Replicating a fundamental finding in psycholinguistics: syntactic priming

A Preprint

Kyle Mahowald
University of California, Santa Barbara

Ariel James
Macalaster College

Richard Futrell
University of California, Irvine

Melodie Yen
University of California, Los Angeles

Idan Blank
University of California, Los Angeles

January 11, 2021

Abstract

We attempt to replicate a fundamental finding in psycholinguistics: syntactic priming in language production, using the written sentence completion paradigm. The basic claim of syntactic priming is that producing or processing a particular syntactic structure X (e.g., the double object dative: "Lu baked me a cake.") makes a language user more likely to then use that same structure X than some alternative structure Y (e.g., prepositional phrase dative: "Lu baked a cake for me."). It has been claimed that syntactic priming provides evidence for abstract syntactic representation, and there have been hundreds of studies exploring syntactic priming in a variety of modalities, languages, paradigms, and social settings. Here, we attempt to replicate two syntactic priming studies: a classic priming construction (dative priming; Pickering and Branigan, 1998) with and without a repeated lexical element between the prime and target (to replicate the so-called "lexical boost effect") and a less widely studied construction in the priming literature on high vs. low attachment priming in relative clauses (Scheepers, 2003; Desmet and Declercq, 2006; Loncke et al., 2011). We will place these results in the context of the larger priming literature and consider these results in light of a recent syntactic priming meta-analysis (Mahowald et al., 2016), as well as comment on the relationship between meta-analysis and replication more broadly.

1 Introduction

We conduct two replication attempts of a fundamental finding in psycholinguistics: syntactic priming in language production. Syntactic priming is a phenomenon whereby producing or processing a particular syntactic structure X (e.g., double object dative: "Lu baked me a cake.") increases a speaker's likelihood of then producing that same structure X compared to some alternative structure Y (e.g., the prepositional object dative: "Lu baked a cake for me.) when describing a new event.[1]

In the traditional set-up (Bock, 1986), a participant is primed with a **prime sentence** and then asked to produce a **target sentence**. The method for eliciting the target sentence varies: the two most popular methods are picture description and written sentence completion. While neither has been the focus of large-scale pre-registered replications, in this study, we focus on replicating syntactic priming using the written sentence completion paradigm.

In the written sentence completion paradigm, a participant is first asked to complete a prime sentence stem that is designed to elicit a particular structure X in one priming condition, or a different stem designed to elicit an alternative structure Y in another priming condition. For instance, the two structures might constitute the dative alternation: the double object (e.g., "The student gave the professor an apple") vs. the prepositional phrase (e.g., "The student gave an apple to the professor"). Thus, in one condition, the prime could be "The student gave the professor ...", which

[1] For helpful discussions, we thank Ted Gibson and Shravan Vasishth.

participants are likely to complete with a nominal (e.g., “an apple.”); in the other condition, the prime could instead be “The student gave the apple ...”, which participants are likely to complete with a prepositional phrase (e.g., “to the professor.”). Next, following this priming trial, participants are asked to complete a target sentence stem in which they can plausibly use either structure X or Y (e.g., “The farmer gave....”). The prediction is that, depending on the prime condition, participants will use X more often after an X prime and Y more often after a Y prime.

The idea that syntactic representations can be primed has had major consequences for psycholinguistic theory. Broadly, syntactic priming has been used as evidence for the abstract nature of syntactic representations: when processing a sentence, the representation of sentence structure being constructed is not inseparable from the specific event being described (i.e., the particular meaning conveyed); rather, this representation can be shared (i.e., generalize) across different sentences as long as their particular events share some broad properties. This claim for abstract representation has been particularly prominent in strands of research that explore how priming can occur between different constructions with a common underlying structure (Bock and Loebell, 1990; Loebell and Bock, 2003) or across languages (Hartsuiker et al., 2004). Nonetheless, others have claimed that syntactic priming is not necessarily evidence for syntactic abstraction and may instead reflect priming of specific words (even, e.g., a shared function word across sentences), specific thematic structures, information structure, or discourse functions (Ziegler et al., 2019; Bidgood et al., 2020).

Syntactic priming has become as much a method in psycholinguistics as it is a finding. In fact, Branigan and Pickering (2017) argued that it could become linguists’ main experimental paradigm for understanding syntactic structure, replacing acceptability judgments. Indeed, syntactic priming studies typically no longer ask “does priming exist?” They instead ask questions such as: how does syntactic priming interact with lexical material? (Pickering and Branigan, 1998; Melinger and Cleland, 2011), how does syntactic priming decay over time? (Kaschak, 2007; Kaschak et al., 2011b; Hartsuiker et al., 2008), how is syntactic priming modulated by conceptual content? (Bunger et al., 2013), how is syntactic priming affected by various linguistic and non-linguistic disorders? (Ferreira et al., 2008; Haarmann and Kolk, 1991; Slocombe et al., 2013; Verreyt et al., 2013), how does priming work across languages? (Kim and McDonough, 2007; Bernolet et al., 2013), what is the developmental trajectory of syntactic priming children? (Bidgood et al., 2020; Messenger et al., 2011, 2012), and even how does priming interact with female fertility? (Coyle and Kaschak, 2012). This list of priming is not exhaustive; for reviews, see Pickering and Ferreira (2008); Mahowald et al. (2016).

The widespread use of syntactic priming in psycholinguistics suggests strong confidence in the basic paradigm. But, across psychology, priming studies of processes other than language production have come under intense scrutiny (see the 2014 special issue of *Perspectives on Psychological Science*, which focused on behavioral priming: Cesario, 2014). The failure to replicate priming studies has emerged in various social and cognitive experimental settings (Pashler et al., 2012; Harris et al., 2013; Doyen et al., 2012), as well as within psycholinguistics (Brand et al., 2003; Peter et al., 1990; Martin and Jensen, 1988; Steele, 2014). These failures are part of a larger reckoning with replicability issues in psychology, arising from a combination of factors such as publication bias (Ioannidis et al., 2014), low statistical power (Button et al., 2013; Gelman and Carlin, 2016; Simonsohn et al., 2014), and researcher degrees of freedom (Gelman and Loken, 2013; Simmons et al., 2011). (Open Science Collaboration, 2015). Ramscar et al. (2015) argue, on theoretical grounds, that priming effects across psychology are unlikely to replicate.

Given these concerns, the robustness of syntactic priming in particular ought to be evaluated. Thus, a meta-analysis of syntactic priming in production examined over 70 papers published between 1986 and 2015 and established effect sizes and confidence bounds on various versions of such priming (Mahowald et al., 2016). Whereas it found that basic syntactic priming is a robust phenomenon, it additionally found that certain constructions and tasks showed bigger priming effects than others. Moreover, many priming studies are likely underpowered when they are not examining priming as a main effect but are instead studying some interaction (e.g., how the size of the priming effect is moderated in the presence of some other variable like speaker bilingualism, time lag between prime and target, etc.). An exception is the moderating effect of repeated content words across prime and target, which increases syntactic priming; this "lexical boost" effect is large and robust.

The extensive body of published work on syntactic priming in production, and the consistency of the methods used across labs, made it an ideal candidate for a meta-analysis. Now, to go beyond what can be learned from meta-analyses, the next critical step for assessing the state of the literature is a large-scale, pre-registered replication. Such replications can indeed provide novel information: Kvarven et al. (2020) explicitly compare replications and meta-analyses and find that 12 out of 15 of their examples show significant differences between meta-analytic effect sizes and replication-derived effect sizes.

Hence, a well-powered, pre-registered replication of syntactic priming, using modern statistical analysis methods, is called for. The current work adopts this approach, and will attempt to replicate two different kinds of effects from the syntactic priming literature:

Exp. 1. Each subject first completes a prime sentence from:	*Then a target sentence:*	*Predicted completions:*
DO same verb: The driving instructor gave the learner [the permit.]	**The consultant gave …**	…the business a lesson.
DO diff. verb: The driving instructor handed the learner [the permit.]		…the business a lesson.
PO same verb: The driving instructor gave the certificate [to the learner.]		…a lesson to the business.
PO diff. verb: The driving instructor handed the certificate [to the learner.]		…a lesson to the business.

Exp. 2. Each subject first completes a prime sentence from:	*Then a target sentence:*	*Predicted completions:*
HA: Amara observed the rehearsal of the actors which was [on Tuesday].	**They congratulated the winners of the tournament that …**	…took place last week.
LA: Amara observed the rehearsal of the actors who were [trained at Yale].		… had been playing tennis for years.

