Linguistic Research 42(1): 1-30
DOI: 10.17250/khisli.42.1.202503.001

Exploring reward effects in experimental syntax

Rui P. Chaves
(University at Buffalo, The State University of New York)

Chaves, Rui P. 2025. Exploring reward effects in experimental syntax. *Linguistic Research* 42(1): 1-30. Experimental Syntax research often involves participants reading or listening to disconnected sentences which are often unusual in some way. Such tasks are rather artificial and involve assigning numbers to sentences, or pressing keys to reveal the next word on a screen. Some participants may lack motivation to process these sentences in a typical way, given these unusual settings, which can lead to data that may not reflect normal language processing. Specifically online and offline measures of comprehension may be affected by not only task demands but also participant motivation. In the present study we manipulate the amount of reward for completing such tasks in order to examine how it impacts (if at all) the experimental outcomes. This is an important question to explore as there are currently no compensation standards in experimental linguistics, with some studies paying subjects with rates above the minimum wage, and others offering course credit instead of financial compensation. The present paper uses reward magnitude as a proxy to motivation to perform a task. The results suggest that reward incentives can impact outcomes, but only subtly so, at least for the populations tested in this study. Additionally, there is a significant degree of variation across studies, suggestive of participants sometimes deploying strategies to maximize the chances of completing the task appropriately, but in the process create artificial patterns in the data. **(University at Buffalo, The State University of New York)**

Keywords adaptation, satiation, reward, acceptability, English, self-paced reading, garden paths, islands

1. Introduction

Linguistic input noisy and ambiguous, requiring individuals to rely on probabilistic expectations to navigate uncertainty. Research has shown that this adaptive process occurs during the comprehension of challenging constructions, such as garden path sentences, and is likely a form of implicit learning. Reward mechanisms enhance

learning by promoting synaptic plasticity and motivating engagement. Experiments have demonstrated that increased rewards improve participants' motivation and strategic adaptation in language tasks. Despite its significance, participant compensation in psycholinguistic research remains inconsistent, with rates varying widely across studies. This inconsistency likely impacts motivation and learning outcomes, underscoring the need for standardized practices. The present exploratory work aims to provide a closer look at the effect of reward on sentence acceptability, as measured by a Likert scale study, and on on-line sentence processing, through the lens of self-paced reading studies. In Section 1 we provide an overview of past research on adaptation and reward, and in Section 2 we detail a number of experiments designed to measure how perceived reward affects linguistic processing. As we shall see, the effect is subtle, but measurable. This suggests that low motivation or reward can lead to failure to detect an effect or difference that actually exists (Type 2 error). On the other hand, too high a reward may also give rise to strategies that interfere with the natural processing of linguistic stimuli in an experimental context, including scenarios in which an effect is detected when none exists (Type 1 error), because of strategies that participants developed in order to secure a high compensation.

2. Background

It is uncontroversial that sensory input is noisy and ambiguous, and that individuals respond to the challenges created by uncertainty and variation by resorting to probabilistic expectations (Anderson 1990; Gigerenzer et al. 1999; Newell and Simon 1972). For example, infants already exhibit the ability integrate prior knowledge and expectations about human actions with new evidence provided by the environment (Xu and Kushnir 2013), and use new evidence to modify their expectations (Brandone et al. 2014). Differences in reward sensitivity and reward-directed behavior likely depend on the nature of rewarding stimuli, and biological and socio-cultural factors. For example, sex differences in learning have been observed in young primates (Bachevalier and Hagger 1991; Clark and Goldman-Rakic 1989; Goldman et al. 1974) and parallel those of children between 15 to 30 months of age (Overman et al. 1996). Although men and women tend to exhibit similar learning profiles–see Grissom and Reyes (2019) for a recent overview–there is evidence for slight differences. For example,

females are subtly more sensitive to frequent negative feedback than males (Ambrase et al. 2021; Lewis et al. 2023), slightly faster in learning to avoid punishment, and better than men in learning from positive (but not negative) feedback in probabilistic selection tasks (Evans and Hampson 2015). In the present studies we will control for such demographic factors.

Linguistic input is particularly noisy, ambiguous and variable across individuals and contexts, and speakers must adjust their expectations in order to create useful heuristic predictions. Indeed, it is known that comprehenders create expectations about upcoming words (Altmann and Kamide 1999; Arai and Keller 2013; Creel et al. 2008; DeLong et al. 2005; Kutas and Hillyard 1984; Metzing and Brennan 2003), about upcoming lexical categories (Gibson 2006; Levy and Keller 2013; Tabor et al. 1997), and about syntactic structures (Fine et al. 2013; Fine and Jaeger 2013; Fine et al. 2010; Kamide and Mitchell 1997; Lau et al. 2006; Levy 2008; Levy et al. 2012; MacDonald et al. 1994; Staub and Clifton Jr. 2006; Wells et al. 2009), among other modalities of input.

In particular, there is much work showing adaptation when processing garden path sentences, as measured by self-paced reading tasks (Farmer et al. 2014; Fine et al. 2013; Fine and Jaeger 2013; Fine et al. 2010; Prasad and Linzen 2019; Stack et al. 2018; Zervakis and Manuka 2013), and evidence for adaptation in island violations, as measured by sentence acceptability tasks (Chaves and Dery 2014, 2019; Chaves and Putnam 2020; Clausen 2011; Do and Kaiser 2017; Francom 2009; Goodall 2011; Hiramatsu 2000; Hofmeister 2015; Lu et al. 2021; Snyder 1994, 2000, 2021), and by priming and self-paced reading tasks (Chaves and Dery 2019; Do and Kaiser 2017). Adaptation has been detected after increased exposure to unusual American English vernacular constructions (Blanchette et al. 2024), as well as to moderately ungrammatical sentences, as measured be the attenuation of event-related potentials (Yano et al. 2021). Such adaptation his arguably a form of learning, which can be modeled in terms of Bayesian belief updating (Fine et al. 2013, 2010; Lu et al. 2021; Wells et al. 2009) or as error-based implicit learning using connectionist models (Chang et al. 2006, 2012). However, learning in humans is not driven purely by frequency, as intrinsic factors (curiosity, interest, or enjoyment of making the right prediction), and extrinsic factors (external rewards, or punishment) play a crucial role. It follows that syntactic adaptation should be sensitive to the predicted reward, not just to structural frequency and task adaptation. For example, Christianson et al. (2022)

argue participants in a psycholinguistic experiment might not be fully engaged or motivated to complete the tasks, potentially leading to unreliable data that does not accurately reflect typical language processing because their level of interest in the experiment is low. There is some experimental support for this view. In a self-paced reading task Chaves and Malone (2022) found presentation order evidence suggesting that increased reward led participants to disambiguate garden path sentences faster than those participants in a control group as the experiment progressed.

