



A failure to replicate rapid syntactic adaptation in comprehension

Caoimhe M. Harrington Stack¹ · Ariel N. James² · Duane G. Watson¹

Published online: 12 April 2018
© Psychonomic Society, Inc. 2018

Abstract

Language comprehension requires successfully navigating linguistic variability. One hypothesis for how listeners manage variability is that they rapidly update their expectations of likely linguistic events in new contexts. This process, called *adaptation*, allows listeners to better predict the upcoming linguistic input. In previous work, Fine, Jaeger, Farmer, and Qian (*PLoS ONE*, 8, e77661, 2013) found evidence for syntactic adaptation. Subjects repeatedly encountered sentences in which a verb was temporarily ambiguous between main verb (MV) and reduced relative clause (RC) interpretations. They found that subjects who had higher levels of exposure to the unexpected RC interpretation of the sentences had an easier time reading the RC sentences but a more difficult time reading the MV sentences. They concluded that syntactic adaptation occurs rapidly in unexpected structures and also results in difficulty with processing the previously expected alternative structures. This article presents two experiments. Experiment 1 was designed as a follow-up to Fine et al.'s study and failed to find evidence of adaptation. A power analysis of Fine et al.'s raw data revealed that a similar study would need double the items and four times the subjects to reach 95% power. In Experiment 2 we designed a close replication of Fine et al.'s experiment using these sample size guidelines. No evidence of rapid syntactic adaptation was found in this experiment. The failure to find evidence of adaptation in both experiments calls into question the robustness of the effect.

Keywords Sentence processing · Syntax · Adaptation · Replication

Successfully comprehending language requires that language users accommodate a wide range of variability in the linguistic signal. Speakers vary in how they prefer to articulate sounds, which words they choose, and which syntactic structures they select. How do listeners successfully understand language when the input is so variable? One proposal is that listeners adapt: They alter their expectations about the input on the basis of past experiences and the current context (Fine, Jaeger, Farmer, & Qian, 2013; Kleinschmidt, Fine, & Jaeger, 2012; Norris, McQueen, & Cutler, 2016; Xiang & Kuperberg, 2015). This approach has been used to explain syntactic-processing effects, in particular (Fine & Jaeger, 2013; Fine et al., 2013; Jaeger & Snider, 2008; Kleinschmidt & Jaeger, 2015). The claim is that listeners learn the frequency of the syntactic structures that occur in a given context. They have difficulty processing unexpected structures, but have less

difficulty processing structures that are more likely. Critically, the more one encounters a structure, the easier that structure is to process.

In this article, two experiments are presented that test whether syntactic adaptation to a low-frequency syntactic structure is possible within a single experimental session. The first study was originally designed as a follow-up to Fine et al. (2013), which reported adaptation to a difficult, low-frequency syntactic structure (reduced relative clauses) within a single experimental session. We were interested in exploring whether these adaptation effects are sensitive to the context in which the initial learning occurs. Not only were no effects of context found, but also no evidence that readers adapted to reduced relative clauses at all. This led us to conduct a power analysis of the original Fine et al. study to determine the sample size that would be needed to reach 95% power in a new study. The second experiment was an attempt to replicate the original findings of Fine et al. using more items and more subjects, with the goal of evaluating the size of the originally reported adaptation effect.

Revisiting this particular syntactic adaptation effect is important for two reasons. First, interest in the possibility of rapid linguistic adaptation has grown in the past several years

✉ Caoimhe M. Harrington Stack
c.stack@vanderbilt.edu

¹ Vanderbilt University, Nashville, TN, USA

² Macalester College, St. Paul, MN, USA

and has spurred a number of studies (Farmer, Fine, Yan, Cheimariou, & Jaeger, 2014; Fine & Jaeger, 2013; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Myslin & Levy, 2016; Yildirim, Degen, Tanenhaus, & Jaeger, 2013), including Experiment 1 below. Given this interest, understanding the size of adaptation effects may help us better understand the role it plays in language processing. Second, if it is the case that different syntactic structures elicit different degrees of adaptation, or that different syntactic structures vary in their propensity for eliciting any adaptation at all, we can build better models of syntactic adaptation. These issues are discussed in more depth in the [General Discussion](#).

Fine et al. (2013) explored whether syntactic adaptation can occur within a single experimental session. Subjects were given a reading task in which they read one sentence at a time. The critical sentences used verbs that were temporarily ambiguous between a main verb reading (MV) and a reduced relative clause reading (RC). Two ambiguous example sentences, along with their unambiguous counterparts, are shown below.

- (1a) MV-Ambiguous: The experienced soldiers warned about the dangers before the midnight raid.
- (1b) MV-Unambiguous: The experienced soldiers spoke about the dangers before the midnight raid.
- (1c) RC-Ambiguous: The experienced soldiers warned about the dangers conducted the midnight raid.
- (1d) RC-Unambiguous: The experienced soldiers who were told about the dangers conducted the midnight raid.

Sentences like Example 1a are typically easier to read than sentences like Example 1c. In both sentences, “warned” is initially parsed as the main verb of the sentence. Whereas this initial parse is correct in Example 1a, in Example 1c the parse must be revised or reranked once the reader reaches the true main verb of the sentence—“conducted” (Ferreira & Clifton, 1986; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994; Rayner, Carlson, & Frazier, 1983; Trueswell & Tanenhaus, 1994; Trueswell, Tanenhaus, & Garnsey, 1994, and many others). Readers tend to slow down when a temporarily ambiguous sentence has an unexpected continuation, and this slowdown tends to occur at the disambiguating portion of the sentence. In the case of Example 1c, this occurs at “conducted.”

Fine et al. (2013) made two predictions about these kinds of syntactic structures. The first was that increased exposure to reduced relative clauses would facilitate processing. If syntactic beliefs are rapidly updated, increased exposure to reduced relatives would lead the language system to expect more of these sentences, making them easier to read. This should be reflected in smaller ambiguity effects, the difference in reading times between ambiguous and unambiguous sentences of

the same type, as readers gain more experience with the structure. The second prediction was that as exposure to RCs increased and exposure to MVs decreased, it would become more difficult to read MV sentences. Fine et al. found that subjects with more exposure to RC sentences showed reduced ambiguity effects to those sentences and increased ambiguity effects to MV sentences. Below, we attempt to replicate these effects.

Experiment 1

In this first experiment, whether or not adaptation is sensitive to the environmental contexts in which language is encountered is investigated. Previous work has suggested that readers rapidly adapt to the statistics of their environment, adjusting their expectations for encountering particular syntactic structures to reflect the current context even when those statistics contrast with the readers’ broader experience (Fine et al., 2013). But how long does this learning persist, and is it constrained by the context in which it is learned? We were interested in seeing whether readers used the same statistics across sessions if the context of those sessions matched. If syntactic adaptation is a result of a rational and efficient information-updating process, then readers should retain information about the context and use that information to determine which statistics to deploy.

In this experiment, the subjects’ physical contexts were manipulated. Participants were seated in one of two rooms and completed a self-paced reading session designed to elicit adaptation, following Fine and Jaeger (2016). Participants then returned a day later to complete a second session of self-paced reading with a new set of sentences. This second session took place in either the same room or a different room. The questions of interest were (a) whether the adaptation achieved on Day 1 would persist until Day 2, and (b) whether evidence of adaptation would be stronger when subjects returned to the same room rather than a different room. Unexpectedly, subjects did not show evidence of adaptation on either day, failing to replicate the previous results (Fine & Jaeger, 2016; Fine et al., 2013). The nonreplication of the basic syntactic adaptation effect rendered it impossible to address the effect’s interaction with context. The experiment is described below.