Figure 1: Example prime and target sentences from each of the two experiments.

- A replication of syntactic priming using one of the most common alternations (the dative), as well as its interaction with lexical overlap between prime and target (the "lexical boost"): In order to replicate Experiments 1 and 2 from Pickering and Branigan (1998), we will use a dative sentence completion task in a 2×2 design, manipulating prime (prepositional phrase dative vs. direct object dative) and lexical boost (i.e., whether the same verb vs. different verbs appear in the prime and the target). Among the most influential syntactic priming studies (cited over 1000 times), Pickering and Branigan (1998) was one of the key pieces of evidence for developing the idea of the lexical boost.
- A replication of a less-studied priming construction (relative clause (RC) attachment height): We will use a 2-condition design (high vs. low attachment) to manipulate RC attachment height, conceptually replicating Scheepers (2003). Here, the two syntactic structures both have a complex noun phrase such as “the scientist with the telescope”, which includes two nouns, one of which (*scientist*) is at a higher position in the syntactic tree compared to the other (*telescope*). The two structures differ in whether a relative clause modifies (attaches to) the higher noun (e.g., “I saw the scientist with the telescope that was teaching a class.”) or the lower noun (e.g., “I saw the scientist with the telescope that had a big lens.”). Scheepers (2003) found that, whereas high attachment was overall dis-preferred, its production likelihood could be increased via priming. We will conduct this experiment in English, unlike the original study, which was in German. Scheepers (2003) is typically taken as strong evidence in favor of abstract syntactic priming, and the phenomenon of priming high vs. low attachment has even been used to claim the existence of cross-domain priming between language and arithmetic (Scheepers et al., 2011).

Together, these replications will explore a varied subspace of the priming literature, and help shed light on the robustness and nature of a major psycholinguistic phenomenon. By framing these results and quantifying them in terms of the existing Mahowald et al. (2016) meta-analysis, we also hope to shed light on the relationship between meta-analysis and replication.

As we show below in our sample size calculations, we expect this replication attempt to substantially improve our best estimates of the effect size and reliability of syntactic priming. It is, to our knowledge, the first high-powered, pre-registered replication of syntactic priming.

All results should be interpreted with care. If we observe a statistically robust effect of syntactic priming, that would be evidence that this paradigm produces such a result even when well powered and pre-registered. But it would not mean that all theoretical implications that have been claimed for syntactic priming are verified. At the same time, if we fail to replicate the phenomenon, that would significantly muddy the empirical landscape of syntactic priming studies. In that case, we would encourage further work to clarify that landscape, and we would urge researchers not to unquestioningly accept that the paradigm of written sentence completion produces robust priming results. But it would not constitute a definitive proof that syntactic priming does not exist. Rather, it might be the case that priming is a robust phenomenon in other paradigms and/or with other constructions.

2 Experiment 1

This experiment is a replication of Experiments 1 and 2 from Pickering and Branigan (1998, hereafter PB), which used a written sentence completion task to investigate dative priming. PB’s Experiment 1 used a 2×2 design. The first factor was prime condition—double object construction (DO) or prepositional object construction (PO). For DO primes, such as “The racing driver showed the helpful mechanic...” in the sentences below, a sentence is likely to be completed with a double object construction (e.g., “her malfunctioning car.”). For PO primes, such as “The racing driver showed the torn overall...” in the sentences below, the likely completion is a prepositional phrase (e.g., “to her friend.”). A target sentence stem (e.g., “The patient showed...”) can then plausibly be completed using either construction. The second factor was lexical overlap—whether the verb in the target sentence stem is the same as, or different from, the verb in the prime.

PB’s sample item for Experiment 1:

Prime Type Type	Primes	Target
PO Prime; Same Verb:	The racing driver showed the torn overall…	
DO Prime; Same Verb:	The racing driver showed the helpful mechanic…	The patient showed…
PO Prime; Different Verb:	The racing driver gave the torn overall…	
DO Prime; Different Verb:	The racing driver gave the helpful mechanic…	

Table 1: Sample item for Experiment 1. For a given item, a participant sees only one of the prime sentence stems and then the target sentence stem.

This experimental design yields 4 conditions, created by crossing prime condition with lexical overlap. Each experimental item thus had 4 versions. Each participant completed stems from all four conditions, but saw only one version of each of the experiment’s 32 items (8 per condition). PB found a clear priming effect when the verb was the same, and a numerical ("marginally significant") priming effect when the verb was different. They thus conducted Experiment 2, which had only the "Different Verb" condition, but instead of having a single prime appear before each target, two primes appeared before each target. PB reasoned that this change would increase the size of the priming effect, and they indeed found a numerically larger—and significant—priming effect. Their Experiments 3 through 5 ran other variants that all focused on the "Same Verb" condition, but with different manipulations (varying number and tense), all of which found significant priming effects as visualized in Figure 2.

Since then, the written sentence completion task has been used to study dative priming, as well as its moderators, in a variety of ways (Kantola and van Gompel, 2011; Kaschak, 2007; Kaschak et al., 2006; Coyle and Kaschak, 2012; Kutta and Kaschak, 2012; Kaschak et al., 2011a). Nonetheless, the classic dative priming effect without lexical overlap has been studied relatively infrequently with this paradigm and, to our knowledge, a formal large-scale replication of these findings has never been conducted. In Figure 2, we plot effect sizes from smaller-scale studies that closely mimicked the classic paradigm, with 95% confidence intervals estimated using the methods in Mahowald et al. (2016). Branigan et al. (1999) studied the effect of different kinds of intervening material between primes and targets, and found that the priming effect dissipates if even a single sentence intervenes between them. Moreover, Corley and Scheepers (2002) conducted an online version of the original study, but found a weaker effect in both the "Same Verb" and "Different Verb" conditions. In contrast, Pickering et al. (2002) explicitly compared the priming manipulation to baseline primes that did not privilege either construction, and found effects more consistent with the original PB finding. Cleland and Pickering (2006) used the dative sentence completion paradigm across different modalities (e.g., spoken primes and written targets) to test whether priming occurred across modalities, and indeed found modality-independent effects (for this experiment, Figure 2 only includes data from the classic, written-only version of the study, excluding tasks with intervening materials).

Broadly, across these studies, dative priming effects appear strong when there is lexical boost; however, priming without such lexical boost has rarely been evaluated, and this effect appears weak and perhaps not significant (with the exception of PB Experiment 2). Moreover, the number of participants and items used to detect these effects in the absence of lexical boost means that, by the standards recommended in Mahowald et al. (2016), all such studies have been relatively underpowered: 36 participants 32 items (PB Experiment 1), 64 participants 12 items (PB Experiment 2), and 54 participants 32 items (Corley Scheepers 2002). Even the smallest number of participants run (n=200) in our replication exceeds that number. In addition, as discussed in Mahowald et al. (2016) and below, power can be much improved when the number of items is larger. We use 48 items—more than any of the previous versions. Thus, we believe this to be, by far, the highest powered replication of syntactic priming of the dative alternation employing this widely used paradigm.

As in the past studies, the main findings we anticipate are that, when a prime is PO-biased, the target completion is more likely to be PO; when a prime is DO-biased, the target completion is more likely to be DO. We will preregister and attempt to replicate two hypotheses concerning this priming effect: (a) that it occurs when the main verb is shared between the prime and target stems; and (b) that it occurs when the main verb is different between prime and target. We will also analyze whether the priming effect is larger when the verb is the same than when it is different, by comparing the coefficients from (a) and (b) in a Bayesian regression model.

