but also by non-prototypical animacy mappings (see also related results in Chapter 5 of Peter, 2015). Finally, some studies (Ivanova et al, 2012; Chapter 6 in Peter, 2015) have examined whether structural priming and PS also appears with non-grammatical sentences (including non-alternating verbs followed by the dispreferred structure). While these latter studies have produced mixed results, such an approach is a promising direction for extending the scope of measurable PS effects.

2.2. Literature search

In line with the PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009), we performed a systematic search of the Google Scholar, ProQuest, Scopus and Web of Science databases to identify all potentially relevant studies. We used the following search strings: "prime surprisal" OR "surprisal-sensitive persistence" OR "prime verb surprisal". Searching for records on Google Scholar including the expression "prime surprisal" was by far the most exhaustive method, identifying 38 out of the 44 records we considered for our review. The last literature retrieval effort was conducted in September 2023.

2.3. Inclusion criteria

As well as 3 entries excluded because the full text was unavailable for retrieval, we have used the following criteria to assess the eligibility of each study in the articles found during our literature search:

1. The entry discussed an analysis of a specific dataset (4 review articles excluded).
2. The entry discussed a dataset that is different from the previously included datasets (13 duplicate entries excluded).
3. The study described meets our PS definition. (11 entries that did not meet our PS definition were excluded: 4 papers did not measure syntactic priming, two did not contrast the level of priming based on surprisal three did not measure the surprisal based on verb-structure frequency, one study featured artificial languages and the last excluded study featured primes and targets from different languages).

These criteria were assessed by the first author as they applied to each entry. After the exclusion of 31 papers we included 13 papers in the final review. We extracted 24 entries from the included papers. Some papers featured multiple studies (e.g., Jaeger & Snider, 2013) and some studies featured multiple participant groups of different ages (e.g., [Omitted for peer-review]; Peter et al., 2015) or of different language backgrounds (e.g., Kaan & Chun, 2018). In these cases we created a separate entry for each study and each participant group (e.g., native speakers and L2 learners or 3- to 4-year-old children and adults) to allow for the comparison of the PS effect in these different groups.

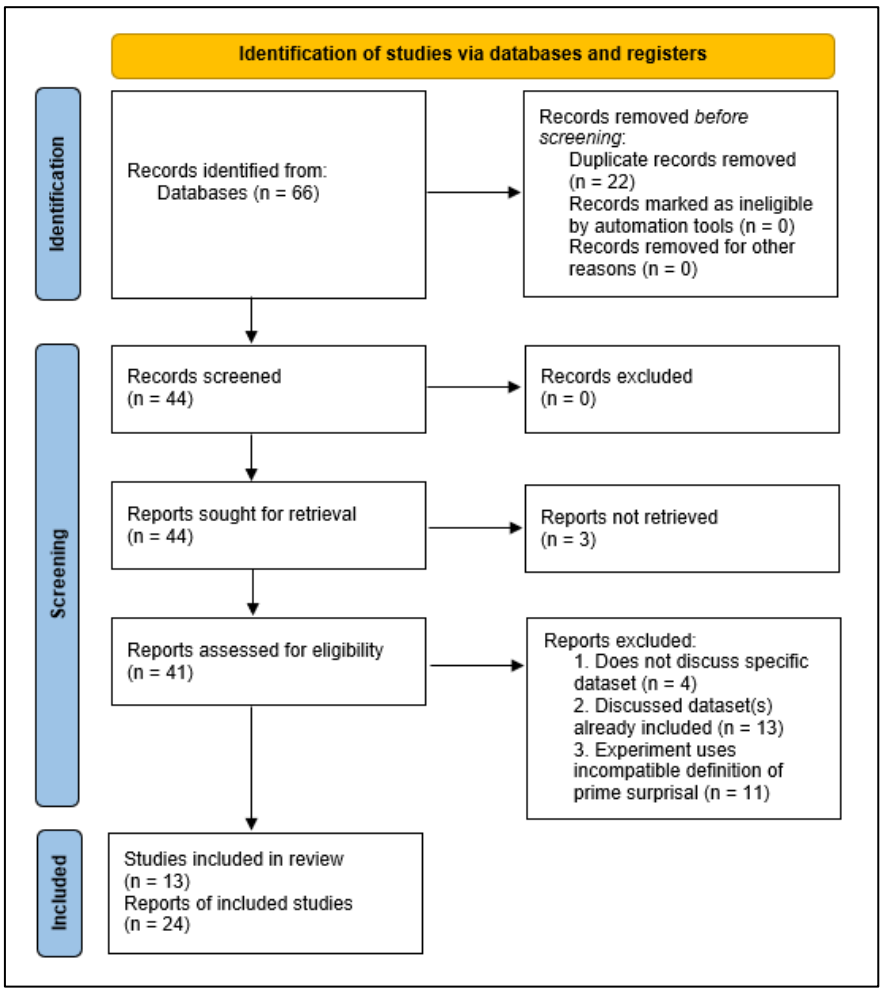

Figure 1: PRISMA flow diagram specifying the study selection process adapted from Page et al (2021).

2.4. Points of comparison

To assess what factors might influence the PS effect in our dataset, we will discuss information we extracted from each study for the moderators listed below.

