Journal of Memory and Language 138 (2024) 104532

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

A meta-analysis of syntactic priming experiments in children

Shanthi Kumarage[a,b,*], Seamus Donnelly[b], Evan Kidd[a,c]

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

ARTICLE INFO

Keywords:
Syntactic priming
Meta-analysis
Language acquisition
Syntax acquisition

ABSTRACT

A substantial literature exists using the syntactic priming methodology with children to test hypotheses regarding the acquisition of syntax, under the assumption that priming effects reveal both the presence of syntactic knowledge and the underlying nature of learning mechanisms supporting the acquisition of grammar. Here we present the first meta-analysis of syntactic priming studies in children. We identified 37 eligible studies and extracted 108 effect sizes corresponding to 76 samples of 2,378 unique participants. Our analysis confirmed a medium-to-large syntactic priming effect. The overall estimate of the priming effect was a log odds ratio of 1.44 (Cohen's d = 0.80). This is equivalent to a structure that occurs 50 % of the time when unprimed occurring 81 % of the time when primed. Several variables moderated the magnitude of priming in children, including (i) within- or between-subjects design, (ii) lexical overlap, (iii) structural alternation investigated and, (iv) the animacy configuration of syntactic arguments. There was little evidence of publication bias in the size of the main priming effect, however, power analyses showed that, while studies typically have enough power to identify the basic priming effect, they are typically underpowered when their focus is on moderators of priming. The results provide a foundation for future research, suggesting several avenues of enquiry.

Introduction

Grammar is a core component of language, and thus its acquisition has long been of interest in language acquisition research, where debate about the representational nature and innateness of syntactic representations has been a key battleground. One important method in the toolkit of developmental psycholinguists is *syntactic priming*, which describes the tendency for an individual to use the same grammatical structure they have previously used or heard. For instance, if a child hears the English passive sentence *Ernie was hugged by Bert* and soon after produces another passive such as *Big Bird was surprised by Mr Snuffaluffagus*, they are said to be primed. Far from being mere imitation, priming in this instance reveals the abstract nature of grammatical representations, since the two sentences do not share overlap in their open class lexical items (Branigan & Pickering, 2017). Furthermore, priming has been argued to be a form of implicit learning (Chang et al., 2000, 2006; Dell & Chang, 2014), and is thus potentially reflective of a *mechanism* underlying language acquisition.

Given its ability to reveal both representational knowledge and learning, the syntactic priming methodology has become a prominent one in the field (see papers in Messenger, 2022). In the current paper we present the first systematic review and meta-analysis of experimental papers that have investigated syntactic priming in developmental populations. In what follows we review the past research, highlighting the trends in the literature that we then quantitatively examine.

Past developmental syntactic priming research

One key question in developmental priming research is the abstractness of syntactic representations, tested by manipulating overlap in lexical content between prime and target trials (Branigan et al., 2005; Branigan & McLean, 2016; Foltz et al., 2015; Kumarage et al., 2022; Peter et al., 2015; Rowland et al., 2012; Savage et al., 2003). Priming in the absence of lexical overlap indicates the existence of abstract syntactic representations, whereas priming restricted to trials where prime and target share lexical content, such as a verb (i.e., *Ernie was hugged by Bert* primed *Bird Bird was hugged by Mr Snuffaluffagus*), indicates syntactic knowledge may be more item-based, therefore bearing upon a hotly debated issue in the field (e.g., Ambridge & Lieven, 2015; Fisher, 2002; Tomasello, 2000; Tomasello & Abbot-Smith, 2002).

---

* Corresponding author at: School of Medicine and Psychology, The Australian National University, Canberra 2601, Australia.
E-mail address: shanthi.kumarage@anu.edu.au (S. Kumarage).

https://doi.org/10.1016/j.jml.2024.104532
Received 16 June 2023; Received in revised form 15 April 2024; Accepted 16 April 2024

Thus, a focus on syntactic priming in development inevitably leads to the question of *when* abstract priming effects emerge, which in the literature has informed arguments over the early or late emergence of core grammatical knowledge (e.g., Bencini & Valian, 2008; Huttenlocher et al., 2004).

The distinction between abstract priming and lexically-based priming also bears upon a broader issue concerning the architecture of the language production system. In the mature adult system, where abstract knowledge can be assumed, priming effects are typically larger when the prime and target sentences share open-class lexical content. This is known as the *lexical boost*. Activation-based production architectures attribute both the abstract priming and lexical boost effects to a single mechanism driven by the residual activation of connected lexical and structural knowledge (Pickering & Branigan, 1998), and therefore predict a lexical boost to emerge in development once abstract knowledge has been acquired. Priming as residual activation also predicts that priming is a relatively fleeting event that may not lead to representational change in the linguistic system. In contrast, implicit learning accounts of priming propose that abstract priming and the lexical boost are attributable to separable implicit and explicit memory systems, respectively (Chang et al., 2006, 2012). Since explicit memory processes exhibit a relatively protracted developmental trajectory, the lexical boost is predicted to emerge after abstract priming (Chang et al., 2012; Rowland et al., 2012). Notably, implicit learning accounts also predict that abstract priming leads to representational change via the same mechanism involved in language development. The most prominent of these accounts invokes *error-based learning* (Chang et al., 2006). Specifically, during comprehension, the syntactic processor predicts upcoming input based on syntactic representations that are sensitive to frequency distributions in the input. It compares its predictions to the actual input and responds to prediction errors by updating the weights of its syntactic representations. In syntactic priming experiments, unexpected syntactic structures (e.g., low frequency structures such as the English passive) encountered in primes result in an updating of syntactic weights that increases the likelihood of their later production, thus leading to priming. In experimental terms, the implicit learning account of priming predicts priming effects to have long-term resonance (see Bock & Griffin, 2000). Several studies in the developmental literature have investigated the long-term nature of syntactic priming (Branigan & McLean, 2016; Fazekas et al., 2020; Hsu, 2019; Kidd, 2012; Messenger, 2021), with at least one study showing that it can persist for up to a month (Savage et al., 2006).

An analysis of syntactic priming across development also bears upon the question of whether and how syntactic representations change across developmental time. As a theory of syntactic acquisition, Chang et al.'s (2006) computational Dual-path model, which learns syntactic representations via implicit learning, differs from traditional nativist (Bencini & Valian, 2008; Messenger & Fisher, 2018) and traditional lexicalist (Savage et al., 2003; Tomasello, 2003) accounts of syntax acquisition. Research has compared these theories by examining the developmental trajectories of the abstract priming effect and lexical boost in both cross-sectional (abstract priming: Garcia & Kidd, 2020; Hsu, 2019; both effects: Peter et al., 2015; Rowland et al., 2012) and longitudinal designs (both effects: Kumarage et al., 2022). Notably, whereas nativist accounts predict no appreciable change in priming magnitude across development because representations are largely unaffected by frequency, lexicalist approaches predict an increase because initially lexically-specific representations become more abstract and thus more primeable with experience. Different still, priming as implicit learning predicts that priming will initially increase as children acquire the necessary representations to be primed and then decrease across development, since error-based learning is strongest when representations are weak (for more discussion see Kumarage et al., 2022).

Researchers have manipulated several other prominent variables in the syntactic priming literature. For example, manipulating whether children must repeat the prime sentence before they describe a target aims to investigate whether their syntactic representations are shared across production and comprehension (Gámez & Shimpi, 2016; Huttenlocher et al., 2004; Shimpi et al., 2007). Other studies have tested how children's syntactic representations interface with semantic information by manipulating the animacy or thematic roles of nouns in prime and target sentences (Bidgood et al., 2021; Buckle et al., 2017; Messenger, Branigan, McLean, & Sorace, 2012; Vasilyeva & Gámez, 2015). The past literature is not limited to priming in typically-developing monolingual children. Research in clinical developmental populations has investigated whether syntactic priming is observed in these populations (e.g., children with Developmental Language Disorder (DLD), Leonard et al., 2000; children who stutter, Anderson & Conture, 2004) and whether it is associated with defining clinical features. For example, research on children with Autism Spectrum Disorder (ASD) has investigated if priming is used as a mechanism of conversational alignment and whether it differs from children without ASD (Allen et al., 2011; Hopkins et al., 2016). Other research on children with DLD has investigated whether priming, as a marker of implicit learning, is compromised in comparison to children without DLD (Garraffa et al., 2015, 2018). Crosslinguistic priming studies have investigated whether children share syntactic representations between languages (Gámez & Vasilyeva, 2020; Vasilyeva et al., 2010; Wolleb et al., 2018). For instance, does *Ernie was hugged by Bert* prime *Big Bird è stato sorpresso dal Signor Snuffaluffagus* in Italian-English bilinguals?

In summary, 20 years of research on syntactic priming in children has investigated a range of theoretical questions regarding the representation and acquisition of syntax. A past meta-analysis of syntactic priming in adults aggregated evidence to provide key summary findings: abstract syntactic priming decays slowly, lexical overlap provides a large boost to the effect, and this boost decays quickly but is more pronounced in a speaker's second language (Mahowald et al., 2016). The substantial literature in children now warrants a similar quantitative summary, which we present in this paper.

The current study

This study reports (i) a summary effect of syntactic priming in typically-developing developmental populations, (ii) sources of variation in the effect, (iii) an analysis of publication bias, and (iv) a power analysis. At its simplest, meta-analysis involves combining the results from many studies to produce a summary effect (Borenstein et al., 2009). In this case, studies reliably find evidence for syntactic priming in children, so we expect to find an overall effect. We report the magnitude of the summary effect computed in the meta-analysis. Additionally, we were interested in the sources of variation in the magnitude of syntactic priming. More advanced meta-analytic techniques allow the investigation of moderator variables, although researchers must be cautious not to ask more of the data than can be obtained from them (Viechtbauer, 2008). We identified which moderators can be reasonably investigated based on the studies available, including both experimentally manipulated variables and researchers' methodological choices, and investigated their influence on syntactic priming in children. We also investigated whether there is evidence for publication bias; that is, the inflation of the estimated effect due to unpublished null results missing from the sample of studies (Borenstein et al., 2009). Finally, we ran a power analysis to estimate the number of participants and items required to detect both the main priming effect and interaction effects, given the field's focus on moderators of syntactic priming. In the process, we summarise the state of the field: which questions are being investigated and how? Does how we run studies have an impact? And what do we have evidence for and where it is lacking?

Meta-analysis

Data availability

Our data sheet of coded and extracted information and analysis scripts are available on the Open Science Framework (OSF; https://osf.io/k6z8g/).

Literature search

Search strategies

The literature search was conducted using three strategies designed to identify as many syntactic priming studies on children as possible. First, a database search was conducted of PsycInfo, Scopus, and Web of Science using search terms designed to maximise the reach of the search. The record needed to contain *synta** *gramma** or *structur** within three words of *priming*, *alignment* or *persistence* and contain *child** or *develop** or *infan**. The search was conducted within English language articles (peer-reviewed journals, books, book chapters and conference proceedings) from 1986, the year of Bock's seminal paper, to February 2023. Secondly, we recorded references from the language acquisition sections of relevant reviews (Atkinson, 2022; Branigan & Pickering, 2017; Pickering & Ferreira, 2008). Finally, we searched for forward citations of the first priming studies in children using the forward citation tools on Scopus and Web of Science (Huttenlocher et al., 2004; Savage et al., 2003). The search was first conducted in November 2019 then replicated and updated in February 2023.

Search results from the three database searches were exported as*.ris* files and imported into Zotero (Corporation for Digital Scholarship, 2023), as were the forward citation search results. Backward citations from relevant reviews were manually entered into Zotero. Title and abstract screening was conducted within Zotero by tagging studies as relevant or not. Studies identified as directly related to syntactic priming were then exported and entered into the MetaLab *Decision Spreadsheet* template (Bergmann et al., 2018). Studies were then assessed against the selection criteria within this spreadsheet. The selection criteria are displayed in Table 1 and described below.

Selection criteria

We focused on the typical development of early syntactic knowledge. Therefore, studies were limited to those testing children under 13 years of age with no history of language or other developmental disorders. Control groups from studies focusing on children with developmental or language disorders were included (e.g., Foltz et al., 2015). We only included studies reporting a syntactic priming experiment as defined by our design criteria. We excluded corpus analyses (3.a.), studies of non-syntactic outcomes or using non-syntactic primes (3.b.), elicitation and baseline production studies (3.c.), novel word or structure studies (3.d.), and training studies (3.e.).

Furthermore, we only included studies of spoken production priming, and thus excluded studies that used written production or primed comprehension (e.g., Thothathiri & Snedeker, 2008; van Beijsterveldt & van Hell, 2009). Comprehension and production processing, and the production processes for written and spoken language are likely to differ, particularly in developmental populations. Therefore, excluding these studies ensures consistency in the processing mechanisms underlying the effects. Similarly, we limited our analyses to within-language priming, and so excluded studies of cross-linguistic priming such as Vasilyeva et al. (2010) and Gámez and Vasilyeva (2020), and the crosslinguistic condition in Wolleb et al. (2018). We made this decision for the following reasons. Practically, there are only a small number of crosslinguistic priming studies in developmental populations, and including them had the potential to cloud any effects we found in the data because crosslinguistic priming inevitably involves other important variables that need to be controlled (e.g., relative proficiency of bilingual children, presence or absence of surface word order across target structures). Theoretically, positing representational overlap across languages raises several questions regarding the cognitive architecture of language in bilinguals (Hartsuiker & Bernolet, 2017; Van Dijk et al., 2022; van Gompel & Arai, 2018), which was beyond the scope of our project.

Finally, we excluded storytelling interventions aimed at increasing children's production of a structure (e.g., Vasilyeva et al., 2006). While these no doubt involve priming, they are difficult to compare to most syntactic priming studies, which investigate the effect of individual primes or a single block of primes on production, whereas interventions investigate sustained exposure to primes over longer time periods

Table 1
Selection criteria.

Category	Criteria
1. Document type	a. Journal papers, book chapters, proceedings papers b. English language
2. Participants	a. Children under 13 years of age b. No history of developmental or language disorders i. Control groups were included c. Sample independent from any other study
3. Design	a. Must be experimental not observational i. No corpus analyses b. Must investigate priming of a structural alternation i. Dependent variable is the choice between two structures within a syntactic alternation 1. No studies investigating the effect of syntactic primes on other outcomes (e.g. lexical decision-making, grammaticality judgements, speech disfluency) 2. No morphosyntactic alternations (e.g., provision of auxiliaries) or non-alternating structures (e.g., intransitive/transitive) ii. Independent variable is the syntactic structure of prime sentences 1. No single word, rhythmic, or arithmetic primes c. Must include a baseline and primed condition i. No elicitation studies with primed condition only ii. Baseline condition could comprise unprimed targets or targets primed with an alternate structure d. Must not use novel words or structures i. Investigating existing syntactic abilities not ability to generalise e. Must not provide feedback on sentence production i. Investigating implicit priming effect not explicit learning f. Outcome variable must be *spoken* production of a sentence/structure g. Primes and targets must be produced in the same language h. Target responses must be measured after no more than one block of primes i. No storytelling interventions

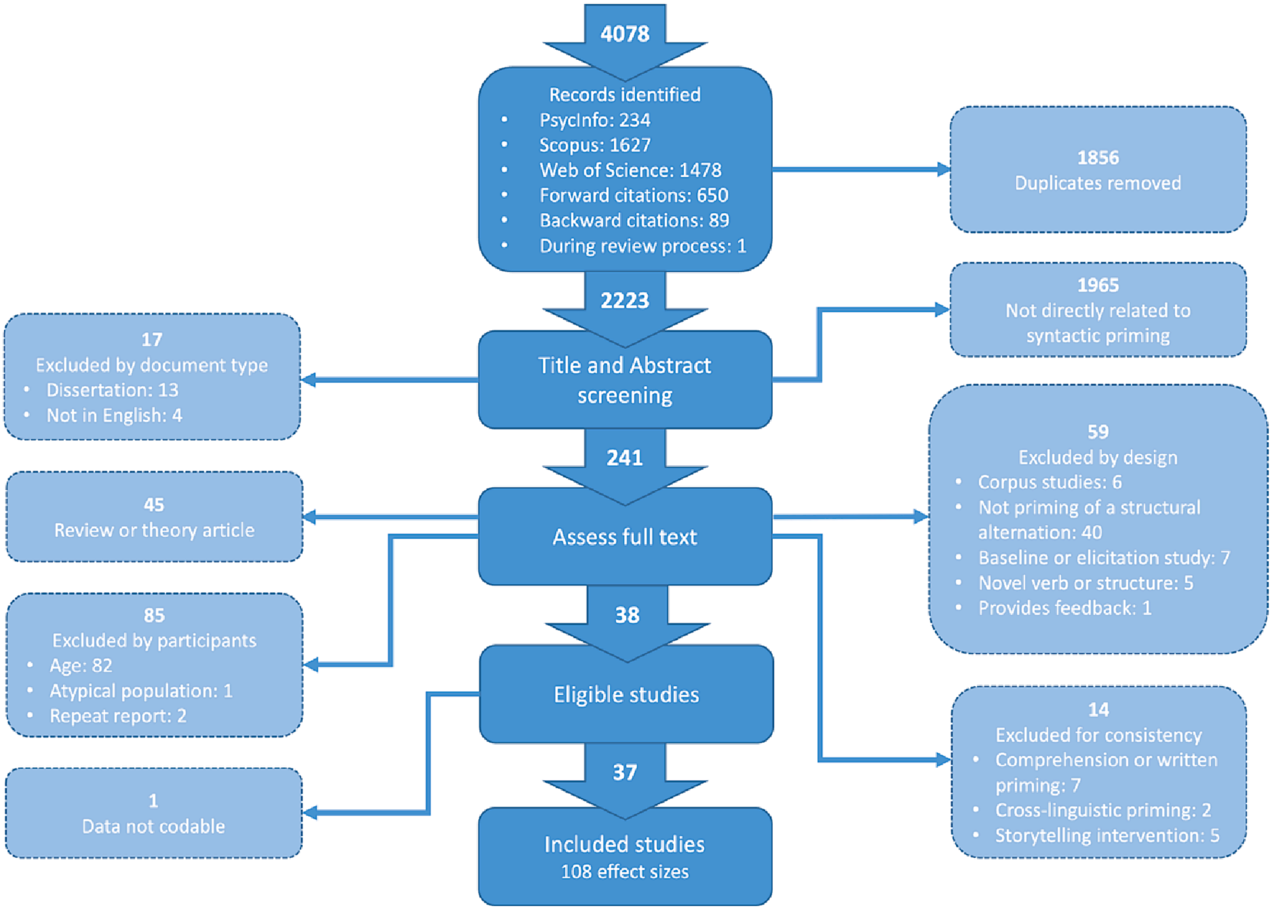


Fig. 1. Flowchart showing literature search.

(sometimes weeks).

As illustrated in Fig. 1, following the application of selection criteria, 38 eligible studies were identified, of which 37 could be analysed. Data from one study was reported such that it could not be coded and the data were not available from the author. From the 37 included studies, 108 effect sizes were extracted. The list of included studies is available in Appendix A.

Coding procedures

Information was extracted separately for each experimental condition, i.e. each observation, within a study (e.g., lexical overlap vs. no lexical overlap). For the primed and unprimed conditions of each observation, we extracted: the number of participants and items, the proportion and/or number of dependent (e.g. passive), alternate (e.g. active), and *other* responses. The dependent, or primed, structure was defined as the less frequently produced structure in the alternation (typically coded as 1 in the dependent variable at the study level). Table 2 describes the five structural alternations included in the meta-analysis, including the dependent and alternate structure for each. When frequencies were graphed but not reported, we estimated the number/proportion of dependent, alternate and *other* responses using WebPlotDigitizer (Rohatgi, 2022). In two cases, we calculated the frequencies from raw data available online (Fazekas et al., 2020; Garcia & Kidd, 2020). We contacted the author/s if the data was still not able to be extracted. We also recorded whether studies reported proportions or numbers of only dependent and alternate responses or also reported *other* response frequencies by condition. Note that when studies scored responses under both a lax and strict coding scheme (e.g., Bencini & Valian, 2008; Kumarage et al., 2022), we coded those scored under strict coding.

We coded each observation for several variables. Some were experimental variables that have been proposed to influence the size of the syntactic priming effect in children and others were methodological variables that differed between the included studies. Table 3 summarises the variables that were included as moderators in our analyses, as well as language studied. It displays the levels of each variable, and how many observations fell into each level. Where observations from a study differed across a moderator that we did not analyse, they were combined into a single observation.

Moderators

Within- vs. Between-Subjects designs. Unlike in adult studies, developmental researchers have often used between-subjects designs, comparing primed and unprimed participants rather than conditions. Firstly, as Shimpi et al. (2007) state, requiring children to complete trials from only one condition reduces the demands of the experimental task. Secondly, researchers have raised concerns that primes from one condition will influence responses in the other (Gámez & Vasilyeva, 2015). It is common in developmental syntactic priming studies to include few or no filler items. This is another way of reducing task demands but may increase the likelihood of interference across trials. Given interference cannot occur in between-subjects designs, we expect a larger priming effect in these than in within-subjects designs. We used treatment contrasts, with within-subjects designs set as the baseline, or reference, level (0) and between-subjects as the treatment level (1). One final issue to note is that primed and unprimed conditions will be correlated in within- but not between-subjects designs. We explain how we control for this in our statistical methods section.

Table 2
Description of structural alternations included in meta-analysis.