Method

Subjects

A total of 33 young adults participated in the study for payment.

Materials

Participants took part in two experimental sessions on two separate days. In each experimental session, subjects read a total of 120 sentences presented via a moving-window self-paced reading paradigm. Each sentence was followed by a yes/no comprehension question presented in its entirety. Of the 120 sentences in each session, 40 were critical items and 80 were fillers.

The critical items included verbs such as *warned* that had the potential to create an ambiguity, as illustrated in Examples 1a–1d above. For each critical sentence, the first verb could plausibly be the main verb (MV), as in Example 1a, or the beginning of a reduced relative clause (RC), as in Example 1c. For all critical items, the sentence was ultimately disambiguated in favor of the RC reading. The items used in this study were different from those used in Fine et al. (2013).

Forty verbs were selected to create the critical items. Two items were created for each verb, so that subjects would read 40 unique verbs on Day 1 and then read the same 40 verbs again on Day 2, in different sentences. This resulted in 80 critical items. An ambiguous and an unambiguous sentence were created for each critical item, resulting in 160 critical sentences. For the ambiguous sentences, the initial verb was temporarily ambiguous between an MV and an RC interpretation, and the ambiguity was not resolved until readers reached the second verb in the sentence (*conducted*, in Example 1c). In the unambiguous sentences, the ambiguity was avoided by using a nonreduced relative clause, as in Example 1d. In each experimental session, half of the critical items the subjects read were temporarily ambiguous, and half were unambiguous.

Filler sentences were created that did not include a potential MV/RC ambiguity. The 160 critical sentences and 160 filler sentences resulted in a total of 320 sentences to be used across the entire experiment. However, since each subject encountered only half of the critical items, each subject read only 240 unique sentences across the entire experiment.

Procedure

Participants completed two experimental sessions, reading 40 unique critical sentences in each session, half of which were RC-Ambiguous and half of which were RC-Unambiguous. The sessions were separated by approximately 24 hours. Participants completed the first session in either Context A (Room A with Experimenter A) or Context B (Room B with Experimenter B). The second session was performed either in the same context as the first or in the other context, and this was crossed with initial context. Thus, the experiment had a 2×2×2 factorial design, with day (1 or 2), context (same or different), and ambiguity (ambiguous or unambiguous) as factors.

Context A Room A was a standard lab room within a larger lab space in the basement of the Psychology building at the University of Illinois Urbana-Champaign. Experimenter A was a Black female in her mid 20s.

Context B Room B was a soundproof recording booth within a small room off of the main corridor on the 7th floor of the Psychology building. The keyboard for the subjects' computer was within the sound booth, and the monitor of the computer was just outside of the booth on the other side of a small window. Experimenter B was a White male in his late 20s.

Analyses

Three models were used to analyze three measures of interest. The dependent measure in all models was the mean length-corrected reading time in the disambiguating region of the critical sentences, defined as the main verb and all subsequent words excluding the final word (*conducted the midnight*, in Examples 1c and 1d above). All analyses were carried out using multilevel mixed-effects regression, executed in R with the lme4 package. Where applicable, the contrasts for fixed effects were set to –1 and 1.

The first question of interest was whether the ambiguous sentences would take longer to read than the unambiguous sentences. To answer this question, the results from both days were analyzed by predicting the mean length-corrected reading times, with time, ambiguity, item order, and the interactions between these variables as fixed effects. The maximal random effects justified by the data were included.

The second question of interest was whether we would find an adaptation effect, which would be evidenced by an Ambiguity × Trial Order interaction, such that the ambiguity effect would diminish as trial order increased. This model predicted the mean length-corrected reading times on Day 1, with room location, ambiguity, item order, and the interactions between these variables as fixed effects. The maximal random effects justified by the data were included.

The last question of interest was whether context would modulate adaptation. If readers use the physical context in which they encounter language to track exposure to syntactic structures (and adjust their expectations about what structures they are likely to encounter), we should observe a three-way Ambiguity × Trial Order × Context interaction on Day 2, such that the adaptation effect would be larger for subjects in the same context than in the different context. This model predicted the mean length-corrected reading times on Day 2, with context (same or different), ambiguity, item order, and the interactions between these variables as fixed effects. The maximal random effects justified by the data were included.

Results and discussion

Figure 1 shows the mean uncorrected reading times across all regions of the sentence for the ambiguous and unambiguous conditions, collapsed over day and context. There was an overall effect of ambiguity in the critical region, such that reading times were slower in the ambiguous sentences. This effect held across both days (Day 1: $\beta = 19.06$, $SE = 5.77$, $p < .001$; Day 2: $\beta = 14.40$, $SE = 6.28$, $p < .05$). This replicated previous work (e.g., Ferreira & Clifton, 1986; MacDonald, Just, & Carpenter, 1992; Rayner et al., 1983).

Trial order did not significantly modulate the effect of ambiguity on either day (Day 1: $\beta = -0.45$, $SE = 0.45$, $p = .32$; Day 2: $\beta = -0.16$, $SE = 0.38$, $p = .67$), failing to replicate previous findings that had shown reduction in ambiguity effects across experimental trials (Fine & Jaeger, 2016; Fine et al., 2013).

Context on Day 1 did not predict reading times ($\beta = 4.73$, $SE = 5.12$, $p = .36$), which suggests that the context manipulation did not produce any unintended baseline differences. Of critical importance in the present study was whether context (same or different) predicted reading effects on Day 2. Given that we found no evidence for adaptation on Day 1, unsurprisingly, there was no evidence of an Ambiguity \times Trial Order \times Context interaction on Day 2 ($\beta = 0.24$, $SE = 0.38$, $p = .67$). However, subjects' overall speed-up across trials was greater on Day 2 than on Day 1 (i.e., a Day \times Trial Order interaction; $\beta = 1.18$, $SE = 0.23$, $p < .001$), and while the ambiguity effect was numerically smaller on Day 2, the Ambiguity \times Day interaction was not significant ($\beta = 2.55$, $SE = 4.03$, $p = .53$).

Thus, we failed to find evidence that context interacts with the rate of adaptation. Of course, a null result can occur for any number of reasons. One possibility is that the present study simply did not have the power to reliably detect adaptation effects, with only 33 subjects. In addition, Experiment 1 contained a higher filler-to-critical trial ratio than had the original Fine et al. (2013) study, which might have made it more difficult for subjects to track RCs as they progressed through the experiment. We therefore ran a power analysis on Fine et al.'s Experiment 2, with the goal of estimating the approximate effect size for the adaptation to MV/RC ambiguities. We outline this analysis below.

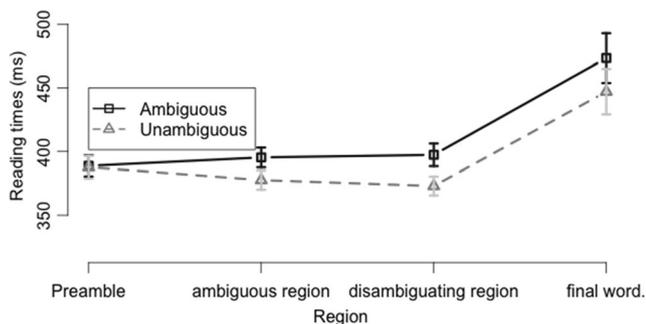


Fig. 1 Mean by-region reading times. The reading times at *who were* in the unambiguous sentences are not shown in this figure. Error bars give the 95% confidence intervals of the means

Power Analysis

Introduction

The overall strategy was to estimate the effect size of the critical comparisons in Fine et al. (2013). We then conducted power simulations to estimate the numbers of items and subjects that would be necessary to attain 95% power in a similar experiment. Below we describe the details of this analysis.