1. Study language
2. Featured construction
3. Study type
4. Participant age group
5. Participant language
6. Number of participants
7. Number of items
8. Overall number of observations
9. Operationalisation of PS
10. Source of surprisal statistics

11. Statistical test used
12. Did the study find PS?

Data related to all papers found by the literature search (including their exclusion criteria if excluded) can be found on the project's OSF site: https://osf.io/cpd43/ . We also coded the studies for year of publication, source of study (journal or conference title or dissertation), whether the study was published in a peer reviewed outlet, modality of prime, target task, lag between prime and target, lexical overlap between prime and target and whether the study detected structural priming. This information can also be found on the project's OSF site, but we do not discuss these factors in detail here.

Based on what we found across these categories, we will discuss four main topics. First, we will consider how often PS was reported in our dataset. Second, we cover the analysis strategies employed across the studies included. Third, we discuss how diverse the approaches were in terms of factors such as the variety of languages, participants, and the targeted syntactic structures. Lastly, we consider the issue of statistical power and the reliability of the reported results.

Author(s)	Year	Outlet	Study language	Construction type	Study type	PPT Age Group	PPT Language	Result	Structural priming
Jaeger & Snider	2008	Proceedings of the 30th annual conference of the Cognitive Science Society	English	Active-passive alternation	Reanalyses of corpus of conversational speech	Adults	Native English	PS, inconclusive	NA
Bernolet & Hartsuiker	2010	Cognition	Dutch	Dative alternation	Production priming	Adults	Native Dutch	PS, inconclusive	Y
Fine & Jaeger	2013	Cognitive Science	English	Dative alternation	Comprehension experiment reanalysis	Adults	Native English	PS, significant	N
Jaeger & Snider	2013	Cognition	English	Dative alternation	Reanalysis of corpus of conversational speech	Adults	Native English	PS, significant	N
Jaeger & Snider	2013	Cognition	English	Dative alternation	Reanalysis of production priming	Adults	Native English	PS, significant	Y
Jaeger & Snider	2013	Cognition	English	Dative alternation	Production priming	Adults	Native English	PS, significant	Y
Fernandes	2015	Dissertation	Portuguese	Locatives	Self-paced reading task	Adults	Native European Portuguese	PS, significant	N
Peter et al	2015	Journal of Memory and Language	English	Dative alternation	Production priming	Children (3-4)	3-4 year old children	PS, significant	Y
Peter et al	2015	Journal of Memory and Language	English	Dative alternation	Production priming	Children (5-6)	5-6 year old children	PS, significant	Y
Peter et al	2015	Journal of Memory and Language	English	Dative alternation	Production priming	Adults	Adults	PS, inconclusive	Y
Peter	2015	PhD dissertation	English	Dative alternation	Production priming with memory task	Adults	Monolingual English	No evidence for PS	Y
Perdomo	2017	PhD dissertation	English	Directional phrasal verb constructions, continuous vs discontinuous construction	Production priming	Adults	Native English	PS, significant	N
Perdomo	2017	PhD dissertation	English	Directional phrasal verb constructions, continuous vs discontinuous construction	Production priming	Adults	Native Mandarin, English learners	No evidence for PS	N
Perdomo	2017	PhD dissertation	English	Directional phrasal verb constructions, continuous vs discontinuous construction	Production priming	Adults	Native Spanish, English learners	PS, inconclusive	N
Kaan & Chun	2018	Bilingualism: Language and Cognition	English	Dative alternation	Production priming	Adults	American English	No evidence for PS	N
Kaan & Chun	2018	Bilingualism: Language and Cognition	English	Dative alternation	Production priming	Adults	Korean L2, English learners	No evidence for PS	N
Fazekas	2020	PhD dissertation	English	Dative alternation	Production priming	Adults	Native British English	PS, inconclusive	N
Fazekas	2020	PhD dissertation	English	Active-passive alternation	Production priming	Adults	Native British English	No evidence for PS	Y
Fazekas	2020	PhD dissertation	English	Dative alternation	Production priming	Children (5-6)	Native British English	PS, inconclusive	Y
Fazekas	2020	PhD dissertation	English	Dative alternation	Production priming	Adults	Native American English	PS, significant	Y
Fazekas et al	2020	Royal Society Open Science	English	Dative alternation	Production priming	Adults	Native British English	PS, significant	N

Fazekas et al	2020	Royal Society Open Science	English	Dative alternation	Production priming	Children (5-6)	Native British English	No evidence for PS	Y
Arai & Van Gompel	2022	Quarterly Journal of Experimental Psychology	English	Dative alternation	Production priming	Adults	Native British English	No evidence for PS	Y
Darmasetiyawan et al	2022	Collabra: Psychology	English	Active-passive alternation	Online production priming	Adults	Native English	No evidence for PS	Y

Table 1: Summary table of the studies included

262

3. Results

3.1. Prevalence of PS effects

In order to see how consistently the PS effect is observed, we will first discuss how often it was detected in our sample. Most studies in our dataset measure PS by assessing the interaction of prime structure and verb bias in a dataset including both alternating structures (e.g. both DODs and PDs). As such, we coded the PS effect as significant if a study reported a significant prime structure - verb bias (or verb bias match) interaction showing enhanced priming after more surprising primes. Based on this criterion, we have categorized the studies using two different schemes. Firstly, we simply split the studies that found a significant PS effect from those that did not. Secondly, we have further split the studies that did not detect significant PS into two categories by adding an additional "Inconclusive" category. This category includes studies that only analysed the data following the different prime structures separately and found PS after one structure but not the other. For instance, Jaeger and Snider (2008) measured the magnitude of priming after passive and active sentences separately in the Treebank corpus and found that surprising passives led to enhanced priming effects, but surprising actives did not. The "Inconclusive" group also includes studies which found an effect in the expected direction (larger priming after surprising structures) but this effect did not reach significance (see e.g., Peter et al., 2015). The goal of this further split was to allow for the possibility that the PS effect did not reach significance in some studies due to either low statistical power or the specific statistical analyses chosen (e.g. no assessment of the prime structure and verb-bias interaction).