There is currently no standard for the compensation of participants in psycholinguisics experiments. For example, Fine et al. (2010) compensated participants with course credit, Fine and Jaeger (2013) paid participants $10, Stack et al. (2018) paid $4, Dempsey et al. (2020) paid $3, and Prasad and Linzen (2021) paid $6.51 per hour. It is possible that these participants experienced different levels of motivation and focus while performing this task, and this influenced the probability of learning regularities in the items. The present paper aims to shed some light on the effect that reward magnitude has on acceptability and self-paced reading tasks, and invite more research and discussion about what the compensation standards should be, and what pitfalls are there, if any.

3. Experiments

3.1 Garden path sentence acceptability

Garden path sentences like (1) provide a well-understood domain for which to study the effect of reward. Such sentences are not easy to process because they at first appear to be ungrammatical.

(1) a. The old man the boat.
(local noun/adjective ambiguity)
b. The horse raced past the barn fell.
(local relative / main clause ambiguity)
c. While Anna dressed the baby slept on the bed.
(local subject/object ambiguity)

If reward promotes attention and learning, then comprehenders should recover from garden path effects faster when offered a higher reward.

3.1.1 Method

Participants

We collected data from 100 self-reported native speakers of English with IP addresses originating from the United States, which we recruited through Prolific's crowdsourcing marketplace (https://www.prolific.co/). The data from 22 participants was excluded because their accuracy scores on comprehension questions (see below) were lower than the 75% threshold. The remaining 78 participants had a mean accuracy level of 84% (SD = 0.06).

Design and materials

In this between-subjects design, participants were randomly assigned to one of two groups: a baseline control group and a bonus group. All participants were told before starting that a comprehension question accuracy below 65% could prevent compensation. Participants in the bonus group were additionally informed that comprehension question accuracy above 75% would result in an extra $4.8, for a total of $7.2. Since the average completion time was 11 minutes, this corresponded to an hourly rate of $13 and $39, respectively.

The experimental items consisted of 16 pairs of sentences like those in (2), counterbalanced across two lists. In the garden path condition there was a classic subject/object ambiguity (Christianson et al. 2001; Ferreira and Henderson 1990; Frazier and Rayner 1982; Jacob and Felser 2016) in which an optionally intransitive verb is at first interpreted as being transitive, and in the non-garden path condition the ambiguity is avoided by the presence of a comma.

(2) a. After the Mayor visited(,) the patients were moved to different rooms.
b. While the kids studied(,) their parents were bribing teachers at school.
c. As the boys bathed(,) the dog was busy catching a mouse.
d. Although the soldiers saluted(,) the flag was not on the pole outside.

To maximize the garden path effect, the subordinate verbs in region 4 came

from a subset of verbs from Ferreira and Henderson (1991) and Staub (2007) that had the highest proportion of transitive uses relative to intransitive uses, according to both (Gahl et al. 2004) and to a corpus study using the Corpus of Contemporary American English (Davies 2008). The strong bias for the transitive use causes a closure parse which is known to be susceptible to priming, as reflected by decreased reading times (Noppeney and Price 2004; Traxler 2015), and physiologically by attenuated neural responses (Noppeney and Price 2004).

All sentences in either condition were followed by a comprehension question, to which the correct answer was "yes" half of the time. After submitting each answer participants were informed about whether their selection was correct or not. For example, after rating (2a) participants were asked *The Mayor paid a visit after the patients were moved. True or False?*

These 16 items were pseudo-randomized and interspersed with 16 distractors, of which there were three types-ungrammatical, grammatical, and grammatical followed by a comprehension question—as illustrated in (3).

(3) a. *Because a bank, now owns this building the rents are going up.
b. *After arranging the bookshelf we realized a novel missing.
c. As the detective arrived the suspects escaped undetected.
d. As the doctors rested in the cafeteria the patients were re-examined.
[The patients were re-examined at the cafeteria by the doctors.
True or False?]

The distractors contained the same prepended adverbs (plus the adverbs 'because', 'if' and 'whenever'), evenly distributed, and a variety of verbal structures different from the items. Across items and distractors, no two stimuli featured the same verb in order to avoid priming effects caused by verb repetition (Fine and Jaeger 2016; Traxler and Pickering 2005). Although all participants in the experiment saw the same stimuli, no two participants saw the same order of stimuli.

Procedure

Presentation of the experimental stimuli was done on a web-based interface using PCIbex (Zehr and Schwarz 2018), such that participants completed the experiments remotely and did not directly interact with the researchers. At the beginning of each

experiment session, participants first read an IRB-approved consent form and checked a box to indicate their consent to participate. Participants were instructed to judge how natural each sentence was, by giving it a rating from 1 (very unnatural) to 7 (very natural). Participants completed four practice trials at the beginning of each experiment to familiarize themselves with the task. Each experiment immediately followed the practice trials.

3.1.2 Results

To account for the possibility of different participants using the Likert scale differently, we transformed the ratings into z-scores, by participant, before conducting any testing. A Linear Mixed Effects (LMER) model was fit with z-score ratings as the dependent variable and with condition (garden path vs. non-garden path) as the predictor.[1] It revealed that non-garden path sentences are overall rater higher than the garden path counterparts, as expected (β = 0.77, SE = 0.1, t = 7.26, p < 0.0001). Next, an LMER model with z-score ratings as the dependent variable and with an interaction between Group and Condition as the predictor was significant (β = 0.21, SE = 0.09, t = 2.24, p = 0.024), suggesting that the effect of sentence condition depends on the reward size: the bonus participants rated non-garden path items higher than expected, as illustrated in Figure 1.

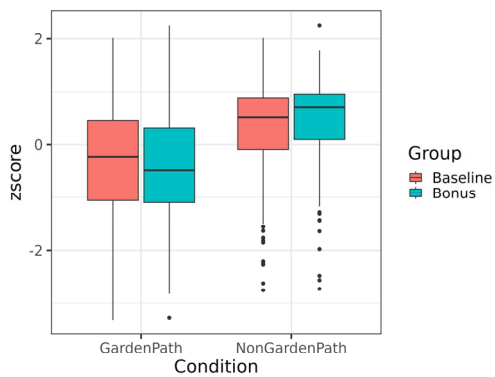

Figure 1. Overall z score distribution per condition type

1 In all models reported in this work, variations of each model was ran so that the intercept was allowed to be adjusted by items, list, subjects, age, and sex, in order to account for random effects. When LMER models could not converge with multiple random effects, they were tested separately, unless stated otherwise. The models reported here are those with the best fit.