Data

The original data from Experiment 2 of Fine et al. (2013) provided the basis for the power simulations. The subjects in Fine et al.'s experiment were assigned to either the treatment condition or the control condition. In the treatment condition, they were presented with sentences in three blocks. In the first block, subjects were exposed to the more difficult reduced relative clause structures. In the second block, they again read reduced relatives, and Fine et al. found that the size of the ambiguity effect was reduced. In the third block, subjects read ambiguous sentences with MV continuations. Fine et al. found that subjects experienced more difficulty with these structures in Block 3, presumably because they had been relatively less frequent than RC structures within the experiment. The control condition differed from the treatment condition only in that Block 1 consisted of nonambiguous filler sentences. Because these subjects were not exposed to the more difficult RC structures in Block 1, they exhibited larger ambiguity effects in Block 2 and no increased difficulty in reading ambiguous sentences with MV continuations in Block 3.

In total, 77 subjects contributed reading time data to the present analyses, and the critical items were divided into three blocks. Subjects were assigned to either the filler-first group ($n = 38$) or the RC-first group ($n = 39$). The items differed between groups only in Block 1; the RC-first group read 16 RC sentences, whereas the filler-first group read 16 filler sentences. In Blocks 2 and 3, all subjects read ten critical sentences, which were RC sentences (Block 2) and MV sentences (Block 3). In each block, half of the critical sentences were ambiguous and half were unambiguous. The dependent measure was the length-corrected reading time for each of the three words in the disambiguating region.

Replication of the regression analysis

The data were run through the analysis procedure reported in the original article, with one critical difference. Fine et al.'s (2013) original analysis included items that the subjects had answered incorrectly. However, the

central question of both Fine et al.’s and the present study was whether subjects comprehend low-frequency sentences better as they encounter them more frequently over time. If the questions are not answered correctly, it is impossible to know whether comprehension has improved. Thus, the regression analysis and the subsequent power analysis that we present below are based on only correct trials. These regression models were run to address each of the three research questions in the experiment. The present analyses were run in RStudio (version 0.98.1103) using the lme4 package (version 1.1-7). Where applicable, the contrasts for factors were set to 1 and – 1.

Model 1: Is there a larger ambiguity effect on MVs for the RC-first group in Block 3? The first regression model only examined Block 3, resulting in 2,152 total data points (77 subjects × 10 items in Block 3 × 3 words, minus trials dropped for inaccuracy or reaction time [RT] outliers). Length-residualized reading times were predicted, with ambiguity condition (ambiguous or unambiguous), group (RC- or filler-first), and their interaction as fixed effects. The original model from Fine et al. (2013) included the maximal random-effects structure that allowed the model to converge; in this case, this included a random intercept and an ambiguity slope by subjects, as well as a random intercept by items. The model structure is given below:

$$RT \sim \text{Ambiguity} * \text{Group} + (1 + \text{Ambiguity} | \text{Subject}) + (1 | \text{Item})$$

The results of the model, and those reported in the original article, are presented in Table 1. The beta estimates in Fine et al. (2013) were rounded to the nearest whole number. As can be seen in Table 1, when incorrect trials were excluded from the analysis, the Ambiguity × Group interaction was no longer significant.

Table 1 Estimated regression coefficients for Model 1 in the present replication and as reported in the original study (Fine et al., 2013)

	Replication With Original Data (Correct Trials Only)		Fine et al. (2013) (All Trials)	
	Beta	p Value	Beta	p Value
Ambiguity	8.17	.003	8	< .05
Group	3.55	.32	4	.3
Ambiguity × Group	3.92	.14	5	< .05

Model 2: Does the ambiguity effect for the RC-first group diminish from Block 1 to 2? This regression model looked only at data from the RC-first group, resulting in 2,735 total data points (39 subjects × [16 + 10] items in Blocks 1 and 2 × 3 words, minus trials dropped for inaccuracy or RT outliers). Length-residualized reading times were predicted, with ambiguity condition (ambiguous or unambiguous), block (1 or 2), and their interaction as fixed effects. Following the original analysis strategy, block was treated as a continuous variable and was then centered, effectively creating a weighted contrast that accounts for the greater number of trials in Block 1. The original model from Fine et al. (2013) included the maximal random-effects structure that allowed the model to converge; in this case, this included a random intercept, ambiguity slope, and block slope by subjects, and a random intercept and ambiguity slope by items. The model structure is given below:

$$RT \sim \text{Ambiguity} * \text{Block} + (1 + \text{Ambiguity} + \text{Block} | \text{Subject}) + (1 + \text{Ambiguity} | \text{Item})$$

The results of the model, and those reported in the original article, are presented in Table 2. The beta estimates in Fine et al. (2013) are rounded to the nearest whole number. The original analysis and reanalysis were largely similar.

Table 2 Estimated regression coefficients for Model 2 in the present replication and as reported in the original study (Fine et al., 2013)

	Replication With Original Data (Correct Trials Only)		Fine et al. (2013) (All Trials)	
	Beta	p Value	Beta	p Value
Ambiguity	20.07	.00	20	< .05
Block	– 63.05	.00	– 63	< .05
Ambiguity × Block	– 9.17	.23	– 9	.2

Model 3: In Block 2, is the ambiguity effect for RCs smaller for the RC-first group than the Filler-first group? This regression model examined the data from all subjects in Block 2, resulting in 2,000 data points (77 subjects × 10 items in Block 2 × 3 words, minus trials dropped for inaccuracy or RT outliers). Length-residualized reading times were predicted, with group (RC- or filler-first) and ambiguity condition (ambiguous or unambiguous) as fixed effects. Following Fine et al. (2013), random slopes and condition intercepts were included for both subjects and items. The model structure is given below:

$$RT \sim \text{Group} * \text{Ambiguity} + (1 + \text{Ambiguity} | \text{Subject}) + (1 + \text{Ambiguity} | \text{Item})$$

The results of the model, and those reported in the original article, are presented in Table 3. The beta estimates in Fine et al. (2013) are rounded to the nearest whole number. Again, the reanalysis and original analysis yielded similar results.

Reanalysis results The results of the reanalysis that included only trials on which questions were answered correctly largely matched the original analysis. They diverged only for Model 1: when only correct trials were analyzed, there was no longer a significant Ambiguity × Group interaction, suggesting that the RC first group did not differ from controls in reading times for MV continuations. It is possible that the influence of the inclusion of incorrectly answered trials in the analysis outcome was theoretically meaningful. Therefore, in Experiment 2 we will discuss analyses that include and exclude incorrect trials.

Power simulation The results of the reanalysis of correctly answered trials were used to estimate the sample size needed to run a new experiment of a similar nature. Statistical power was estimated through simulation. The analysis code was

based on an R script generated by the MLPowSim software package (Browne, Lahi, & Parker, 2009). With this software, after simulated data are produced, regression models equivalent to those in the original analyses are run. Finally, power is calculated in two ways. First, the zero/one method calculates power as the proportion of times that the null hypothesis (that the estimated beta is zero) is rejected after 1,000 simulations. Second, the standard error method estimates power by first averaging the standard errors for the estimated effects after 1,000 simulations and then plugging that average as the SE into the following formula:

$$z_{\text{power}} = (\text{beta} / \text{SE}) - z_{\text{critical}}$$

where *beta* is the prespecified coefficient of interest (e.g., 3.92 for the interaction in Model 1), *z_{critical}* is the standard z-value associated with the specified significance level (set to *z_{critical}* = 1.64 and *α* = .05), and *z_{power}* is the estimated power. This procedure is repeated for the specified numbers of subjects and items in the simulated data. Separate power analyses were run for each of the three models of interest.