When examining the studies based on the first categorisation, we found that the majority of studies do not report significant PS. In our sample, only 10 out of 24 (41.67%) of studies found significant PS effects. Introducing the third, inconclusive category paints a somewhat more complex picture. Here we find that, in addition to the 10 studies that found significant PS, a further 6 studies reported inconclusive results. Three studies reported PS in a subset of the data, and three studies showed a non-significant effect in the same direction as would be caused by the expected PS effect. Thus, it is possible that participants in these additional six studies were sensitive to the prime's surprisal, but due to insufficient statistical power or the statistical analyses chosen the PS effect was not detected. This possibility requires further investigation. Crucially, however, eight out of 24 (33.34%) of the studies in our sample did not report any results that are in line with a PS effect.

Regardless of the categorisation we use, these results clearly show that PS does not always appear when the magnitude of syntactic priming is contrasted after more versus less surprising sentences. In the following sections, we will discuss the different experimental circumstances under which PS does and does not appear. First, we will discuss how these results are shaped by the analysis strategies employed.

3.2. Analysis strategies

The analysis strategies chosen naturally affect the results described across each study. Despite the relative novelty of the PS method, there is a general consensus in the tests

used to assess PS. All production priming studies in our data have used variations of logistic mixed effect models (Baayen, Davidson & Bates, 2008; Jaeger, 2008) to analyse their binary outcome measures. The slight variation in this strategy is seen in the two earliest studies (Jaeger & Snider, 2008 and Bernolet & Hartsuiker, 2010), which analysed their data separately after the two alternative prime structures. All later studies included both prime structures in their models and assessed the interaction of a prime structure and a surprisal predictor.

However, despite the broad agreement on the most appropriate statistical tests to use to analyze PS studies, consensus is absent regarding how in how a key metric – the surprisingness of the prime – is quantified. There is significant variation in how surprisingness is defined, measured and even entered in the models. We have added this information to each study on the excel sheet on OSF, but due to this large variation it was not possible to make concrete comparisons based on this.

The first source of variation we discuss is how the verb's subcategorization bias is defined. Studies can compute an alternation verb bias, only taking into account the two meaning-equivalent structural alternatives (e.g. give appears in DOD/give appears in all datives). However, it is also possible to compute an overall verb bias, in which case all possible subcategorization frames of the verb are considered (e.g. give appears in dative / give appears in any structure; see Jaeger and Snider 2008 for more discussion). When choosing between these two alternatives, it is crucial to consider the relevant study design. For instance, if participants see a picture or video before the prime sentence, this might reliably constrain their structure prediction to one of the meaning-equivalent structural alternatives (e.g. DOD or PD). In this case, the more typically used alternation verb bias might be the most fitting measure. However, if the potential structure predictions are not constrained, the overall verb bias (allowing for all possible structure competitions) might be a better choice.

Another source of variability across PS studies is the source of the surprisal measures and the period over which surprisingness is determined. The period can either include previous experience in the study (e.g., Fernandes 2015), or a metric that aims to approximate participants overall previous language experience (e.g., Peter et al., 2015). Measuring previous experience within a study is relatively straightforward, as authors readily have access to all the relevant data. It is significantly more challenging to find a measure that approximates participants' overall linguistic experience; thus, there is much more variation in how this is done. The two key ways in which this measurement is typically obtained is either via a norming study or corpus-based estimates. Norming studies can vary further depending on the task they use. For instance they can use picture description, forced choice, sentence completion or sentence judgement tasks. These are, of course, all viable avenues in themselves, but the overall approach is hardly standardised.

While the source of corpus-based estimates is relatively clear, how these estimates translate to appropriate surprisal metrics is less straightforward. Corpus-based estimates tend to operationalise surprisingness in terms of a verb's bias towards one construction versus a nearly-synonymous rival. For example, give is usually treated as a DO-biased verb because corpus data shows that it appears approximately 80% of the time in a DO structure (e.g., Ambridge et al., 2018). Yet the vast majority of the time, the DO construct in question lacks an overt subject and uses a pronoun (e.g., Give me that!), with potentially limited relevance to understanding the surprisingness of the longer, more formulaic sentences used as stimuli (e.g., The man gave the woman a present/The man gave a present to the woman). This disparity between the corpus-based surprisal estimates

used in PS studies and the actual sentence stimuli they are applied to may have contributed to the variability in the results.

Finally, even when the level of surprisingness is determined, authors can still make different measurement choices. For instance, studies can operationalise predictability as either a graded or an overall measure (see e.g., Arai and Van Gompel,2022 versus Peter et al.,2015). Overall predictability is a binary measure, where structures are either classified as surprising (due to a mismatch between verb and structure) or as predictable (matching verb and structure). Graded predictability operates with a continuous surprisal measure, where each verb has an assigned score depending on how likely is it to appear with a given structure (based on the previously-chosen metric). It is worth noting that while it is not advised to dichotomise continuous variables (e.g. Royston, Altman & Sauerbrei, 2006), some studies might have found PS effects with this method by contrasting extreme cases: very surprising versus very predictable sentences. In this case a dichotomous variable might still have captured a large part of the relevant variance.