Separate LMER models, fit on each of the sentence conditions to predict z-scores from the group variables indicate that the participant group had no effect on garden path ratings (β = −0.1, SE = 0.07, t = −1.4, p = 0.14), and approached significance on non-garden path ratings (β = 0.1, SE = 0.06, t =1.77, p = 0.07). The latter result suggests a trend in which increased reward boosts acceptability of uncontroversially grammatical sentences. Including list, sex or age as random effects did not yield qualitatively different results.

Bayesian Mixed Effects Linear Regression models (Bürkner 2017) with the same structure as our LMER models were also ran, in order to check whether including multiple random effects simultaneously made a difference. We used default (flat) priors, to keep the model very conservative, and checked it for convergence after fitting it with four chains and 6,000 iterations, half of which were the warm-up phase, by verifying that the R-hat values were close to 1, and visually inspecting the chains. Participant group came very close to conventional significance levels: β = −0.13, Error = 0.08, Credible interval = [-0.25, 0.01], Evidence Ratio = 15.74, P(β > 0) = 0.94. Although the credible intervals technically include zero, this is by a very slim margin, which is non-negligible value which should not be summarily dismissed by attempts to dichotomize significance. Stronger results could be obtained with more data and more informative priors. Different permutation of the random effects did not have an impact on the results. In a second BRMS model, participant group had an effect on non-garden path ratings (β = 0.11, Error = 0.06, Credible interval = [0.01, 0.21], Evidence Ratio = 23.24, P(β > 0) = 0.96). We conclude that there is evidence suggestive of an admittedly very subtle trend for higher reward leading to polarization of ratings, whereby lower acceptability sentences are rated lower than expected, and higher acceptability sentences are rated higher than expected. We will keep to the LMER models for the remainder of this study, as they are more widely used and will suffice for our purposes.

Next, we looked at the effect of presentation order on acceptability ratings. Did rating change as the experiment progressed? Probing for an interaction between Condition, Group and Order in predicting z scores, with sentence and participant as random slopes, we found no evidence of an effect (β = 0.01, SE = 0.02, t = 0.76, p = 0.4). Separate LMER models, fit on each of the conditions to predict z-scores from the interaction between group and order variables found no effect on garden path ratings (β = −0.008, SE = 0.01, t = −0.4, p = 0.61), nor on non-garden path

ratings (β = 0.008, SE = 0.01, t = 0.67, p = 0.53). Including Sex and/or Age as random effects did not yield qualitatively different results. The results in Figure 2 illustrate the findings.

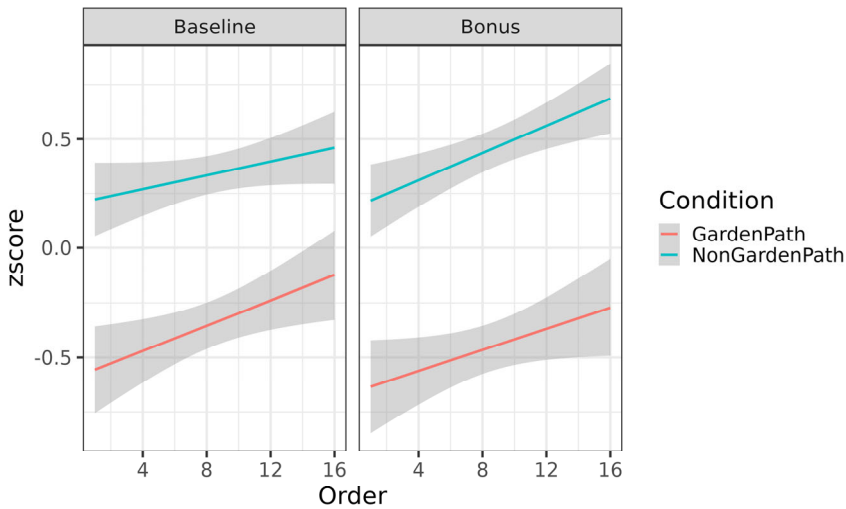

Figure 2. Mean z score ratings as a function of presentation order

3.1.3 Discussion

Nagata (1987a,b, 1988) hypothesized three possible effects of repeated exposure on acceptability judgments: leniency (sentences seem more acceptable; related to habituation), stringency (repeated exposure might highlight more issues in sentences, leading to stricter judgments), and polarization (confidence in initial judgments might increase, with good sentences rated better and bad sentences worse). The evidence described above is consistent with polarization in the reward condition, resulting from increased confidence about the acceptability of each sentence type. The ratings of non–garden path sentences were boosted by the bonus group, and there is some evidence that the ratings of the garden path sentences were curtailed. If this is true, then increased reward may cause ratings to be more extreme, if only slightly.

Null effects should be interpreted with caution, as it is possible that a very small impact of reward on acceptability ratings exists but was undetectable in the current study. It is also worth considering that the null effect may reflect a ceiling effect

in motivation. If the baseline level of reward was already sufficient to engage participants fully, increasing the reward might not have provided additional incentive. Alternatively, participants may not have consciously linked the reward amount to their ratings, as the task required them to focus on linguistic judgments rather than performance-based outcomes. As it stands, if such an effect exists, it does not appear to be very strong, given the small effect sizes.

3.2 Garden path self-paced reading

Chaves and Malone (2022) reports a self-paced reading experiment in which 100 participants recruited via the Amazon Mechanical Turk marketplace were asked to read garden path sentences of the same kind as (2), and answer comprehension questions. A sample of the items is shown in (4). This time there was no disambiguated condition, and all items were garden path sentences.

(4) a. After $_1$| the $_2$| Mayor $_3$| visited $_4$| the $_5$| patients $_6$| were $_7$| moved$_8$| to $_9$| different $_{10}$| rooms. $_{11}$|
[The Mayor paid a visit after the patients were moved. True or False?]
b. While $_1$| the $_2$| customers $_3$| ate $_4$| some $_5$| food $_6$| was $_7$| cooking$_8$| on $_9$| the $_{10}$| grill. $_{11}$|
[The customers ate only after all the cooking was done. True or False?]

All participants saw exactly the same sentences, but they were randomly assigned to either the control group or the bonus group. All participants were informed that the experimenters might not be able to compensate them if their comprehension accuracy dipped significantly below 70%. The individuals in the bonus group saw additional text and instructions informing them that if their comprehension question accuracy was above 75%, they would receive a bonus of $4.80, for a total of $7.2. The results suggest that participants in the bonus exhibited a faster rate of reduction in reading times in the spill-over region, after the disambiguating main verb in region 7. In other words, the reading times for region 8 decreased at a sharper rate for the control bonus than for the control group. In what follows we describe a replication of this study.