Structural alternation	Dependent structure	Alternate structure
Passive	Passive	Active
	Ernie was hugged by Bert	*Bert hugged Ernie*
Dative	Double-object dative	Prepositional dative
	Elmo gave Big Bird a cat	*Elmo gave a cat to Big Bird*
MandarinSVO/ba	ba	SVO
	*Xiaogou **ba** xiamao baozhu –**le***	*Xiaogou baozhu –**le** xiaomao*
	*Little dog **ba** little cat hug-tight **PFV***	*Little dog hug-tight **PFV** little cat*
	A little dog hugged a little cat tightly	*A little dog hugged a little cat tightly*
Relative clause	Relative clause	Adjective-noun
	The car that is red	*The red car*
Tagalog Symmetrical Voice (SV)-transitive	Patient-initial	Agent-initial
	Agent voice	Agent voice
	*H < **um** > ahabol **ng** babae **ang** bata*	*H < **um** > ahabol **ang** bata **ng** babae*
	*<**AV** > chase **NSBJ** woman **SBJ** child*	*<**AV** > chase **SBJ** child **NSBJ** woman*
	The child is chasing a woman	*The child is chasing a woman*
	Patient voice	Patient voice
	*H < **in** > ahabol **ang** babae **ng** bata*	*H < **in** > ahabol **ng** bata **ang** babae*
	*<**PV** > chase **SBJ** woman **NSBJ** child*	*<**PV** > chase **NSBJ** child **SBJ** woman*
	The child is chasing the woman	*The child is chasing the woman*

Note: The SVO/ba alternation occurs only in Mandarin and the Symmetrical-voice transitive alternation only in Tagalog. The three other alternations were tested in multiple languages. Example for SVO/ba adapted from Hsu (2014) and example for SV-transitive adapted from Garcia & Kidd (2020). Linguistic glosses: perfective aspect (PFV), agent voice (AV), patient voice (PV), subject (SBJ), non-subject (NSBJ).

Baseline. Most studies compare to a baseline of responses following primes in the alternate structure, but some compare to a baseline of responses produced with no prime, typically collected in a preceding phase before priming begins. At present, there is no evidence regarding whether this difference in methodology affects the magnitude of the priming effect. However, Bencini and Valian (2008) reported that children in a no prime condition produced no passives, whilst those in the active priming condition produced two. Therefore, a larger priming effect may be expected when comparing to a no prime baseline. We used treatment contrasts, with the *alternate prime* condition as the reference level (0) and *no prime* as the treatment level (1).

Animacy. The animacy of verb arguments can favour or disprefer the dependent structure. For example, a preference to put humans in sentence-initial positions means the human-patient nonhuman-agent is the canonical passive form (e.g., *the man was bitten by the dog*; see Bock et al., 1992). In the dative, a preference for animate arguments to precede inanimate ones means the canonical double object dative involves an animate recipient and inanimate theme (e.g., *The boy gave the girl a present*). The influence of animacy on priming was demonstrated by Vasilyeva and Gámez (2015), who found that the animacy of arguments moderated abstract passive priming: sentences with an animate patient and inanimate agent were subject to greater priming. Most studies have controlled for animacy by using the same animacy configuration in primes and targets: all animate arguments (e.g., passive: Kumarage et al., 2022; e.g., dative: Rowland et al., 2012), all inanimate arguments (e.g., Savage et al., 2003) or an animacy configuration favouring the dependent structure (e.g., passive: Branigan & McLean, 2016; e.g., dative: Fazekas et al., 2020). Others have used materials with a mix of configurations in both primes and targets (e.g., Huttenlocher et al., 2004). We could not code for particular animacy configurations, given they are specific to particular structures. Instead, we created three broad categories: *controlled animacy* used either all animate or all inanimate arguments, *uncontrolled animacy* used a mixture of configurations, and *favourable animacy* used a configuration favouring the dependent structure. We set *controlled animacy* as the reference level and used treatment contrasts to compare the other two conditions to this baseline. We expect greater priming when animacy favours the dependent structure. It is not clear whether not controlling animacy will reduce or increase priming.

Prime repetition. We coded whether studies required children to repeat the prime sentence before producing their target sentence using treatment contrasts (*prime not repeated,* 0; *prime repeated,* 1). Huttenlocher et al. (2004) found equivalent priming with and without prime repetition in 4–5-year-olds, concluding that syntactic representation is shared across comprehension and production. Shimpi et al. (2007) found no priming effect in 3-year-olds unless they repeated the prime, whereas 4-year-olds did not need to repeat the prime to demonstrate a priming

Table 3
Moderator variables: coding, levels and distributions.

Variable	Coding	Level	N observations
Within- vs. between-subjects design	0	Within-subjects	68
	1	Between-subjects	40
Baseline	0	Alternate prime	94
	1	No prime	14
Animacy	0	Controlled	57
		Animacy not relevant – Relative clause(Included as 0 not NA as otherwise automatically excluded from model)	4
	1	Uncontrolled	24
	2	Favours dependent structure	23
Prime repetition	0	Prime not repeated	76
	1	Prime repeated	32
Lexical overlap	0	No overlap	85
	1	Overlap	23
Structure	0	Passive	63
	1	Dative	22
	2	SVO-ba (Mandarin)	13
	3	Relative clause	4
	4	Symmetrical Voice Transitive (Tagalog)	6
Age		In months as z-score	$M = 59.35$ $SD = 18.15$
Lag	0	Alternating design	74
	1	Blocked design	27
	2	Long blocked design	7
Language		English	77
		German	2
		Italian	4
		Mandarin	13
		Norwegian	2
		Russian	1
		Spanish	3
		Tagalog	6

effect. This suggests prime repetition may facilitate abstract priming, with the effect weakening as children develop. We therefore included an interaction between prime repetition and age in the *meta*-analysis.

Lexical overlap. In adults, lexical overlap between prime and target sentences (e.g., shared verb or noun) reliably boosts priming (Mahowald et al., 2016). In children, evidence for the lexical boost is more mixed. Some studies have found a lexical boost at a young age (Branigan et al., 2005; Branigan & McLean, 2016; Savage et al., 2003), whilst others have only found the effect in older children (Kumarage et al., 2022; Rowland et al., 2012) or not at all (Foltz et al., 2015; Peter et al., 2015). The developmental trajectory of the lexical boost has implications for theories of syntactic acquisition. A lexical boost effect that increases over development is in line with the prediction that lexically-based priming reflects developing explicit memory processes rather than implicit learning (Chang et al., 2012; Rowland et al., 2012), whilst a decreasing effect is in line with a transition from lexically-based to abstract syntactic representation (Savage et al., 2003; Tomasello, 2003). We coded whether or not there was lexical overlap between prime and target using treatment contrasts (*no lexical overlap,* 0; *lexical overlap,* 1) and included an interaction with age.

Structure. We found five structural alternations in the included studies: passive/active, double object dative/prepositional dative, Mandarin SVO/ba, relative clause/adjective-noun and patient-initial/agent-initial in Tagalog's symmetrical voice transitive (see Table 2 for details). Since syntactic priming is stronger for infrequent structures (Ferreira, 2003; Jaeger & Snider, 2013; Kaschak et al., 2011), the strength of the priming effect may vary by structure. We used the passive alternation as the reference level with which to compare other structures because it is the most researched alternation and likely to be well estimated. Given children acquire different syntactic structures at different ages, the age at which children have abstract knowledge and therefore show abstract priming of a structure is likely to vary. We therefore included an interaction between structure and age.

Age. A variety of developmental trajectories for abstract syntactic priming have been proposed. For example, Nativist accounts typically propose stable priming effects over development (Bencini & Valian, 2008; Messenger & Fisher, 2018). Error-based learning in the Chang et al. (2006) model predicts that abstract priming can decrease as children become better at predicting the prime sentences (lower error; Peter et al., 2015). However, the Chang et al. model also implicitly learns its syntactic representations from word sequences, so priming can also increase during an early developmental period as structures become more abstract, since abstract structures increase the transfer of changes on the prime to the target. This is similar to the predictions of lexicalist accounts such as Tomasello (2003). As mentioned above, age may also interact with prime repetition, lexical overlap and structure. We coded the average age in months of the sample for each observation. This value was converted to a z-score for our analyses.

Lag. The observation of priming at long distances between prime and target led to the theory that priming is a form of implicit learning (Bock & Griffin, 2000). We coded three categories of lag between prime and target: alternating design, blocked design, and long blocked design. Studies which alternate between prime and target sentences are the most common. Several studies use a blocked design, where primes are presented as a block before children describe a block of targets. Hsu (2019) found that a blocked design showed a larger priming effect than an alternating design. Relatively few studies in children have investigated the length of time for which priming endures. Only three studies contained conditions with long blocked designs with a delay of an hour (Hsu, 2019), a week (Kidd, 2012; Savage et al., 2006) or a month (Savage et al., 2006) between primes and targets. We excluded the lag 2 condition from Branigan and McLean (2016) and Garraffa et al. (2018), where two sentences intervened between prime and target. To include multiple effect observations from the same sample of participants we required a reported or estimated correlation between observations, which was not available in this case. As with animacy and structure, we dummy coded this variable, with alternating design set as the reference level.

Table 4
Example cell count data.

	Dependent response	Alternate response
Primed	*a*	*b*
Unprimed	*c*	*d*

Other variables

We identified other potential moderator variables that, for several reasons, were not included in our analyses. Firstly, in some cases, there were too few observations of one level of the variable to make a reasonable comparison. For instance, only one study investigated priming in a second language (Gámez & Vasilyeva, 2015), ruling out a comparison between priming in a first and second language. Secondly, some variables would be suitable for a meta-analysis of a particular structure but not across structures. For example, several studies have investigated the influence of verb-type (i.e. agent-patient, experiencer-theme, theme-experiencer; Bidgood et al., 2020; Messenger, Branigan, McLean, & Sorace, 2012) on priming of the active–passive alternation; however, this variable cannot be generalised across the other structural alternations so could not be included. Thirdly, other variables were precluded from inclusion because of confounds. Task type was confounded with structure, with stem completion typically used in studies of the dative alternation and picture description for other structures. Similarly, for priming paradigm, the *bingo* game was used almost exclusively in studies of the dative alternation, whilst studies of other structures used the *snap* game or picture description.[1] We recorded the language in which the study was conducted but could not include it as a moderator or random effect. In the case of Mandarin and Tagalog, language is confounded with structure (SVO/ba, symmetrical voice transitive), preventing its inclusion as a moderator. In addition, most studies were conducted in English, with very few observations from other languages, reflecting a common bias in the field (Kidd & Garcia, 2022). A random effects structure is unlikely to be sensibly estimated in this case and so we did not include random effects by language. Finally, we could only include a limited number of moderators (as a rough rule, Borenstein et al. (2009) recommend approximately 10 observations per moderator). We coded but did not analyse the influence of: the number of confederates (as in Mahowald et al., 2016), the number and frequency of filler items, and lax vs. strict coding of responses. This information is included in the Data Collection Form accessible on our OSF site.

Statistical methods

We ran our analyses in R (version 4.3.2; R Core Team, 2022). We used *tidyverse* packages for data manipulation (version 2.0.0; Wickham et al., 2019) and the *metafor* package for *meta*-analysis functions (version 4.4–0; Viechtbauer, 2010). All data and code is available on our OSF site (https://osf.io/k6z8g/).

Effect size

The first step in meta-analysis is to compute a common effect size across studies (Borenstein et al., 2009). We chose to compute a log odds

---

[1] We thank an anonymous reviewer for suggesting the analysis of priming paradigm. Whilst we could not include it as a moderator due to this confound we discuss its potential impacts in the discussion.

ratio (LOR) as the effect size for each observation based on raw cell count data. The LOR is appropriate for binary outcome data, such as in syntactic priming studies, where participants respond with either the dependent or alternate structure (e.g., passive or active: Jaeger, 2008). In addition, studies reported a mixture of ANOVAs and logistic mixed models in their results section. These are difficult to combine due to the different statistics than can be extracted from them. Lastly, LORs taken directly from reported mixed models would not be equivalent if those models used different fixed effects or random effects structures. Calculating a LOR from cell count data, which is routinely reported in syntactic priming studies, combats these issues.

We calculated the LOR using the *escalc* function in the *metafor* package. This takes the number of responses in each condition (as in Table 4) and calculates the LOR using Equation 2.[2] This calculation cannot handle 0 scores in Table 4 so values of *a, b, c* and *d* were adjusted beforehand using Smithson and Verkuilen's (2006) approach. The proportion of dependent responses in each condition was adjusted using Eq. (1), where N is the number of participants in that condition. This proportion was then multiplied by N to calculate a and c, and $(1 - \mathrm{P}_{\mathrm{adjusted}})$ was multiplied by N to calculate b and d.

The *escalc* function also calculates the associated variance, or sampling variance, for each LOR effect size using Eq. (3). We adjust this *escalc* calculated sampling variance for within-subjects designs and for two definitions of sample size. We next explain these issues in accounting for sampling variance, as well as our approach to accounting for two other forms of variance.

$$P_{adjusted} = \frac{P(N-1)+0.5}{N} \tag{1}$$

$$LOR = \ln\left(\frac{ad}{bc}\right) \tag{2}$$

$$V_{LOR} = \frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d} \tag{3}$$

Variance: Accounting for complex data structure using multilevel meta-analysis

The nature of the developmental syntactic priming literature leads to three sources of variance in effect sizes: sampling variance, between-study variance, and within-study variance. We conducted a multilevel meta-analysis to account for all three sources of variance, as detailed below.

Sampling variance. The *escalc* function automatically calculates the LOR and its associated sampling variance, or sampling error, as per Equation (3). Sampling variance is used to weight studies, or more accurately, observations, by their precision, giving more weight to more precise observations (Borenstein et al., 2009). It is mostly influenced by sample size and reflects sampling error in the estimation of an effect (Borenstein et al., 2009).

Multiple responses per participant. Log odds ratios have typically been used in the meta-analysis of clinical trials where each outcome comes from a separate participant (e.g., cured vs. not cured). In syntactic priming studies, participants provide multiple responses per condition. Calculating sampling variance using $\mathrm{N}_{\mathrm{responses}}$ ignores the dependence between responses from the same participant. However, calculating sampling variance using $\mathrm{N}_{\mathrm{participants}}$ dramatically reduces the power of the analysis and does not account for the true number of responses. In line with Mahowald et al. (2016), we calculated sampling variance using both $\mathrm{N}_{\mathrm{responses}}$ and $\mathrm{N}_{\mathrm{participants}}$. To do this, we followed the procedure described in the *Effect size* section twice, first multiplying $\mathrm{P}_{\mathrm{adjusted}}$ by $\mathrm{N}_{\mathrm{responses}}$ to calculate the adjusted values of *a, b, c* and *d,* and second using $\mathrm{N}_{\mathrm{participants}}$. Our results section reports results from analyses using both forms of variance.

Within- vs. Between-Subjects designs. The use of a within-subjects design also affects precision: the *primed* and *unprimed* conditions are likely to be correlated due to individual participants' tendency to produce the dependent structure. In other words, comparing participants to themselves reduces error and increases precision. Sampling variance should be calculated accordingly (Morris & DeShon, 2002).

Therefore, we adjusted our sampling variance estimates for within-subjects studies using the Becker-Balagtas method (Becker & Balagtas, 1993) as described in Stedman et al. (2011). Equation (4) can be used instead of Equation (3) to calculate sampling variance if we know p, the correlation between production of the dependent structure in the primed and unprimed conditions (Stedman et al., 2011). The correlation, p, is used to calculate s (Equation (6)), and s to calculate Δ (Equation (5)). Then Δ and n, the total number of responses, can be used to calculate an adjustment to the sampling variance (Equation (4)).

$$V_{LOR} = \frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}-\frac{\Delta}{2n} \tag{4}$$

$$\Delta = n^2\left(\frac{ns-ab}{abcd}\right) \tag{5}$$

$$s = \frac{p\sqrt{abcd}+ab}{n} \tag{6}$$

The correlation, p, between production of the dependent structure in the primed and unprimed conditions is not reported in syntactic priming studies. However, raw data available to the authors from the *Canberra Longitudinal Child Language* study (Donnelly et al., in press; Kidd et al., 2018; Kumarage et al., 2022) allowed us to estimate it at $r = .35$. Appendix B details how correlations between conditions were estimated. Using $p = .35$, the adjusted sampling variance was calculated for studies with within-subjects designs. All analyses (both $\mathrm{N}_{\mathrm{responses}}$ and $\mathrm{N}_{\mathrm{participants}}$) use sampling variance corrected for within-subjects designs.

Between-Study variance. Traditional meta-analysis assumes that each study estimates the same true effect size with differences between studies due only to sampling error (Borenstein et al., 2009). However, effects are likely to vary between studies due to factors outside of control or moderators not included in analysis (Borenstein et al., 2009). Including a random effects structure allows us to instead assume that there is distribution of true effects, which differ due to sampling error and heterogeneity in effect sizes. We therefore included random intercepts by observation nested within study. This adds the assumption that effect sizes vary both within- and between-studies due to heterogeneity, not just sampling error.

Within-Study variance. This meta-analysis includes 108 observations from 37 studies. Ignoring the dependence between effect sizes from the same study is a common mistake in meta-analyses (e.g., 9 of the 20 most highly cited meta-analyses in the exercise science field of strength and conditioning did not account for this dependence; Kadlec et al., 2022). Treating correlated effect sizes as independent observations is problematic because it assumes they contribute independent information, which can inflate the strength of evidence for an effect and assigns more weight to studies with multiple observations (Borenstein et al., 2009). A multilevel meta-analysis allows the inclusion of multiple dependent effect sizes within a single analysis. Several types of dependence between observations were present in our data set and we describe how we accounted for each below.

Independent groups. Multiple observations from independent samples within the same study result from manipulating moderators between-subjects. For example, comparing different age groups, or assigning children to either the prime repetition or no repetition condition.

---

[2] We excluded *other* responses from our calculation of the LOR. However, Appendix C compares results including and excluding *other* responses for a subset of studies that reported the frequencies of *other* responses.

Table 5
Estimated correlations between priming conditions.

Structure	Overlap	Correlation estimate
Same	Same	1
	Different	.20
Different	Same	.15
	Different	.06

Despite having independent samples, effects from the same study are likely to be subject to the same uncontrolled variables or moderators. Specifying random effects that are nested by study allows us to account for the possibility that independent effects from the same study may be more similar than independent effects from different studies.

Shared comparison groups. When two or more experimental conditions are compared to the same baseline, the observations in those conditions will be correlated. We specified which group of participants within a study was primed and unprimed in each observation using the *grp1* and *grp2* arguments in *metafor*'s *vcalc* function, and the number of participants or responses in each group using *w1* and *w2*. The *vcalc* function uses this information to calculate the correlation between observations and the variance–covariance matrix used in multilevel meta-analysis (Viechtbauer, 2010).

Multiple outcomes. Several studies manipulated lexical overlap or target structure (passive vs. dative) within-subjects, leading to multiple dependent outcomes. Treating these observations as independent assumes the correlation between them is 0, despite them coming from the same sample of participants. Another common approach is to use an average effect size and variance. However, this implicitly assumes that the correlation between the effects is 1, or that they do not contribute any extra information as separate observations (for detailed discussion see Borenstein et al., 2009). It also prevents the analysis of moderators in studies where they are manipulated within-subjects, and therefore more precisely estimated. Borenstein et al. (2009) recommend using a plausible correlation over assuming an extreme of either 0 or 1.

We estimated correlations between priming with and without lexical overlap, and between passive and dative priming (see Table 5). Appendix B details how these were calculated from priming data from the *Canberra Longitudinal Child Language* project (Donnelly et al., in press; Kidd et al., 2018; Kumarage et al., 2022). Using these estimates, *vcalc* was used to calculate a variance–covariance matrix for dependent observations, which was specified in our meta-analysis.[3]

Multiple timepoints. Some studies measured priming at more than one timepoint. Branigan and Messenger (2016) conducted two priming sessions one week apart. They reported the correlation between these sessions which we were able to specify in our meta-analysis. Kumarage et al. (2022) conducted four priming sessions over 18 months. The data from this study were available to calculate correlations between the sessions, which we specified in our meta-analysis. Kidd (2012), Fazekas et al. (2020), and Savage et al. (2006) instead primed children at only one time but recorded target responses at more than one timepoint after this. The correlations between the test, posttest (immediately after a prime phase), and long posttest (one week later) timepoints from Kidd (2012) were provided by the author. The data from Fazekas (2020) were available to calculate correlations between the baseline, test, and two posttest (immediately after test) conditions. We specified the provided and calculated correlations from these two studies in our meta-analysis and used an average of these correlations as an estimate for the correlation between timepoints for Savage et al., (2006; see Appendix B). Using this estimate of .242, *vcalc* was used to calculate the variance–covariance matrix for this study, which was then specified in our meta-analysis.

Table 6
Results of multilevel meta-analytic models without moderators.

	N responses		N subjects	
Estimate (log *odds ratio*)	1.44		1.27	
	1.20 | 1.69		1.01 | 1.53	
	$p < .001^{***}$		$p < .001^{***}$	
Odds ratio	4.24		3.55	
Cohen's d	0.80		0.70	
Q	698.15		149.23	
	df = 107		df = 107	
	$p < .001^{***}$		$p = .005^{**}$	
Variance	0.675		0.823	
Sampling variance	0.102	15.12%	0.504	61.22%
I^2	0.573	84.88%	0.319	38.78%
Between-study		48.27%		25.33%
Within-study		36.61%		13.45%

Results

We report results for two different meta-analytic models because we calculated sampling variance in two ways. Using the number of responses as sample size ignores the fact that, unlike in clinical trials, participants in priming studies provide multiple responses, which will depend on an individual's tendency to produce the target structure. Using the number of participants as the sample size is more conservative but ignores the extra information provided by collecting multiple responses from each participant. All models included random effects of observation nested within study and a variance–covariance matrix specifying reported or estimated correlations between observations.

Overall effect size and heterogeneity

We first ran models without moderators to estimate an overall effect size and the heterogeneity in observations. For both models, profile-likelihood plots indicated the variance components could be estimated and DFBETA values revealed no influential observations. Residuals were normally distributed for the first model and showed slightly less variance than expected for the second.