When considering extreme cases, it is also worth noting that with the statistical power most existing studies had, it is possible that only such cases (very surprising versus very predictable sentences) could be successfully contrasted. But for PS to reflect the kind of learning mechanism proposed by EBL theories – and for the paradigm to be of sustained utility in exploring these theories – it is not sufficient for it to remain detectable only in extreme, item-based edge cases. Future research will have to contrast PS after various levels of surprisal in appropriately powered studies.

It is also worth considering that verb-based PS can in principle be measured in any structural priming study, even if the study was not originally set up to test this effect. This is possible because almost all sentences include verbs and a syntactic structure, for which verb-structure frequencies can be obtained (for instance from a corpus, norming study or based on previous experience in the experiment). Indeed, the majority of the 24 studies in our sample were either the reanalysis of a previous, non-PS focused study (5 studies) or an exploratory analysis in a study set up to test another research question (8). While using exploratory PS analyses to target data from existing studies in such a way is a promising avenue for learning more about the PS effect, the conclusions based on such results are necessarily less strong. When studies are not set up directly to test PS, the stimuli, experimental set-up and verb choice are likely to reflect other experimental questions and can make a possible PS effect less strong and thus harder to detect.

3.3. Variety

In this section, we will review the studies in our sample based on the different languages, participant groups, grammatical structures, and modalities they examine.

3.3.1. Stimuli language

Similarly to structural priming studies in general (see Mahowald et al., 2016), PS studies chiefly feature English sentences. In fact, only two studies in our sample featured materials from a language other than English. Both Bernolet and Hartsuiker's (2010) production priming study looking at the dative alternation in Dutch and Fernandes' (2015) comprehension priming study looking at Portuguese locatives found significant PS

effect with adult speakers. While it is noteworthy that these two studies suggest that PS effects can be detected in languages other than English, further studies are needed.

While this is hardly an uncommon issue in language development research (Kidd & Garcia, 2022) further investigation is vital here. For one, if a PS effect does indeed reflect the existence of a key underlying learning mechanism per EBL accounts, then it should be detectable across all languages. Furthermore, the ambiguity of English-language results suggests that a wider examination of individual languages is needed to better understand and delineate when and how the effect can be detected. Given that English features only a limited number of grammatical forms that can be operationalised effectively in an experimental context, research in other languages will be necessary to map out the phenomenon more fully.

3.3.2. Participant groups

A key characteristic of EBL theories is that they propose a life-long learning mechanism, meaning that they can be meaningfully assessed with studies featuring adult participants. However, to assess the learning mechanism when it is the most active, it is particularly crucial to test these theories among cohorts at the early stages of language acquisition, in which we would expect the learning process to be heightened. We therefore assessed how many PS studies feature groups learning their first or second language. While the majority of our sample (20 out of 24 studies) were conducted among adult native speakers, we also found four studies including child participants and two studies with L2 speakers.

Both of these participant groups showed mixed results. The two studies measuring PS in second language learners included English directional phrasal verb constructions with either Mandarin or Spanish speakers learning English. Neither of these studies found significant PS effects. It is worth noting that the main goal of these studies was not to measure the PS effect, but rather the acquisition of phrasal verbs in L2 learners of English. It is therefore possible that they were not ideally set up to measure the PS contrast.

When it comes to the child studies, only two out of four studies found significant effects of PS. While two studies found no PS with 5- to 6-year-old children in a dative production study ([Omitted for peer-review]), Peter and colleagues (2015) found significant PS effects with both 3- to 4- and 5- to 6-year-old children. At first glance these results seem hard to reconcile, especially as all four studies examined PS with English datives using a similar paradigm. However, it is worth noting that while the main target of Peter and colleagues' study was PS in different age groups, the main goal of the other studies was not to detect the PS effect. [Omitted for peer-review] study was an underpowered pilot study, while [Omitted for peer-review]'s study measured PS only as a partially between-subject variable. Thus, the studies where the main focus was detecting PS in different age groups (Peter et al, 2015) did find the predicted effects, while those with a different focus that only looked at PS in exploratory analyses [Omitted for peer-review]did not.

Importantly, one of the main goals of Peter and colleagues' paper was to compare the magnitude of the PS effect in different age groups: 3- to 4-year-olds, 5- to 6-year-olds and adults. This allowed them to assess another important prediction of EBL theories: that the PS effect is larger for younger participants, so long as they had already acquired knowledge of the verb-structure frequencies relevant to the PS effect. EBL theories predict stronger PS (as well as priming) effects in younger age groups because – due to their limited exposure to the language itself – their linguistic representations are less stable

and more malleable. Thus these representations would be expected to shift more in response to the error signal resulting from incorrect predictions, leading to larger priming and learning effects. Peter and colleagues did indeed find that the PS effect was the largest in the youngest participant group included in the study, and gradually decreased in magnitude across the two older groups.

While it is promising that participant groups other than adult native speakers are included in some PS studies, it is difficult to draw firm conclusions based on such a limited number of studies, some of which only assessed PS in exploratory analyses. More studies focusing specifically on learners are necessary to get a comprehensive picture of how PS works as a measure of the proposed EBL mechanism.