Power analysis For each of the three models, the coefficient of primary theoretical interest was the interaction effect. The estimated power for each of the three interaction effects, plotted across combinations of subjects and items, is shown below.

Results and discussion

The results of the simulated power analysis suggest that to achieve sufficient power, a replication of Fine et al. (2013) would require more items and subjects than in the original experiment. Fine et al. included 77 subjects. The curves marked with circles indicate the numbers of items used in Fine et al. To achieve 80% power in Model 1, which modeled power for the Group × Ambiguity interaction in Block 3, almost 480 subjects would need to be run (see Fig. 2). In Model 2, which modeled power for the Block × Ambiguity interaction for the RC-first subjects, approximately 280 subjects would need to be run (see Fig. 3). Finally, in Model 3, which modeled power for the Group × Ambiguity interaction in Block 2, 360 subjects would be needed (see Fig. 4).

Because the power analysis showed that a new experiment should be run with more subjects and items than Fine et al. (2013) had originally included, a higher-powered near replication of Fine et al.’s study was conducted. This near replication was designed to reach 95% power (Model 1). To reach this goal, the number of critical items used by Fine et al. was doubled in order to lower the number of subjects needed to reach adequate power, as shown by the lines marked with squares in Figs. 2, 3, and 4. In the end, 72 critical items and 423 subjects were used. Importantly, when the power simulation was run with data that included incorrect responses

Table 3 Estimated regression coefficients for Model 3 in the present replication and as reported in the original study (Fine et al., 2013)

	Replication With Original Data (Correct Trials Only)		Fine et al. (2013) (All Trials)	
	Beta	<i>p</i> Value	Beta	<i>p</i> Value
Group	− 6.87	.06	− 7	< .05
Ambiguity	19.84	.01	19	< .05
Group × Ambiguity	− 5.60	.09	− 5	.08

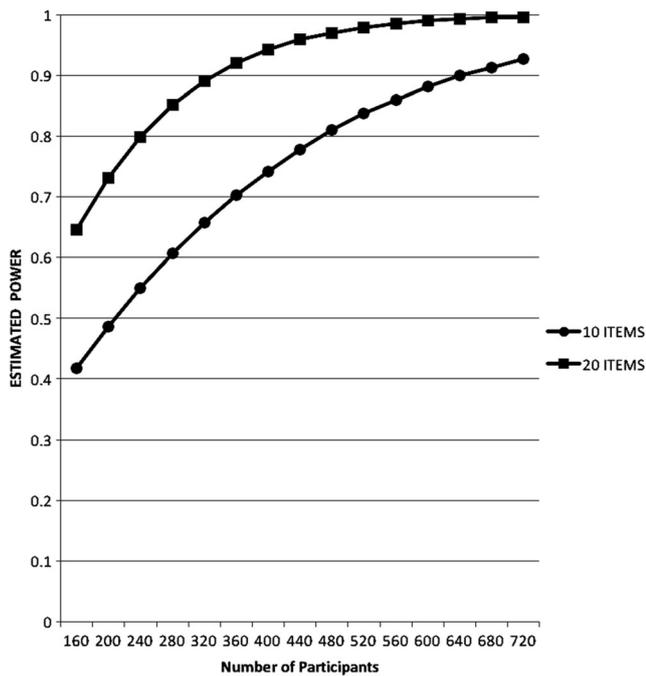


Fig. 2 Model 1: Group \times Ambiguity interaction effect in MV sentences, Block 3. The numbers of subjects needed (x -axis) to reach estimated power levels (y -axis) are shown for both the number of items used in the original Fine et al. (2013) study (circles) and if the items were doubled (squares)

(which matched Fine et al.'s, 2013, data inclusion criteria), we found that the number of subjects needed to achieve 95% power in Model 1 was 280 subjects if the number of items were doubled. Given this, the numbers of critical items and

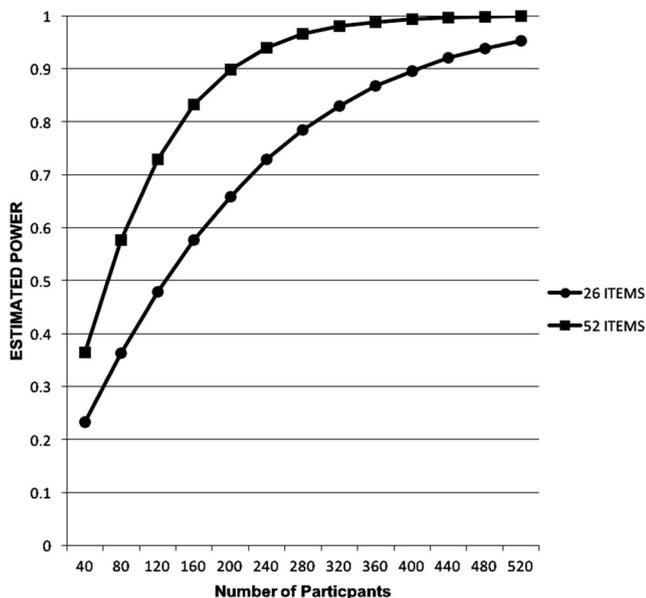


Fig. 3 Model 2: Block \times Ambiguity interaction effect for the RC-first group across Blocks 1 and 2. Numbers of subjects needed (x -axis) to reach estimated power levels (y -axis) are shown for both the number of items used in the original Fine et al. (2013) study (circles) and if the items were doubled (squares)

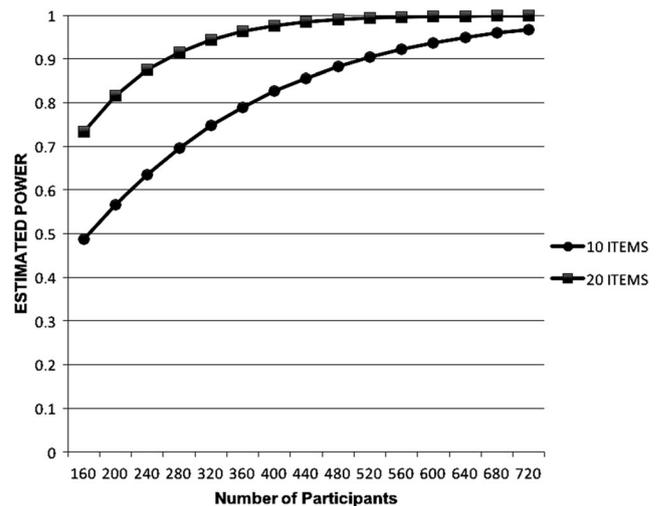


Fig. 4 Model 3: Group \times Ambiguity interaction effect in RC sentences, Block 2. The numbers of subjects needed (x -axis) to reach estimated power levels (y -axis) are shown both for the number of items used in the original Fine et al. (2013) study (circles) and if the items were doubled (squares)

subjects in Experiment 2 should provide enough power to detect an effect, regardless of whether the data are analyzed with or without the incorrectly answered items included. The replication is discussed below.

Experiment 2

Method

Subjects Fine et al. (2013) had recruited 80¹ subjects from the University of Rochester. Subjects were paid \$10 for their participation.