Table 6 displays the results of these models. In both, the overall priming effect is significant with odds ratios larger than one, indicating greater odds of producing the dependent structure in the primed than unprimed condition. The magnitude of this effect is medium-to-large: when converted to odds ratios, the LORs translate to an effect of 3.6 to 4.2 times greater odds of the dependent structure in the primed than unprimed condition, or a Cohen's *d* of 0.70——0.80 (multiplying the LOR by $\frac{\sqrt{3}}{\pi}$ converts it to Cohen's *d*; Borenstein et al., 2009). Fig. 2 shows a forest plot of all included effect sizes and the estimated overall effect from the first model.

We also observe significant heterogeneity relative to overall variance ($Q_1(107) = 698.15$, $p < .001$; $Q_2(107) = 149.23$, $p < .01$). That is, there is significant variance not explained by sampling error. Estimates of I^2 indicate this remaining heterogeneity is substantial in the first model, at about 85 % of the observed variance. In the second model, we used a more conservative estimate of sampling variance and the remaining heterogeneity reduced to 39 % of observed variance. An investigation of theorised moderators of the priming effect is warranted in both cases; however, the power to detect heterogeneity differs between our two models. This difference in power can be attributed to the more and less conservative calculations of sampling variance because lower power of individual studies can reduce overall power to detect heterogeneity (Borenstein et al., 2009).

Moderator analyses

We next ran models that included our identified moderator variables:

---

[3] Two studies manipulated animacy within subjects (Buckle et al., 2017; Vasilyeva & Gámez, 2015). Since we had no estimate of the correlation between dependent animacy conditions, these conditions were combined into a single observation per sample coded as uncontrolled animacy.

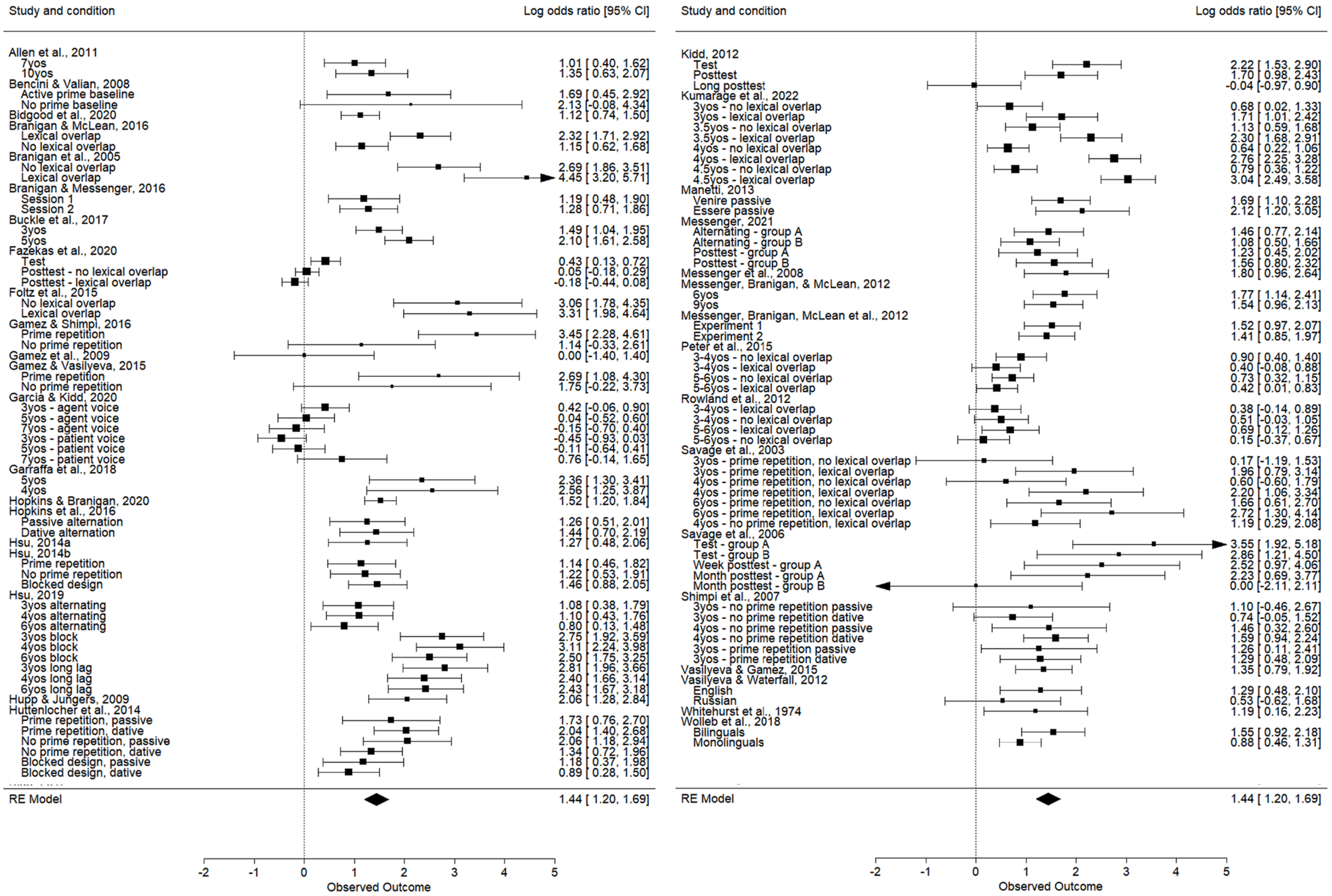


Fig. 2. Forest plot for meta-analysis using N responses.

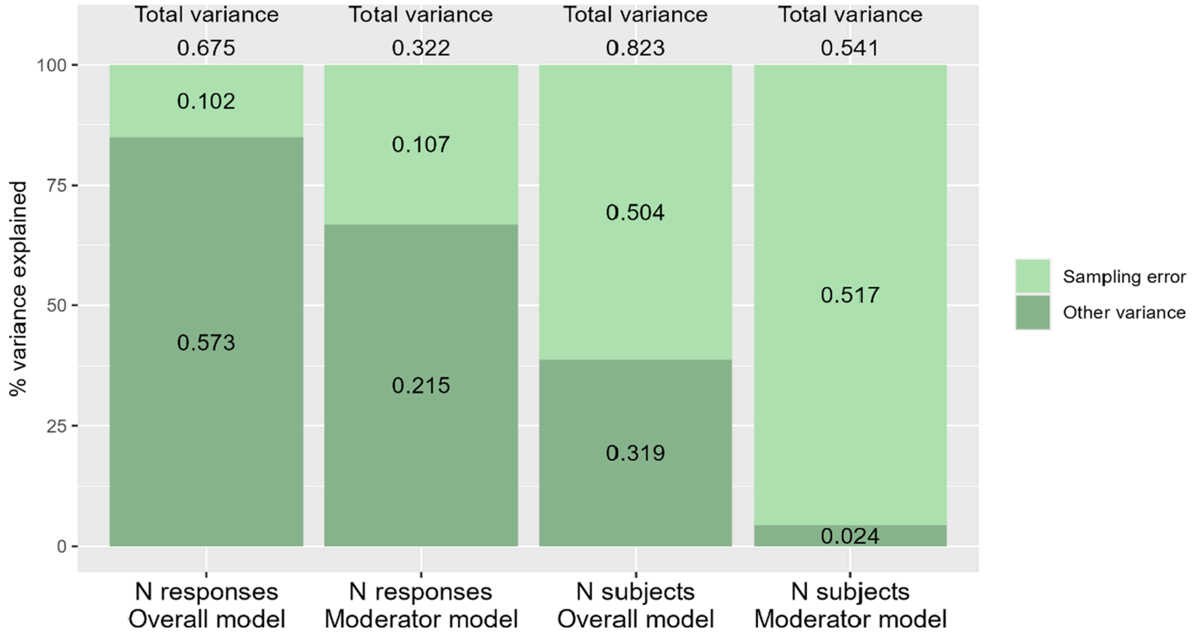


Fig. 3. Proportion of sampling variance and non-sampling variance in models including and excluding moderator variables.

Table 7
Results of *meta*-analyses including moderator variables.

	N responses			N subjects		
	β	CI	p	β	CI	p
Intercept	0.94	0.56 | 1.31	<.001***	1.03	0.66 | 1.39	<.001***
Between subjects	1.02	0.57 | 1.47	<.001***	1.02	0.35 | 1.70	.003**
Baseline	−0.34	−0.88 | 0.19	.207	−0.22	−0.95 | 0.52	.563
Animacy						
Uncontrolled	−0.18	−0.72 | 0.36	.489	0.03	−0.73 | 0.78	.938
Favourable	0.54	0.06 | 1.02	.028*	0.38	−0.16 | 0.92	.159
Prime repetition	0.28	−0.07 | 0.63	.117	0.06	−0.37 | 0.49	.785
Lexical overlap	0.77	0.39 | 1.15	<.001***	0.74	0.28 | 1.19	.002**
Structure						
Dative	−0.71	−1.04 | −0.38	<.001***	−0.85	−1.26 | −0.44	<.001***
SVO-ba	0.27	−0.34 | 0.89	.364	0.16	−0.48 | 0.80	.611
RC	1.84	0.27 | 3.41	.024*	1.8	−0.91 | 4.50	.179
SV-Transitive	−0.92	−1.69 | −0.15	.021*	−1.03	−1.66 | −0.41	.003**
Age	0.02	−0.15 | 0.19	.835	0.07	−0.16 | 0.32	.566
Lag						
0 vs block	−0.14	−0.52 | 0.24	.461	−0.19	−0.76 | 0.30	.484
0 vs long	−0.25	−0.92 | 0.41	.454	−0.14	−1.23 | 0.95	.797
Prime repetition*Age	0.23	−0.17 | 0.63	.263	0.01	−0.51 | 0.50	.978
Lexical overlap*Age	−0.09	−0.58 | 0.40	.704	−0.26	−0.87 | 0.34	.389
Structure*Age						
Dative	0.03	−0.27 | 0.33	.842	0.07	−0.35 | 0.50	.735
SVO-ba	−0.2	−0.68 | 0.27	.398	−0.23	−0.89 | 0.43	.497
RC	−0.4	−3.61 | 2.81	.795	−0.22	−5.79 | 5.35	.936
SV-Transitive	0.06	−0.36 | 0.49	.764	−0.03	−0.55 | 0.49	.906

design (between- vs. within-subjects), baseline (alternate structure prime, no prime), animacy (uncontrolled, controlled, favourable), age, prime repetition (and interaction with age), lexical overlap (and interaction with age), structure (passive, dative, relative clause, SVO-ba, SV-Transitive; and interaction with age), and lag (alternating, blocked, long-blocked). Residuals for both models showed less variability than expected, likely due to slight overfitting after including moderators of the effect. Profile-likelihood plots showed that variance components were adequately estimated. We identified the second post-test condition of Fazekas et al. (2020) as an outlier with large influence over parameters in the model by examining DFBETAS values and running models with and without the outlier. We report results excluding this outlier.

The moderator models successfully explained remaining heterogeneity. The test of moderators was significant for both models (F (19,17) = 5.22, p < .001; F(19,17) = 4.42, p < .01), indicating that a significant portion of heterogeneity is explained by predictors in the model. This is illustrated in Fig. 3, which shows that non-sampling variance is reduced after including moderators. After accounting for moderators there was significant residual heterogeneity in the $\text{N}_{\text{responses}}$ model, Q(87) = 230.79, p < .001, but not the $\text{N}_{\text{subjects}}$ model, Q(87) = 53.55, p = .998.

Table 7 displays the results of the moderator models. The intercept is significant in both models, indicating a significant priming effect in the baseline-coded condition of all our moderators. That is, the model predicts a significant priming effect for a within-subjects study of the passive in 59.35-month-olds, which compares to an active-primed baseline,

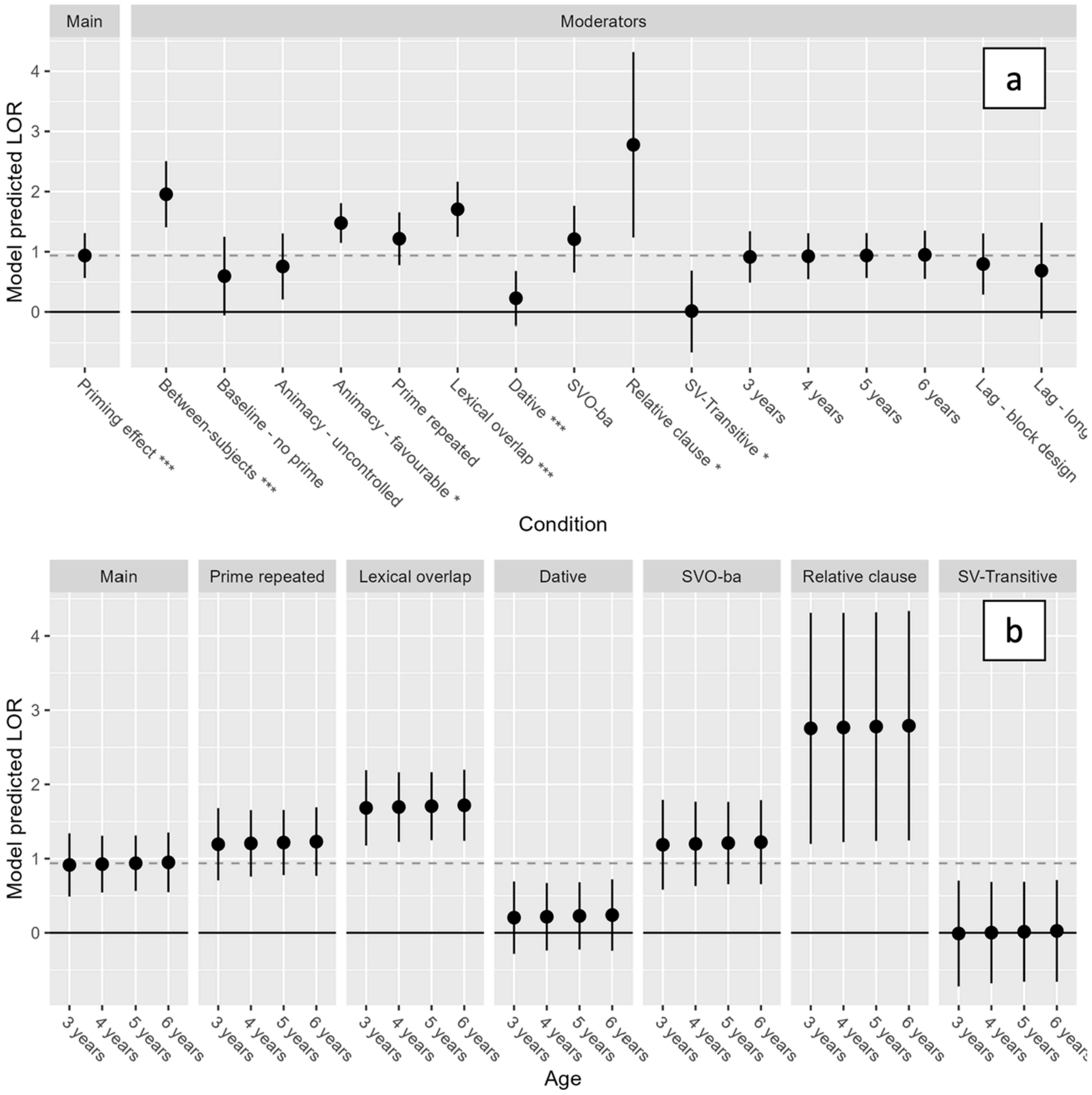


Fig. 4. Size of model-predicted priming effect under different experimental conditions for (a) main effects and (b) interaction effects.

controls for animacy, has no lexical overlap, no prime repetition and no lag between prime and target. Fig. 4 shows the model-predicted LOR at treatment levels of moderators for the model using $N_{responses}$ (Fig. 4a: main effects, Fig. 4b: interaction effects). For example, the predicted priming effect for within-subjects studies (baseline level, 0) is shown by the Main effect (intercept), whilst the predicted priming effect for between-subjects studies (treatment level, 1) is shown under Moderators.

The priming effect was significantly larger in studies using between-subjects designs. We adjusted for the increased precision of within-subjects designs, where the primed and unprimed conditions are correlated due to individual rates of producing the dependent structure, using the Becker-Balagtas method. Therefore, a remaining difference between study designs suggests that there is a factor besides this correlation which differs between them (Morris & DeShon, 2002).

Controlling animacy did not have a significant effect on the size of the priming effect. However, using animacy configurations that favoured the dependent structure did significantly increase the size of the priming effect in the more powerful ($N_{responses}$ as sample size) model.

Studies with lexical overlap between prime and target showed larger priming effects than those with no lexical overlap. We note that including the second posttest condition of Fazekas et al. (2020) reduced the size of the lexical overlap effect for both models and produced a significant lexical overlap*age interaction in the $N_{subjects}$ model, with a decreasing lexical boost over development. This observation is unusual in combining lexical overlap with a long lag between prime and target and found no priming. The reduced priming effect is likely due to the lexical boost being short-lived (Branigan & McLean, 2016) rather than the older age of the sample. Since removing this single observation eliminates the interaction effect, we interpret the evidence to support a non-significant interaction effect.

The structural alternation significantly affected the size of the priming effect. Studies of the dative alternation reported smaller effects than those of the passive. In fact, Fig. 4a shows that the model-predicted priming effect is no longer significant in dative studies, with the null effect of 0 contained within the 95 % confidence interval. Studies of the relative clause reported larger effects than the passive in the more powerful model; however, there were only two studies of this structure. Similar caution should be applied when interpreting the non-significant model-predicted priming effect in observations of the Tagalog SV-Transitive alternation: there were 6 observations from 6 samples, all from the same study.

There were no effects of baseline type (no prime or alternate prime), prime repetition, lag between prime and target, or age. There was also no significant interaction between age and other moderators: Fig. 4b shows that the effects of prime repetition, lexical overlap and structure were constant across age.

Assessing publication bias

Studies with significant results or larger effect sizes are more likely to be published than those with smaller or null results (Dickersin, 2005; Franco et al., 2014). In meta-analyses, synthesising the results of a biased sample of studies can then lead to spurious findings. This is common in psychology, with a recent estimate that 60 % of meta-analyses in psychology overestimate the evidence for an effect (Bartoš et al., 2023). For an example in psycholinguistics, the widely accepted bilingual advantage in executive functioning has more recently been attributed to publication bias (de Bruin et al., 2015; Lehtonen et al., 2018). In the syntactic priming literature, Mahowald et al. (2016) found their meta-analysis of studies in adults was not overly influenced by publication bias. We expected a similar finding if the child literature is comparable and given that in the developmental context null results can themselves be of interest (e.g., Savage et al. (2003) reported 3-year-olds were not primed in the absence of lexical overlap).

Funnel plots aid in detecting publication bias by depicting the relationship between study precision and effect size. More precise studies with smaller standard errors tend to cluster around the estimated effect size at the top of the plot. In a sample of studies without publication bias, smaller studies with larger standard errors towards the bottom of the plot will be symmetrically distributed around the estimated effect size. However, if studies suffer from publication bias, funnel plot asymmetry is observed, wherein small studies with null effects are missing from the bottom left. Funnel plot asymmetry should be assessed statistically, not just visually. We used a test conceptually similar to Egger's regression test and Peters' regression test (Egger et al., 1997; Peters et al., 2006). It is not possible to execute these tests in a multi-level meta-analysis containing dependent sampling variances.[4] Instead, we added the inverse of sample size to the meta-regression as a moderator. If study precision, as indexed by inverse sample size, significantly predicts the size of the overall effect, we can conclude there is funnel plot asymmetry. We use inverse sample size rather than the variance or standard error of the LOR because these are already mathematically dependent on the size of the LOR (Peters et al., 2006).

Fig. 5 shows funnel plots for models including and excluding moderators, with a circle indicating where we may expect missing observations to be in asymmetrical plots. Funnel plots for the models without moderators show a somewhat asymmetrical distribution of effects when using number of responses as sample size, and more so when using the number of participants as sample size (Fig. 5a and b). This is confirmed by marginally significant funnel plot asymmetry for the first model, but significant asymmetry for the second. Asymmetry can indicate publication bias, but also heterogeneity in observations (Sterne et al., 2011). In both plots, observations do not narrow around the overall estimate with increased precision. This can also indicate substantial heterogeneity in effect sizes (Sterne et al., 2011). For example, Kidd (2012) had a large sample size but reported an observation of no priming in the long post-test condition (one week after priming), Branigan et al. (2005) and Foltz et al. (2015) had small sample sizes but investigated priming in the relative clause, finding large priming effects. The manipulation of lag and the structure may be better predictors of the size of the effect than sample size in these cases.

Funnel plots for the models including moderators plot standard error against residual value rather than observed outcome – thus taking moderators into account. Both funnel plots are fairly symmetrical (see Fig. 5c and d) and the inverse of sample size did not significantly predict effect size in either model containing moderators. This suggests there is no evidence for publication bias in the priming literature once accounting for different manipulations in studies of different sizes.

Power analysis

Following Mahowald et al. (2016), we conducted several simulations to estimate the power of observing a significant priming effect with and without lexical overlap at incrementally increasing values of participants and items. To make our estimates more relevant to the child language literature, we made a few different decisions than Mahowald et al. (2016). They simulated experiments in which each participant (P) saw each word (W) once, resulting in a data frame with P*W rows. A 'word' here corresponds mostly to a verb, since the majority of priming studies test argument structure alternations, and it is this sense in which we use it. Because child languages studies often use a smaller set of verbs than adult studies, it is common for participants to see each word twice or more. Thus, we simulated data where each participant saw each word twice, resulting in data frames of P*W*2 rows. However, we only simulated random effects by subject and by item (i.e., verb), and not by the interaction between subject and item. While it is very possible that this standard deviation is non-zero in this population, these effects would be very difficult to estimate given the sample sizes of most child language studies, and models estimating them would be very unlikely to converge (especially with only 2 unique values per crossing of participant and item). Following Mahowald et al. (2016), we removed 20 % of observations at random to simulate missing data.