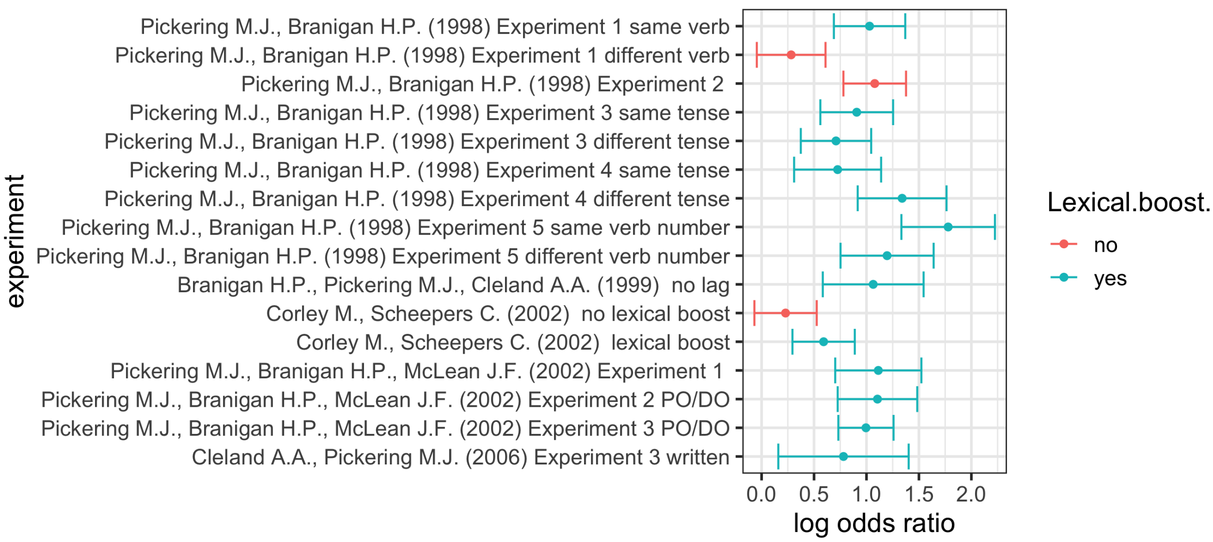


Figure 2: Results from past studies of dative priming that have used a comparable design to our Experiment 1. Effect sizes are log-odds ratios for producing the primed structure (vs. the alternative structure), and error bars are variance estimates from Mahowald et al. (2016)). We will plot our results alongside these.

2.1 Materials

To the extent possible, we use the items that PB used, which appear in the appendix. Because we combine their Experiments 1 and 2, we needed more materials and so wrote a few additional sentences. Moreover, PB's Experiments 3-5 replaced problematic items from Experiments 1 and 2 (namely, those that generate many completions classified as using neither DO nor PO structures, i.e., "Other" responses). Thus, we drew on materials from those latter 3 experiments to create a set of 48 items. We also edited the items to use American, rather than British, English and updated some of the content to reflect cultural changes over the decades since the original study. Our set of materials is available in our OSF repository: https://osf.io/vprg5/.

2.2 Procedure

The experiment is run online using Ibex Farm, with participants recruited through either Amazon's Mechanical Turk or Prolific. Sentence stems will be presented on the screen (and completed by the participant) one at a time. Participants will control when they proceed to the next screen, but they will not be able to go back and modify their previous answers.

2.3 Scoring

Scoring will be done as described in the Scoring section of PB. Each sentence completion will be coded as a PO completion, DO completion, or Other. We will also have a category "NA," which is distinct from Other and will be used if a sentence is not a semantically coherent or syntactically grammatical completion (i.e., the participant did not follow instructions).

Scoring will be blind: target items will be coded separately from prime items, such that the coder will not know which target is paired with which prime.

For convenience, we include the scoring methodology below from Pickering and Branigan (1998):

For each experimental fragment, the first legible response made by the participant was scored as a PO, a DO, or Other. Prime completions were scored as POs if the completion contained a beneficiary noun phrase which was the object of the preposition to and as DOs if the completion contained a patient (or theme) noun phrase. To be scored as either a PO or a DO completion, the verb provided in the fragment could not form part of a phrasal verb (e.g., The architect handed the latest plan over to the builder). All other prime completions were scored as Others. Note that scoring was based on participants' actual completions. For example, if a participant completed a DO-inducing prime fragment as a PO (e.g., completing The mother gave the baby with to her husband), the completion was scored as a PO.

Target completions were scored as POs if the verb provided in the fragment was immediately followed by a noun phrase which acted as the patient or theme and then by a prepositional phrase beginning with to, which acted as the beneficiary. Target completions were scored as DOs if the verb was immediately followed by a noun phrase which acted as the beneficiary and then by a noun phrase which acted as the patient (or theme). To be scored in either category, a completion had to have a grammatical alternative in the other category, where the order of the patient and beneficiary was reversed. Additionally, the verb provided in the fragment could not form part of a phrasal verb. All other target completions were scored as Others.

2.4 Analysis plan

2.4.1 Exclusions

We will exclude participants who did not complete the task as instructed, as evidenced by more than 50% of their sentences being coded as NA.

We will also exclude participants who did not complete the task in earnest, measured as participants who (a) supplied one-word answers for more than 50% of completions or (b) supplied the same answer for 8 or more items.

We exclude Other responses from our analysis and focus only on the key PO and DO target responses. Following PB, we only analyze target items for which the corresponding prime items are completed with the expected prime (i.e., DO completions for DO primes, PO for PO primes).

2.4.2 Effect size

We will treat the proportion of DO responses (out of DO and PO responses) as the key measure. Following Mahowald et al. (2016), we will compute the effect size as a log odds ratio, comparing the log odds of DO responses after DO primes to after PO primes.

2.4.3 Analysis

To assess significance, we will run a Bayesian linear logistic regression mixed effect model, using the R package `brms`.

Our key model will be a maximal mixed effects model (Barr et al., 2013), with random intercepts and slopes for participants and items. There will be three fixed effect predictors: `Prime.Boost`, `Prime.NoBoost`, and `Boost`. The first two, `Prime.Boost` and `Prime.NoBoost`, represent the priming effect (DO completions for targets following DO primes vs. PO primes) in the "Same Verb" and "Different Verb" conditions, respectively; in other words, the priming effect will be nested within each boost condition (Schad et al., 2020). The third predictor (`Boost`) is the effect of lexical overlap (i.e., DO completions following both DO and PO primes in the "Same Verb" condition, contrasted with such completions following both PO and DO primes in the "Different Verb" condition). The contrast coding is as shown in Table 2.

Following Schad et al. (2020), the key analyses will compare the full model to each of two null models that are fit with the same prior and random effect structures but are missing one of the fixed effect terms. The first null model contains all terms except for the fixed effect term Prime.NoBoost; thus, comparing the full model to this simpler model tests the size of the priming effect when there is no lexical boost (i.e., "Different Verb" condition). The second null model instead removes the `Prime.NoBoost` term and thus allows us to estimate the gain in model fit associated with the priming effect in the "Same Verb" condition.

To perform these model comparisons, we use Bayes Factors, following best practices in Schönbrodt and Wagenmakers (2018). Roughly, A Bayes Factor quantifies the support for the full model relative to the simpler models.

- Full model:

Boost	Prime Condition	`Prime.Boost`	`Prime.NoBoost`	`Boost`
Same verb	PO	.5	0	.5
Same verb	DO	-.5	0	.5
Different verb	PO	0	.5	-.5
Different verb	DO	0	-.5	-.5

Table 2: Contrast coding for Experiment 1, where the prime effect is nested within the verb match.

```
TargetCompletionIsPO ~
Prime.Boost + Prime.NoBoost + Boost +
(1 + Prime.Boost + Prime.NoBoost + Boost|Subject) +
(1 + Prime.Boost + Prime.NoBoost + Boost|Item)
```

- Null Model 1:

```
TargetCompletionIsPO ~
Prime.Boost + Boost +
(1 + Prime.Boost + Prime.NoBoost + Boost|Subject) +
(1 + Prime.Boost + Prime.NoBoost + Boost|Item)
```

- Null Model 2:

```
TargetCompletionIsPO ~   Prime.NoBoost + Boost +
(1 + Prime.Boost + Prime.NoBoost + Boost|Subject) +
(1 + Prime.Boost + Prime.NoBoost + Boost|Item)
```

We make our code available in our pre-registered OSF repository: https://osf.io/vprg5/.

We choose to focus our hypothesis testing on the key question of whether there is a priming effect with, and without, lexical boost. But we also will explore whether the priming effect is larger in the "Same Verb" (boost) condition than in the "Different Verb" (no boost) condition, by comparing the posterior draws for the two nested priming coefficients and evaluating how often `Prime.Boost` is bigger than `Prime.NoBoost`. This value will estimate the probability that $\beta_{\texttt{Prime.Boost}}$ is bigger than $\beta_{\texttt{Prime.NoBoost}}$.

Our effect size will be the size of the coefficients `Prime.Boost` and `Prime.NoBoost`. This corresponds, as in Mahowald et al. (2016), to the log odds ratio associated with the priming effect and should be directly comparable to the effect sizes in that meta-analysis.

2.4.4 Sample size and stopping procedures

We follow the procedure for "Sequential Bayes factor with maximal n: SBF+maxN," as described in Schönbrodt and Wagenmakers (2018). As described in detail below, We will run up to 5 batches of 128 participants each (ensuring 500 non-excluded participants using our procedure described), always using the 48 experimental items described above.