3.2.1 Methods

Participants

We collected data from 110 self-reported native speakers of English with IP addresses originating from the United States, recruited through Prolific. These subjects were different from those that participated in Experiment 1 above. The data from 7 individuals was excluded from the analysis as their accuracy scores on comprehension questions were lower than the 75% threshold. The remaining 78 participants had a mean accuracy level of 84% (SD = 0.05).

Participants were randomly assigned to either the control group or the bonus group. All participants were informed that the experimenters might not be able to compensate them if their comprehension accuracy dipped significantly below 70%, although in practice no participants were excluded from compensation. The individuals from the bonus group saw additional text and instructions informing them that if their comprehension question accuracy was above 75%, they would receive a bonus of $4.80, for a total of $7.2, as in the original study. The participants from each group saw the same stimuli.

Design and materials

The experimental items consisted of 16 garden path sentences from Chaves and Malone (2022), very similar to the items in Experiment 1 above. These 16 items were pseudo-randomized and interspersed with 16 distractors, illustrated in (5), also taken from Chaves and Malone (2022). Across items and distractors, no two stimuli contained the same verb, as to avoid priming effects caused by verb repetition. Although all participants in the experiment saw the same stimuli, no two participants saw the same order of stimuli.

(5) a. Though $_{1}$| the $_{2}$| bus $_{3}$| driver $_{4}$| missed $_{5}$| a $_{6}$| street $_{7}$| Sue $_{8}$| was $_{9}$| at school $_{10}$| on time. $_{11}$|
[Sue brought a child home after the bus missed its stop. True or False?]

Procedure

Subjects read sentences in a self-paced moving window display (Just et al. 1982), using the self-paced reading mode of the PCIbex platform (Zehr and Schwarz 2018). Three

practice trials were conducted before the experiment started. Each trial began by presenting a sequence of dashes representing the non-space characters in the sentences. Pressing the space bar caused the dashes corresponding to the first region to 18 be replaced by words. Subsequent presses revealed subsequent regions, while the previous region reverted to dashes. Reading times between each pair of button presses were recorded. All stimuli items were followed by a Yes/No comprehension question probing the lingering of the initial interpretation. The correct answer was "yes" half of the time, and after submitting each answer participants were informed about whether their selection was correct or not. The stimuli were pseudo-randomized so that no two participants saw the items in the same order and no more than two critical items were allowed to immediately follow each other.

Participants took 15 minutes on average to complete the experiment, meaning that Control group participants were paid at an hourly rate of about $9.4 while the Bonus group participants were paid at a $28.8 hourly rate. Unlike Chaves and Malone (2022), we collected demographic information about the participants, namely age and sex, as collected by Prolific as participants sign up to work on that crowd-sourcing marketplace.

3.2.2 Results

All observations with reading times lower than 100ms and longer than 2000ms were removed, excluding 3% of all observations. A LMER models, with residual reading time as the dependent variable and group (bonus vs. baseline) as the predictor (allowing the intercept to be adjusted by Sentence), were fit separately for region 7 (the disambiguating region), region 8 (spill-over region), and region 8 (post-spill-over). Results indicated that the residual reading times for the bonus group were significantly different from those in the baseline group (reference group) as shown in Table 1.

Table 1. Effect of bonus on residual reading times in regions 7–9

Region	β_{group}	SE	t	p
7	66	30.13	2.19	0.02
8	-77	22.58	-3.44	0.0006
9	39	10.7	3.6	0.0002

These results suggest the bonus group was slower in regions 7 and 9 than the baseline

control group, but was faster in the spill-over region, as Figure 3 illustrates.

If Sex is added as a random effect in addition to sentence the results are qualitatively the same, albeit more polarized for the first two regions of interest. Namely, for region 7 we have β_{Group} = 91 (p = 0.002), region 8 we have β_{Group} = −89 (p = 0.0001), and region 9 β_{group} = 26 (0.01). If group and sex are instead taken as interacting factors in region 8, the LMER model is significant ($\beta_{groupXsex}$ = −120, SE = 46, t = −261, p = 0.009), suggesting that the reading time of males was faster than that of females, in the bonus condition.

For completeness, a plot with the mean residual times for all sentence regions is provided in Figure 4. The behavior of the two groups of participants was generally the same, except that the bonus group slowed down at region 7 and accelerated at region 8. This is consistent with these participants being more attentive and taking greater care to perform the task.

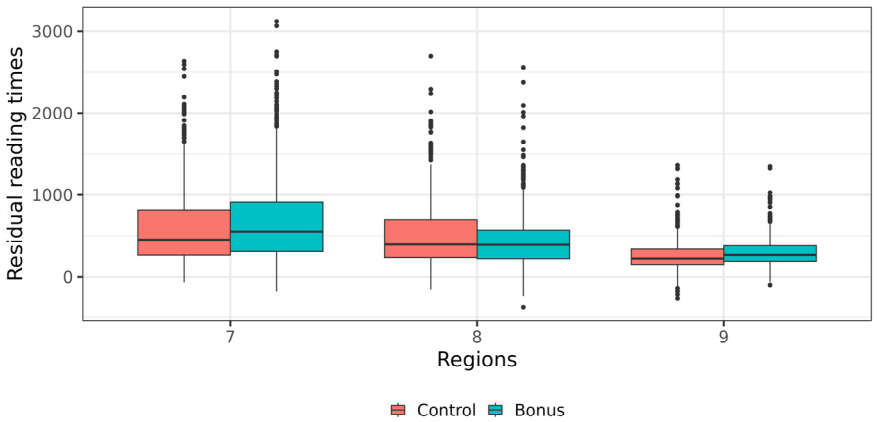

Figure 3. Effect of repeated exposure and reward differential in region 8

Probing for an interaction between group and presentation order in predicting RTs, with sentence as a random effect, there was no significance in any of the three regions of interest, even when Sex is included as a random effect:

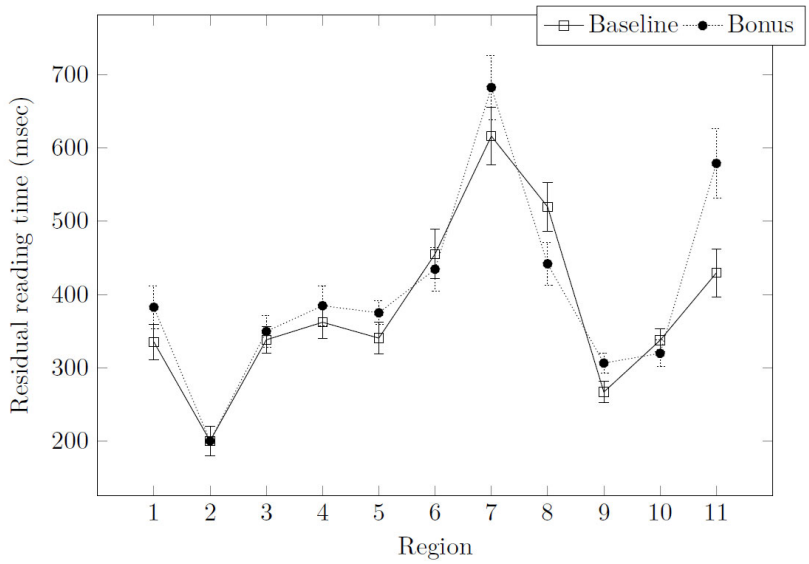

Figure 4. Mean residual reading times for all sentence regions

Table 2. Effect of bonus and presentation order in regions 7–9

Region	$\beta_{groupXorder}$	SE	t	p
7	66	30.13	2.19	0.02
8	-77	22.58	-3.44	0.0006
9	39	10.7	3.6	0.0002

As illustrated in Figure 5, all conditions saw a decrease with similar rates, albeit it is clear that males took longer overall in the baseline (control) condition.

3.2.3 Discussion

This experiment failed to replicate the main finding of Chaves and Malone (2022), which was a sharper rate of reading time reduction for the bonus group (as compared to the baseline group) in the spill-over region 8. In the latter study, participants in the two groups exhibited a very different processing profile, as illustrated in Figure 6. Overall, increased reward led participants in the Chaves and Malone (2022) study to take longer to read the sentences across most regions, perhaps a reflection of

strategies geared to completing the comprehension questions successfully. The wider error bars in the bonus condition, in particular, are a consequence of adaptive changes in reading time as the experiment progressed.

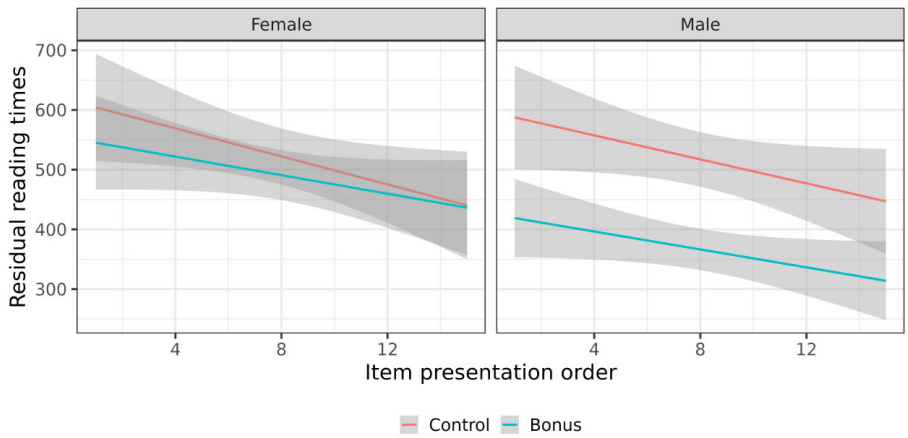

Figure 5. Effect of repeated exposure and reward in spill-over region 8

In the present study, in contrast, there is no difference in overall reading times across the two groups, perhaps reflecting a population difference between Prolific and Amazon Mechanichal Turk subjects, and the spread of the error bars is comparable across both conditions and regions. Despite not replicating the original result, we observed a different kind of evidence for adaptation to garden path sentences as a function of reward: faster reading times in the spill-over region for the bonus group, as compared to the baseline. In other words, the individuals in the bonus used cues in the input to strategically predict the upcoming structure and recover from the garden path faster than the individuals in the baseline group.

3.3 Subject island self-paced reading

In this next experiment we investigate whether constructions which are more difficult to process than the foregoing garden path effect elicit different online strategies as a function of reward. For this purpose turn to Subject Islands, which are considered to be among the strongest constraints on English extraction. Whereas it is usually possible to extract an NP from an NP complement, as in (6), extracting an NP from

an NP subject as in (7) is much harder, and traditionally regarded as categorically impossible (Chomsky 1973, 1977; Huang 1982; Kayne 1981; Lasnik and Saito 1992) until only more recently (Chomsky 2008; Haegeman et al. 2014; Kluender 2004).

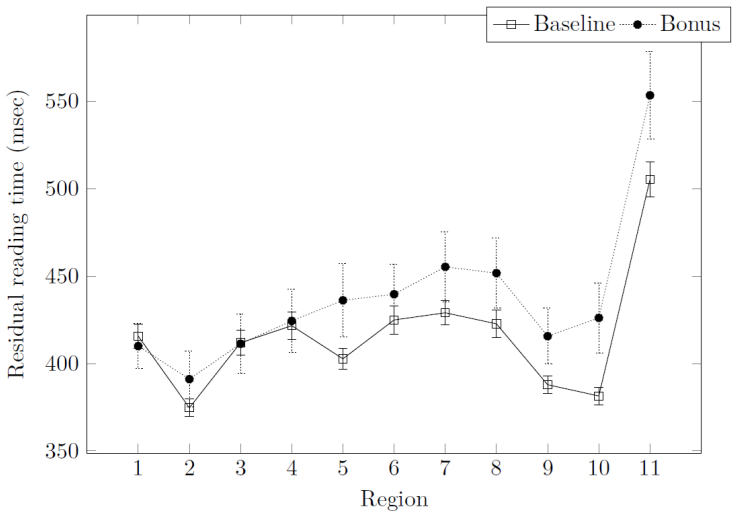

Figure 6. Mean residual reading times (Chaves and Malone 2022)

(6) Who did you hear [stories about _____]? (Chomsky 1973:86)

(7) a. *Who did [stories about _____] terrify John? (Chomsky 1973:92)
b. *a man who [pictures of ___] are on the table (Chomsky 1986:31,61)

A number of experimental studies have confirmed that interrogative Subject Island violations are graded and that with mere repetition their acceptability can improve significantly (Chaves and Dery 2014; Clausen 2011; Francom 2009; Hiramatsu 2000; Lu et al. 2021) and are processed faster a function of increased exposure (Chaves and Dery 2019; Chaves and Putnam 2020). Other studies, however, have not found clear evidence for such amelioration effect (Crawford 2011; Snyder 2000; Sprouse et al. 2013). Such null results can result of a number of experimental and design factors, such as an insufficient number of exposures (e.g. no effect was detected by Hiramatsu (2000) with 4 exposures, but an effect was found with 7), the quality of the items (Hofmeister and Sag 2010), and statistical methodology. With respect to the latter, Lu et al. (2024) conducted a meta-analysis that included data from several published

studies, and found significant amelioration in Subject Islands for the experiments in Sprouse (2009), even though the original study reported null effects. Taken together, all extant evidence suggests that that interrogative Subject Islands are quite difficult–but not impossible–to process, given sufficient exposure in ideal conditions.