Mahowald et al. (2016) estimated random effect standard deviations from their validation model, a generalized linear mixed model fit to raw data they had available. We suspect between-participant and between-item heterogeneity in child priming studies will be larger than that typically observed adult priming studies, given that child language samples often contain wide age ranges and participants with varying levels of linguistic proficiency. We therefore considered two empirically-based random effects structures. To do so, we first collected the by-participant and by-item random intercepts and random slopes (for priming) from eight analyses reported in two recent developmental studies of priming, one focused on the passive (Kumarage et al., 2022) and one focused on the dative (Donnelly et al., in press). We took the mean of each of these standard deviations as our average scenario and the highest standard deviation as a high variability scenario. We simulated (and estimated) uncorrelated random effects for each scenario.

Power analyses for detecting a priming effect

Given that children produce the relevant grammatical structures at very low rates, we chose a value of −2 as our intercept (corresponding to a probability of .12). For models of main effects, we simulated from the models in Equations (7) and (8).

Model without lexical overlap:

$$Pr(Structure) = logit^{-1}\left(-2 + \tau_{1i} + \lambda_{1j} + Prime^{*}\left(0.94 + \tau_{2i} + \lambda_{2j}\right) \right) \tag{7}$$

Model with lexical overlap:

$$Pr(Structure) = logit^{-1}\left(-2 + \tau_{1i} + \lambda_{1j} + Prime^{*}\left(1.71 + \tau_{2i} + \lambda_{2j}\right) \right) \tag{8}$$

We assume a priming effect of 0.94 and 1.71 respectively, which are the estimated priming effects taken from our meta-analysis including moderators. Importantly, as the active/passive alternation was the reference level, these effect sizes estimate passive priming. These simulations correspond to a within-subjects design, where prime was coded as +/-.5, τ is a matrix of by-participant random effects, and λ is a matrix of by-item random effects. Random effects values for the two scenarios (average and high variability) are shown in Table 8. Overall, then, we

---

[4] For discussion of this issue by Wolfgang Viechtbauer, see https://stats.stackexchange.com/q/155875.

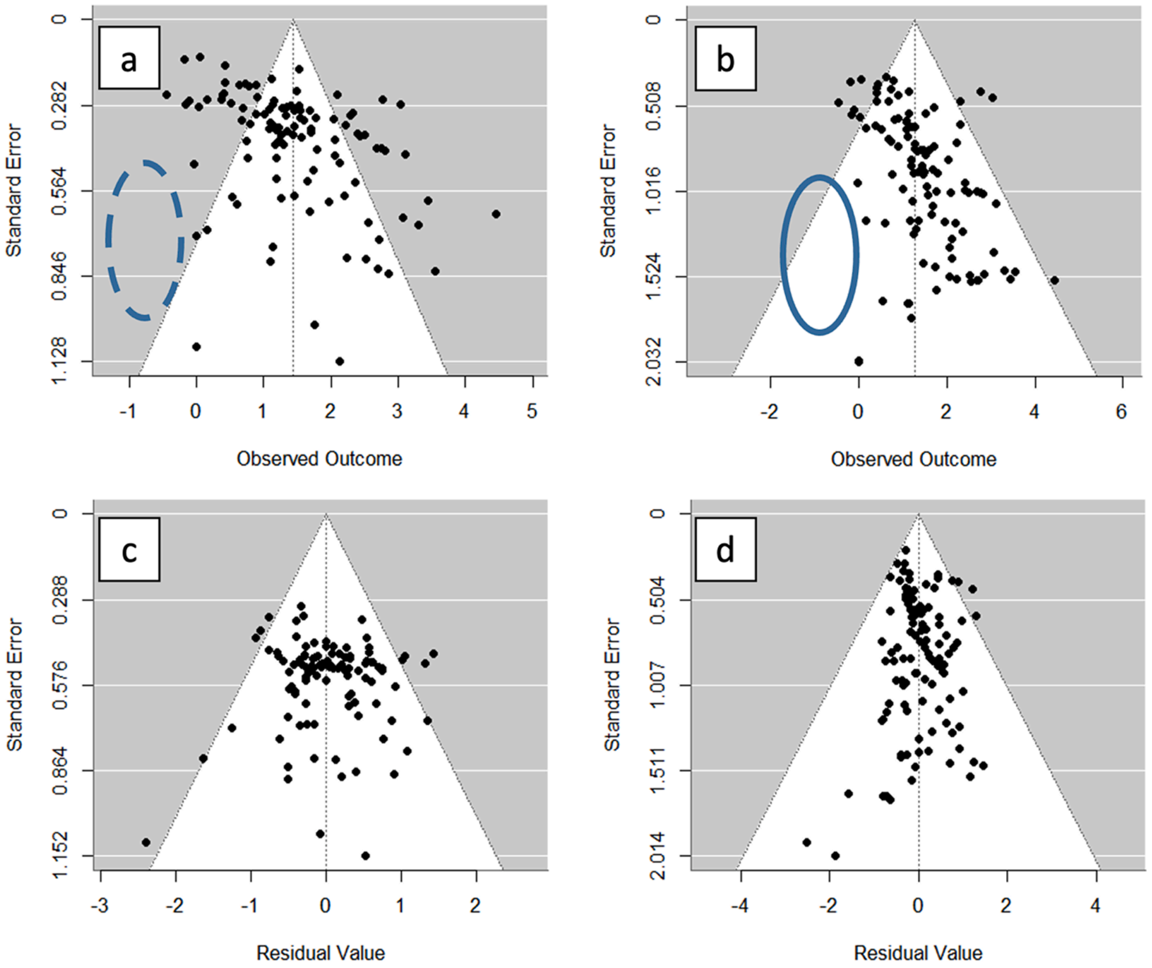


Fig. 5. Funnel plots for meta-analysis models (a) without moderators, using $N_{responses}$, (b) without moderators, using $N_{subjects}$, (c) including moderators, using $N_{responses}$, and (d) including moderators, using $N_{subjects}$.

considered 4 scenarios (2 random effects specifications with each main effect specified above). We simulated 1000 data sets for each crossing of participant number (20, 30, 40, 60, 80, 100 and 200) and item number (6, 8, 10, 12, 18, 24, 30), fitting the above model to each data set. Note that each item corresponded to two trials, not one, since we assumed most developmental priming studies would repeat words across trials.

One challenge in power analysis with mixed models is handling non-convergent models. While it is advisable to consider the full random effect structure implied by the design (Barr et al., 2013), such models are often empirically unidentified, resulting in non-convergent model with unreliable parameter estimates (Bates et al., 2015). This is likely exacerbated in developmental priming research, where between-participant variability is high and mean productions of the target structure are low. Because these data were simulated, we knew a priori that the random effect structure specified in the model was the correct one, and we included all models in our main power calculations. However, we also report on the number of non-convergent models (See Appendix D).[5]

Table 8
Values of random effects for power simulations of the priming effect.

Variability	τ_1	τ_2	λ_1	λ_2
Average	1.85	0.81	1.07	0.77
High	2.20	1.24	1.92	1.59

We plot the power for detecting a significant abstract priming effect (with and without overlap) at average and high levels of between-participant and between-item heterogeneity in Fig. 6. The average priming study in the present meta-analysis included 49.27 subjects and 9.08 items. Our analyses suggest that this as an adequate number of participants and items to detect a significant priming effect with lexical overlap, but results in less power to observe significant effect without lexical overlap (∼.67) when between-participant and between-item variability are average. In the latter case, relatively small increases in the number of items and participants would yield power close to or above .8. However, when between-participant and between-item heterogeneity are especially high, the average sample size is close to adequately powered for effects with lexical overlap (∼.75) but considerably underpowered to detect priming effects without lexical overlap (∼.36). It will be especially challenging to reach sufficient levels of power for detecting priming effects without overlap under such high levels of variability, given the number of items needed (24 or 30 words corresponding to 48 or 60 trials). However, reasonable increases in the number of participants and items may yield sufficient power to detect a priming effect with lexical overlap. At the same time, we think these high-variability estimates are overly conservative (given the large

---

[5] A small number of iterations (roughly 4.5% for simulations of both priming effects and interactions) produced implausible standard errors ($<.1$) or, in very rare cases implausible effect sizes (> 5 or < -5 on the logit scale). Because these estimation errors would lead to significant results, we removed these iterations from power calculations. We report on the number of iterations producing implausible values for each scenario (range 0—16% of iterations) in Appendix D. We have also posted power calculations with these values included to file Supp 1 on the OSF site.

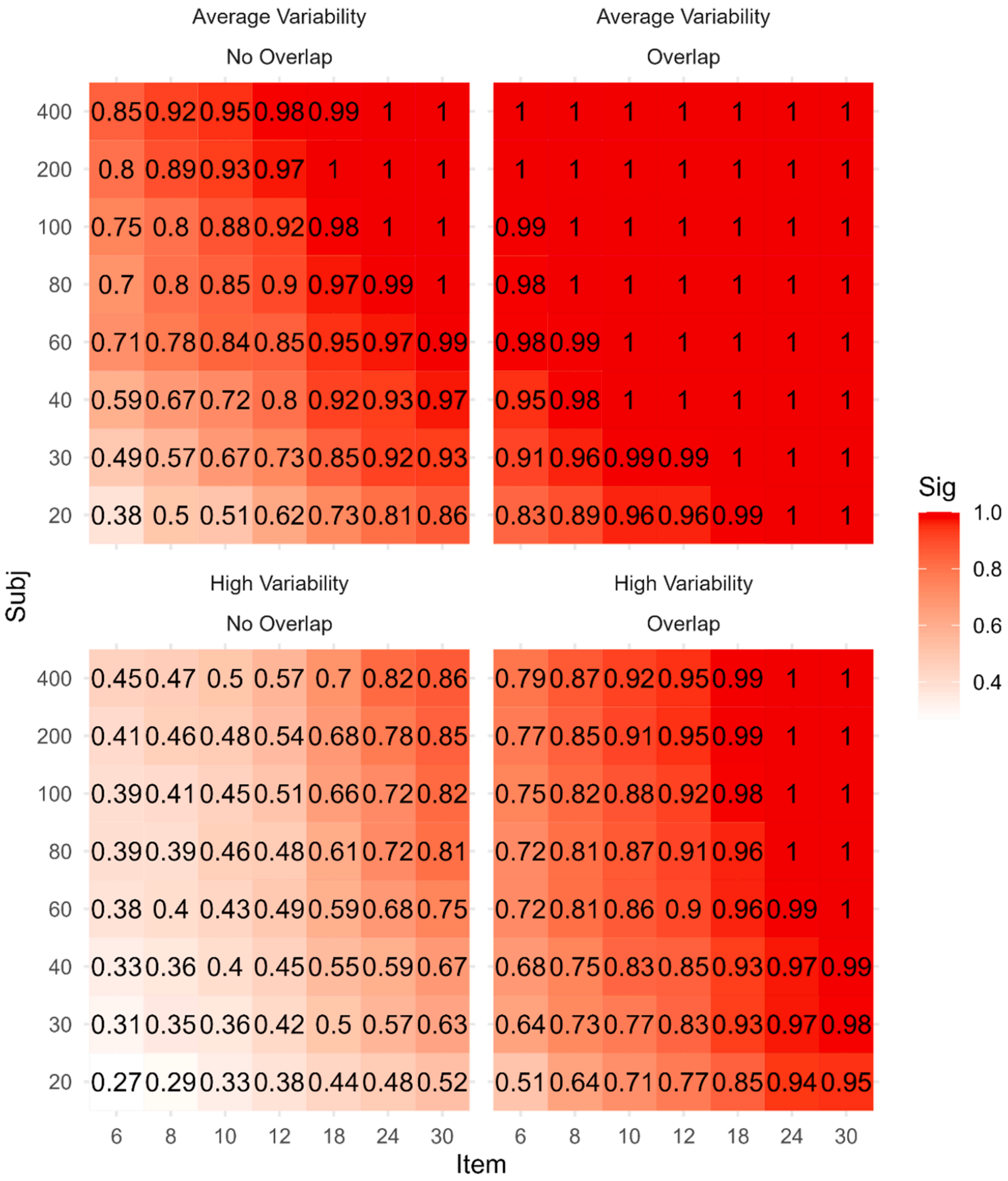


Fig. 6. Power estimates for priming effect with and without lexical overlap at average and high levels of heterogeneity.

random effect standard deviations) and might be expected in sample with (a) a very wide age range of participants or linguistic abilities (as in studies of clinical populations), and (b) constructions that have very strong verb biases, such that some verbs consistently elicit one construction or the other. Nonetheless, our results suggest that while the average sample size in the field is likely to be sufficiently powered to observe priming effects, power estimates are affected by the size of random effects, and researchers should carefully consider the degree of between-participant and between-item heterogeneity when designing their studies.

Power analyses for detecting an interaction

We next conducted simulations testing the power for detecting interactions of various magnitudes between prime and some other factor *B*. We included four random effects by random factor: a random intercept, a random slope for prime, a random slope for *B* and a random slope for the interaction. We used the average by-participant and by-item random slopes for prime from the previous simulations. As it was unclear what the random slope on the interaction should be, we divided the random slopes for the priming effect by 2, under the assumption that individual differences in the interactions would generally be smaller than individual differences in priming. We also included by-participant and by-item random effects for the main effect of *B*, though we assumed its fixed effect was 0. We simulated these random effects because *lme4* (Bates et al., 2015) estimates these by default when including random slopes for prime*B, and if their true value was 0, the model will likely not converge.

We simulated from the model in Equation (9) where prime and *B* were coded as .5, −.5 and *M* was the magnitude of the interaction effect.

Table 9
Simulated interactions of priming effects and factor B.

Interaction	*B* level	Prime level		Priming effect at level of *B*		Difference in priming effect	
		.5	−.5	Additive	Ratio	Additive	Ratio
Small (0.4)	.5	.193	.071	.122	2.72	.044	1.42
	−.5	.164	.086	.078	1.91		
Medium (0.8)	.5	.201	.065	.136	3.09	.079	1.92
	−.5	.151	.094	.057	1.61		
Large (1.6)	.5	.244	.054	.19	4.52	.175	4.00
	−.5	.127	.112	.015	1.13		

Model with interaction effect:

$$Y = logit^{-1}\left(-2+\tau_{1i}+\lambda_{1j}+Prime^*\left(0.94+\tau_{2i}+\lambda_{2j}\right)+\left(\tau_{3i}+\lambda_{3j}\right)^*B + \left(M+\tau_{4i}+\lambda_{4j}\right)^*B^*Prime\right) \quad (9)$$

$M_{\text{small}} = 0.4;\ M_{\text{medium}} = 0.8; M_{\text{large}} = 1.6$

We considered interactions of three magnitudes, 0.4, 0.8 and 1.6 on the logit scale, corresponding to a small, medium and large effect. As interactions on the logit scale can be difficult to interpret (especially since the variables for prime condition and the interaction are on different scales), we present the difference (and ratio) between priming effects at each level of *B* for each of the three interaction sizes considered on the probability scale (see Table 9). For each of the three simulated interaction sizes (0.4, 0.8, and 1.6), the crossed factors of Prime and *B* indicate the probability of producing the relevant construction. The column *Priming effect at level of B (Additive)* presents the priming effect as a difference score (e.g., Pr(Passive | Passive Prime) – Pr(Passive | Active Prime)) at each level of *B*. The column *Priming effect at level of B (Ratio)* presents the priming effect at each level of *B* as a ratio (e.g., Pr(Passive | Passive Prime)/Pr(Passive | Active Prime)). The columns *Difference in priming effect* represent the difference in priming effects across levels of *B* (as a difference score and a ratio). From these numbers, we can see that an interaction of 0.4 indicates a 4.4 percentage point difference in the priming effect across conditions, an interaction of 0.8 indicates a 7.9 percentage point difference in the magnitude of the priming effects across conditions, and an interaction of 1.6 indicates a 17.5 percentage point difference. The latter scenario reflects a situation where participants produce a large priming effect in one condition (being roughly 4.5 times more likely to produce the target structure after being primed by it than being primed by an alternative structure) and almost no priming effect in the other condition.

We simulated 100 data sets at each crossing of number of participants and number of items considered in the previous analyses,[6] and report on the proportion of significant interactions in each. In addition, we calculated the Type M and Type S error rates for each simulation (Gelman & Carlin, 2014). In low powered studies, only extreme results are significant. As a result, samples that do reach significance in low powered studies may substantially overestimate the magnitude of the true effect or even have the wrong sign. The Type M error rate is the ratio of the average observed effect size to the true effect size amongst statistically significant studies. The Type S error rate is the proportion of statistically significant effects that have the wrong sign (Gelman & Carlin, 2014). Power, Type M and Type S Error rates for each effect size are displayed in Figs. 7, 8, and 9 respectively. Note that the Type S error rate for large effect sizes was 0 at all sample sizes, so we omit this figure.

The results reveal that current samples sizes are likely too small to detect interactions with sufficient power. Detection of small interactions in particular may be very resource-intensive (requiring for example 400 participants paired with 30 items and 60 trials to reach a power of .8). Detection of medium interactions may be more achievable with larger than typical sample sizes (N = 80 to 100 with 18 items and 36 trials). Detection of large interactions with adequate power will require modest increases in the number of participants and items (for example, to 60 participants or to 12 items and 24 trials). These results also reveal that statistically significant small-to-medium sized interactions may represent substantial over-estimates of the true effect sizes (for example, at current sample sizes, significant results may over-estimate true small effects by a factor of 3.31 and medium effects by a factor of 1.99) and, particularly in the case of small interactions, may reflect a non-trivial number of sign errors (for example, 10 % of significant tests of small interactions observed in samples of 40 participants and 8 items had the wrong sign).

Discussion

In this paper we ran the first meta-analysis of syntactic priming studies in children. We found evidence for a medium-to-large main effect, with substantial heterogeneity but with no evidence for publication bias. However, the publication bias analysis did not apply to moderator effects and our power analyses suggest that the average developmental priming study is likely to be underpowered, especially concerning the detection of moderators. Once aggregating studies, the factors that significantly influenced the size of a study's effect were: (i) within- or between-subjects design, (ii) lexical overlap, (iii) structural alternation studied, and (iv) the animacy of verb arguments. We discuss each of these results in turn.

Size and significance of the effect

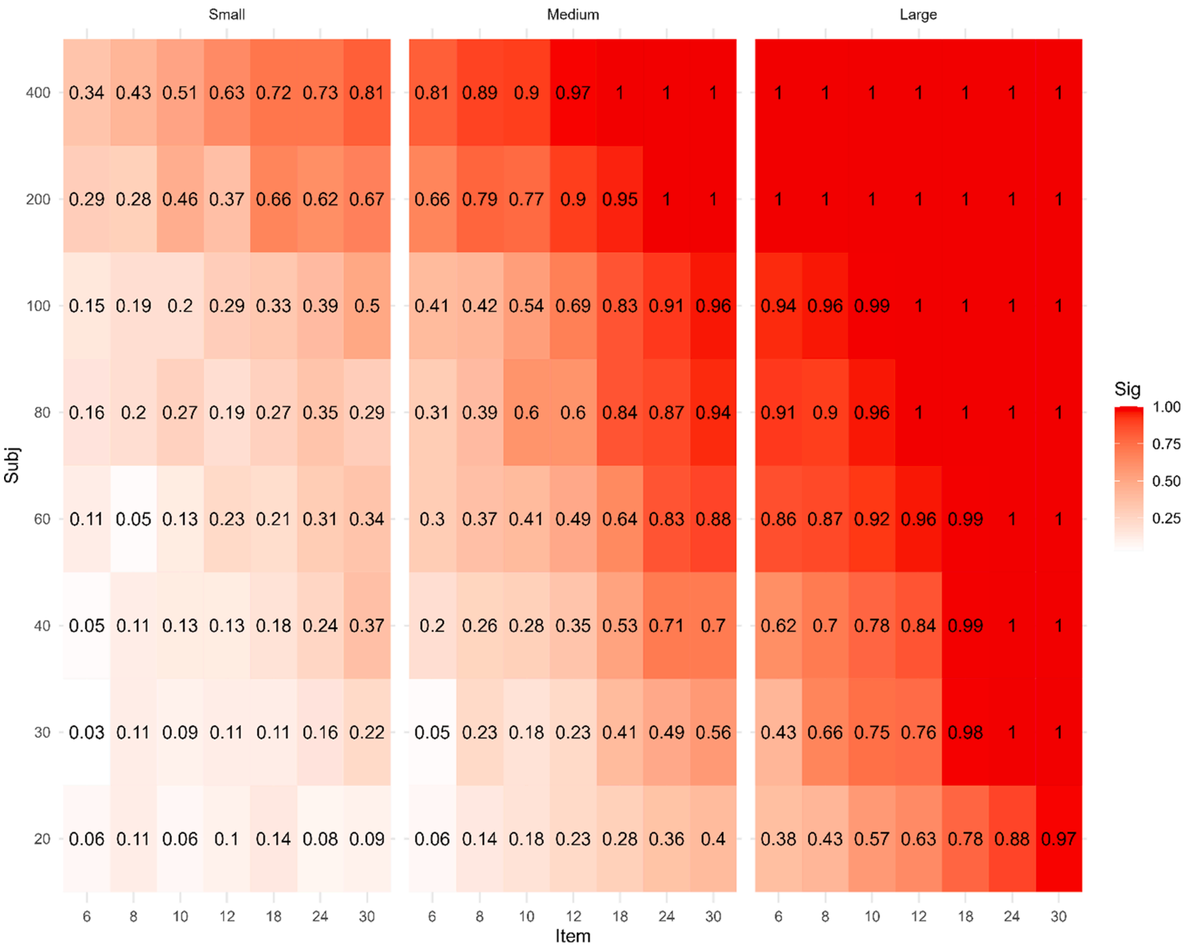
We found a significant overall syntactic priming effect, estimated at a log odds ratio of 1.42, or more conservatively estimated as 1.25. Converted to Cohen's *d* this is a medium-to-large effect of 0.78 or 0.69. Bergmann et al. (2018) found a median effect size of Cohen's $d = 0.45$ (range 0.12 – 1.24) across 12 meta-analyses of effects in language acquisition studies conducted with children aged 0–5 years. Our finding places the syntactic priming effect within the typical range of effect sizes in language acquisition, although priming is typically investigated in slightly older children. In comparison to adults, the effect is numerically larger than the effect reported in Mahowald et al.'s (2016) meta-analysis of syntactic priming studies in adults. After controlling for moderators (e.g. lexical overlap and between-subjects design, which both significantly increased the size of the effect), in the $N_{\text{responses}}$ model, the estimated log odds ratio was 0.94 [95 % CI, 0.58, 1.29] or Cohen's $d = 0.52$. In comparison, after controlling for moderators (including lexical overlap), Mahowald et al. (2016) reported a log odds ratio of 0.52 [95 % CI, 0.22, 0.82] or Cohen's $d = 0.29$ across syntactic priming studies in adults.