We will start with an initial double batch of 256 participants. Because some participants will be excluded, we conservatively assume that we obtain 100 per batch (although we may obtain more, and so any batch can be expected to have approximately but not exactly 100 participants). Because each batch will have a maximum of 128 participants but will invariably be smaller due to exclusions, if, after a particular batch, the total number of non-excluded participants is fewer than 100, we will run additional mini-batches (of 32 participants) until we have at least 100 non-excluded participants in that batch.

After every such batch, we will run our statistical analysis. A Bayes Factor is greater than 6 or smaller than 1/6 would be treated as sufficient evidence (in favor or against the full model, respectively), in which case we will stop running further batches. If we reach 5 batches and the Bayes Factor is still not greater than 6 or smaller than 1/6, we will conclude that the experiment was inconclusive.

Because the fillers for Experiment 1 are, in fact, the items for Experiment 2 (described below) and vice versa, we will actually run new batches until both Experiment 1 and Experiment 2 have passed this threshold. That is, if Experiment 1 has a Bayes Factor of 10.0 after 200 participants but Experiment 2 has a Bayes Factor of only 2.0, we will run another batch for both experiments. The final analysis will always include all non-excluded participants.

More than 5 batches of participant would be difficult or impossible to obtain, given limits to our time and budget. Below, we justify why this design is likely to provide significant insight into the robustness and replicability of dative priming.

2.4.5 Justification for sample size and design analysis

Because we use a Bayesian approach with a pre-specified stopping criterion, a traditional power analysis does not apply to our scenario. But, to assess the feasibility of our plan and evaluate the space of likely outcomes, we conduct a Bayesian design analysis, whereby we simulate experiments using the stopping procedure above (sampling 100 participants at a time, analyzing the results and, if inconclusive, sampling another 100 participants) with a set of pre-chosen design characteristics estimated based on the data reported in Mahowald et al. (2016); Corley and Scheepers (2002); Pickering and Branigan (1998).

In order to simulate experiments, we need to decide on a set of underlying parameters that we can then use to simulate data. To do so, we sample the priming effect sizes from the following distributions: Normal(.88, .10) for the boost ("Same Verb") condition, and Normal(.29, .05) for the no-boost ("Different Verb") condition. We sample these estimates from normal distributions to reflect the inherent uncertainty over the possible effect sizes. We set the third fixed effect, i.e., the "boost" term, to 0 because there is no theoretical reason to expect that PO completions for target sentence stems are more or less likely when the verb is repeated (vs. changed) between prime and target. The intercept is set to -.33, estimated from Corley and Scheepers (2002).

We also simulate participant and item effects by sampling from a Normal(0, 1.5) for the intercepts, and a Normal(0, .14) for the slopes corresponding to the two priming effects (i.e., Prime.NoBoost and Prime.Boost). These parameters were obtained by estimating the participant-specific and item-specific effects for relevant experiments (dative, written sentence completion experiments) in Mahowald et al. (2016). We set the third random slope, Boost, as well as the off-diagonal covariance components, to 0 because we expect them to be very small.

In our simulation, we randomly delete 40% of the data, to reflect the fact that some trials will be excluded because the completions fall into the Other category. We derived this ballpark figure from examining the exclusions in Corley and Scheepers (2002).

A simulation is only as good as the data-generating process used to simulate. To that end, we do some predictive checks on our simulation, comparing (as seen in Figure 3) the results of the simulation to the real data from Corley and Scheepers (2002) Experiment 1. In particular, the first two plots in the figure, (a) and (b), show the distribution of proportion of DO completions over participants and items for the real data (top left panel) compared to 8 randomly chosen simulations. By inspection, the simulations appear to provide a good fit. Plots (c) and (d) in the figure show the priming effect (the difference in proportion of DO completions of targets following DO primes minus the proportion such target completions following PO primes) by participant (c) and item (d). [2]

Having tested the ecological validity of our simulation framework, we then perform the following procedure:

1. Simulate data using the simulation procedure.
2. Fit two Bayesian models: the Full model and Null model 1, as described in our analysis plan.
3. Compute Bayes factors.
4. Apply stopping criteria described in the analysis plan. If necessary, simulate more data and repeat until stopping criteria is reached.

When fitting the Bayesian models in this design analysis, we use the following priors (which are somewhat more informative than the priors we use for the main analysis in order to increase convergence), in brms notation:

```
priors <-c(set_prior("normal(-.3, .1)", class = "Intercept"),
                  set_prior("normal(0, .5)", class = "b"),
                  set_prior("normal(0, .05)", class = "sd"),
                  set_prior("lkj(2)", class = "L"),
                  set_prior("normal(0, .5)", class = "sd", group="subj", coef="Intercept" ),
                  set_prior("normal(0, .5)", class = "sd", group="item", coef="Intercept" ))
```

Note that, for the design analysis only, the intercept is “informative” (with a mean at -.3) in order to aid model fitting.

[2]Based on visual inspection of the by-participant priming effects, it seems that the real data have a slightly bigger peak around .50 compared to the simulated data. Decreasing the variance for the by-participant priming effect would smooth the peak, but would also reduce the number of participants who show near-categorical priming effects (see the peak near 1.00 in the real data). We suspect this term would best be modeled by a two-step model that first samples whether a participant has 1.00 and then, if not, draws from a normal. But, because the fit here is relatively good as is, we chose to avoid the increased computational cost associated with such a model. We hope that these modeling decisions will be the subject of future work.

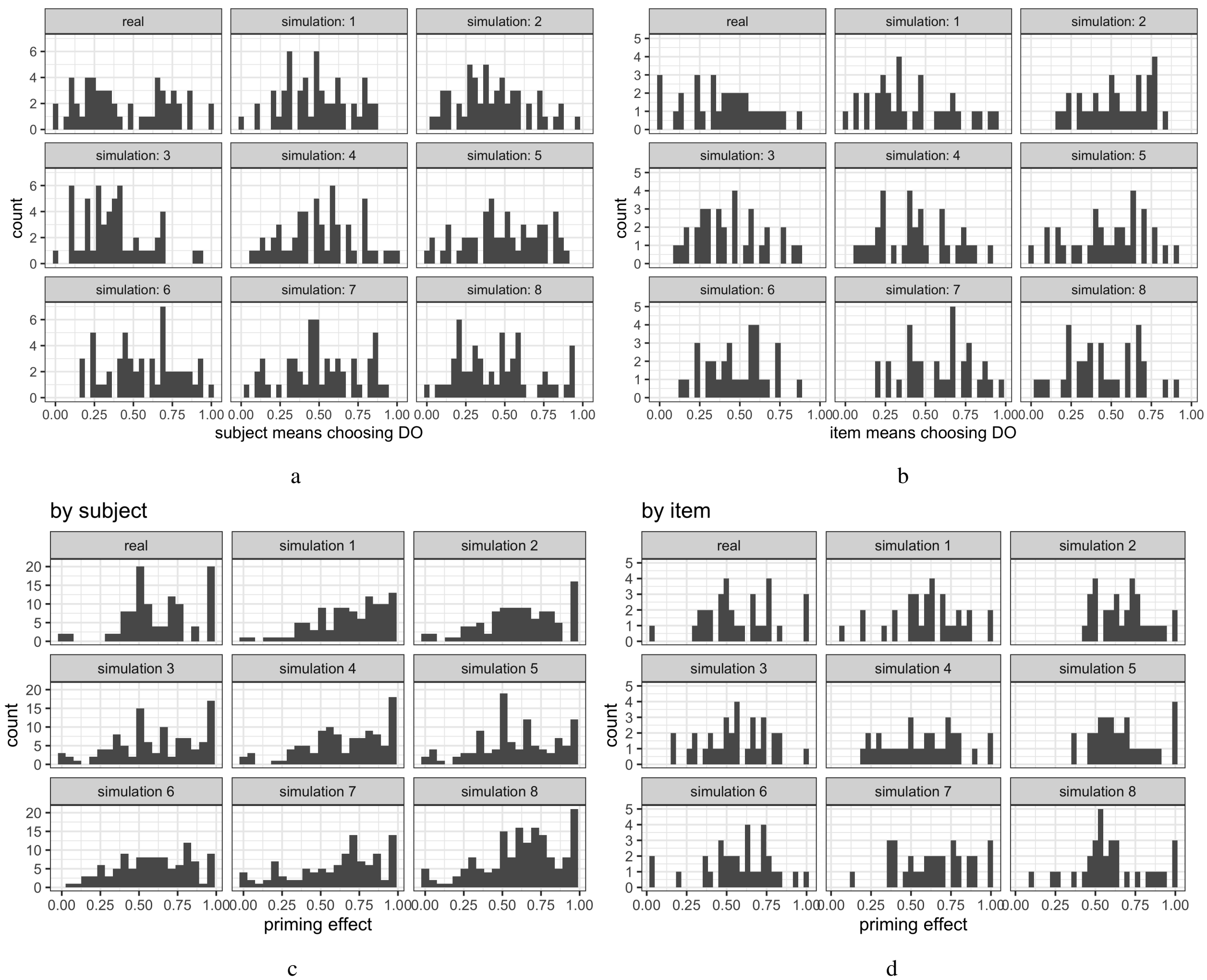


Figure 3: (a) For the data from Corley et al., the mean probability of choosing DO by subject, compared to 8 simulated baselines using our choice of parameters. (b) For the data from Corley et al., the mean probability of choosing DO by item, compared to 8 simulated baselines using our choice of parameters. (c) The size of the priming effect (difference in proportions choosing DO between conditions) per subject in real vs. simulated data. (d) The size of the priming effect (difference in proportions choosing DO between conditions) per item in real (top-left) for simulated data.