3.3.3. Modality

While most studies in our dataset examined the PS effect in production priming studies, we also found two studies that assessed PS in comprehension. Fine and Jaeger (2013) reanalysed eye-movement data from Thothathiri and Snedeker's (2008) dative comprehension priming experiment and found significant PS based on the first (but not the second) prime sentence. Fernandes (2015) also found significant PS in a moving-window self-paced reading task including Portuguese locatives. Similarly to the previous categories, there are not enough studies that look at PS in comprehension to draw strong conclusions or comparisons when it comes to the modality of the PS effect. However, it is promising that PS in comprehension has already been demonstrated in two studies featuring entirely different methodologies and languages.

3.3.4. Variety in Structures

Another key way in which we can attempt to compare PS studies is through the syntactic structure they feature in their stimuli. Structural priming studies in general tend to favour the dative and active-passive alternations (Mahowald et al., 2017) and this tendency is also found in PS studies. The majority of studies (17 out of 24) included the dative alternation, but we also found three studies including the English active-passive alternation (Darmasetiyawan et al., 2022; [Omitted for peer-review]; Jaeger & Snider, 2008). The remaining four studies contained English directional phrasal verb constructions (Perdomo, 2017) or Portuguese locatives (Fernandes, 2015). Here, we will concentrate on the dative and active-passive studies as our search did not uncover sufficient studies featuring other structures.

The most salient observation in connection with dative and active-passive studies is that while the appearance of PS varies with datives, no active-passive study in the whole dataset reported significant PS (but see Jaeger and Snider's (2008) work, who found significant PS after passive, but not active primes). Based on this contrast, it is worth considering whether the active-passive alternation might be a worse candidate for PS than the dative alternation. In a PS paper featuring both datives and passives, [Omitted for peer-review] argue that while the dative is the perfect candidate alternation to show PS, the active-passive is not. They suggest that a key difference between the two structures is the relative location of the verb (that sets up the surprisal effect in verb-based PS studies) and the structure decision point (where it becomes clear to the listener which alternative structure they are hearing, and therefore where surprisal may be expected to happen). In dative sentences, verbs always appear early, thus they can set up an expectation that is

then either confirmed or violated by the subsequent structure. For instance, in the sentence "Lisa gave a ball to Bart." The DOD-biased verb *give* sets up an expectation of the DOD structure, which is then violated when participants hear a PD structure. The situation is very different in passive sentences where the verb typically does not precede the structure decision point. For instance, in the sentence "Lisa was scared by Bart", when participants hear *was* it indicates that they are hearing a passive sentence before the active-biased verb *scared* follows. It is possible that it is essential for verb-based PS studies to feature the verb first, with the structure revealed second.

Another potential reason why a PS effect is not observed in active-passive studies is the relative surprisingness of the structural alternatives. As Darmasetiyawan and colleagues (2022) note, the passive structure is extremely surprising regardless of the identity of the verb. For instance, it only constitutes around 1% of all verb uses in the British National Corpus (based on the frequencies reported in Ambridge et al., 2016). This would mean that all verb+passive combinations are inherently extremely surprising, making it harder to detect the small differences in these already high surprisal levels. In contrast, the distribution of DOD and PD datives is much more equal.

Overall, in the current dataset the only structure that showed PS in more than one study is the dative alternation, while all active-passive studies failed to find significant PS. In addition to replicating the above effects in well-powered studies, it is therefore crucial to explore PS in further structures. A good candidate for this could be the locative structure, in which the verb always precedes the structure decision point in English and both structural alternatives are relatively common. As noted above, PS using locatives has been successfully observed in Portuguese (Fernandes, 2015), lending further credence to the notion that English-language studies would represent a useful next step for research in this area.

3.4. Statistical power

Naturally, to obtain trustworthy results about any phenomenon it is crucial to assess whether the studies examining it had sufficient statistical power. Our initial intention was to conduct a meta-analysis on the data from the studies in our sample to determine the effect size for PS overall and across different groups, as well as to assess whether the existing studies had sufficient statistical power. Unfortunately, a meta-analysis has proven not to be possible based on the available data. Most papers included in our sample did not report all the metrics (e.g., standard errors, standard deviations, or exact p-values) that are necessary to compute standardised effect sizes, and the scope of the current project did not allow for requesting raw data from the authors of the featured publications as an alternative.

To gain some speculative insight into the effect sizes PS studies yield and whether the current studies are sufficiently well-powered to detect such a contrast, we collected the average participant and item numbers and effect size estimates in log odds ratio changes from production priming studies that reported such a metric for the interaction between prime structure and level of surprisal. This comparison included 18 studies. The studies and the corresponding estimates are reported on the project's OSF site. Based on these values we determined the average PS effect to be .6 in difference in log odds ratio. We also computed separate effect sizes for studies including datives: 0.84 and two active-passive structures: 0.11. The 11 adult dative studies showed an average .71 effect size, while the 4 child dative studies a larger effect of 1.14. We have also determined the

average participant number to be 89.11 (ranging from 24 to 392) and average item number to be 24.71 (ranging from 16 to 80) in our studies.