Moderators of the effect

As expected, given the range of design choices and variables manipulated in syntactic priming experiments, there was significant heterogeneity in the overall syntactic priming effect. Even after including moderators in the analysis, there was significant

---

[6] Simulations of interactions were considerably more time-intensive than simulations for main effects, so we reduced the number of iterations.



Small

Subj	6	8	10	12	18	24	30
400	0.34	0.43	0.51	0.63	0.72	0.73	0.81
200	0.29	0.28	0.46	0.37	0.66	0.62	0.67
100	0.15	0.19	0.2	0.29	0.33	0.39	0.5
80	0.16	0.2	0.27	0.19	0.27	0.35	0.29
60	0.11	0.05	0.13	0.23	0.21	0.31	0.34
40	0.05	0.11	0.13	0.13	0.18	0.24	0.37
30	0.03	0.11	0.09	0.11	0.11	0.16	0.22
20	0.06	0.11	0.06	0.1	0.14	0.08	0.09

Medium

Subj	6	8	10	12	18	24	30
400	0.81	0.89	0.9	0.97	1	1	1
200	0.66	0.79	0.77	0.9	0.95	1	1
100	0.41	0.42	0.54	0.69	0.83	0.91	0.96
80	0.31	0.39	0.6	0.6	0.84	0.87	0.94
60	0.3	0.37	0.41	0.49	0.64	0.83	0.88
40	0.2	0.26	0.28	0.35	0.53	0.71	0.7
30	0.05	0.23	0.18	0.23	0.41	0.49	0.56
20	0.06	0.14	0.18	0.23	0.28	0.36	0.4

Large

Subj	6	8	10	12	18	24	30
400	1	1	1	1	1	1	1
200	1	1	1	1	1	1	1
100	0.94	0.96	0.99	1	1	1	1
80	0.91	0.9	0.96	1	1	1	1
60	0.86	0.87	0.92	0.96	0.99	1	1
40	0.62	0.7	0.78	0.84	0.99	1	1
30	0.43	0.66	0.75	0.76	0.98	1	1
20	0.38	0.43	0.57	0.63	0.78	0.88	0.97

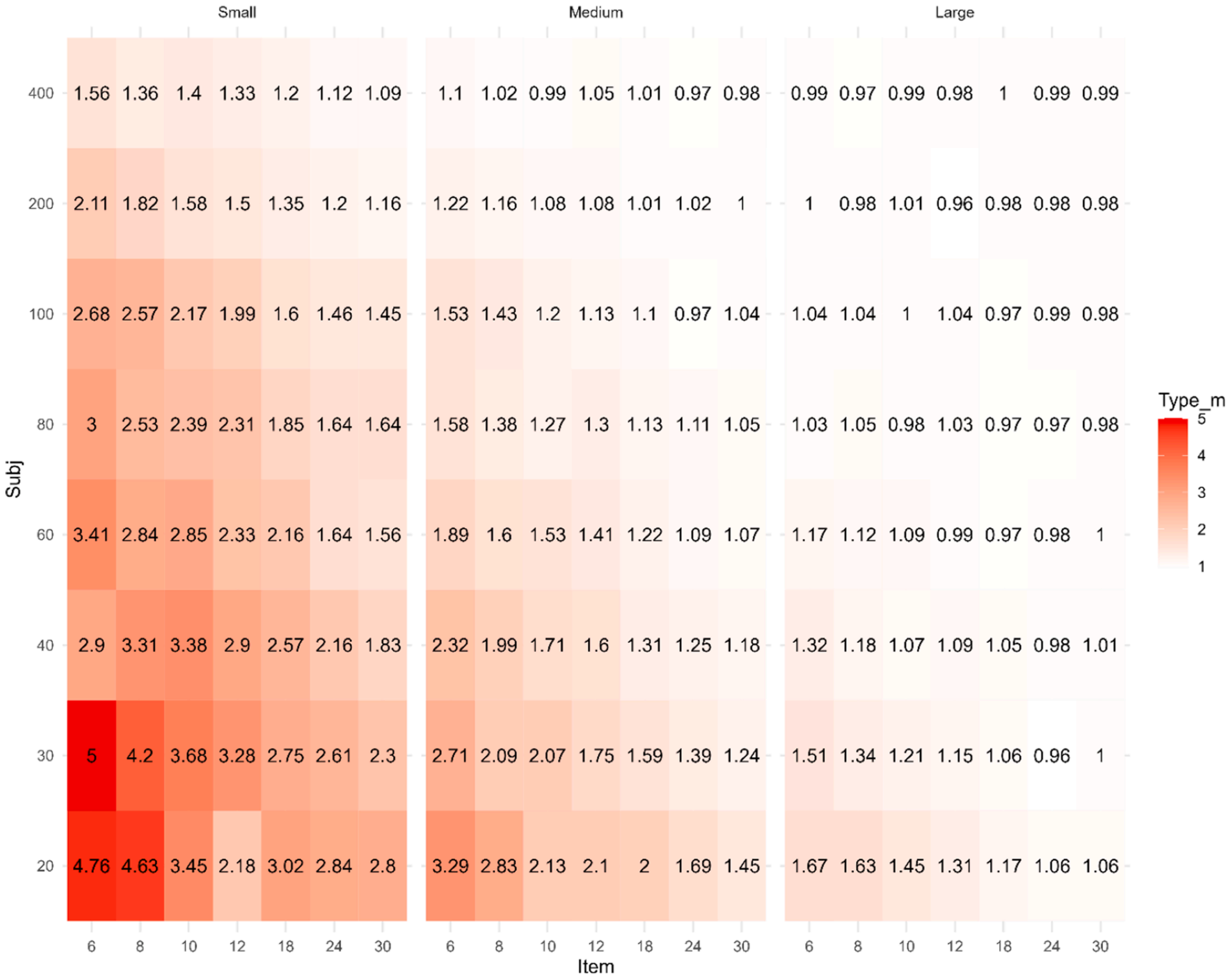
Fig. 7. Power estimates for detecting small, medium, and large interactions.

heterogeneity in effect sizes for the $N_{responses}$ model, suggesting our list of moderators was not exhaustive. Besides lexical overlap, the significant moderators in our meta-analysis tend not to be manipulated variables but ones that vary across studies, so despite issues with low power we remain confident these findings are not greatly affected by publication bias. In our discussion of moderators of the effect, we follow Goodman (1991) in assuming that greater insight can be gained by combining thoughtful analysis with the quantitative results of a meta-analysis. That is, when the summary result differs from the results of individual studies, we should not accept either finding without consideration of why they differ. The value of meta-analysis is not only the summary results, but also the systematic consideration of the current state of evidence in a research field.

Abstract and lexically-based priming

Of particular importance to theories of syntactic acquisition is evidence regarding the emergence of abstract syntactic knowledge in comparison to lexically-based syntactic knowledge. Our findings support the early abstraction of syntax with no reliance on lexically-based representation for the most frequently studied passive structure. We found both a significant abstract priming effect (significant intercept) and lexical boost effect (significant lexical overlap effect). That is, lexical overlap increased the magnitude of a study's reported syntactic priming effect but was not necessary to observe one (cf. main priming effect to priming under lexical overlap in Fig. 4a). The findings reflect a growing body of evidence that lexically-based priming does not precede abstract syntactic priming in English-speaking children (Kumarage et al., 2022; Peter et al., 2015; Rowland et al., 2012; but c.f. Savage et al., 2003), and therefore evidence against a transition from lexically-based to abstract syntactic representation (Savage et al., 2003; Tomasello, 2003). However, we also found significant effects of the structural alternation studied. For example, we did not find significant abstract priming for the dative alternation. As we later discuss, syntactic priming research is heavily skewed towards the active/passive structural alternation and generalising beyond the current state of the literature is therefore difficult.

Concerning the developmental trajectory of each effect, age did not moderate the magnitude of either abstract syntactic priming or the lexical boost (Fig. 4b). Several factors limit the interpretation of this result: limited power for studies to detect an interaction effect, an inability to detect non-linear effects, and the wide age ranges of included studies. A stable lexical boost effect contrasts with accepted findings that the effect increases over development. The highest quality evidence to date comes from longitudinal data, which found an increasing lexical boost effect (Kumarage et al., 2022). Most cross-sectional data also support this finding (Peter et al., 2015; Rowland et al., 2012; cf. Savage et al., 2003). However, we found a stable effect and an almost identical estimate of the impact of lexical overlap to that found in adults (Table 7: 0.74; Mahowald et al., 2016, p.8.: 0.76). We note that although it is a key point of contention, the study of lexical influence in syntactic priming in children is at an early stage. Eight studies have manipulated lexical overlap (Branigan et al., 2005; Branigan & McLean, 2016; Fazekas et al., 2020; Foltz et al., 2015; Kumarage et al., 2022; Peter et al., 2015; Rowland et al., 2012; Savage et al., 2003), and two have investigated priming only under lexical overlap (Buckle et al., 2017; Savage et al., 2006). All but one were conducted in English. Notably, the inclusion of a single outlier (Fazekas et al., 2020: long lag lexical overlap condition) produced a significant decreasing lexical boost effect in our analysis,



Subj	Small							Medium							Large						
Item	6	8	10	12	18	24	30	6	8	10	12	18	24	30	6	8	10	12	18	24	30
400	1.56	1.36	1.4	1.33	1.2	1.12	1.09	1.1	1.02	0.99	1.05	1.01	0.97	0.98	0.99	0.97	0.99	0.98	1	0.99	0.99
200	2.11	1.82	1.58	1.5	1.35	1.2	1.16	1.22	1.16	1.08	1.08	1.01	1.02	1	1	0.98	1.01	0.96	0.98	0.98	0.98
100	2.68	2.57	2.17	1.99	1.6	1.46	1.45	1.53	1.43	1.2	1.13	1.1	0.97	1.04	1.04	1.04	1	1.04	0.97	0.99	0.98
80	3	2.53	2.39	2.31	1.85	1.64	1.64	1.58	1.38	1.27	1.3	1.13	1.11	1.05	1.03	1.05	0.98	1.03	0.97	0.97	0.98
60	3.41	2.84	2.85	2.33	2.16	1.64	1.56	1.89	1.6	1.53	1.41	1.22	1.09	1.07	1.17	1.12	1.09	0.99	0.97	0.98	1
40	2.9	3.31	3.38	2.9	2.57	2.16	1.83	2.32	1.99	1.71	1.6	1.31	1.25	1.18	1.32	1.18	1.07	1.09	1.05	0.98	1.01
30	5	4.2	3.68	3.28	2.75	2.61	2.3	2.71	2.09	2.07	1.75	1.59	1.39	1.24	1.51	1.34	1.21	1.15	1.06	0.96	1
20	4.76	4.63	3.45	2.18	3.02	2.84	2.8	3.29	2.83	2.13	2.1	2	1.69	1.45	1.67	1.63	1.45	1.31	1.17	1.06	1.06

Fig. 8. Type M errors for small, medium, and large interactions (ratio of average significant effect size to true effect size).

suggesting this interaction effect was not estimated with great certainty. We are therefore cautious in interpreting our finding that there is no change in the lexical boost effect over development.

Several factors should be considered when interpreting our finding of a developmentally stable abstract priming effect. Firstly, participants must know the structure before they can be primed (Kidd, 2012; Kumarage et al., 2022). Whilst Chang et al. (2006) predict a decreasing abstract priming effect with increasingly stable representations, they first predict an increasing priming effect as the model learns representations abstract enough to observe priming (p.261.). The linear age term in our meta-analysis would not in principle be able to detect this inverse U-shaped trajectory. However, in line with this proposal, Kumarage et al. (2022) only observed a decrease in priming from 36 to 54 months when analysing data from children who were confirmed to know the passive at 36 months (and when excluding an item effect). As further evidence, our meta-analysis did not find significant priming of datives, an alternation that takes much longer to gain productive mastery over (Donnelly et al., in press), or priming of agent-patient order in the Tagalog Symmetrical Voice transitive, which is more complex than the transitive in European languages and thus seems to be acquired over a longer developmental trajectory (Garcia et al., 2020; Garcia & Kidd, 2020; Kidd & Garcia, 2022). Secondly, any developmental effects, but especially an increase then decrease, will be difficult to observe in samples containing a range of proficiency levels, such as those in syntactic priming studies, where age ranges are typically about 18 months (Messenger, 2022; observations included in our analysis had an average age range of 16.58 months). Indeed, cross-sectional studies have found a range of developmental trajectories (increasing: Peter et al., 2015; decreasing: Rowland et al., 2012; stable: Hsu, 2019; Messenger, 2021; Messenger, Branigan, & McLean, 2012; Messenger, Branigan, McLean, & Sorace, 2012). Meta-analysis is essentially a cross-sectional analysis of observations of syntactic priming and so faces these same issues. Thus, any developmental trends that might exist could be obscured by differences across studies in design and participant characteristics. Overall, the developmental trajectory of priming effects is of great theoretical interest; however, the current evidence base does not provide conclusive support for any linear effects and is limited in not examining non-linear effects. We strongly encourage future research that is longitudinal and carefully considers both non-linear effects and the range of proficiency levels in the sample.

Comparing across a larger age gap, from children to adults, the implicit learning account predicts a decrease in priming magnitude (Chang et al., 2006). Interestingly, our estimate of the syntactic priming effect in children was numerically larger than that in adults (Mahowald et al., 2016). Three factors complicate the comparison between these estimates. Even when comparing across a larger age gap, if our estimate contains a mixture of developmental stages associated with varied priming magnitudes, it will not be an accurate point of comparison. Secondly, although we have compared estimates that take into account moderators of priming (e.g., greater use of between-subjects and blocked designs in children), the moderators we included and therefore the variables controlled for were not the same across the two meta-

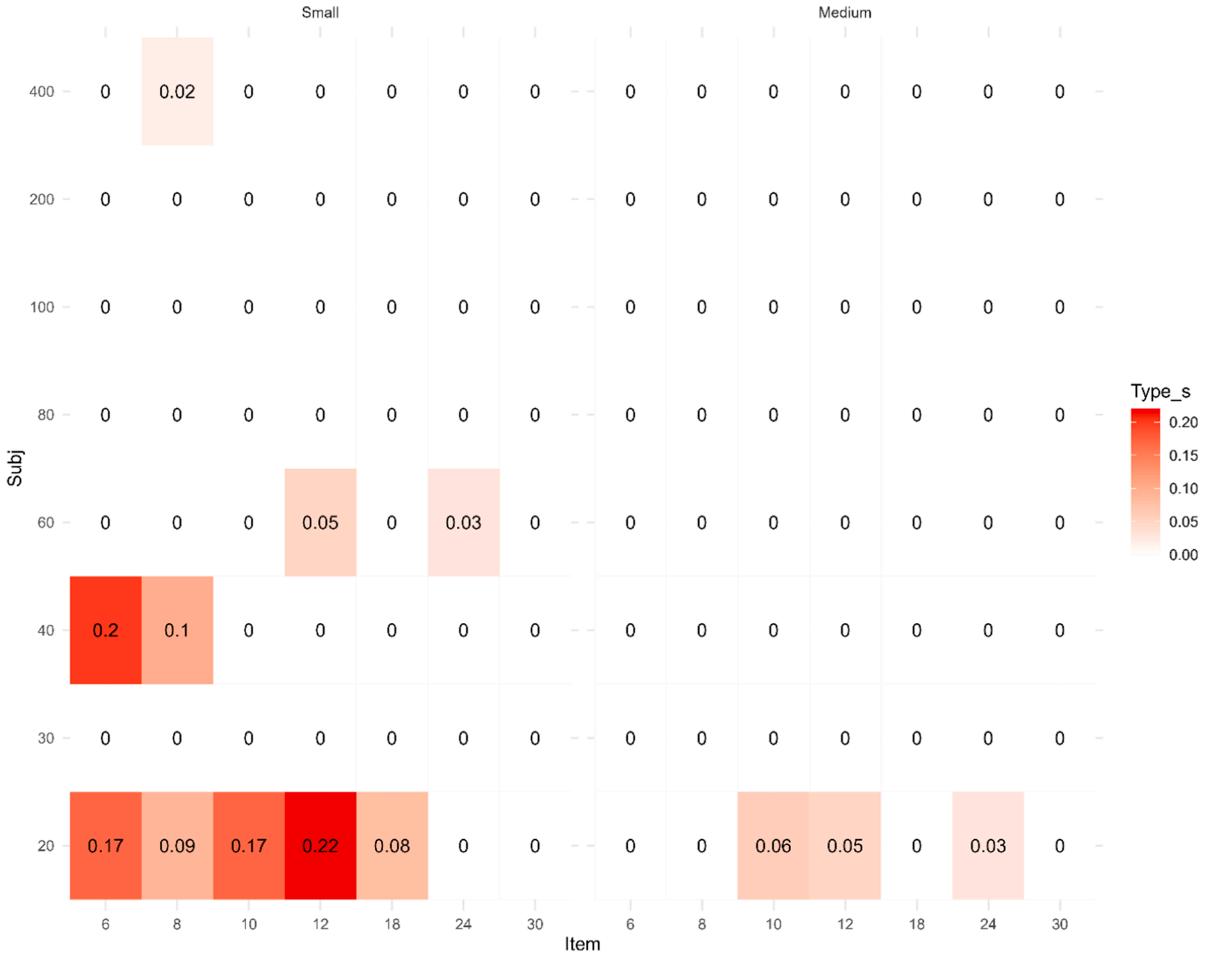


Fig. 9. Type S errors for small and medium interactions (likelihood of getting an effect with the wrong sign)..

analyses. Finally, design differences remain between child and adult priming studies that could influence priming magnitude. For example, studies in children use fewer fillers and whilst the dative is the most frequently studied alternation in adults, the active–passive alternation is more frequently studied in children. Thus, although a comparison with Mahowald et al. (2016) suggests some support for a decrease in abstract priming across age, which is predicted by the implicit learning account, the two literatures differ on many different dimensions that mean a direct comparison, at this point in time, is not warranted. Future studies that compare multiple age groups and adults across multiple alternations using the same method are required.

Explicit vs implicit processes

An important takeaway from our findings is that study design choices can have large impacts on the magnitude of the syntactic priming effect. These impacts are likely to reflect insights into syntactic priming. Researchers' aims of reducing task demands and interference between trials have led to a higher frequency of between-subjects designs in the developmental literature than in the adult literature. Choosing to include only within-subjects designs would result in excluding a large amount of information from the literature (40 of 108 observations, see Table 3), so we chose to combine design types in our analysis. Morris and DeShon (2002) describe three main issues with combining effect size estimates from within- and between-subjects designs. Two are statistical: a common metric must be used as the effect size, and variance estimates must be calculated using design-specific methods. In order to satisfy these conditions, we used the log odds ratio and the Becker-Balagtas method to adjust variance estimates for within-subjects designs. The third issue is conceptual: do the two designs estimate the same effect? Whilst both designs answer the question of whether presenting primes increases the frequency of that structure in subsequent speech, the large difference in effect size suggests additional factors. Messenger (2021) points out that between-subjects designs model only one structure rather than providing varied input. Although child directed language has a degree of 'burstiness' (Lester et al., 2022), the kind of flooding that occurs in between-subjects priming designs does not realistically reflect the input.

We suggest that this flooding may lead to explicit processes that point children in the direction of using the structure modelled and artificially inflate the implicit priming effect. Firstly, children may invoke explicit memory. Chang et al. (2006) propose that implicit learning effects underlie long-term priming, whilst the immediate lexical boost effect relies on explicit memory. Hartsuiker et al. (2008) show evidence supporting this prediction: abstract priming was persistent, whilst the lexical boost was not. They also proposed that both implicit and explicit processes are required to explain such findings. Interestingly, in their data, the prime remaining visible on-screen after the target was presented also appeared to increase the magnitude of priming. This suggests that not only lexical overlap but other factors can act as retrieval cues, which facilitate the explicit memory mechanism and increase the syntactic priming effect size. If cues to explicit memory can facilitate priming, part of the effect is likely attributed to explicit memory. Unvaried input in between-subjects designs may act as a retrieval cue, facilitating explicit memory processes and increasing priming. Secondly, as suggested by an anonymous reviewer, such input may play a normative role, indicating to children that the structure used by the adult experimenter is the one that *should* be used in the context of the task. Given the focus on syntactic priming as implicit learning in children, we recommend the consideration of potential explicit sources of priming when designing studies. For example, the use of games such as *bingo* and *snap* makes the task more enjoyable for children, but may introduce some reliance on explicit memory for the prime structure.

Prime repetition could feasibly act as a retrieval cue in syntactic priming studies but, as in adults (Mahowald et al., 2016), was not a significant effect in the meta-analysis. Several studies of less experienced speakers have directly manipulated prime repetition and found an effect (L2 speakers: Gámez & Vasilyeva, 2015; 3yos: Shimpi et al., 2007; late acquired structure: Gámez & Shimpi, 2016). Other studies directly manipulating the variable have found no effect (Hsu, 2014; Huttenlocher et al., 2004). We included the interaction between prime repetition and age, which was not significant. Though the wide age ranges within studies make the effect more difficult to interpret, this result does not support prime repetition acting as a retrieval cue only for less experienced speakers. This finding also supports shared representation between comprehension and production (Bock et al., 2007).