We record the output of the final Bayes factors attained using this procedure, as well as the number of participants ultimately run to attain those Bayes Factors. We also obtained, for each simulation, the probability that the boost priming effect was greater than the no-boost priming effect.

Because these models are computationally costly to fit in repeated simulations, our design analysis does not compare the full model to Null Model 2 (which removes the boost priming term, testing dative priming in the "Same Verb" condition) in our design analysis (although we will in the final analysis). But, because we expect the priming effect to be much larger in the boost condition, the sample size calculations here (for the weaker, no-boost condition) serve as a conservative estimate of our power for exploring the boost condition.

Bayes factors depend on the choice of prior, and the design analysis here of course depends on the choice of parameters used to simulate our data. While we believe these are realistic, a different choice of parameters could paint a very different picture. An advantage of using the Bayesian approach is that, even if we have overestimated the effect size or underestimated the variance, our resultant analysis will quantify the strength of evidence obtained.

In Figure 5, we plot the Bayes factors obtained in our simulations. In 85% of our simulations, we obtained a Bayes Factor > 6. 13% of the time, we obtained a Bayes Factor < 1/6 (which would cause us to arbitrate in favor of the simpler,

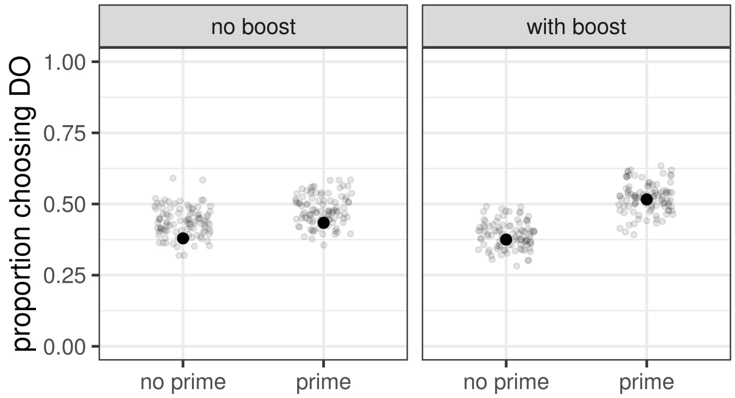


Figure 4: Simulation results (gray) for proportion of target DO completions for each of the four experimental conditions in Experiment 1. Simulations were based on data from Corley and Scheepers (2002) (black).

null model, i.e., against a dative priming effect). And 2% of the time, we found no conclusive effect (1/6 < BF < 6) even after running 500 participants.

In addition, our ability to detect the lexical boost effect in these simulations was extremely high. Namely, the priming effect coefficient nested inside the boost ("Same Verb") condition was larger than that nested inside the no-boost ("Different Verb") condition in all of our simulations. And, when we computed the probability that the boost coefficient was greater than the no-boost by comparing posterior draws ($P(\beta_{\texttt{boost}} > \beta_{\texttt{no.boost}})$, the probability was above .95 in 98% of our simulations.

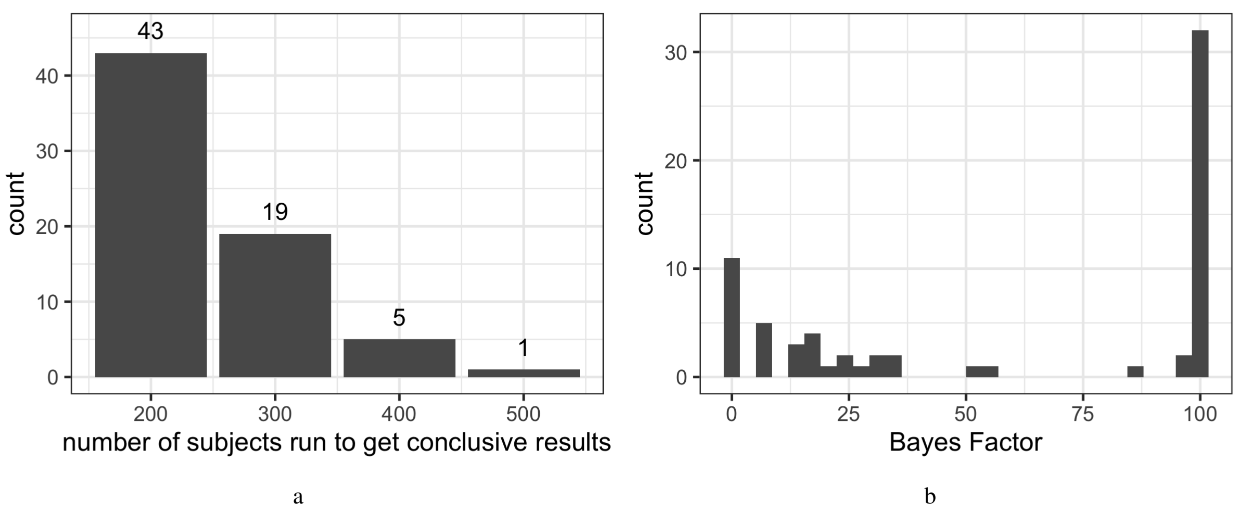


Figure 5: (a) Distribution of the number of subjects run across simulations until convergence was reached. (b) Distribution of Bayes Factors (with all Bayes Factors greater than 100, truncated to 100), across simulations.

2.4.6 Justification for prior

We use semi-informative priors (Gelman et al., 2004; Vasishth et al., 2018). Following Vasishth et al. (2018), we perform prior predictive checks in order to set realistic priors. Because our model is a logistic regression model, we operate in logit space and values like -2.0 and 2.0 (which correspond to roughly .12 and .88 in proportion/probability space) are actually quite extreme, and so values more extreme than that for any of our parameters are unlikely. (For instance, a priming effect of 2.0 in this set-up would be implausibly large.) In `brms` notation, we set the prior as follows:

```
priors <-c(set_prior("normal(0, .5)", class = "Intercept"),
```

```
                  set_prior("normal(0, .5)", class = "b"),
                  set_prior("normal(0, .05)", class = "sd"),
                  set_prior("lkj(2)", class = "L"),
                  set_prior("normal(0, .5)", class = "sd", group="subj", coef="Intercept" ),
                  set_prior("normal(0, .5)", class = "sd", group="item", coef="Intercept" ))
```

That is, both the intercept and the fixed effect coefficients have a normal prior with mean 0 and standard deviation 0.5 (setting most of the distribution mass to fall between -2.0 and 2.0).

For random intercepts, we set the prior to be a truncated normal, with mean 0 and standard deviation 0.5. For the random slopes, we use a tighter prior, a truncated normal with mean 0 and standard deviation 0.05. We choose this prior because, based on an analysis of the raw data analyzed in Mahowald et al. (2016). Broadly, we believe that, while our random intercepts may be relatively large, the random slopes are likely to be small because there is relatively little item and subject-consistency. The LKJ prior on the correlation matrix is widely used in Bayesian statistics, and we follow existing literature in using an LKJ prior of 2.