This then allows us to compare the participant and item numbers in the studies in our dataset to the sample size recommendations in Mahowald and colleagues' syntactic priming meta-analyses (2016). According to Mahowald's sample size recommendations, the participant and item numbers in our studies would only yield over 90% statistical power to detect an interaction if the interaction coefficient was 1 in difference in log odds ratio in a PS study. This average effect size is only approximated in child dative studies, where the sample sizes were significantly below average (in the four included child studies the average item number was 17 (overall average 24) while the average participants number was 58.25 (overall average 89)). Naturally, this is a crude attempt to assess statistical power, but the large gap between suggested sample sizes in Mahowald and colleagues' work and those featured in our paper allows for the preliminary conclusion that the studies in this sample are typically underpowered.

4. Discussion

Despite being a relatively new method, the PS paradigm has already shown huge potential for addressing key questions in psycholinguistic research, especially concerning prediction's role in language acquisition. However, also due to its relatively recent emergence, there are still many open questions regarding how this method can be used and what the wider pattern of results signifies. In particular, there is a lot left to learn about the specific circumstances the PS effect appears in, and the best ways to analyse these results. In turn, these questions limit how far existing results can be interpreted as support for various hypotheses and how this paradigm is suited to address additional questions in the future.

To gain a clear picture of how much we now know about the PS effect, we conducted a systematic review of PS studies and offered an overview of their scope and variety (in terms of languages, participant groups, the modality they examine, the structures they feature and the statistical methods they use). The key conclusion we can draw based on the studies in our review is that the PS effect is not yet well established. A majority of studies in our sample do not report significant PS effects and based on our preliminary calculations studies typically did not have sufficient statistical power to reliably detect the PS interaction. Furthermore, PS effects also do not appear consistently in subgroups of studies (e.g., studies with children or including datives), although there are certain categories of studies that report PS more often than others. This leads to the inevitable conclusion that more research is needed in order to definitively establish the existence and scope of the PS effect.

We also examined diversity in our dataset in terms of stimuli language, age and language background of the participants and whether the study targeted production and/or comprehension. However, across most factors examined, we did not find much variation in our data: most papers include English sentences, adult native English participants and/or assess comprehension to production priming. There are some exceptions to this pattern: for instance, two studies looking at different languages (Dutch and Portuguese: Bernolet and Hartsuiker, 2008 and Fernandes, 2015), two studies looking at comprehension to comprehension priming (Fernandes, 2015; Fine & Jaeger, 2013) and a small number of studies targeting L2 learners (Kaan & Chun, 2018; Perdomo, 2017) and children ([Omitted for peer-review]; Peter et al., 2015). However, in most categories there was not enough variety to make substantive within-group comparisons.

There were two categories for which we found more variation: age group and featured syntactic structures. When it comes to syntactic structures, PS seems to be almost impossible to detect with active-passive structures, but features somewhat more reliably in studies targeting datives. However, as many dative studies are exploratory and/or underpowered, more comprehensive dative studies are still needed to solidify findings regarding this effect.

When considering adult versus child participants, PS is not detected consistently across either group, but the effect sizes are larger with children compared to adult studies. This difference is also supported by Peter and colleagues' (2015) work comparing PS in three different age groups within one study and which found that PS effects decrease with age. Given the key expectation of error-based theories that children's linguistic states are more malleable, and therefore more susceptible to change via surprisal, this result is suggestive but not in itself definitive, especially given the relatively low number of child studies available.

Crucially, this dataset also suggests that the appearance of a PS effect cannot be directly equated with linguistic predictions in a given study. For instance, in our dataset there is no reliable PS effect found in studies with English active-passive sentences, even though studies using other experimental paradigms suggest that adult native English speakers make linguistic predictions in such sentences (see e.g. Heilbron and colleagues,2022). This in turn indicates that we cannot be confident that PS always appears when participants are making predictions (and creating error signals) in specific studies. This is something that needs to be considered when evaluating both existing evidence and the future of studies utilising PS-based paradigms, as discussed in further depth below.

Next, we will summarize our recommendations for future research using the PS paradigm. It is worth emphasizing that these studies would also broadly benefit from practices associated with open and replicable science, such as pre-registration, attempts to replicate previous results, and open data, materials and code (see e.g. Crüwell et al., 2018; Nelson, Simmons & Simonsohn, 2018; Kathawalla, Silverstein & Syed, 2021). Furthermore, especially as many studies in our sample seem to be underpowered, utilizing practices that prevent inconclusive results due to insufficient statistical power is especially important. In addition to carrying out accurate power calculations and using more accurate surprisal estimates for the sentences involved, we also recommend considering Bayesian analyses methods, potentially combined with pre-specified stopping rules (see Dienes, 2014 & 2016; Bürkner, 2017). A key advantage of these approaches is that (unlike frequentist methods) they can determine when a non-significant result provides support for the null-hypothesis as opposed to when it indicates data insensitivity. Furthermore, when using Bayesian approaches, pre-specified stopping rules can be set both for and against the null-hypothesis. These approaches can ensure sufficient statistical power with the minimal participant numbers needed, which is especially useful when working with harder-to recruit populations such as children. Next, we will turn to suggestions more specific to PS studies.

First, it is useful to differentiate between two kinds of studies featuring the PS paradigm. The first category of studies aims to better understand the PS effect itself, including questions such as under which circumstances we should expect PS to occur if the participants in the study are engaging in linguistic prediction. The second kind of study aims to use PS as a tool to examine a theoretically important question, such as whether a specific group of participants engage in predictive learning (as detected by PS). A study can be informative for both categories, but the intention when designing the study needs to be clear, not just to the eventual audience for the research, but also in terms of shaping how exactly the study seeks to utilise the paradigm. It is worth noting that these two

research strands have significant synergy – demonstrating the connection between PS and error-based learning enhances the significance of studies targeting PS in new contexts. Conversely, demonstrating that PS does operate across languages and structures strengthens the claim that it reflects a broader predictive learning mechanism in line with EBL theories.