In the present replication, 481 American subjects were recruited via Amazon Mechanical Turk. Of these, 58 were excluded due to experimenter error (i.e., a mistake in the consent form). In total, data from 423 subjects were analyzed, with 210 subjects in the RC-first group and 213 subjects in the Filler-first group. The subjects were paid \$4 for completing the experiment.

Materials Fine et al. (2013) had modified the sentences from MacDonald et al. (1992). All critical sentences were either RC or MV sentences, as can be seen in Examples 2a–2d below (repeated from Example 1). Half of all the critical items were ambiguous, as in Examples 2a and 2c, in which the verb *warned* is temporarily ambiguous as to whether the sentence will be resolved as RC or MV. The other half of the items were

¹ Our power analysis analyzed data from 77 subjects on the basis of the raw data provided by Fine et al. (2013). However, Fine et al. (2013) reported recruiting 80 subjects for their study and we report that number here.

either unambiguous RC sentences, as seen in Example 2d, or MV sentences, as seen in Example 2b.

- (2a) MV-Ambiguous: The experienced soldiers warned about the dangers before the midnight raid.
- (2b) MV-Unambiguous: The experienced soldiers spoke about the dangers before the midnight raid.
- (2c) RC-Ambiguous: The experienced soldiers warned about the dangers conducted the midnight raid.
- (2d) RC-Unambiguous: The experienced soldiers who were told about the dangers conducted the midnight raid.

In total, subjects read 71 sentences over three blocks. In Blocks 1 and 2, subjects read only RC and filler items. In Block 3, subjects read only MV and filler items. The verbs in the RC sentences in Block 1 were unique. However, all of the RC sentences in Block 2 included verbs that overlapped with five verb pairs used in Block 1. The ambiguous MV sentences in Block 3 included the same verbs that had been used in Block 2, but the unambiguous MV sentences had entirely unique verbs. Filler sentences were created so that they did not contain verbs that overlapped with those used in the MV/RC ambiguity manipulation.

The materials used in this article's replication were modified from those of Fine et al. (2013), as is discussed below. To double the items to reach .95 power as indicated by the power analysis, additional critical and filler items were constructed. Additionally, the RC items were modified so that the unambiguous condition used the same verbs as the ambiguous RC items, as in Example 4b below. This differs from the construction of the Fine et al. (2013) items, in which the yoked ambiguous and unambiguous items included different verbs. This was done to reduce variability across conditions. Additionally, the critical items had variable disambiguating regions, as opposed to those of Fine et al., in which each critical sentence had a disambiguating region of three words. However, analyses of the Experiment 2 results did not differ when we analyzed the entire length of the variable disambiguating region as compared to analyzing just the first three words after the ambiguous region, so we report the former analysis below.

- (3a) MV-Ambiguous: The aging professors warned about the midterm just before fall break.
- (3b) MV-Unambiguous: The aging professors spoke about the midterm just before fall break.
- (4a) RC-Ambiguous: Several angry workers warned about low wages decided to file complaints.
- (4b) RC-Unambiguous: Several angry workers who were warned about low wages decided to file complaints.

In total, the subjects read 142 sentences over three blocks. In Blocks 1 and 2, subjects read only RC and filler items. In

Block 3, subjects read only MV and filler items. Sixteen verbs were presented to the subjects who read RC sentences in Block 1. Each of these verbs was seen twice by a subject within Block 1, once as an ambiguous RC and once as an unambiguous RC, but never as an MV, and always in differing items. In Block 2, all RC sentences included verbs that overlapped with ten verb pairs used in Block 1. Again, subjects saw each verb twice, once as an ambiguous RC and once as an unambiguous RC. Finally, in Block 3, the ten ambiguous MV sentences used the same verbs that had been used in Block 2, whereas the ten unambiguous MV sentences included unique verbs. Filler sentences were created so that they did not contain verbs that overlapped with those used in the MV/RC ambiguity manipulation.

Two counterbalanced experimental lists were created for the control and experimental groups, so that the ambiguity of items was counterbalanced between subjects. The lists were counterbalanced using a Latin square design, such that the sentences that were ambiguous in List 1 were unambiguous in List 2, and vice versa. Half of the critical items were ambiguous and half were unambiguous for all subjects.

Item errors After running the replication, errors were discovered in 19 sentences (14 fillers and five critical items). Seventeen of these errors were due to discrepant items across lists, in which sentences with slight variations in wording were used across different lists. The remaining two errors were grammatical. Of the 19 items affected, five were critical items. One of these critical items was an RC that occurred in Block 2. The other four errors were MV sentences found in Block 3. The analyses presented below yield the same results, regardless of whether or not these items were included. We present the data with these items excluded, although we refer to the full item set when discussing the experimental design.

Procedure Fine et al. (2013) had randomly assigned subjects to either the RC-first group or the Filler-first group. The experiment was split into three blocks, but from the perspective of the subject it was one continuous experiment. The subjects in the RC-first group read 16 RCs in Block 1, ten RCs and 20 fillers in Block 2, and ten MVs and 15 fillers in Block 3. The subjects in the Filler-first group read 16 fillers in Block 1, and otherwise their blocks were identical to those of the RC-first group. Subjects read sentences in the lab in a word-by-word self-paced reading task. Each trial began with a series of dashes representing all non-space characters on the screen. Subjects were instructed to press the spacebar in order to view each word. After a word was read, subjects moved to the next word, which turned the prior word back into dashes. The duration between spacebar presses was recorded. Each sentence was followed by a yes/no comprehension question, to which “yes” was the correct answer half of the time.

In the present replication, the task was a word-by-word self-paced reading task hosted online on Ibex Farm. The task itself was otherwise identical to that used in Fine et al. (2013). Prior to beginning the experiment, subjects were given instructions about how to read the sentences by pressing the spacebar. They then received two practice sentences followed by two comprehension questions, to ensure that they had practice with the experimental design.

The experiment was divided into three blocks, but from the subject's viewpoint it was one continuous block. In total, subjects read 144 sentences, including the two practice items. Subjects were randomly assigned to one of four lists, which were generated by counterbalancing the ambiguity conditions. Two of the lists were for the Filler-first conditions, and two were for the RC-first conditions. The RC-first group read 32 RC sentences in Block 1. The filler-first group read 32 filler items instead. Blocks 2 and 3 were identical for the two groups in terms of the number and type of items that were read. Block 2 consisted of 20 critical items that had RC readings. Block 2 also included 40 filler items. Block 3 consisted of 20 MV sentences and 30 fillers. Item order in the lists was pseudorandomized such that at least one filler was interposed between critical items within each block.

Analysis and results

Fine et al. (2013) conducted three analyses investigating syntactic adaptation. These analyses were referred to as Questions 1, 2, and 3. Question 1 asked whether the MV ambiguity effect in Block 3 was larger for the RC-first group than in the Filler-first group. Question 2 asked whether the RC ambiguity effect for the RC-first group was reduced from Block 1 to Block 2. Question 3 asked whether the ambiguity effect in Block 2 for the RC-first group was smaller than that for the Filler-first group.