Because Bayes factors are sensitive to the choice of prior, we will also report a secondary analysis with a less informative prior as a sensitivity check:

```
priors <-c(set_prior("normal(0, 2)", class = "Intercept"),
                  set_prior("normal(0, 2)", class = "b"),
                  set_prior("normal(0, 2)", class = "sd"),
                  set_prior("lkj(2)", class = "L"),
                  set_prior("normal(0, 2)", class = "sd", group="subj", coef="Intercept" ),
                  set_prior("normal(0, 2)", class = "sd", group="item", coef="Intercept" ))
```

2.4.7 Exploratory Analyses

We will also run additional exploratory analyses to be determined. These will focus, in part, on how our priming effect fits in with the expected effect predicted by previous meta-analyses of syntactic priming.

2.5 Pilot Study and Results

We ran a pilot study identical to the study we plan to run. The goal of the pilot study was to assess the mechanics of the experiment. We collected data from 8 participants, of which 7 provided usable data. The average time to complete the experiment was 88.9 minutes, and we paid participants at a rate of $12 per hour.

For the target completions, 49.7% were Other completions (roughly in line with our 40% estimate). 98.5% of primes were completed with the expected completion.

We noticed that target completions using the verb "sell" led to noticeably more Other completions than other verbs. However, to maintain consistency with the original study, and because we could not find a compelling alternative, we chose to retain "sell".

For the final experiment, we also fix a few minor mistakes: one sentence appeared without its full completion and two others contained small typos.

The Bayes factors obtained are consistent with what we expected: the pilot study provides far too little data to draw conclusions. The pilot study BF is 1.12 for the priming effect with no boost, and 1.00 for priming with boost–consistent with favoring neither the Full model nor the simpler, Null models. We provide an analysis script in our OSF repository.

3 Experiment 2

Our Experiment 1 is a replication of a classic priming construction, the dative alternation. Next, we turn to an effect which is less well-attested in the literature: priming of high vs. low relative clause (RC) attachment. The original finding of this phenomenon is reported in Scheepers (2003). Across three experiments in German, he used prime sentences including the form NP-of-NP-RC (e.g. "...the waiter of the actress, who..."), which triggered completions of the RC with either high or low attachment (i.e., modification of either *waiter* or *actress*), depending upon the gender marking of the relative pronoun. Following these primes, target fragments had a similar structure, but ended in an ambiguous relative pronoun that could attach to either noun phrase:

Scheepers' Experiments 1 and 2 demonstrated RC attachment priming: completing targets by using a high-attachment RC was more likely after high-attachment primes than after low-attachment primes. Importantly for Scheepers' study,

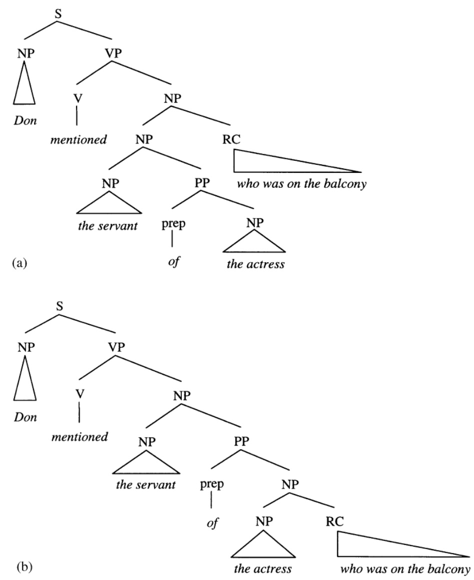


Fig. 1. (a) High attachment of the relative clause (RC). (b) Low attachment of the RC.

Figure 6: Figure from Scheepers (2003), showing a syntactic analysis of high and low relative clause attachment. The strong interpretation of syntactic priming claims that the underlying structure—here, the abstract phrase structure tree—is primed as either high or low.

Sentence Type	German	English
LA Prime	Die Assistentin verlas den Punktestand der Kandidatin, die ...	The assistant announced the score [masc, sing] of the candidate [fem, sing] that [fem, sing] ...
HA Prime	Die Assistentin verlas den Punktestand der Kandidatin, der ...	The assistant announced the score [masc, sing] of the candidate [fem, sing] that [masc,sing] ...
Baseline Prime	Die Assistentin verlas den Punktestand der Kandidatin, bevor ...	The assistant announced the score [masc, sing] of the candidate [fem, sing] before...
Target	Der Rentner schimpfte über die Autorin der Flugblätter, die...	The pensioner railed about the author [fem, sing] of the fliers [neut,plur] that [?]

Table 3: Sample items from Scheepers (2003), with English translations.

this priming effect indicated that syntactic priming does not depend on particular lexical items shared across primes and targets (cf. the function word *to* in prepositional datives of our Experiment 1); it also indicated that priming does not depend on activating particular context-free phrase structure rules in isolation (cf. a prepositional phrase rule in prepositional datives in our Experiment 1), because high and low attachment completions could both be generated by the same set of rules (albeit in different orders). Rather, because high and low attachments differ in their underlying syntactic tree structure, the observed priming effect indicated that language users represented these structures, and that these representations persist across sentences. Experiment 3 in Scheepers' study provided further evidence that syntactic tree structure, rather than some non-syntactic factor, was responsible for priming, because priming of RC attachment was not induced by adverbial phrases referring to either the first (high) or second (low) noun phrase (e.g. "...the waiter of the actress, when she...").

Later work has demonstrated RC attachment priming cross-linguistically (Desmet & Declercq, 2006) and cross-structurally (Loncke, Van Laere, & Desmet, 2011). Specifically, after demonstrating that RC attachment priming obtains when Scheepers' German materials are translated into Dutch (Experiment 1), Desmet and Declerq tested whether Dutch-English bilinguals demonstrate RC attachment priming across languages (Experiment 2). Native Dutch speakers, proficient in English, completed Dutch prime sentence stems and English target sentence stems; the cross-linguistic priming effect was strikingly similar to the within-language effect obtained in their Experiment 1 (see Figure 7).

Loncke and colleagues (2011) designed a study to test two different explanations for the RC attachment priming results observed in the previous experiments. One possibility is that attachment height is primed independent of the particular

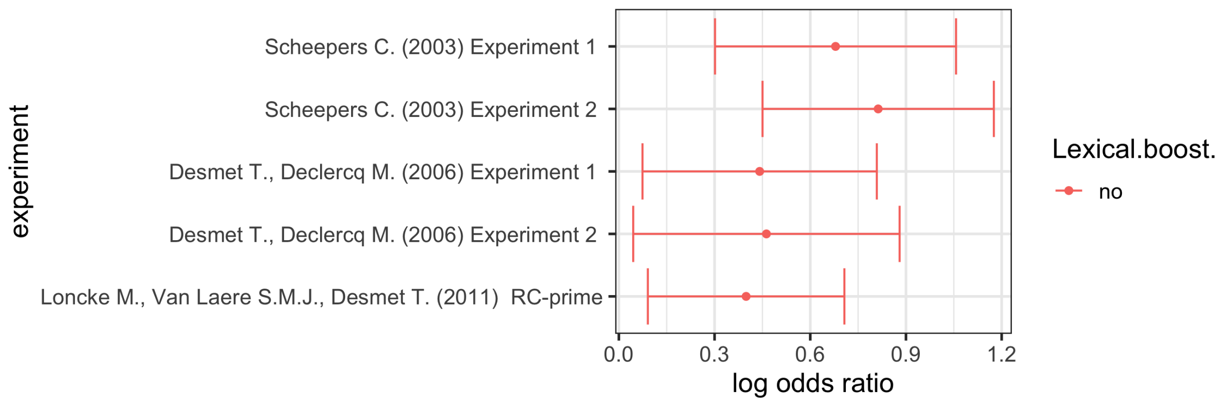


Figure 7: Effect size and error bars (using variance estimates from Mahowald et al. (2016) for the Experiment 2 studies. We will plot our results alongside these.

type of modifying phrase being attached; another possibility is that the priming depends on a specific syntactic structure, such that RC attachment would not be primed by, e.g., a prepositional phrase. Dutch speakers completed stems that primed a high or low attachment of either a prepositional phrase (PP) or RC. Critically, Loncke and colleagues found that PP attachment primed RC attachment, supporting the first of their hypotheses. Note that, because the (Mahowald et al., 2016) meta-analysis excluded cross-structure priming, Figure 7 only reports the results of the RC attachment prime conditions, which replicates the previous results.

We chose RC attachment as the focus of Experiment 2 for several reasons: the demonstration of RC attachment priming bears on theories of syntactic priming mechanisms generally; this case is relatively understudied and, to our knowledge, untested within English; and it provides an example of a smaller (i.e., weaker) priming effect to replicate, relative to that of our Experiment 1. We will conduct a replication of Scheepers (2003) Experiment 1, in which participants were presented with equal numbers of high and low attachment primes, but we are not including his baseline prime condition. Comparisons to baseline were not included in our meta-analysis, and so we chose to exclude them here to reduce the session length.