One obvious future direction for studies targeting the scope of PS is to expand the syntactic structures and languages the effect can be detected in. In terms of languages, if the PS effect is really as universal as claimed, it has to show in a wide variety of languages and structures. As far as we are aware, PS has only been targeted in 3 languages (English, Dutch and Portuguese) so far. Thus, the possibilities in terms of languages are limitless. In terms of syntactic structures, there remains scope even within English-language studies to target structures, such as locatives, that are more rarely tested but might be expected to work experimentally.

Furthermore, in order to confirm existing results – namely that PS can appear with dative structures in English speakers – it is also important to overcome several persistent methodological problems, such as low sample sizes and potentially inaccurate surprisal measurements. This could be done via large scale studies that determine sample sizes based on accurate power calculations and use more accurate surprisal estimates for the sentences involved, considering the relevant study design and the specific sentences featured in the study (see section 3.2.). One such effort by [Omitted for peer-review], Blything and Ambridge is already underway, and uses frequencies generated via Large Language Models on the exact sentences featured in the study to compute more precise surprisal estimates.

Another inherent limitation of the current PS literature stems from the lack of on-line measurements. While the PS paradigm can directly address potential changes in language production depending on the predictability of the input, behavioural observation methods do not give us any detailed insight into what the processing differences are between surprising and predictable sentences that lead to increased repetition or learning. Future work combining the PS method with on-line measures such as EEG or eye-tracking could help us map the location and nature of the processing differences guiding these results. Neurological measures would be particularly desirable here, as they provide precise temporal (and some spatial) information on sentence processing (see [Omitted for peer-review]).

In addition, as PS effects can be computed based on any structural priming study, there is useful scope for a meta-analysis drawing on a large pool of existing data. In principle, such an analysis could be done on a collection of any raw structural priming data that includes the sentences used in the study. Even if not yet included in the materials, surprisal estimates can be computed, such as by using norming or language model data to then assess the magnitude of the priming depending on the level of surprisal of the prime sentence. This kind of reanalysis approach has already been used in individual PS studies (e.g. Jaeger & Snider, 2008; Fine & Jaeger, 2013) and could be extended to a larger scale meta-analysis as well.

Another potential avenue for future research is using PS as a tool to measure whether error-based learning occurs in specific groups of people (e.g., children of various ages, see Peter et al.,2015) or in specific circumstances (e.g., in cumulative, delayed contexts such as [Omitted for peer-review]). These studies can be informative about whether such groups engage in EBL. In this case, we suggest a different approach to the one described above. As mentioned earlier, it is problematic if a variation of the PS paradigm (or indeed any experimental method testing any specific effect) is used with new participant

groups or in situations when we cannot yet be relatively confident that the measure is actually capable of reliably detecting the effect of interest, at least in a typical population. Thus, we suggest that – in addition to using well-supported participant numbers and surprisal measurements – studies that plan on measuring EBL via verb-based PS in new populations use syntactic structures where the verb consistently precedes the structure and the level of surprisal is balanced between structures. For instance, the English dative (or a similar structure in other languages) is a good candidate for new PS studies as it is the structure that currently has the most consistently supportive evidence of PS and one that also fits the above criteria.

Furthermore, when testing PS with new participant groups, it helps the interpretation of the results if a group of typical adults (who are generally easier to recruit) are also tested. If PS reliably appears in the typical but not the new group, the difference is likely to come from the differences between the participant groups rather than the underlying approach, allowing for stronger conclusions that the new group likely does not engage in predictive learning in this context. In contrast, if neither group show PS, it is more likely that the set-up did not allow for detecting PS in the first place.

5. Conclusions

The emerging evidence surrounding PS paints a mixed picture. On one hand, it remains a promising avenue for experimentally targeting the learning mechanism proposed by error-based learning theories. Some early results – such as detecting PS with English datives and higher effect sizes among children – are indeed in line with these theories' expectations. However, the difficulty in detecting the PS effect in a range of structures beyond datives and the limited number of languages covered by existing work makes any claim that PS is intrinsically connected to predictive processes (let alone learning) premature. We propose that alongside addressing methodological issues such as low sample sizes and inconsistent surprisal measures, future research should better distinguish between studies seeking to expand our knowledge of the contexts in which PS occurs (particularly regarding syntactic structures, learning level and languages) and studies seeking to use PS as a tool to examine predictive learning. The latter studies should focus chiefly on methods and structures in which a reliable PS effect has been established previously to avoid the risk of ambiguous results in the absence of an effect.

Author Contributions: **[**Omitted for peer-review**]**

Funding: This research received no external funding.

Informed Consent Statement: Not applicable as no new data requiring ethical approval was collected for this work.

Data Availability Statement: **[**Omitted for peer-review**]**

Conflicts of Interest: The authors declare no conflict of interest.

References

Arai, M., & van Gompel, R. P. (2022). Lexically independent representation of the monotransitive structure. Quarterly Journal of Experimental Psychology, 75(9), 1773-1789.