Structural alternation

Our findings regarding the structure studied raise several important points. The first regards the relative frequencies of structures in a language. Though there were only two studies of the relative clause/adjective-noun alternation, it showed a much larger priming effect than the passive. One explanation for this potential difference could be a larger difference in frequencies of adjective-noun vs. relative clauses as compared to actives vs. passives. That is, adjective-noun combinations are highly preferred in languages like English, whereas a relative clause structure may be more marked, even than the passive. Secondly, as previously mentioned, children must have acquired the structure to demonstrate priming of it (Kidd, 2012). The subject relative clause is early-acquired and the larger priming effect could additionally be attributed to the fact that most if not all children in the studies had sufficient mastery to display priming compared to only some in the case of the passive. At the other end of the scale, smaller and non-significant priming of the Tagalog symmetrical voice transitive may reflect the protracted acquisition of this comparatively complex set of structures (for evidence of restrictions on priming with the Tagalog symmetrical voice system see Garcia et al., 2023). Related to the acquisition of structure is the acquisition of structural *alternation*. Whereas the English active/passive alternation occurs across a broad category of transitive verbs, children must learn restrictions on which verbs participate in the dative alternation (Gropen et al., 1989). Learning such restrictions may delay the demonstration of syntactic priming. That is, even when children have acquired the double-object dative for particular lexical items (e.g. "*Give me_*"), they may not have abstracted this frame, and therefore the alternation to other lexical items. The dative, in both forms, is acquired at least as early as the passive in English (Campbell & Tomasello, 2001; Marchman et al., 1991), yet we found significant abstract priming for the active/passive alternation but not the dative alternation. Therefore, an important consideration in syntactic priming studies is the age of participants and the age of acquisition of both the structure and structural alternation.

As a final point regarding the varying effect of structural alternations, we note that of the studies included in this meta-analysis 71 % were conducted in English, and 58 % on the active/passive alternation (see Table 3). Therefore, using findings of syntactic priming studies in children to support broad claims about syntax acquisition requires caution. For example, in English there are few structures that allow a patient-focus (e.g., the passive and object cleft), whereas languages that have more flexible word order, such as Russian and Spanish, permit the use of alternative structures that achieve a functionally equivalent outcome (e.g., the Spanish middle voice, or object fronting in Russian). These alternatives can also be primed by passives, suggesting that the locus of priming is not purely syntactic (Gámez et al., 2009; Vasilyeva & Waterfall, 2012). Overall, the dominance of English as a target language is consistent with the broader sampling bias found in the language acquisition literature (Kidd & Garcia, 2022). Given our finding that priming effects vary with structure (and language), researchers should be careful when considering how their findings from one language generalise across languages, if at all.

Other study design variables

Animacy configurations moderated the priming effect, with canonical animacy patterns resulting in a higher priming effect. There has been a long-standing tendency in the psycholinguistic literature to neutralise animacy cues in experimental contexts, under the assumption that such cues are not revealing about purely syntactic processes. This tendency is not restricted to priming studies; for instance, it is common in studies of syntactic processing (e.g., relative clause comprehension, Gibson, 2000; see Kidd et al., 2007; Mak et al., 2002). Our finding that priming is higher in instances of canonical animacy configurations suggests that children's structural knowledge is best revealed when their input-based expectations about syntactic-semantic correlations for different argument roles are met.

Two other design features of the developmental priming studies did not have an influence on priming. Firstly, comparing to a no-prime baseline was uncommon and did not impact on the magnitude of priming, and neither did using a blocked rather than alternating design (c.f., Hsu, 2019). However, given that between-subjects studies showed considerably larger priming than within-subjects studies, a study which measures baseline responses, exposes participants to primes of only one structure, and then measures primed responses is subject to the issues considered by Messenger (2021) and in our section on explicit vs. implicit processes.

Finally, a long lag between primes and targets did not affect the syntactic priming effect. However, there were few and heterogenous observations of priming under this condition (1 hour lag: Hsu, 2019; 1 week lag: Kidd, 2012; 1 week to 1 month lag: Savage et al., 2006). The 95 % confidence interval for the predicted priming effect in this condition overlapped with both no priming effect and the main priming effect (Fig. 4a). This confirms that the effect was not precisely estimated and so conclusions about long-term priming in children are preliminary.

Considering variability

The assumptions of meta-analysis focus on moderators of the size of the effect. However, variability in the effect can also change over development. Kumarage et al. (2022) found considerable variation in 3-year-olds ability to be primed but more reliable priming at older ages. This kind of effect can tell us something about acquisition but is not well captured in a meta-analysis, such as the link between priming and proficiency. We encourage future research, particularly longitudinal research, to consider other impacts of moderators. Given the expense of longitudinal research and our power analyses, this might require multi-lab studies such as those conducted by the *Many Babies* consortium (Frank et al., 2020).

Power in the syntactic priming literature

The power analysis showed that syntactic priming studies are generally underpowered, especially for the types of questions they usually aim to answer. Whilst studies typically have sufficient power to detect a priming effect under lexical overlap (because this is a large effect), for the abstract priming effect, an average study with about 40 participants and 8 verbs presented twice each has only 67 % power. In a scenario with high heterogeneity between subjects and items, the power of an average study drops even further to 75 % for priming with lexical overlap and 36 % for abstract priming (though this may reflect a lower limit on power, given the high levels of between-participant and between-item heterogeneity assumed). Importantly, our power analyses assume a priming effect of the magnitude of priming in the active/passive alternation. Smaller priming effects would require larger samples to achieve the same power. Also important to note is that whilst the average overall sample is close to 50, average samples per cell are in reality closer to 30. This means that analyses within a cell, for example of each age group separately, have lower power to detect a priming effect.

Although some studies aim to examine whether the priming effect exists at all, many aim to test hypotheses about potential moderators of the effect. Mahowald et al. (2016) found low power and lower evidential value for interactions than the main effects in adults. Correspondingly, we found that current studies in children are seriously underpowered for this purpose, precluding the detection of all but large interaction effects. These results are consistent with recommendations from statisticians that interactions may require between 4 and 16 times more data to estimate than main effects (Gelman et al., 2020, pp. 301-304). This threatens the validity of research findings, for several reasons. Low power can result in inconclusive results and high Type II error rates, where we fail to reject the null hypothesis. Additionally, in low powered studies, significant results are not representative of the true effect size, resulting in the possibility that studies could overestimate the underlying effect – Type M error – or even find an effect in the opposite direction – Type S error. Our simulations suggest that significant results at current sample sizes may overestimate the magnitude of true effect sizes. When combined with publication bias, this is a serious problem. Meta-analysis can mitigate Type II errors by aggregating results to gain greater power, but if the sample of studies is not in fact representative, it will be aggregating exaggerated or invalid findings. We did find that, once accounting for moderators, there is no evidence for publication bias in the priming effect. However, this does not imply whether the effects of moderators themselves are subject to publication bias. This raises concerns regarding the quality of current evidence for moderator effects: given power estimates, reported significant effects may be overstated. Future research requires much higher power than currently achieved, but, encouragingly, more recent studies do contain substantially larger samples than earlier research (e.g., Fazekas et al., 2020; Garcia & Kidd, 2020; Messenger, 2021).

Researchers typically conceive of increasing power by increasing sample size. However, we are attempting to generalise across the language from a selection of items, as well as across the population from a selection of participants. Thus, another way to increase power is to increase items, which is often overlooked. In most alternations tested (excluding the N-Adj/RC alternation), the relevant number of items in a priming study is the number of target verbs presented to participants rather than the number of trials. This is because, all other things being equal, we would expect responses to vary more strongly by verb, which differ in their associations with various grammatical structures. This poses a problem in developmental studies: there are fewer verbs that are well known by children and appropriate for experimental use, especially for a structure like the dative (although for verb-specific effects in the active–passive alternation see Kumarage et al., 2022). However, if possible, we encourage researchers to consider increasing the number of verb types (vs. tokens) as well as the number of participants. For the abstract priming effect, an increase from 8 to 12 unique verbs requires 40 participants for 80 % power, while using 8 unique verbs requires 80 participants for 80 % power (assuming average heterogeneity).

While our power analyses yielded important information about the existing evidence base, they necessarily rely on several assumptions which may by imperfect representations of reality. First, we chose an intercept value of −2 on the logit scale, which corresponds to an average probability of producing the target construction of .12. We believe this is justifiable given that children often produce the target construction at near-floor levels (indeed large numbers often produce 0 instances of the target construction; see Donnelly et al., in press; Kumarage et al., 2022). However, it may be easier to reliably detect moderation of the priming effect if the baseline rate of production is higher. Second, consistent with the widespread use of mixed models, we assumed Gaussian random effects. In reality, this is very unlikely to be true for participants. In empirical studies, we have observed that large numbers of children often fail to produce the target structure at all, with one possible explanation being that they have not acquired it (Donnelly et al., in press; Kumarage et al., 2022), meaning that the true distribution is probably bimodal. It is unclear what the consequences of this would be for power estimation.

Reporting recommendations

Our meta-analysis included a wide range of studies that studied different structures and languages, and which made a range of methodological choices. Accordingly, our analysis could be considered a broad quantitative overview of the literature. Meta-analyses could be applied in the future to more specific questions. For example, meta-analyses of particular structural alternations could investigate more specific moderators like particular animacy configurations and thematic role structures in passives (Messenger, Branigan, McLean, & Sorace, 2012; Vasilyeva & Gámez, 2015), or verb bias effects in datives (Peter et al., 2015). To aid in this endeavour we suggest reporting the following information. Firstly, a table of cell count data separated by condition is required. That is, the number of dependent, alternate and *other* responses in the primed and unprimed condition at each level of each variable manipulated in the study (e.g., Table 1 in Branigan & McLean, 2016). We recommend reporting raw numbers instead of or in addition to proportions: proportions can be calculated from raw numbers but if excluded trials are not reported by condition, the reverse is not always true. Most researchers excluded *other* responses from their analysis but we still recommend reporting their raw frequencies. Regarding this analysis decision, see Appendix C for a comparison between meta-analyses run with and without *other* responses. Including *other* responses appears to numerically reduce the priming effect and potentially the power to detect moderators. Secondly, for within-subjects studies, we recommend reporting the correlation between dependent structure production in the primed and unprimed condition for each level of the variable manipulated in the study. We report how we calculated an estimate of this correlation in Appendix B. Thirdly, when moderators are manipulated within-subjects, reporting the correlation between priming in different experimental conditions allows dependent observations to be included in a meta-analysis. We report how we calculated an estimate for dependent observations at different levels of lexical overlap, and structure studied in Appendix B. Finally, the recent move towards making raw data accessible online allows researchers to calculate both these and other potentially useful statistics themselves (Fazekas et al., 2020; Garcia & Kidd, 2020).

Conclusion

In this paper we reported, to our knowledge, the first meta-analysis of syntactic priming in children. We found evidence of a medium-to-large priming effect, which appears to be influenced by but not dependent upon several moderating variables, including lexical and semantic factors of content words and methodological choices made by different researchers. Therefore, like in adults (Mahowald et al., 2016), syntactic priming is a robust though variable phenomenon in children. Additionally, it fares well as a reliable effect against other meta-analysis estimates in other domains of language development (Bergmann et al., 2018). These features make it an important empirical phenomenon bridging language acquisition and adult language processing. However, studies using syntactic priming to answer theoretical questions involving moderators of the effect are currently seriously underpowered, limiting the reliability of their findings. At the same time, we found that the current evidence base is limited in several ways, which prevents us from generalising from the data to acquisition in general. Notably, we found that the majority of studies focused on English and the active/passive structural alternation, suggesting that investigating a wider array of languages and structures is an important priority (Atkinson, 2022; Kidd & Garcia, 2022). Additional future directions include the careful investigation of developmental effects in priming and the nature and persistence of long-term priming effects.

Author Note

This research was supported by the Max Planck Society and the Australian Research Council (CE140100041: CI Kidd).

CRediT authorship contribution statement

Shanthi Kumarage: Conceptualization, Investigation, Formal analysis, Visualization, Writing – original draft. **Seamus Donnelly:** Formal analysis, Visualization, Writing – review & editing. **Evan Kidd:** Conceptualization, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

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 paper.

Appendix A

References for included papers

Allen, M. L., Haywood, S., Rajendran, G., & Branigan, H. (2011). Evidence for syntactic alignment in children with autism. *Developmental Science*, *14*(3), 540–548. https://doi.org/10.1111/j.1467–7687.2010.01001.x.

Bencini, G. M. L., & Valian, V. V. (2008). Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language*, *59*(1), 97–113. DOi: 10.1016/j.jml.2007.12.007.

Bidgood, A., Pine, J. M., Rowland, C. F., & Ambridge, B. (2020). Syntactic representations are both abstract and semantically constrained: Evidence from children's and adults' comprehension and production/priming of the English passive. *Cognitive Science*, *44*(9). DOi: 10.1111/cogs.12892.

Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult-child dialogue. *Journal of Memory and Language*, *91*, 141–157. DOi: 10.1016/j.jml.2016.02.002.

Branigan, H. P., McLean, J. F., & Jones, M. (2005). A blue cat or a cat that is blue? – Evidence for abstract syntax in young children's noun phrases. In A. Brugos, M. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 109–121). Cascadilla Press.

Branigan, H. P., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition*, *157*, 250–256. DOi: 10.1016/j.cognition.2016.09.004.

Buckle, L., Lieven, E., & Theakston, A. L. (2017). The effects of animacy and syntax on priming: A developmental study. *Frontiers in Psychology*, *8*. DOi: 10.3389/fpsyg.2017.02246.

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*, *7*(11), 180877. DOi: 10.1098/rsos.180877.

Foltz, A., Thiele, K., Kahsnitz, D., & Stenneken, P. (2015). Children's syntactic-priming magnitude: Lexical factors and participant characteristics. *Journal of Child Language*, *42*(4), 932–945. DOi: 10.1017/S0305000914000488.

Gámez, P. B., & Shimpi, P. M. (2016). Structural priming in Spanish as evidence of implicit learning. *Journal of Child Language*, *43*(1), 207–233. DOi: 10.1017/S0305000915000161.

Gámez, P. B., Shimpi, P. M., Waterfall, H. R., & Huttenlocher, J. (2009). Priming a perspective in Spanish monolingual children: The use of syntactic alternatives. *Journal of Child Language*, *36*(2), 269–290. DOi: 10.1017/S0305000908008945.

Gámez, P. B., & Vasilyeva, M. (2015). Increasing Second Language Learners' Production and Comprehension of Developmentally Advanced Syntactic Forms. *Language Learning and Development*, *11*(2), 128–151. DOi: 10.1080/15475441.2014.894873.

Garcia, R., & Kidd, E. (2020). The acquisition of the Tagalog symmetrical voice system: Evidence from structural priming. *Language Learning and Development*, *16*(4), 399–425. DOi: 10.1080/15475441.2020.1814780.

Garraffa, M., Coco, M. I., & Branigan, H. P. (2018). Impaired implicit learning of syntactic structure in children with developmental language disorder: Evidence from syntactic priming. *Autism & Developmental Language Impairments*, *3*, 2396941518779939. https://doi.org/10.1177/2396941518779939.

Hopkins, Z. L., & Branigan, H. P. (2020). Children show selectively increased language imitation after experiencing ostracism. Developmental Psychology, 56(5), 897–911. DOi: 10.1037/dev0000915.

Hopkins, Z., Yuill, N., & Keller, B. (2016). Children with autism align syntax in natural conversation. *Applied Psycholinguistics*, *37*(2), 347–370. DOi: 10.1017/S0142716414000599.

Hsu, D.-B. (2014a). Mandarin-speaking three-year-olds' demonstration of productive knowledge of syntax: Evidence from syntactic productivity and structural priming with the SVO-ba alternation. *Journal of Child Language*, *41*(5), 1115–1146. DOi: 10.1017/S0305000913000408.

Hsu, D.-B. (2014b). Structural priming as learning: Evidence from mandarin-learning 5-year-olds. *Language Acquisition: A Journal of Developmental Linguistics*, *21*(2), 156–172. DOi: 10.1080/10489223.2014.884571.

Hsu, D.-B. (2019). Children's adaption to input change using an abstract syntactic representation: Evidence from structural priming in Mandarin-speaking preschoolers. *Frontiers in Psychology*, *10*. DOi: 10.3389/fpsyg.2019.02186.

Hupp, J. M., & Jungers, M. K. (2009). Speech priming: An examination of rate and syntactic persistence in preschoolers. *British Journal of Developmental Psychology*, *27*(2), 495–504. DOi: 10.1348/026151008X345988.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language*, *50*(2), 182–195. DOi: 10.1016/j.jml.2003.09.003.

Kidd, E. (2012). Individual differences in syntactic priming in language acquisition. *Applied Psycholinguistics*, *33*(2), 393–418. DOi: 10.1017/S0142716411000415.

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*, 1–26. DOi: 10.1016/j.jml.2022.104374.

Manetti, C. (2013). On the production of passives in Italian: Evidence from an elicited production task and a syntactic priming study with preschool children. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th Boston University conference on language development online supplement* (pp. 1–16). Boston University.

Messenger, K. (2021). The persistence of priming: Exploring long-lasting syntactic priming effects in children and adults. *Cognitive Science*, *45*(6). DOi: 10.1111/cogs.13005.

Messenger, K., Branigan, H., McLean, J., & Sorace, A. (2008). English-speaking children's early passives: Evidence from syntactic priming. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings of the 32nd Annual Boston University Conference on Language Development, Vols 1 and 2* (pp. 275–286). Cascadilla Press.

Messenger, K., Branigan, H. P., & McLean, J. F. (2012). Is children's acquisition of the passive a staged process? Evidence from six- and nine-year-olds' production of passives. *Journal of Child Language*, *39*(5), 991–1016. DOi: 10.1017/S0305000911000377.

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*(4), 568–587. DOi: 10.1016/j.jml.2012.03.008.

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. DOi: 10.1016/j.jml.2014.12.002.

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. DOi: 10.1016/j.cognition.2012.06.008.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science*, *6*(5), 557–567. DOi: 10.1111/1467-7687.00312.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2006). Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds. *Language Learning and Development*, *2*(1), 27–49. DOi: 10.1207/s15473341lld0201_2.

Shimpi, P. M., Gámez, P. B., Huttenlocher, J., & Vasilyeva, M. (2007). Syntactic priming in 3- and 4-year-old children: Evidence for abstract representations of transitive and dative forms. *Developmental Psychology*, *43*(6), 1334–1346. DOi: 10.1037/0012-1649.43.6.1334.

Vasilyeva, M., & Gámez, P. B. (2015). Exploring interactions between semantic and syntactic processes: The role of animacy in syntactic priming. *Journal of Experimental Child Psychology*, *138*, 15–30. DOi: 10.1016/j.jecp.2015.04.009.

Vasilyeva, M., & Waterfall, H. (2012). Beyond syntactic priming: Evidence for activation of alternative syntactic structures. *Journal of Child Language*, *39*(2), 258–283. DOi: 10.1017/S0305000911000055.

Whitehurst, G. J., Ironsmith, M., & Goldfein, M. (1974). Selective imitation of the passive construction through modeling. *Journal of Experimental Child Psychology*, *17*(2), 288–302.

Wolleb, A., Sorace, A., & Westergaard, M. (2018). Exploring the role of cognitive control in syntactic processing: Evidence from cross-language priming in bilingual children. *Linguistic Approaches to Bilingualism*, *8*(5), 606–636. https://doi.org/10.1075/lab.17002.wol.

Appendix B

Calculating correlations

We provide the analysis code for calculating correlations on our OSF page. To estimate required correlations, we mostly used raw data from the *Canberra Longitudinal Child Language* study (Donnelly et al., in press; Kidd et al., 2018; Kumarage et al., 2022). This consisted of longitudinal priming data for the passive alternation with and without lexical overlap at 4 timepoints (36 m, 42 m, 48 m, and 54 m) and for the dative alternation with and without lexical overlap at 3 timepoints (42 m, 48 m, 54 m). Compared to other studies in the literature, the study has a large sample. Over 100 children participated in the study, although not all provide data at every time point in the priming tasks.

Correlation between primed and unprimed conditions for within-subjects studies

We calculated the proportion of dependent responses (either passive or double object dative) produced by each participant at each timepoint, for each structural alternation in the lexical overlap and abstract priming conditions. We then calculated the correlation between production in the primed and unprimed conditions for each condition at each timepoint (see Table B1). We used the average of these 14 correlation coefficients as the estimated correlation between dependent structure production in the primed and unprimed condition.

Table B1
Correlations between primed and unprimed conditions in the Canberra Longitudinal Child Language study.

Condition			Correlation
36 *m*	Passive	Lexical overlap	.40
		No lexical overlap	.17
42 *m*	Passive	Lexical overlap	.07
		No lexical overlap	.38
	Dative	Lexical overlap	.39
		No lexical overlap	.30
48 *m*	Passive	Lexical overlap	.24
		No lexical overlap	.45
	Dative	Lexical overlap	.53
		No lexical overlap	.61
54 *m*	Passive	Lexical overlap	.19
		No lexical overlap	.41
	Dative	Lexical overlap	.49
		No lexical overlap	.27
Average			.35

Correlation between lexical overlap conditions

We subtracted the proportion of dependent responses in the unprimed condition from the proportion in the primed condition to calculate a priming effect for each participant in each lexical overlap condition, for each structural alternation at each timepoint. We then calculated the correlation between priming with and without lexical overlap for each structural alternation at each timepoint (see Table B2). We used the average of these seven correlation coefficients as our estimate.

Table B2
Correlations between lexical overlap conditions in the Canberra Longitudinal Child Language study.