3.1 Materials

Materials are available in our OSF repository. Because previous studies of high attachment priming were not done in English, we adapted materials for our purposes. Whereas studies in German and Dutch could use a rich gender+number agreement system as a cue to priming, English has is more constrained in its morphological markings, such that deciphering whether participant's completion uses high vs. low attachment can be challenging. Therefore, in each of our items, one of the two nouns is singular and the other plural. This lets us, in some cases, use verb number agreement as a cue, but only if the participant uses the third person present tense, or the past *was*/*were*. We also make sure that each sentence has one animate noun (the plural) and one inanimate noun (the singular) so as to aid in disambiguating whether attachment is high or low based on the semantic content of the completions.

Varying animacy raises a possible confound: because one might expect a baseline preference to attach the relative clause to the animate noun (as explored in more detail in Yen et al., in prep), it is possible that the preference to attach to an animate (or inanimate) noun can itself be subject to priming. To avoid this confound, we counterbalance the items such that half the targets have the animate noun first and half have the inanimate noun first. Primes always have the order inanimate (singular) before animate (plural).

Below, we give example items with brackets showing attachment in the primes (brackets do not appear in the actual experiment). To force the completions of the primes to have the desired RC attachment structures, we use a verb that agrees in number with either the high, singular noun (*was*) or the low, plural noun (*were*), as well as a pronoun consistent with the animacy of the noun (*which* or *who*). Because the nouns in the primes always differ in animacy and number, the combination of the pronoun and the verb strongly encourages a completion with either the high or low noun.

For targets, we use the ambiguous relative pronoun *that* and do not supply a verb.

For each participant, half of the targets will have the animate noun first, and half will have the inanimate noun first; each target will be completed once, following either a low-attachment or a high-attachment prime. Across

Prime Type Type	Primes	Target
Prime; Low Attachment:	Fran corrected the manuscript of [the authors who were]…	Terry inspected the lawn of the parents that…
Prime: High Attachment:	Fran corrected [the manuscript of the authors which was]	

Table 4: Sample item 1 for Experiment 2, showing the variant that has a target with the inanimate noun first.

Prime Type Type	Primes	Target
Prime; Low Attachment:	We criticized the viewpoint of [the journalists who were]…	They congratulated the winners of the tournament that…
Prime: High Attachment:	We criticized [the viewpoint of the journalists which was]	

Table 5: Sample item 2 for Experiment 2, showing the variant that has a target with the animate noun first.

items, each participant will complete primes of both types of attachment (i.e., the syntactic priming manipulation is within-participant).

3.2 Procedure

The procedure is the same as in Experiment 1, with the written sentence completion task run on Ibex Farm. Because the items from Experiment 1 serve as fillers for Experiment 2 and vice versa, the experimental setting and procedures are the same.

3.3 Scoring

Each prime sentence will be scored as high attachment (HA), low attachment (LA), or Other. Prime sentences that elicit the desired attachment are considered "correct." Target sentences are also scored as HA, LA, or Other. Following Scheepers (2003), scoring relies on a combination of syntactic cues (number agreement) and plausibility (it is, in many cases, easy to tell whether the clause is attached to the animate or inanimate item). We analyze only targets which follow "correctly" completed primes. For cases in which two blind annotators disagree on the preferred reading, the completion will be marked as Other. We exclude target responses coded as Other from our analysis and focus only on the key HA and LA target responses. As in Experiment 1, items can also be coded as NA if they are not possible completions of the sentence (i.e., the participant has made a mistake or did not follow the instructions for the task).

3.4 Analysis plan

3.4.1 Exclusions

We will exclude participants who did not complete the task as instructed, as evidenced by more than 50% of their sentences being coded as NA. (Other responses are counted as valid, along with HA and LA responses, for purposes of this exclusion criterion.)

3.4.2 Effect size

We will treat the proportion of HA responses (out of HA and LA responses) as the key measure. Following Mahowald et al. (2016), we will compute the effect size as a log odds ratio, comparing the log odds of HA responses after HA primes to after LA primes.

3.4.3 Analysis

To estimate the priming of RC attachment, we will run a Bayesian linear logistic regression mixed effect model, using the R package `brms`, following the same approach as in Experiment 1.

Our key model will be a maximal mixed effects model (Barr et al., 2013), with random effects and slopes for participants and items. There will be one fixed predictor: `Prime` (coded .5/-.5 for high and low attachment primes, respectively).

The key comparison is to compare the full model to a simple, Null model without a fixed effect of prime, using the Bayes factor.

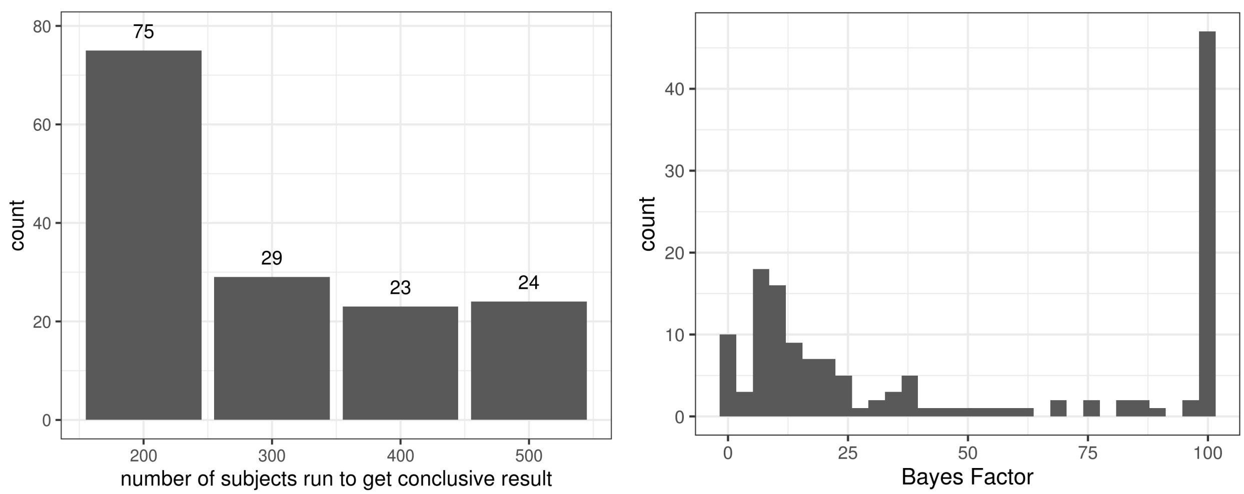
- Full model:
  `TargetCompletionIsHA ∼ 1 + Prime + (1 + Prime|Subject) + (1 + Prime|Item)`
- Null Model 1:
  `TargetCompletionIsHA ∼ 1 + (1 + Prime|Subject) + (1 + Prime|Item)`

3.4.4 Sample size and stopping procedures

As in Experiment 1, we follow the procedure for "Sequential Bayes factor with maximal n: SBF+maxN," as described in Schönbrodt and Wagenmakers (2018). Once again, will have 48 items and will run up to 500 participants. More than 500 participants would be difficult or impossible given budgetary constraints.

In our simulations, we sample the effect size from a Normal(.2, .05) distribution. We use the same participant and item variance parameters as in Experiment 1. As in Experiment 1, in our simulation we randomly delete 40% of the data to reflect the fact that some trials will be excluded because the completions fall into the Other category. We follow the same batch procedure and stopping procedure as in Experiment 1.

We ran 150 simulations. In 91% of them, the Bayes Factor was greater than 6, as shown in Figure 5. Most of the time, we had sufficient evidence to draw a conclusion after 200 or 300 participants. But, as seen in Figure 8a, there were some simulations in which we had to run the maximum 500 participants (and, in 7% of cases, even 500 participants were insufficient).



a b

Figure 8: (a) The number of subjects run across simulations until convergence was reached. The leftmost bar shows that, in 75 of the simulations run, convergence was reached after 200 participants. (b) Distribution of Bayes factors (with all Bayes factors greater than 100, truncated to 100), across simulations.

Across the 5 studies displayed in Figure 7, there are a total of 182 participants and so, even our minimum number of participants will double the existing data on relative clause attachment priming. And it will, for the first time, extend the work to English.

3.4.5 Exploratory Analyses

We will also run additional exploratory analyses to be determined.