Baayen, R. H., Davidson, D. J., Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. Journal of Memory and Language, 59, 390–412.

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming Cognition, 114(3), 455-461.

BNC Consortium. (2007). British national corpus. Oxford Text Archive Core Collection.

Bock, K. (1986). Syntactic persistence in language production. Cognitive Psychology, 18, 355-387.

Bürkner, P. C. (2017). Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:1705.11123.

Bovolenta, G., & Marsden, E. (2021). Expectation violation enhances the development of new abstract syntactic representations: Evidence from an artificial language learning study. Language Development Research, 193-243.

Bovolenta, G., & Marsden, E. (2023). The effect of verb surprisal on the acquisition of second language syntactic structures in adults: An artificial language learning study. Applied Psycholinguistics.

Buckle, L., Lieven, E., & Theakston, A. L. (2017). The effects of animacy and syntax on priming: A developmental study. Frontiers in psychology, 8, 2246.

Chang, F., Dell, G. S., Bock, K. (2006). Becoming syntactic. Psychological review, 113, 234. (doi:10.1037/0033-295X.89.1.1)

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J., ... & Schulte-Mecklenbeck, M. (2018). 7 easy steps to open science: An annotated reading list.

Darmasetiyawan, I. M. S., Messenger, K., & Ambridge, B. (2022). Is Passive Priming Really Impervious to Verb Semantics? A High-Powered Replication of Messenger Et al.(2012). Collabra: Psychology, 8(1), 31055.

Dell, G. S., Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. Phil. Trans. R. Soc. B, 369, 20120394. (doi:10.1098/rstb.2012.0394)

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. Frontiers in psychology, 5, 781.

Dienes, Z. (2016). How Bayes factors change scientific practice. Journal of Mathematical Psychology, 72, 78-89.

[Omitted for peer-review]

[Omitted for peer-review]

Ferreira, V. S., 2003. The processing basis of syntactic persistence: We repeat what we learn. Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver, Canada.

Fernandes, E. G. (2015). Syntactic priming as a window into the representational and experiential basis of syntactic processing in comprehension.

Fine, A. B., & Florian Jaeger, T. (2013). Evidence for implicit learning in syntactic comprehension. Cognitive Science, 37(3), 578-591.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of memory and language, 28(5), 501-518.

Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. Proceedings of the National Academy of Sciences, 119(32), e2201968119.

Ivanova, I., Pickering M. J., McLean J. F., Costa A., & Branigan H. P. (2012). How do people produce ungrammatical utterances? Journal of Memoryand Language, 67, 355-370.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, 59, 434–446.

Jaeger, T. F., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In Proceedings of the Cognitive Science Society Conference (pp. 1061-1066).

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, 57–83.

Janciauskas, M. (2017). Bridging between on-line linguistic adaptation and long-term language learning. The University of Liverpool (United Kingdom).

Kaan, E., & Chun, E. (2018). Priming and adaptation in native speakers and second-language learners. Bilingualism: Language and Cognition, 21(2), 228-242.

Kapatsinski, V. 2006. Towards a single-mechanism account of frequency effect. LACUS Forum, 32: 325-335.

Kapatsinski, V. (2007). Frequency, neighborhood density, age-of-acquisition, lexicon size, neighborhood density and speed of processing: Towards a domain-general, single-mechanism account. In Proceedings of the 6th Annual High Desert Linguistics Society Conference, 121-40.

Kathawalla, U. K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. Collabra: Psychology, 7(1), 18684.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research?. First Language, 42(6), 703-735.

Luce, Paul A., Stephen D. Goldinger, Edward T. Auer & Michael S. Vitevitch. 2000. Phonetic priming, neighborhood activation, and parsyn. Perception and Psychophysics 62:615–25.

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. Journal of Cognitive Neuroscience, 26, 2530–2539.

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.

Messenger, K., Branigan, H. P., McLean, J. F., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. Journal of Memory and Language, 66, 568-586.

Moder, Carol L. 1992. Productivity and categorization in morphological classes. Ph.D. diss.: SUNY Buffalo.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine, 151(4), 264-269.

Novembre, G., Keller, P. E. (2011). A grammar of action generates predictions in skilled musicians. Consciousness and cognition, 20, 1232-1243.

Page, M. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Systematic Reviews, 10, 89.

Perdomo, M. (2017). Second Language Production and Processing of Phrasal Verb Constructions: A Priming Study (Doctoral dissertation, University of Florida).

Peter, M. S. (2015). The Role of the Verb in the Development of Syntax: Evidence from the Structural Priming Paradigm (Doctoral dissertation, University of Liverpool).

[Omitted for peer-review**]**

Peter, M. S., & Rowland, C. F. (2019). Aligning developmental and processing accounts of implicit and statistical learning. Topics in cognitive science, 11(3), 555-572.

Rabagliati, H., Gambi, C., Pickering, M. J. (2016). Learning to predict or predicting to learn? Language, Cognition and Neuroscience, 31, 1-12.

Ramscar, M., Dye, M., McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. Language, 89, 760–793.

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in medicine, 25(1), 127-141.

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: what is next?. Trends in Cognitive Sciences.

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three-and four-year-old children. Journal of Memory and Language, 58(2), 188-213.

Urgesi, C., Savonitto, M. M., Fabbro, F., Aglioti, S. M. (2012). Long-and short-term plastic modeling of action prediction abilities in volleyball. Psychological research, 76, 542-560.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.