The same analyses are presented below for comparison. The results of all three analyses are presented visually in Fig. 5 and summarized in Table 4. In each analysis, Fine et al.'s (2013) results are discussed in detail before the results of the present replication are presented. In Fine et al.'s study, reading times below 100 ms and above 2,000 ms were excluded. The data exclusion in the replication was based on similar criteria. As we discussed in the reanalysis of Fine et al.'s data, the original analysis included items that subjects had answered incorrectly. The numbers and figures presented below for our analysis exclude items that were answered incorrectly. However, a post-hoc analysis that included incorrectly answered items revealed no difference in outcomes, so we present the data from only correctly answered trials. Both Fine et al. and the present replication used length-corrected RTs as the dependent measure. This measure was obtained by regressing the raw RTs onto word length as well as by including a by-subject random slope of letter count and a random

intercept for each subject. All analyses discussed below examine the disambiguating region of each sentence. For the purpose of our analyses, it was assumed that $p < .05$ when t values were greater than 2, following Baayen (2008). The results of all three analyses are summarized in Table 4.²

Question 1—Effect of adaptation on the dispreferred structure Question 1 investigated whether the ambiguity effect in Block 3 was larger for the RC-first group than for the Filler-first group. A larger ambiguity effect for the RC-first group would suggest that increased experience with RCs early in the experiment resulted in difficulty reading MVs later in the experiment. Importantly, previous research had demonstrated that readers have an easier time reading MVs than RCs (Ferreira & Clifton, 1986; MacDonald et al., 1992; Rayner et al., 1983). If it were the case that the RC-first group experienced greater difficulty when processing MVs, it would suggest that an a priori expected structure can become unexpected, and therefore more difficult to process.

Fine et al. (2013) predicted length-corrected RTs, with ambiguity, group, and their interaction as fixed effects. The maximal random effects justified by the data were included. Fine et al. found a main effect of ambiguity ($\beta = 8, p < .05, t = 3$), in which ambiguous MVs were read more slowly than unambiguous MVs. They found no main effect of group ($\beta = 4, p = .3, t = 1$). Crucially, Fine et al. found a two-way interaction between ambiguity and group ($\beta = 5, p < .05, t = 2$). The subjects in the RC-first group showed a larger ambiguity effect for MVs than did the subjects in the Filler-first group. Fine et al. concluded that repeated exposure to RCs resulted in a processing cost for MVs.

We ran the same analysis as had Fine et al. (2013). There was no main effect of ambiguity in the present replication ($\beta = 1.28, p > .05, t = 1.58$). Subjects spent similar amounts of time reading the ambiguous and unambiguous MV sentences. We also observed no main effect of group; the subjects in both groups spent similar amounts of time reading the ambiguous and unambiguous MV sentences ($\beta = 1.69, p > .05, t = 1.22$). Crucially, there was also not a significant interaction between ambiguity and group ($\beta = -0.24, p > .05, t = -0.3$). This suggests that the RC-first group did not experience more difficulty reading ambiguous MV sentences than did the Filler-first group.

Question 2—Comparison of ambiguity effects across blocks in the experimental group Question 2 examined whether the ambiguity effect was reduced from Block 1 to Block 2 for the RC-first group. If the ambiguity effect were reduced for the

² Raw reading times are not included because they showed a similar pattern. The de-identified raw reading times are available online at the Open Science Framework, at <https://osf.io>.

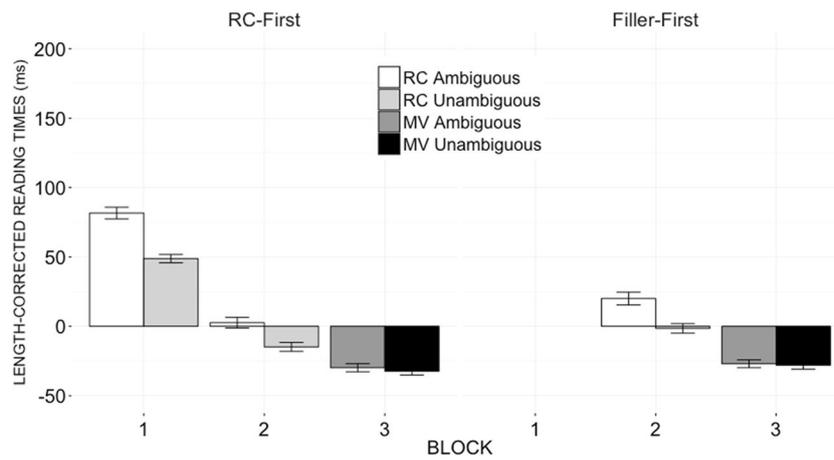


Fig. 5 Results of Experiment 2: Mean length-corrected reading times (y-axis) for both the Filler-first and RC-first groups by block (x-axis)

RC-first group, it would suggest that increased exposure to RCs resulted in an easier time processing this structure.

For this analysis, Fine et al. (2013) predicted length-corrected RTs, with ambiguity, block, and their two-way interaction as fixed effects. The model also included the maximal random-effects structure justified by the data.

Fine et al. (2013) found a significant effect of ambiguity, so that ambiguous RCs were read more slowly than unambiguous RCs ($\beta = 20, p < .05, t = 4$). They also observed a significant main effect of block in which subjects read the sentences in Block 2 more quickly than those in Block 1 ($\beta = -63, p < .05, t = -4$). Finally, they found that the interaction between these two factors was in the predicted direction but did not reach significance ($\beta = -9, p = .2, t = -1.3$). Fine et al. argued that this lack of significance was most likely due to the reduced power that resulted from grouping item order into two blocks, rather than including item order as a continuous variable. To correct for this, a new analysis was run in which the length-corrected RTs for the disambiguating region were

predicted from ambiguity, item order, the interaction between ambiguity and item order, and the log stimulus order. The model also included the maximal random-effects structure justified by the data. Fine et al. now found main effects of ambiguity ($\beta = -39, p < .05$) and log stimulus order ($\beta = -176, p < .05$). Crucially, there was also a significant Ambiguity \times Item Order interaction ($\beta = 2, p < .05$). They concluded that the ambiguity effect in Block 2 was reduced for subjects in the RC-first group due to their increased experience with RC sentences.

The analyses run in the present experiment were the same as those used in Fine et al. (2013). We found a main effect of ambiguity, in which unambiguous sentences were read more quickly than ambiguous ones ($\beta = 14.19, p < .05, t = 6.64$). This was expected, since it replicated the garden-path effect found in previous research on RC sentences (Ferreira & Clifton, 1986; MacDonald et al., 1992; Rayner et al., 1983). There was also a main effect of block, in which subjects read sentences more quickly in Block 2 than in Block 1 ($\beta = -75.95, p < .05, t = -8.28$), which was likely due to increased familiarity with the task. A reliable Block \times Ambiguity interaction also emerged ($\beta = -9.82, p < .05, t = -2.57$).

This analysis replicated one of the findings from Fine et al. (2013). The ambiguity effect appears to get smaller as subjects in the RC-first group proceed through the experiment. However, this by itself is not necessarily evidence for adaptation. It could be that the size of the ambiguity effect interacts with the task, such that as the task becomes more familiar, complexity effects are smaller. This question could be addressed by comparing the reduction in size of the ambiguity effect in the RC-first group to the size of the ambiguity effect in the Filler-first group, who had equal experience with the task (but not with relative clauses). Thus, we turned to Question 3.