3.5 Pilot Study and Results

As in Experiment 1, we collected pilot data from 8 participants, of which 7 provided usable data. 98.5% of primes were completed with the correct prime completion, suggesting the prime sentences behaved as intended. 10.7% of target completions were completed as Other (less than the 40% we accounted for in our model, which means power may be higher that expected), 49.1% with low attachment and 40.2% with high attachment.

One item contained a mistake (it was missing its relative pronoun), which we have fixed for the final version. Three others frequently resulted in completions that were ambiguous in attachment, making it difficult to discern which nouns the participants intended to modify. These three items were edited to make it clearer in the final experiment whether participants produced high- or low-attachment completions.

The Bayes factors obtained by comparing the full model to the simpler, Null model was .48, which means that the null model is somewhat favored. However, we do not treat these pilot data as sufficient for drawing conclusions.

Further details of the pilot data are available in the OSF repository, but we do not recommend drawing any strong conclusions from these results given the small amount of data.

4 Discussion

The discussion will be written after data collection, but we pre-register the following:

- If we see strong evidence in favor of the full models (i.e., we "replicate" the priming effects), we will take that as evidence in favor of syntactic priming as a robust phenomenon in the written sentence completion paradigm. We will, however, urge caution that results not be over-interpreted overly strongly as evidence for abstract syntactic structure since, as Ziegler et al. (2019) and Bidgood et al. (2020) discuss, there are various other possible explanations for the phenomenon.
- If we fail to replicate one or both experiments, we will take that as evidence that researchers should proceed with much more caution in using syntactic priming as it is currently used. But we would again urge caution and will discuss the fact that a single failed, high-powered replication does not mean syntactic priming does not exist in some form, for some constructions, and in some paradigms.
- If our evidence is inconclusive, we will discuss possible ways to study syntactic priming further. We will also discuss the possibility that earlier studies were underpowered, if it turns out that even our large-N study lacks sufficient power.

In all cases, we will quantitatively and qualitatively compare these results to those discussed in Mahowald et al. (2016).

References

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

Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, 127(3):287–306.

Bidgood, A., Pine, J. M., Rowland, C. F., and 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):e12892.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(3):355–387.

Bock, K. and Loebell, H. (1990). Framing sentences. *Cognition*, 35(1):1–39.

Brand, M., Rey, A., and Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory and Language*, 48(2):435–443.

Branigan, H. P. and Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, 40.

Branigan, H. P., Pickering, M. J., and Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, 6(4):635–640.

Bunger, A., Papafragou, A., and Trueswell, J. C. (2013). Event structure influences language production: Evidence from structural priming in motion event description. *Journal of Memory and Language*, 69(3):299–323.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376.

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on psychological science*, 9(1):40–48.

Cleland, A. and Pickering, M. (2006). Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language*, 54(2):185–198.

Corley, M. and Scheepers, C. (2002). Syntactic priming in English sentence production: Categorical and latency evidence from an internet-based study. *Psychonomic Bulletin & Review*, 9(1):126–131.

Coyle, J. M. and Kaschak, M. P. (2012). Female Fertility Affects Men's Linguistic Choices. *PLoS ONE*, 7(2):e27971.

Desmet, T. and Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, 54(4):610–632.

Doyen, S., Klein, O., Pichon, C.-L., and Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind? *PloS one*, 7(1):e29081.

Ferreira, V. S., Bock, K., Wilson, M. P., and Cohen, N. J. (2008). Memory for syntax despite amnesia. *Psychological Science*, 19(9):940–946.

Gelman, A. and Carlin, J. (2016). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, 9(6):641–651.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*. CRC press, Boca Raton, FL.

Gelman, A. and Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no 'fishing expedition'or 'p-hacking'and the research hypothesis was posited ahead of time. *Downloaded January*, 30:2014.

Haarmann, H. J. and Kolk, H. H. (1991). Syntactic priming in broca's aphasics: Evidence for slow activation. *Aphasiology*, 5(3):247–263.

Harris, C. R., Coburn, N., Rohrer, D., and Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PloS one*, 8(8):e72467.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., and 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.

Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, 15(6):409–414.

Ioannidis, J. P., Munafo, M. R., Fusar-Poli, P., Nosek, B. A., and David, S. P. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in cognitive sciences*, 18(5):235–241.

Kantola, L. and van Gompel, R. P. G. (2011). Between- and within-language priming is the same: Evidence for shared bilingual syntactic representations. 39(2):276–290.

Kaschak, M. P. (2007). Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition*, 35(5):925–937.

Kaschak, M. P., Kutta, T. J., and Jones, J. L. (2011a). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, 18(6):1133–1139.

Kaschak, M. P., Kutta, T. J., and Schatschneider, C. (2011b). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, 39(3):381–388.

Kaschak, M. P., Loney, R. A., and Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition*, 99(3):B73–B82.

Kim, Y. and McDonough, K. (2007). Learners' Production of Passives during Syntactic Priming Activities. *Applied Linguistics*, 29(1):149–154.

Kutta, T. J. and Kaschak, M. P. (2012). Changes in task-extrinsic context do not affect the persistence of long-term cumulative structural priming. *Acta Psychologica*, 141(3):408–414.

Kvarven, A., Strømland, E., and Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4):423–434.

Loebell, H. and Bock, K. (2003). Structural priming across languages. *Linguistics*, 41(5; ISSU 387):791–824.

Loncke, M., Van Laere, S. M. J., and Desmet, T. (2011). Cross-Structural Priming: Prepositional Phrase Attachment Primes Relative Clause Attachment. *Experimental Psychology*, 58(3):227–234.

Mahowald, K., James, A., Futrell, R., and Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, 91:5–27.

Martin, R. C. and Jensen, C. R. (1988). Phonological priming in the lexical decision task: A failure to replicate. *Memory & Cognition*, 16(6):505–521.

Melinger, A. and Cleland, A. A. (2011). The influence of sentential position on noun phrase structure priming. *The Quarterly Journal of Experimental Psychology*, 64(11):2211–2235.

Messenger, K., Branigan, H. P., and McLean, J. F. (2011). Evidence for (shared) abstract structure underlying childrenâs short and full passives. *Cognition*, 121(2):268–274.

Messenger, K., Branigan, H. P., McLean, J. F., and Sorace, A. (2012). Is young childrenâs passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language*, 66(4):568–587.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716–aac4716.

Pashler, H., Coburn, N., and Harris, C. R. (2012). Priming of social distance? failure to replicate effects on social and food judgments. *PloS one*, 7(8):e42510.

Peter, M., Lukatela, G., and Turvey, M. (1990). Phonological priming: Failure to replicate in the rapid naming task. *Bulletin of the Psychonomic Society*, 28(5):389–392.

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

Pickering, M. J., Branigan, H. P., and McLean, J. F. (2002). Constituent Structure Is Formulated in One Stage. *Journal of Memory and Language*, 46(3):586–605.

Pickering, M. J. and Ferreira, V. (2008). Structural priming: A critical review. 134(3):427–459.

Ramscar, M., Shaoul, C., Baayen, R. H., and Tbingen, E. K. U. (2015). Why many priming results don't (and won't) replicate: A quantitative analysis. *Unpublished manuscript*.

Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110:104038.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition*, 89(3):179–205.

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., and Viskupova, I. (2011). Structural Priming Across Cognitive Domains: From Simple Arithmetic to Relative-Clause Attachment. 22(10):1319–1326.

Schönbrodt, F. D. and Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review*, 25(1):128–142.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, page 0956797611417632.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-Curve and Effect Size Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science*, 9(6):666–681.

Slocombe, K. E., Alvarez, I., Branigan, H. P., Jellema, T., Burnett, H. G., Fischer, A., Li, Y. H., Garrod, S., and Levita, L. (2013). Linguistic Alignment in Adults with and Without Aspergerâs Syndrome. *Journal of Autism and Developmental Disorders*, 43(6):1423–1436.

Steele, K. M. (2014). Failure to replicate the mehta and zhu (2009) color-priming effect on anagram solution times. *Psychonomic bulletin & review*, 21(3):771–776.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71:147–161.

Verreyt, N., Bogaerts, L., Cop, U., Bernolet, S., De Letter, M., Hemelsoet, D., Santens, P., and Duyck, W. (2013). Syntactic priming in bilingual patients with parallel and differential aphasia. *Aphasiology*, 27(7):867–887.

Ziegler, J., Bencini, G., Goldberg, A., and Snedeker, J. (2019). How abstract is syntax? evidence from structural priming. *Cognition*, 193:104045.