Condition		Correlation
36 *m*	Passive	.06
42 *m*	Passive	.13
	Dative	.21
48 *m*	Passive	.19
	Dative	.25
54 *m*	Passive	.36
	Dative	.17
Average		.20

Correlation between passive and dative priming

We again used the proportion of dependent responses in the unprimed condition subtracted from the proportion in the primed condition to calculate a priming effect for each participant in each condition. The correlation between this priming effect in the passive and dative alternation was calculated at both levels of lexical overlap for the three timepoints where both alternations were tested (see Table B3). We used the average of these six coefficients as our estimate.

Table B3
Correlations between passive and dative priming in the Canberra Longitudinal Child Language study.

Condition		Correlation
Lexical overlap	42 *m*	.24
	48 *m*	.17
	54 *m*	.24
No lexical overlap	42 *m*	.15
	48 *m*	.21
	54 *m*	−.14
Average		.15

Correlation between passive and dative priming under different lexical overlap conditions

Using the calculated priming effects for each participant, we calculated the correlation between priming at different levels of both structure and lexical overlap (see Table B4). We used the average of these six coefficients as our estimate.

Table B4
Correlations between passive and dative priming under different lexical overlap conditions in the Canberra Longitudinal Child Language study.

Condition		Correlation
Dative overlap*passive abstract	42 *m*	.11
	48 *m*	.32
	54 *m*	.01
Passive overlap*dative abstract	42 *m*	−.07
	48 *m*	.09
	54 *m*	−.10
Average		.06

Correlation between timepoints

The correlation estimate required by *vcalc()* in *metafor* is of the correlation between passive production at adjacent timepoints (e.g. baseline-test correlation; test-posttest correlation) rather than of the correlation between priming (e.g. correlation between baseline-test and baseline-posttest priming effects). Table B5 displays the appropriate correlations from two studies. Kidd (2012) provided correlations between passive production in different test blocks. Fazekas et al. (2020) made their raw data publicly accessible so we could calculate the proportion of double object datives produced by each participant in each test block and calculate correlations. We excluded the posttest 2 condition from Fazekas et al. (2020) because it is a lexical overlap condition unlike the other conditions in the study. We used the average of these five coefficients as our estimate.

Table B5
Correlations between production of the dependent structure in different test blocks.

Condition		Correlation
Kidd (2012)	Baseline-test	.07
	Test-posttest	.48
	Posttest-long posttest	.01
Fazekas et al. (2020)	Baseline-test	.54
	Test-posttest	.65
Average		.35

Appendix C

Analyses including and excluding other responses

Not all productions in syntactic priming experiments can be classified as the dependent or alternate structure in the alternation being studied. "Other" responses can make up a large proportion of responses (e.g., over 50 % of responses from 3 year olds in Shimpi et al., 2007) or a very small proportion in stem-completion studies (e.g. 0–5 % in Garcia & Kidd, 2020 and; Rowland et al., 2012). Researchers can choose to include *other* responses in the denominator when calculating the proportion of dependent responses or exclude *other* responses and restrict analyses to only structures within the alternation being studied (see Bencini & Valian, 2008 for discussion). A subset of 31 studies reported the frequency of *other* responses in each condition. We ran analyses for the 94 observations from these studies to compare the analysis choice of including vs. excluding *other* responses. The analysis script is available on our OSF page. Table C1 shows that including *other* responses numerically reduces the overall effect. In addition, including *other* responses results in less heterogeneity between effect sizes. Although the amount of sampling variance is similar, the amount attributed to within- and between-study heterogeneity is reduced.

Table C1
Results of overall multilevel meta-analytic models including and excluding other responses.

	N responses		N subjects	
	Other excluded	Other included	Other excluded	Other included
Estimate	1.42	1.22	1.27	1.07
	1.16 | 1.68	1.00 | 1.44	0.98 | 1.57	0.81 | 1.34
	p < .001***	p < .001***	p < .001***	p < .001***
Odds ratio	4.12	3.38	3.57	2.92
Cohen's d	0.78	0.67	0.70	0.59
Q	615.72	539.66	135.34	104.20
	df = 93	df = 93	df = 93	df = 93
	p < .001***	p < .001***	p < .01**	p = .201
Variance	0.673	0.524	0.893	0.776
Sampling	0.108	0.094	0.533	0.544
I^2	0.565	0.430	0.360	0.232

Table C2 shows results for the models using N responses to calculate variance and including moderators. We report results excluding the lexical overlap with lag condition from Fazekas et al. (2020) and the dative priming condition from Hopkins et al. (2016), both of which were identified as outliers with large effects on the results. Regardless of *other* response inclusion, the moderators accounted for a significant proportion of variance (*other* excluded: F(18,12) = 5.91, p < .01**; *other* included: F(18,12) = 5.99, p < .01***) but did not eliminate residual heterogeneity (*other* excluded: Q(73) = 153.26, p < .001***; *other* included: Q(73) = 155.33, p < .001***). Compared to the full sample of studies, additional moderators significantly predict the priming effect in the subset of studies that reported *other* responses. However, we focus on the difference between including and excluding *other* responses. The results are very similar but there is a trend towards numerically smaller model coefficients and less significant effects for moderators when including *other* responses. Therefore, there may be less power to detect moderators of priming when including *other* responses. This may be related to there being less total variance to explain when including *other* responses (Table 1).

Table C2
Results of N responses moderator models including and excluding other responses.

	Excluding other responses			Including other responses		
	β	CI	p	β	CI	p
Intercept	0.87	0.48 | 1.26	<.001***	0.78	0.46 | 1.10	<.001***
Between subjects	1.13	0.67 | 1.58	<.001***	0.85	0.47 | 1.24	<.001***
Baseline	−0.42	−0.94 | 0.10	.112	−0.23	−0.67 | 0.20	.287
Animacy						
Uncontrolled	−0.3	−0.89 | 0.29	.288	−0.22	−0.69 | 0.26	.343
Favourable	0.52	0.00 | 1.03	.049*	0.38	−0.05 | 0.81	.076^
Prime repetition	0.51	0.12 | 0.89	.011*	0.42	0.09 | 0.74	.013*
Lexical overlap	1.13	0.69 | 1.57	<.001***	0.97	0.56 | 1.39	<.001***
Structure						
Dative	−0.84	−1.23 | −0.45	<.001***	−0.77	−1.10 | −0.44	<.001***
SVO-ba	0.53	−0.24 | 1.31	.158	0.44	−0.17 | 1.05	.141

(*continued on next page*)

Table C2 (*continued*)

	Excluding other responses			Including other responses		
	β	CI	p	β	CI	p
RC	1.78	0.37 | 3.19	.018*	1.71	0.45 | 2.97	.012*
SV-Transitive	−0.86	−1.65 | −0.06	.037*	−0.77	−1.39 | −0.15	.019*
Age	0.03	−0.14 | 0.19	.744	0.07	−0.07 | 0.21	.336
Lag						
0 vs block	−0.15	−0.52 | 0.23	.443	−0.07	−0.40 | 0.25	.651
0 vs long	−0.42	−1.06 | 0.23	.2	−0.23	−0.80 | 0.35	.438
Prime repetition*Age	0.49	0.06 | 0.93	.026*	0.45	0.08 | 0.83	.018*
Lexical overlap*Age	0.14	−0.42 | 0.71	.62	0.2	−0.33 | 0.74	.452
Structure*Age						
Dative	−0.46	−0.91 | −0.01	.046*	−0.52	−0.92 | −0.11	.014*
SVO-ba	−0.16	−0.62 | 0.29	.478	−0.22	−0.63 | 0.20	.302
SV-Transitive	0.05	−0.34 | 0.43	.809	0.01	−0.36 | 0.38	.957

Table C3 shows results for the models using N subjects to calculate variance and including moderators. For the N subjects models, the same observation from Fazekas et al. (2020) was excluded but the Hopkins et al. (2016) observation no longer had large impacts on the results so was left in. In both models, the moderators accounted for a significant proportion of variance (*other* excluded: F(18,12) = 5.26, p < .01**; *other* included: F(18,12) = 4.15, p < .01**), with no significant residual heterogeneity (*other* excluded: Q(74) = 35.64, p = 1.000; *other* included: Q(74) = 33.13, p = 1.000). The results of these models should be interpreted with caution: although profile likelihood plots showed clear peaks in estimating variance components, both were estimated at zero. Nevertheless, for most effects, we also see a pattern of numerically smaller effects when including *other* responses.

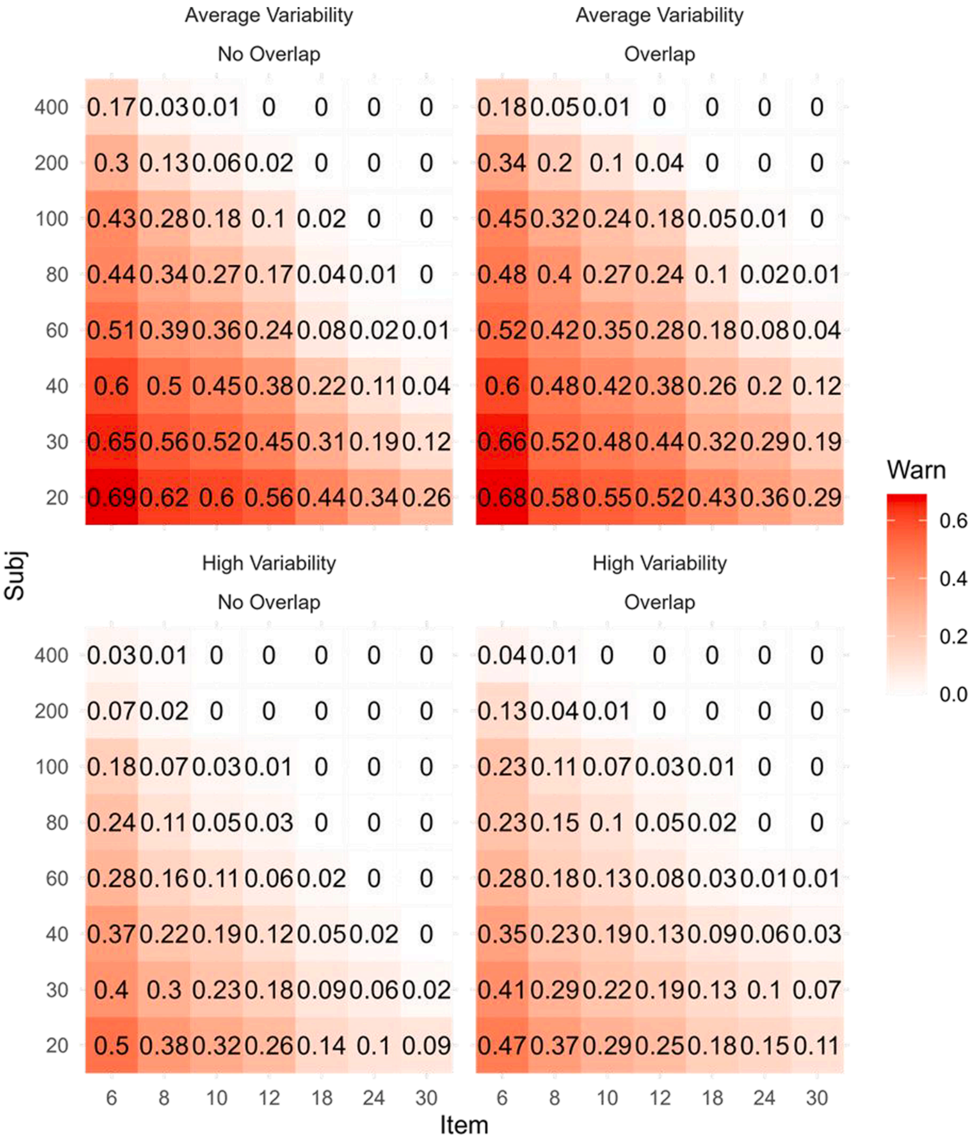
Table C3
Results of N subjects moderator models including and excluding other responses.

	Excluding other responses			Including other responses		
	β	CI	p	β	CI	p
Intercept	0.89	0.51 | 1.27	<.001***	0.72	0.34 | 1.10	.002**
Between subjects	1.15	0.45 | 1.85	.002**	1.04	0.33 | 1.75	.005**
Baseline	−0.22	−0.97 | 0.52	.552	−0.12	−0.89 | 0.64	.749
Animacy						
Uncontrolled	0.13	−0.70 | 0.96	.738	−0.27	−1.11 | 0.58	.507
Favourable	0.41	−0.21 | 1.02	.175	−0.04	−0.59 | 0.66	.898
Prime repetition	0.15	−0.38 | 0.69	.57	0.11	−0.43 | 0.65	.682
Lexical overlap	1.37	0.77 | 1.96	<.001***	1.26	0.67 | 1.86	<.001***
Structure						
Dative	−0.97	−1.44 | −0.50	<.001***	−1.07	−1.55 | −0.60	<.001***
SVO-ba	0.36	−0.43 | 1.14	.345	−0.07	−0.84 | 0.70	.844
RC	1.68	−0.77 | 4.12	.161	1.63	−0.75 | 4.01	.162
SV-Transitive	−0.89	−1.52 | −0.26	.009**	−0.74	−1.37 | −0.12	.024*
Age	0.07	−0.17 | 0.31	.545	0.24	−0.00 | 0.48	.053ˆ
Lag						
0 vs block	−0.23	−0.77 | 0.32	.412	−0.17	−0.72 | 0.38	.533
0 vs long	−0.36	−1.47 | 0.76	.525	−0.12	−1.21 | 0.97	.824
Prime repetition*Age	0.14	−0.43 | 0.71	.623	0.31	−0.27 | 0.88	.289
Lexical overlap*Age	0.12	−0.65 | 0.88	.764	0.16	−0.61 | 0.93	.682
Structure*Age						
Dative	−0.01	−0.47 | 0.45	.96	0.05	−0.41 | 0.52	.82
SVO-ba	−0.21	−0.91 | 0.49	.549	−0.29	−0.96 | 0.37	.384
SV-Transitive	−0.04	−0.55 | 0.47	.88	−0.18	−0.69 | 0.33	.479

Appendix D

Additional power analysis figures

Non-convergent models were included in power analyses because the full random effects structure was known a priori. Fig. D1 reports the proportion of non-convergent models in power analyses for the main priming effects. Fig. D2 reports the same proportion for power analyses of interaction effects. However, models with implausible estimates were excluded from power analyses because they produce significant results which may inflate estimates of power. Figs. D3 and D4 report the proportion of models excluded for this reason for main and interaction effects respectively.



Average Variability

No Overlap

Subj	6	8	10	12	18	24	30
400	0.17	0.03	0.01	0	0	0	0
200	0.3	0.13	0.06	0.02	0	0	0
100	0.43	0.28	0.18	0.1	0.02	0	0
80	0.44	0.34	0.27	0.17	0.04	0.01	0
60	0.51	0.39	0.36	0.24	0.08	0.02	0.01
40	0.6	0.5	0.45	0.38	0.22	0.11	0.04
30	0.65	0.56	0.52	0.45	0.31	0.19	0.12
20	0.69	0.62	0.6	0.56	0.44	0.34	0.26

Average Variability

Overlap

Subj	6	8	10	12	18	24	30
400	0.18	0.05	0.01	0	0	0	0
200	0.34	0.2	0.1	0.04	0	0	0
100	0.45	0.32	0.24	0.18	0.05	0.01	0
80	0.48	0.4	0.27	0.24	0.1	0.02	0.01
60	0.52	0.42	0.35	0.28	0.18	0.08	0.04
40	0.6	0.48	0.42	0.38	0.26	0.2	0.12
30	0.66	0.52	0.48	0.44	0.32	0.29	0.19
20	0.68	0.58	0.55	0.52	0.43	0.36	0.29

High Variability

No Overlap

Subj	6	8	10	12	18	24	30
400	0.03	0.01	0	0	0	0	0
200	0.07	0.02	0	0	0	0	0
100	0.18	0.07	0.03	0.01	0	0	0
80	0.24	0.11	0.05	0.03	0	0	0
60	0.28	0.16	0.11	0.06	0.02	0	0
40	0.37	0.22	0.19	0.12	0.05	0.02	0
30	0.4	0.3	0.23	0.18	0.09	0.06	0.02
20	0.5	0.38	0.32	0.26	0.14	0.1	0.09

High Variability

Overlap

Subj	6	8	10	12	18	24	30
400	0.04	0.01	0	0	0	0	0
200	0.13	0.04	0.01	0	0	0	0
100	0.23	0.11	0.07	0.03	0.01	0	0
80	0.23	0.15	0.1	0.05	0.02	0	0
60	0.28	0.18	0.13	0.08	0.03	0.01	0.01
40	0.35	0.23	0.19	0.13	0.09	0.06	0.03
30	0.41	0.29	0.22	0.19	0.13	0.1	0.07
20	0.47	0.37	0.29	0.25	0.18	0.15	0.11

Fig. D1. Proportions of models producing warning messages for power analyses of priming effect.

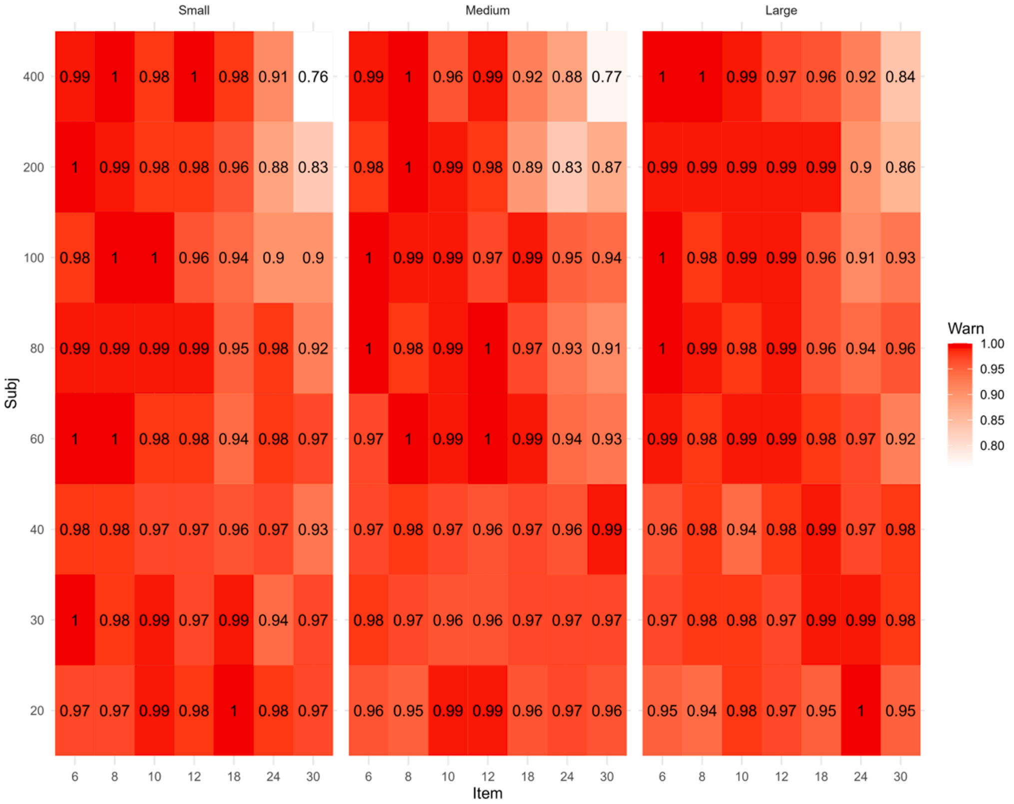
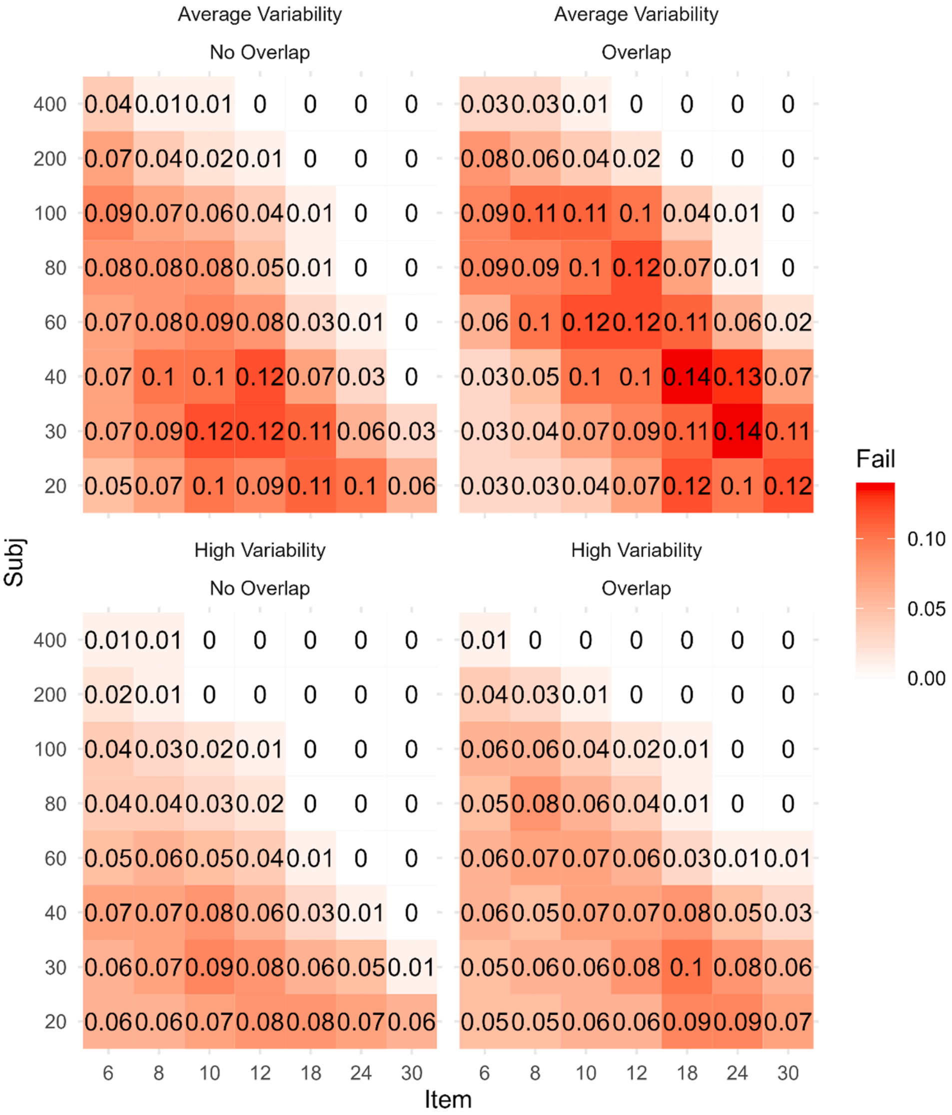


Fig. D2. Proportion of models producing warning messages in power analyses of interactions.