Question 3—Comparison of ambiguity effects across groups in Block 2 Question 3 examined whether the ambiguity effect

Table 4 Summary of Experiment 2 results

	Beta	t Value	p Value
Question 1			
Ambiguity	1.28	1.58	> .05
Group	1.69	1.22	> .05
Ambiguity \times Group	-0.24	-0.3	> .05
Question 2			
Ambiguity	14.19	6.64	< .05
Block	-75.95	-8.28	< .05
Ambiguity \times Block	-9.82	-2.57	< .05
Question 3			
Ambiguity	9.14	4.65	< .05
Group	7.55	4.46	< .05
Ambiguity \times Group	1.25	1.05	> .05

Significant results are in bold

in Block 2 was greater for the Filler-first group than for the RC-first group. Fine et al. (2013) predicted length-corrected RTs from ambiguity, group, and the interaction between the two. The model included the maximal random-effects structure justified by the data. Fine et al. found a main effect of ambiguity ($\beta = 19, p < .05, t = 3$), in which unambiguous RCs were read more quickly than ambiguous RCs. There was also a main effect of group, in which the subjects in the RC-first group had overall faster reading times ($\beta = -7, p < .05, t = -2$). The two-way interaction between ambiguity and group was marginally significant in the expected direction ($\beta = -5, p = .08, t = -1.7$). That is, the ambiguity effect in Block 2 was larger for the Filler-first group than for the RC-first group, suggesting that increased experience with RC sentences decreased the processing difficulty for the RC-first group.

We conducted an analysis similar to that of Fine et al. (2013), but the by-item random slope of ambiguity was excluded in order for the model to converge. We observed a main effect of group, such that the RC-first group read both ambiguous and unambiguous sentences more quickly than the filler-first group ($\beta = 7.55, p < .05, t = 4.46$). There was also a main effect of ambiguity in which unambiguous sentences were read more quickly than ambiguous sentences in both groups ($\beta = 9.14, p < .05, t = 4.65$). The crucial interaction between ambiguity and group was not significant ($\beta = 1.25, p > .05, t = 1.05$), suggesting that the subjects in the RC-first group and the Filler-first group experienced similar difficulties with reading RCs, despite the RC-first group having more experience with that structure.

Thus, these data suggest that, at least for this structure, the size of the ambiguity effect does not seem to vary with the amount of exposure to reduced relative clauses. The apparent reduction in the size of the effect seen under Question 2 does not appear to be related to exposure to relative clauses, since an ambiguity effect of similar magnitude was seen in the control condition.

General discussion

Fine et al. (2013) reported evidence of adaptation to sentences containing reduced relative clauses. They found that the ambiguity effect for RCs was significantly reduced after reading only 16 RC sentences. Furthermore, after reading 26 RC sentences, their subjects had more difficulty processing MV sentences. They concluded that readers rapidly learned the frequencies of the two structures across the experiment, which in turn facilitated RC processing and inhibited MV processing.

However, in the set of experiments presented above, we failed to find evidence of adaptation using a very similar paradigm. There was no evidence of adaptation in Experiment 1. A reanalysis of Fine et al.'s (2013) data showed that the reversal of MV reading times was only present when incorrect trials

were included. A sample size calculation using Fine et al.'s original data revealed that an appropriately powered replication would require hundreds of subjects. Experiment 2 served as a near replication of Fine et al., with 423 subjects and 72 critical items. Unlike Experiment 1, Experiment 2 revealed a reduction in the size of the ambiguity effect for the RC-first group; however, we found no evidence that this reduction was the result of exposure to RCs, for there was no significant difference in the magnitudes of the ambiguity effect in the RC-first group and the control group. We also found no evidence that the RC-first group experienced increased difficulty with ambiguous main verb sentences relative to the controls in Block 3.

An initial concern may be that methodological differences were to blame for the differing results found in Experiment 2 and Fine et al. (2013). It could be that the population on Mechanical Turk was different in some important way from the undergraduate population tested by Fine et al. It is true that there were likely differences in the demographics of the two samples. In general, recruiting subjects from Mechanical Turk results in a more diverse sample than is normally found in university settings (Buhrmester, Kwang, & Gosling, 2011). However, there are reasons to think that this was not the cause of the discrepant results. First, in Experiment 1 we used university students as subjects and still failed to find evidence of adaptation, suggesting that these effects are difficult to find both in the lab and online. Second, prior research has suggested that data collected in the lab and on Mechanical Turk are similar for linguistic tasks such as cloze sentence completion and semantic similarity judgments (Schnoebelen & Kuperman, 2010). Within psycholinguistics, prior experiments have investigated adaptation using Mechanical Turk subjects without raising concerns (Buz, Tanenhaus, & Jaeger, 2016; Fine & Jaeger, 2013; Fraundorf & Jaeger, 2016; Gibson et al., 2017; Liu & Jaeger, 2018; Wittenberg & Levy, 2017).

Another methodological difference between the present study and Fine et al. (2013) is in how the critical verbs were distributed across items. In Fine et al. (2013), different verbs were used for the unambiguous and ambiguous RC sentences, whereas in our Experiment 2 subjects experienced the same verbs across conditions. This was done to reduce variability between the conditions, but it is possible that experiencing the critical verbs in unambiguous conditions diminished expectation for the same verbs in reduced relative contexts. Although this is a possible explanation, it does raise questions. For one thing, it makes the prediction that readers who have been exposed to verbs in either type of relative clause should expect those verbs to appear in more relative clauses in the future. Consequently, these readers should experience more difficulty processing main verb continuations that contain these critical verbs than do readers in the control condition. Yet we saw that in Block 3 of Experiment 2, readers did not slow down when

they encountered main verb continuations. This was affirmed in the reanalysis of the original Fine et al. study, which had used different verbs in the ambiguous and unambiguous conditions. Their data showed no increased difficulty for main verb sentences when we included only correct trials in the analysis. Another possibility is that experiencing a verb in the unambiguous context increased the difficulty of interpreting the same verb in the ambiguous context, which might have reduced our ability to detect adaptation effects. This raises the possibility that adaptation to syntactic structures is linked to their co-occurrence with specific lexical items, which is consistent with the previous literature (e.g., Garnsey et al., 1997; Snedeker & Trueswell, 2004). However, a problem with this approach is that when comprehenders encounter the same verbs in unambiguous contexts, the lack of ambiguity means that no error signal is generated. Thus, it is not clear why learning would occur when reading these sentences, and if it did, why this would diminish learning in the ambiguous sentences. We leave this question for future work.

A third potential concern is that the results in Experiment 2 were analyzed using different exclusion criteria than in Fine et al. (2013). Fine et al. included all trials in their analyses, whereas in Experiment 2 only correctly answered trials were included. As we discussed above, the results reported above for Experiment 2 did not change when all trials were included in a post-hoc analysis. One possible explanation for this discrepancy is that the significant interaction between group and ambiguity in Block 3 of Fine et al. was spurious. Of course, it is also possible that there are theoretically interesting reasons why the inclusion of incorrect trials in the analysis might reveal effects of adaptation, and that we simply did not have the power in the present study to detect it. However, our power analysis that included both correct and incorrect trials showed that to obtain 95% power with double the items, only 280 subjects were required. This suggests that our results were not due to a lack of power. We also leave this question for researchers to investigate in the future.

Another potential explanation for the null results in Blocks 2 and 3 is linked to the increased items in the present experiment as compared to Fine et al. (2013). By doubling the number of items in the replication, we increased the amount of exposure to the RC sentences in the control group in Block 2, effectively reducing the difference in RC expectancy across the two groups. Thus, by the time the subjects in the two conditions entered Block 3, the difference in expectations for MV sentences was not as great as it had been for Fine et al. Similarly, differences between the two groups in the magnitude of the ambiguity effect in Block 2 should have been reduced.