3.3.1 Methods

Participants

We analyzed the data from 117 self-reported native speakers of English with IP addresses originating from the United States recruited through Prolific, with accuracy scores over the 75% threshold. These subjects were different from those that participated in the prior experiments, and their mean comprehension question accuracy was 90% (SD = 0.06). As in prior experiments, participants were randomly assigned to either the control group or the bonus group. All participants were informed that the experimenters might not be able to compensate them with $3.6 if their comprehension accuracy dipped significantly below 65%, although in practice no participants were excluded from compensation. The individuals from the bonus group saw additional text and instructions informing them that if their comprehension question accuracy was above 75%, they would receive a bonus of $5.2, for a total of $8.7. The participants from each group saw the same stimuli.

Design and materials

The experimental items consisted of 16 subject island sentences like (8), pseudo-randomized and interspersed with 32 distractors. As in the foregoing studies, although all participants in the experiment saw the same stimuli, no two participants saw the same order of stimuli.

(8) a. Which book $_1$| did $_2$| the author of $_3$| nearly $_4$| die $_5$| in $_6$| a recent $_7$| car accident? $_8$|
b. Which question $_1$| did $_2$| the answer to $_3$| supposedly $_4$| satisfy $_5$| most $_6$| of $_7$| the philosophers? $_8$|
c. Which bird $_1$| did $_2$| the song of $_3$| almost $_4$| wake up $_5$| the baby $_6$| in $_7$| the crib? $_8$|

The critical region for the experimental items was region 3, since this is the point at which comprehenders would postulate a gap and link it to the wh-phrase in region 1. Region 4 consitently contained an adverb that is consistent both with an upcoming verb and with an upcoming subjectembedded object. For example. In (8a) the string *Which book did the author of nearly* could in principle be continued by a nominal phrase like *Which book the author of nearly twenty books.* This ought to be the *a priori* preferential type of contination, as it does not trigger any Subject Island violation. However, all items in the experiment involved a verb after the adverb, and triggered a Subject Island violation to which participants need to adapt in order to be able to parse the sentence. Thus, regions 4 and 5 are regions of interest as well, since this is where the expectation for an upcoming nominal phrase is contradicted.

The 32 distractors where hetereogenous as illustrated in (9), and half were followed by comprehension questions, as in (9d,e).

(9) a. Which artifact $_1$| does $_2$| the Museum of $_3$| Fine Arts $_4$| wish $_5$| quickly $_6$| purchase? $_7$|

b. Where $_1$| did $_2$| the audience $_3$| see $_4$| the ace $_5$| of $_6$| spades $_7$| reappear? $_8$|.

c. Which offer $_1$| did $_2$| the union $_3$| leaders $_4$| decide $_5$| to $_6$| not $_7$| accept? $_8$|

[The union leaders decided that one of the offers was not good enough. True or False?]

d. Which animal $_1$| do $_2$| the villagers $_3$| allegedly $_4$| eat $_5$| as their $_6$| staple $_7$| food? $_8$|

[The diet of the local inhabitants consists uniquely of grains and herbs. True or False?]

Procedure

The procedure was the same as for Experiment 2. Participants took 12 minutes on average to complete the experiment, meaning that the baseline group was compensated paid at an hourly rate of about $18 while the Bonus group participants were paid at a $47 hourly rate.

3.3.2 Results

All observations with reading times lower than 100ms and longer than 2000ms were removed, excluding 2% of all observations. Regression models with residual reading times as the dependent variable and participant group (baseline vs. bonus) as the predictor found that for regions 3 through 5 the bonus group took longer than the baseline group as shown in Table 3. The addition of random effects did not change the results qualitatively.

Table 3. Effect of bonus group on residual reading times in regions 3 – 5

Region	β_{group}	SE	t	p
3	82.05	21.54	3.8	0.0001
4	111.93	21.99	5.09	< 0.0001
5	40.57	17.61	2.3	0.02

As a way of illustrating the findings, see Figure 7. Here, it is clear that the subjects in the bonus group took overall more time to process the sentence.

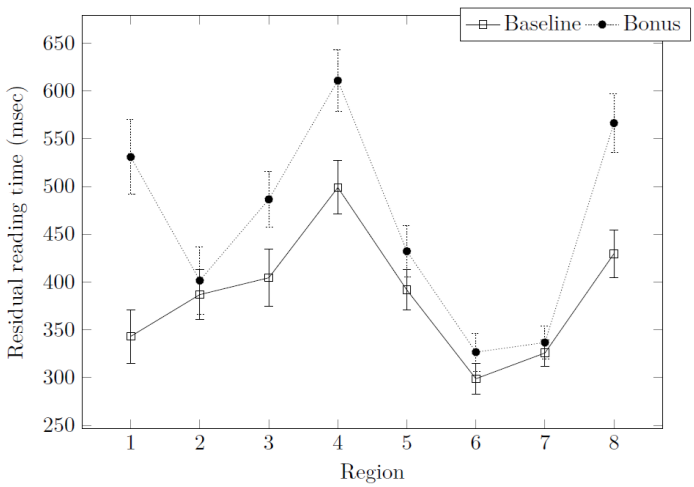

Figure 7. Mean residual reading times for all sentence regions

Finally, LMER models looking at the interaction between participant group and presentation order were not significant for the regions 3 and 4 (even with random slope models), but significant in region 5, with item and age as random effects. This

suggests that participants in the baseline group became faster in this post-spillover region than the bonus participants, as the experiment progressed. The results are in Table 4.

Table 4. Interaction of bonus and presentation on regions 3 – 5

Region	$\beta_{groupXorder}$	SE	t	p
3	3.0	3.0	1.0	0.31
4	1.9	3.1	0.6	0.54
5	5.6	2.37	2.3	0.01

Participants sped up during the experiment in all regions, as shown by Table 5, which reports LMER models with residual reading times as the dependent variable and presentation order as a simple effect. Figure 8 illustrates how spill-over regions 4 and 5 have the sharpest declines, and how the latter shows a subtle difference as a function of reward.