Average Variability
No Overlap

Subj	6	8	10	12	18	24	30
400	0.04	0.01	0.01	0	0	0	0
200	0.07	0.04	0.02	0.01	0	0	0
100	0.09	0.07	0.06	0.04	0.01	0	0
80	0.08	0.08	0.08	0.05	0.01	0	0
60	0.07	0.08	0.09	0.08	0.03	0.01	0
40	0.07	0.1	0.1	0.12	0.07	0.03	0
30	0.07	0.09	0.12	0.12	0.11	0.06	0.03
20	0.05	0.07	0.1	0.09	0.11	0.1	0.06

Average Variability
Overlap

Subj	6	8	10	12	18	24	30
400	0.03	0.03	0.01	0	0	0	0
200	0.08	0.06	0.04	0.02	0	0	0
100	0.09	0.11	0.11	0.1	0.04	0.01	0
80	0.09	0.09	0.1	0.12	0.07	0.01	0
60	0.06	0.1	0.12	0.12	0.11	0.06	0.02
40	0.03	0.05	0.1	0.1	0.14	0.13	0.07
30	0.03	0.04	0.07	0.09	0.11	0.14	0.11
20	0.03	0.03	0.04	0.07	0.12	0.1	0.12

High Variability
No Overlap

Subj	6	8	10	12	18	24	30
400	0.01	0.01	0	0	0	0	0
200	0.02	0.01	0	0	0	0	0
100	0.04	0.03	0.02	0.01	0	0	0
80	0.04	0.04	0.03	0.02	0	0	0
60	0.05	0.06	0.05	0.04	0.01	0	0
40	0.07	0.07	0.08	0.06	0.03	0.01	0
30	0.06	0.07	0.09	0.08	0.06	0.05	0.01
20	0.06	0.06	0.07	0.08	0.08	0.07	0.06

High Variability
Overlap

Subj	6	8	10	12	18	24	30
400	0.01	0	0	0	0	0	0
200	0.04	0.03	0.01	0	0	0	0
100	0.06	0.06	0.04	0.02	0.01	0	0
80	0.05	0.08	0.06	0.04	0.01	0	0
60	0.06	0.07	0.07	0.06	0.03	0.01	0.01
40	0.06	0.05	0.07	0.07	0.08	0.05	0.03
30	0.05	0.06	0.06	0.08	0.1	0.08	0.06
20	0.05	0.05	0.06	0.06	0.09	0.09	0.07

Item

Fig. D3. Proportions of models producing implausibly small standard errors (<.1) or large effect sizes (>5 or < -5) for power analyses of priming effect.

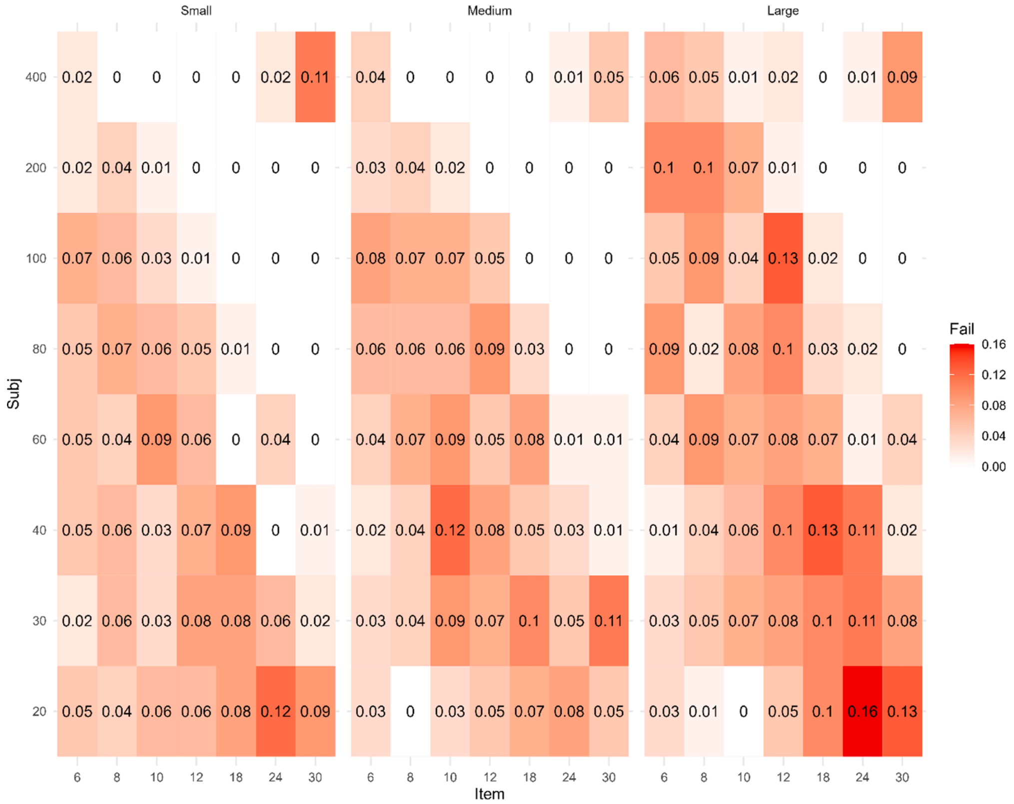


Fig. D4. Proportion of models producing implausibly small standard errors or implausibly large effect sizes for power analyses of interactions.

References

Allen, M. L., Haywood, S., Rajendran, G., & Branigan, H. (2011). Evidence for syntactic alignment in children with autism. *Developmental Science, 14*(3), 540–548. https://doi.org/10.1111/j.1467-7687.2010.01001.x

Ambridge, B., & Lieven, E. (2015). A Constructivist Account of Child Language Acquisition. In *The Handbook of Language Emergence* (pp. 478–510). John Wiley & Sons. https://doi.org/10.1002/9781118346136.ch22.

Anderson, J. D., & Conture, E. G. (2004). Sentence-Structure Priming in Young Children Who Do and Do Not Stutter. *Journal of Speech, Language, and Hearing Research, 47*(3), 552–571. https://doi.org/10.1044/1092-4388 %282004/043 %29

Atkinson, E. (2022). Sticking to what we know: Methodological limitations to generalizability. In K. Messenger (Ed.), *Syntactic priming in language acquisition: Representations, mechanisms and applications* (Vol. 31, pp. 83–106). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.31.05atk.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoš, F., Maier, M., Shanks, D. R., Stanley, T. D., Sladekova, M., & Wagenmakers, E.-J. (2023). Meta-analyses in psychology often overestimate evidence for and size of effects. *Royal Society Open Science, 10*(7), Article 230224. https://doi.org/10.1098/rsos.230224

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models. arXiv.*. https://doi.org/10.48550/arXiv.1506.04967

Becker, M. P., & Balagtas, C. C. (1993). Marginal Modeling of Binary Cross-Over Data. *Biometrics, 49*(4), 997. https://doi.org/10.2307/2532242

Bencini, G. M. L., & Valian, V. V. (2008). Abstract sentence representations in 3-year-olds: Evidence from language production and comprehension. *Journal of Memory and Language, 59*(1), 97–113. https://doi.org/10.1016/j.jml.2007.12.007

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting Replicability in Developmental Research Through Meta-analyses: Insights From Language Acquisition Research. *Child Development, 89*(6), 1996–2009. https://doi.org/10.1111/cdev.13079

Bidgood, A., Pine, J. M., Rowland, C. F., & Ambridge, B. (2020). Syntactic representations are both abstract and semantically constrained: Evidence from children's and adults' comprehension and production/priming of the English passive. *Cognitive Science, 44*(9). https://doi.org/10.1111/cogs.12892

Bidgood, A., Pine, J., Rowland, C., Sala, G., Freudenthal, D., & Ambridge, B. (2021). Verb argument structure overgeneralisations for the English intransitive and transitive constructions: Grammaticality judgments and production priming. *Language and Cognition: An Interdisciplinary Journal of Language and Cognitive Science, 13*(3), 397–437. https://doi.org/10.1017/langcog.2021.8

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

Bock, K., & Griffin, Z. (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//0096-3445.129.2.177

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

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult-child dialogue. *Journal of Memory and Language, 91*, 141–157. https://doi.org/10.1016/j.jml.2016.02.002

Branigan, H. P., McLean, J. F., & Jones, M. (2005). A blue cat or a cat that is blue? - Evidence for abstract syntax in young children's noun phrases. In A. Brugos, M. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 109–121). Cascadilla Press.

Branigan, H. P., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children's sentence production: Evidence for error-based implicit learning. *Cognition, 157*, 250–256. https://doi.org/10.1016/j.cognition.2016.09.004

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences, 40*. https://doi.org/10.1017/S0140525X16002028

Buckle, L., Lieven, E., & Theakston, A. L. (2017). The effects of animacy and syntax on priming: A developmental study. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02246

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied Psycholinguistics, 22*(2), 253–267. https://doi.org/10.1017/S0142716401002065

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

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research, 29*(2), 217–230. https://doi.org/10.1023/A:1005101313330

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass, 6*(5), 259–278. https://doi.org/10.1002/lnc3.337

Corporation for Digital Scholarship. (2023). *Zotero* (6.0.26) [Computer software]. https://www.zotero.org.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science, 26*(1), 99–107. https://doi.org/10.1177/0956797614557866

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). https://doi.org/10.1098/rstb.2012.0394

Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In I. H. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 9–33). Wiley Online Library. https://doi.org/10.1002/0470870168.

Donnelly, S., Rowland, C., Chang, F., & Kidd, E. (in press). A comprehensive examination of prediction-based error as a mechanism for syntactic development: Evidence from syntactic priming. *Cognitive Science*. DOi: 10.1111/cogs.13431.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

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, 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

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition, 82*(3), 259–278. https://doi.org/10.1016/S0010-0277(01)00159-7

Foltz, A., Thiele, K., Kahsnitz, D., & Stenneken, P. (2015). Children's syntactic-priming magnitude: Lexical factors and participant characteristics. *Journal of Child Language, 42*(4), 932–945. https://doi.org/10.1017/S0305000914000488

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Frank, M. C., Alcock, K. J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., Bergelson, E., Bergmann, C., Black, A. K., Blything, R., Böhland, M. P., Bolitho, P., Borovsky, A., Brady, S. M., Braun, B., Brown, A., Byers-Heinlein, K., Campbell, L. E., Cashon, C., & Soderstrom, M. (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Advances in Methods and Practices in Psychological Science, 3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Gámez, P. B., & Shimpi, P. M. (2016). Structural priming in Spanish as evidence of implicit learning. *Journal of Child Language, 43*(1), 207–233. https://doi.org/10.1017/S0305000915000161

Gámez, P. B., Shimpi, P. M., Waterfall, H. R., & Huttenlocher, J. (2009). Priming a perspective in Spanish monolingual children: The use of syntactic alternatives. *Journal of Child Language, 36*(2), 269–290. https://doi.org/10.1017/S0305000908008945

Gámez, P. B., & Vasilyeva, M. (2015). Increasing Second Language Learners' Production and Comprehension of Developmentally Advanced Syntactic Forms. *Language Learning and Development, 11*(2), 128–151. https://doi.org/10.1080/15475441.2014.894873

Gámez, P. B., & Vasilyeva, M. (2020). Shared syntactic representations in balanced bilinguals: Cross-linguistic priming with and without verb overlap. *Language Learning and Development, 16*(1), 89–106. https://doi.org/10.1080/15475441.2019.1689822

Garcia, R., & Kidd, E. (2020). The acquisition of the Tagalog symmetrical voice system: Evidence from structural priming. *Language Learning and Development, 16*(4), 399–425. https://doi.org/10.1080/15475441.2020.1814780

Garcia, R., Roeser, J., & Höhle, B. (2020). Children's online use of word order and morphosyntactic markers in Tagalog thematic role assignment: An eye-tracking study. *Journal of Child Language, 47*(3), 533–555. https://doi.org/10.1017/S0305000919000618

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

Garraffa, M., Coco, M. I., & Branigan, H. P. (2015). Effects of immediate and cumulative syntactic experience in language impairment: Evidence from priming of subject relatives in children with SLI. *Language Learning and Development, 11*(1), 18–40. https://doi.org/10.1080/15475441.2013.876277

Garraffa, M., Coco, M. I., & Branigan, H. P. (2018). Impaired implicit learning of syntactic structure in children with developmental language disorder: Evidence from syntactic priming. *Autism & Developmental Language Impairments, 3*. https://doi.org/10.1177/2396941518779939, 2396941518779939.

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science, 9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

Gibson, E. (2000). The Dependency Locality Theory. In A. P. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium* (pp. 95–126). MIT Press. https://doi.org/10.7551/mitpress/3654.003.0008.

Goodman, S. N. (1991). Have you ever meta-analysis you didn't like? *Annals of Internal Medicine, 114*(3), 244–246. https://doi.org/10.7326/0003-4819-114-3-244

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The Learnability and Acquisition of the Dative Alternation in English. *Language, 65*(2), 203–257. https://doi.org/10.2307/415332

Hartsuiker, R. J., & Bernolet, S. (2017). The development of shared syntax in second language learning. *Bilingualism: Language and Cognition, 20*(2), 219–234. https://doi.org/10.1017/S1366728915000164

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language, 58*(2), 214–238. https://doi.org/10.1016/j.jml.2007.07.003

Hopkins, Z., Yuill, N., & Keller, B. (2016). Children with autism align syntax in natural conversation. *Applied Psycholinguistics, 37*(2), 347–370. https://doi.org/10.1017/S0142716414000599

Hsu, D.-B. (2014). Structural priming as learning: Evidence from mandarin-learning 5-year-olds. *Language Acquisition: A Journal of Developmental Linguistics, 21*(2), 156–172. https://doi.org/10.1080/10489223.2014.884571

Hsu, D.-B. (2019). Children's adaption to input change using an abstract syntactic representation: Evidence from structural priming in Mandarin-speaking preschoolers. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02186

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*(2), 182–195. https://doi.org/10.1016/j.jml.2003.09.003

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

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

Kadlec, D., Sainani, K. L., & Nimphius, S. (2022). With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. *Sports Medicine, 53*(2), 313–325. https://doi.org/10.1007/s40279-022-01766-0

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

Kidd, E. (2012). Individual differences in syntactic priming in language acquisition. *Applied Psycholinguistics, 33*(2), 393–418. https://doi.org/10.1017/S0142716411000415

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes, 22*(6), 860–897. https://doi.org/10.1080/01690960601155284

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language, 42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Kidd, E., Junge, C., Spokes, T., Morrison, L., & Cutler, A. (2018). Individual Differences in Infant Speech Segmentation: Achieving the Lexical Shift. *Infancy, 23*(6), 770–794. https://doi.org/10.1111/infa.12256

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*, 1–26. https://doi.org/10.1016/j.jml.2022.104374

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? *A meta-analytic review. Psychological Bulletin, 144*(4), 394–425. https://doi.org/10.1037/bul0000142

Leonard, L. B., Miller, C. A., Grela, B., Holland, A. L., Gerber, E., & Petucci, M. (2000). Production operations contribute to the grammatical morpheme limitations of children with specific language impairment. *Special Issue: Disorders of Language and Memory: Implications for Cognitive Theory., 43*(2), 362–378. https://doi.org/10.1006/jmla.1999.2689

Lester, N. A., Moran, S., Küntay, A. C., Allen, S. E. M., Pfeiler, B., & Stoll, S. (2022). Detecting structured repetition in child-surrounding speech: Evidence from maximally diverse languages. *Cognition, 221*, Article 104986. https://doi.org/10.1016/j.cognition.2021.104986

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

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The Influence of Animacy on Relative Clause Processing. *Journal of Memory and Language, 47*(1), 50–68. https://doi.org/10.1006/jmla.2001.2837

Marchman, V. A., Bates, E., Burkardt, A., & Good, A. B. (1991). Functional constraints of the acquisition of the passive: Toward a model of the competence to perform. *First Language, 11*(31), 65–92. https://doi.org/10.1177/014272379101103104

Messenger, K. (2021). The persistence of priming: Exploring long-lasting syntactic priming effects in children and adults. *Cognitive Science, 45*(6). https://doi.org/10.1111/cogs.13005

Messenger, K. (Ed.). (2022)*, Vol. 31. Syntactic priming in language acquisition: Representations, mechanisms and applications*. John Benjamins Publishing Company. https://doi.org/10.1075/tilar.31.01mes.

Messenger, K., Branigan, H. P., & McLean, J. F. (2012). Is children's acquisition of the passive a staged process? Evidence from six- and nine-year-olds' production of passives. *Journal of Child Language, 39*(5), 991–1016. https://doi.org/10.1017/S0305000911000377

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*(4), 568–587. https://doi.org/10.1016/j.jml.2012.03.008

Messenger, K., & Fisher, C. (2018). Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax. *Cognition, 178*, 118–132. https://doi.org/10.1016/j.cognition.2018.05.002

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7* (1), 105–125. https://doi.org/10.1037/1082-989X.7.1.105

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

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of Two Methods to Detect Publication Bias in Meta-analysis. *Journal of the American Medical Association, 295*(6), 676–680. https://doi.org/10.1001/jama.295.6.676

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

Pickering, M. J., & Ferreira, V. S. (2008). Structural Priming: A Critical Review. *Psychological Bulletin, 134*(3), 427–459. https://doi.org/10.1037/0033-2909.134.3.427

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rohatgi, A. (2022). *Webplotdigitizer: Version 4.6*. https://automeris.io/WebPlotDigitizer.

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

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science, 6*(5), 557–567. https://doi.org/10.1111/1467-7687.00312

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2006). Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds. *Language Learning and Development, 2*(1), 27–49. https://doi.org/10.1207/s15473341lld0201_2

Shimpi, P. M., Gámez, P. B., Huttenlocher, J., & Vasilyeva, M. (2007). Syntactic priming in 3- and 4-year-old children: Evidence for abstract representations of transitive and dative forms. *Developmental Psychology, 43*(6), 1334–1346. https://doi.org/10.1037/0012-1649.43.6.1334

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Stedman, F. M. R., Curtin, F., Elbourne, D. R., Kesselheim, A. S., & Brookhart, M. A. (2011). Meta-analyses involving cross-over trials: Methodological issues. *International Journal of Epidemiology, 40*(6), 1732–1734. https://doi.org/10.1093/ije/dyp345

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal, 343*, Article d4002. https://doi.org/10.1136/bmj.d4002

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. https://doi.org/10.1016/j.jml.2007.06.012

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74* (3), 209–253. https://doi.org/10.1016/S0010-0277(99)00069-4

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M., & Abbot-Smith, K. (2002). A tale of two theories: Response to Fisher. *Cognition, 83*(2), 207–214. https://doi.org/10.1016/S0010-0277(01)00172-X

van Beijsterveldt, L. M., & van Hell, J. G. (2009). Structural priming of adjective-noun structures in hearing and deaf children. *Journal of Experimental Child Psychology, 104* (2), 179–196. https://doi.org/10.1016/j.jecp.2009.05.002

Van Dijk, C., Van Wonderen, E., Koutamanis, E., Kootstra, G. J., Dijkstra, T., & Unsworth, S. (2022). Cross-linguistic influence in simultaneous and early sequential bilingual children: A meta-analysis. *Journal of Child Language, 49*(5), 897–929. https://doi.org/10.1017/S0305000921000337

van Gompel, R. P. G., & Arai, M. (2018). Structural priming in bilinguals. *Bilingualism: Language and Cognition, 21*(3), 448–455. https://doi.org/10.1017/S1366728917000542

Vasilyeva, M., & Gámez, P. B. (2015). Exploring interactions between semantic and syntactic processes: The role of animacy in syntactic priming. *Journal of Experimental Child Psychology, 138*, 15–30. https://doi.org/10.1016/j.jecp.2015.04.009

Vasilyeva, M., Huttenlocher, J., & Waterfall, H. (2006). Effects of language intervention on syntactic skill levels in preschoolers. *Developmental Psychology, 42*(1), 164–174. https://doi.org/10.1037/0012-1649.42.1.164

Vasilyeva, M., & Waterfall, H. (2012). Beyond syntactic priming: Evidence for activation of alternative syntactic structures. *Journal of Child Language, 39*(2), 258–283. https://doi.org/10.1017/S0305000911000055

Vasilyeva, M., Waterfall, H., Gámez, P. B., Gomez, L. E., Bowers, E., & Shimpi, P. (2010). Cross-linguistic syntactic priming in bilingual children. *Journal of Child Language, 37* (5), 1047–1064. https://doi.org/10.1017/S0305000909990213

Viechtbauer, W. (2008). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift Für Psychologie/Journal of Psychology, 215*(2), 104. https://doi.org/10.1027/0044-3409.215.2.104

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wolleb, A., Sorace, A., & Westergaard, M. (2018). Exploring the role of cognitive control in syntactic processing: Evidence from cross-language priming in bilingual children. *Linguistic Approaches to Bilingualism, 8*(5), 606–636. https://doi.org/10.1075/lab.17002.wol