To address this last prediction, we ran a post-hoc analysis comparing the control group and the experimental group halfway through Block 2. This allowed for a comparison at the

point at which the experimental subjects had been exposed to 42 RCs and the control group had been exposed to only ten RCs. This analysis roughly matched Fine et al.'s (2013) Block 2 analysis, in which the experimental group had been exposed to 26 RC sentences, and the control group only to ten RC sentences. We found that although there was still a main effect of ambiguity ($\beta = -23.87$, $t = -3.94$, $p < .001$), there was no Group \times Ambiguity interaction ($\beta = 6.28$, $t = 1.04$, $p = .150$), suggesting that the failure to find evidence for adaptation in Block 2 was not simply due to doubling the items. Unfortunately, we cannot use a similar post hoc analysis to address this concern for Block 3 effects. Because all the subjects only saw MVs after reading at least 20 RCs, it was impossible to conduct an analysis comparable to the one above. This question can only be resolved by running a future study with the same number of items as Fine et al. It is worth mentioning that to achieve 95% power with the number of items used in the original study, more than 800 subjects would need to be run (according to the power analysis procedure described above, running 800 subjects would result in a power of .948 to detect the Group \times Ambiguity interaction). If syntactic adaptation in this structure exists but the effect size is so small that it can only be detected in a large- N study, the field would need to carefully consider how central syntactic adaptation really is to the language-processing system.

Another potential issue is that the resistance to adaptation seen in this experiment was limited to MV/RC ambiguities, perhaps because reduced relatives are infrequent or because the relative difference in frequency between main verb continuations and reduced relative continuations is so great. However, this explanation would be surprising, given the claims that have been made about language processing in general, and syntactic adaptation in particular. Many models of language processing incorporate prediction error as a way to explain how the language system changes its expectations over time (Chang, 2002; Chang, Dell, & Bock, 2006; Elman, 1990; Miikkulainen & Dyer, 1991; Rohde & Plaut, 1999). Syntactic adaptation theories, in particular, state that this error-driven recalibration occurs at a rapid pace and is especially sensitive to infrequent structures (e.g., Fine et al., 2013; Jaeger & Snider, 2013; Kleinschmidt et al., 2012; Kleinschmidt & Jaeger, 2015). The theory argues that less-frequent structures should generate a larger error signal, which in turn should result in a larger adjustment to the reader's statistics about the environment, which ultimately should result in larger adaptation effects. Thus, although it is possible that MV/RC ambiguities are unique in their resistance to adaptation, it is unlikely that this would be driven by frequency. If some feature of the MV/RC structure makes adaptation more difficult, it is not immediately clear what that feature might be.

One remaining puzzle is understanding why the magnitude of the ambiguity effect in Block 2 was smaller for both the RC-first and Filler-first groups than for the RC group in

Block 1. One possible explanation is that the change in magnitude was due to task demands. In Block 1, subjects had to learn the moving-windows task, and the unusual nature of the task may have amplified the complexity effects in language processing. Thus, for both controls and the experimental group, the size of the ambiguity effect becomes smaller as the subjects become more used to the task. Another possibility, which is related to the first, is that the magnitude of ambiguity effects varies across the RT scale. Because the interaction is removable (see Loftus, 1978) and the reading times in Blocks 1 and 2 are on different points on the RT scale, it is not clear that this interaction is a meaningful one; the same underlying processes may be at work, but they may not map onto RTs in the same way at different reading speeds. Whatever the reason, the fact that the effects are similar for both groups suggests that the attenuation of the effect is not due to adaptation, which highlights the importance of including a control group in this type of study.

Finally, one possible explanation for these results is that adaptation does, in fact, occur, but that it does not do so rapidly. There is evidence for this in the literature. Wells, Christiansen, Race, Acheson, and MacDonald (2009) showed a reduced ambiguity effect for RC sentences after subjects were exposed to 160 RC sentences over the course of three to four weeks. Fine and Jaeger (2011) found similar results in a multiday study. It is possible that the processing system does adapt but requires either a great deal of linguistic evidence or a longer time frame to integrate the new statistics.

Conclusions

Fine et al. (2013) proposed that listeners and readers rapidly adapt to the syntactic distributions in their environment. Although this may be the case, the data presented here suggest that for MV/RC ambiguities, a relatively brief exposure to a difficult structure is not sufficient for adaptation.

Author note This research was supported by the James S. McDonnell Foundation. We thank Alex Fine and Florian Jaeger for providing the data from the original Fine et al. (2013) experiment, providing stimuli to use in the replication, and helpful comments on the results.

References

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Browne, W. J., Lahi, M. G., & Parker, R. M. (2009). *A guide to sample size calculations for random effect models via simulation and the MLPowSim software package*. Bristol, UK: University of Bristol.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3–5. <https://doi.org/10.1177/1745691610393980>
- Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86.
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26, 609–651. https://doi.org/10.1207/s15516709cog2605_3
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234–272. <https://doi.org/10.1037/0033-295X.113.2.234>
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Farmer, T. A., Fine, A. B., Yan, S., Cheimariou, S., & Jaeger, F. (2014). Error-driven adaptation of higher-level expectations during reading. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2181–2186). Austin, TX: Cognitive Science Society
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368. [https://doi.org/10.1016/0749-596X\(86\)90006-9](https://doi.org/10.1016/0749-596X(86)90006-9)
- Fine, A. B., & Jaeger, T. F. (2011). Language comprehension is sensitive to changes in the reliability of lexical cues. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 33, 925–930.
- Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37, 578–591. <https://doi.org/10.1111/cogs.12022>
- Fine, A. B., & Jaeger, T. F. (2016). The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1362–1376.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8, e77661. <https://doi.org/10.1371/journal.pone.0077661>
- Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, 91, 28–58.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93. <https://doi.org/10.1006/jmla.1997.2512>
- Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't Underestimate the Benefits of Being Misunderstood. *Psychological Science*, 28, 703–712.
- Jaeger, T. F., & Snider, N. E. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In B. C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Cognitive Science Society* (pp. 1061–1066). Austin, TX: Cognitive Science Society.
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127, 57–83. <https://doi.org/10.1016/j.cognition.2012.10.013>
- Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In N. Miyake, D. Peebles, & R.P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 599–604). Austin, TX: Cognitive Science Society.
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122, 148–203. <https://doi.org/10.1037/a0038695>
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. (2014). Rapid adaptation in online pragmatic interpretation of

- contrastive prosody. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 791–796). Austin, TX: Cognitive Science Society
- Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*, 55–70. <https://doi.org/10.1016/j.cognition.2018.01.003>
- Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, *6*, 312–319. <https://doi.org/10.3758/BF03197461>
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703. <https://doi.org/10.1037/0033-295X.101.4.676>
- Miikkulainen, R., & Dyer, M. G. (1991). Natural language processing with modular PDP networks and distributed lexicon. *Cognitive Science*, *15*, 343–399.
- Myslin, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition*, *147*, 29–56.
- Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*, 4–18. <https://doi.org/10.1080/23273798.2015.1081703>
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.
- Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67–109. [https://doi.org/10.1016/S0010-0277\(99\)00031-1](https://doi.org/10.1016/S0010-0277(99)00031-1)
- Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, *43*, 441–464.
- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*, 238–299.
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318. <https://doi.org/10.1006/jmla.1994.1014>
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–271.
- Wittenberg, E., & Levy, R. (2017). If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of Memory and Language*, *94*, 254–271. <https://doi.org/10.1016/j.jml.2016.12.001>
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*, 648–672. <https://doi.org/10.1080/23273798.2014.995679>
- Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2013). Linguistic Variability and Adaptation in Quantifier Meanings. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3835–3840). Austin, TX: Cognitive Science Society.