Table 5. Effect of presentation order on regions 3 – 5

Region	β_{order}	SE	t	p
3	-4.67	2.35	-1.98	0.04
4	-15.34	2.38	-6.44	< 0.0001
5	-15.65	1.88	-8.31	< 0.0001

Looking into each participant group more closely the data becomes multimodal, with females in the baseline control group being slower than the females in the bonus group in region 3 (β = −180.97, SE = 22.44, t = −8.06, p < 0.0001), using LMERs with group as a simple effect, and item and age as random effects. In contrast, males in the baseline control group were faster than the males the bonus group in the same region 3 (β = 262.07, SE = 21.17, t = 12.38, p < 0.0001), using again LMERs with group as a simple effect, and item and age as random effects, as these result in the best fit. Although there was no difference between females in the baseline control group and in the bonus group for region 4 (β = 29.76, SE = 25.23, t = 1.06, p = 0.28) and region 5 (β = 0.78, SE = 17.57, t = 0.04, p = 0.9), males in the bonus group were slower than those in the baseline control group for region 4 (β = 184.55, SE = 22.31, t = 8.27, p < 0.0001), and region 5 (β = 66.34, SE = 20.26, t = 3.27, p = 0.001). Figure 9 serves to illustrate. This suggests that the overal higher reading

times for the baseline condition were mostly driven by the male participants.

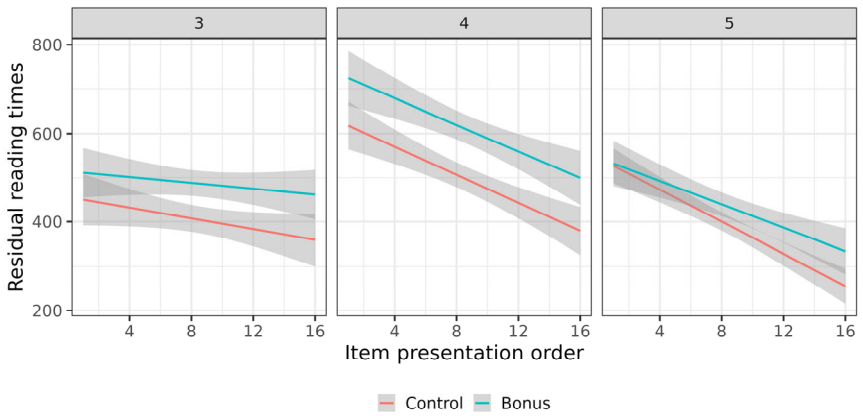

Figure 8. Effect of repeated exposure and reward

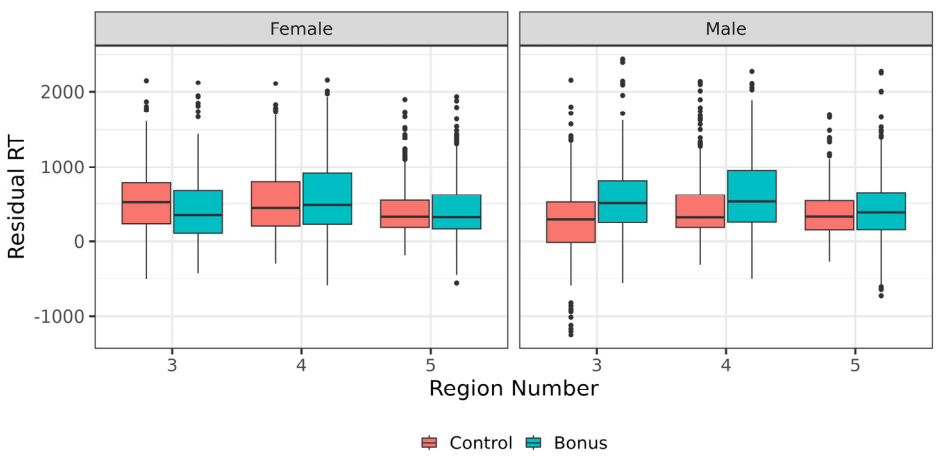

Figure 9. Residual reading times as a function of participant groups

The by-region plot for the female participants is given in Figure 10. Here we see the faster reading times for the bonus group than for the baseline control group, in critical region 3, a difference which vanishes in subsequent regions, as detailed above. This is consistent with a very rapid adaptation to the structure of the items.

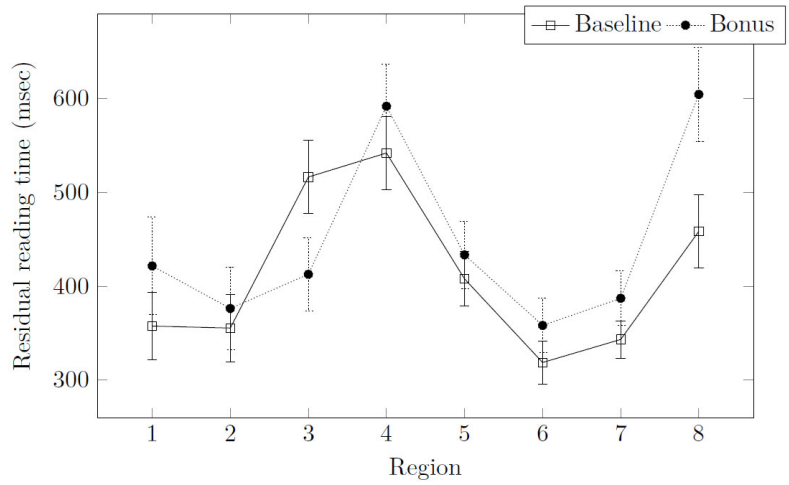

Figure 10. Mean residual reading times for all sentence regions (women)

For completeness the males' by-region plot is provided in Figure 11, which reveals a very different processing profile.

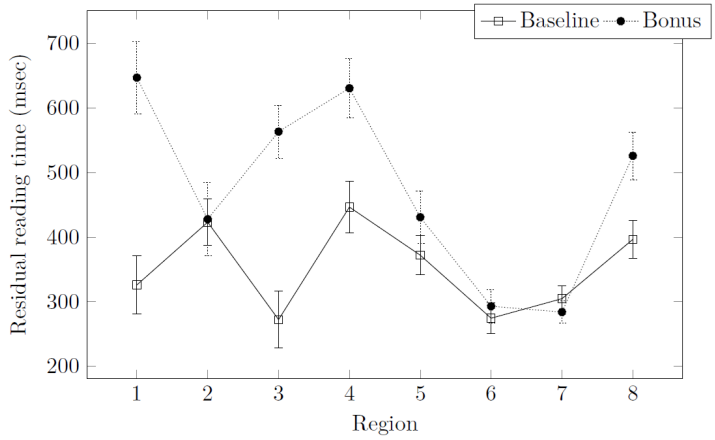

Figure 11. Mean residual reading times for all sentence regions (men)

The latter reveals a very different processing profile, and presumably one in which the participants in the control group deployed strategies to perhaps to ensure success

in the task. This included a glaring slow down in region 1, possibly in anticipation of a difficult sentence to process and risk in failing subsequent comprehension questions. To be sure, the female age average was 40.2 (SD = 12.9, N = 62), and the male average was 39.6 (SD = 11.4, N = 55), a small difference that is nonetheless sufficient to be statistically detectable (t = 2.51, p = 0.01).

3.4 Discussion

In this study we observe two kinds of adaptation effect. One is sentencewide and involves slower reading times relative for the participants with higher expected reward relative to the baseline control participants. This is makes sense, as it suggests that individuals strategized and took more care to perform the task. The other adaptation effect is more interesting and localized on the regions of interest in the sentence, near where the subject island effect is triggered, namely, region 5. Here, reading times decreased more sharply for some of the participants in the baseline control group, than for the bonus group, perhaps reflecting lower motivation levels.

Perhaps the most notable effect, however, is the sharp difference in overall reading times in the critical region 3, which go in different directions according to the sex of the participant. This might be a quirk in the population of this study, or it might reflect subtle differences in how these participants react to different expected reward magnitudes.

4 Conclusion

Experimental syntax research often involves participants processing unusual, disconnected sentences in artificial tasks. These conditions, combined with varying participant motivation, might influence both online and offline comprehension measures, potentially distorting results. This study examines how reward magnitude, used as a proxy for motivation, impacts experimental outcomes. Our results suggest that reward incentives can subtly affect outcomes, with differential rewards leading to more extreme (polarizing) ratings of sentence acceptability (Experiment 1), and very different reading time profiles. Specifically, bonus group participants tend to take longer to perform the task (Experiments 2 and 3), which is consistent with

attention-based strategies, as well as faster recovery from garden path sentences (Experiment 2) and − at least for some participants − faster recovery from island violations (Experiment 3).

Overall, the study highlights the nuanced role of expected reward in experimental syntax tasks and its influence on participant strategies. More research is needed to verify how robust these effects are. As it stands, the findings suggest that reward can have a small but measurable effect on experimental outcomes, both in terms of strategies and more automatic adaptation. Experiments should take to incorporate a commonsense reward protocol in their experiments, such that compensation is at a level that is commensurate to the task, using the average from comparable tasks, or only marginally higher than the average, as extreme values might have a more pronounced impact.

References

Altmann, Gerry T. M. and Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3): 247-264.

Ambrase, Aiste, Carolin A. Lewis, Claudia Barth, and Birgit Derntl. 2021. Influence of ovarian hormones on value-based decision-making systems: Contribution to sexual dimorphisms in mental disorders. *Frontiers in Neuroendocrinology* 60: 100873.

Anderson, John R. 1990. *The adaptive character of thought.* Mahwah, NY: Lawrence Erlbaum Associates, Inc.

Arai, Manabu and Frank Keller. 2013. The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes* 4(28): 525-560.

Bachevalier, Jocelyne and Corrine Hagger. 1991. Sex differences in the development of learning abilities in primates. *Psychoneuroendocrinology* 16: 177-188.

Blanchette, Frances, Erin Flannery, Carrie Jackson, and Paul Reed. 2024. Adaptation at the Syntax–Semantics Interface: Evidence from a vernacular structure. *Language and Speech* 67(1): 140-165.

Brandone, Amanda C., Suzanne R. Horwitz, Richard N. Aslin, and Henry M. Wellman, 2014. Infants‘ goal anticipation during failed and successful reaching actions. *Developmental Science* 17(1): 23-34.

Bürkner, Paul-Christian. 2017. brms: An R package for bayesian multilevel models Using Stan. *Journal of Statistical Software* 80(1): 1–28.

Chang, Franklin, Gary S. Dell, and Kathryn Bock. 2006. Becoming syntactic. *Psychological Review*

113(2): 234-272.

Chang, Franklin, Marius Janciauskas, and Hartmut Fitz. 2012. Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass* 6: 259–278.

Chaves, Rui P. and Jeruen E. Dery. 2014. Which subject islands will the acceptability of improve with repeated exposure? In Robert E. Santana-LaBarge (ed.), Proceedings of the *31st West Coast Conference on Formal Linguistics,* 96–106. Somerville, MA: Cascadilla Proceedings Project.

Chaves, Rui P. and Jeruen E. Dery. 2019. Frequency effects in subject islands. *Journal of Linguistics* 3(1): 1–47.

Chaves, Rui P. and Avery Malone. 2022. Reward-modulated syntactic adaptation in self-paced reading. A poster presented at *the 35th Annual Conference on Human Sentence Processing*. Santa Cruz, CA: University of California Santa Cruz. March 24-26.

Chaves, Rui P. and Michael T. Putnam. 2020. *Unbounded dependency constructions: Theoretical and experimental perspectives.* Oxford: Oxford University Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson and Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Reinhart & Winston.

Chomsky, Noam. 1977. *Essays on form and interpretation*. New York, NY: North-Holland.

Chomsky, Noam. 1986. *Barriers.* Cambridge, MA: The MIT Press.

Chomsky, Noam. 2008. On Phases. In Robert Freidin, David Michaels, Carlos P. Otero, and Maria Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud,* 133-165, Cambridge, MA: The MIT Press.

Christianson, Kiel, Jack Dempsey, Anna Tsiola, and Maria Goldshtein. 2022. What if they're just not that into you (or your experiment)? On motivation and psycholinguistics. *Psychology of Learning and Motivation* 76: 51-88.

Christianson, Kiel, Andrew Hollingworth, John F. Halliwell, and Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 4(42): 368-407.

Clark, Ann and Patricia P. Goldman-Rakic. 1989. Gonadal hormones influence the emergence of cortical function in nonhuman primates. *Behavioral Neuroscience* 103(6): 1287-1295.

Clausen, David R. 2011. Informativity and acceptability of complex subject islands. A poster presented at *the 24th Annual CUNY Sentence Processing Conference.* Stanford, CA. May 16-18.

Crawford, Jean. 2011. Using syntactic satiation effects to investigate subject Islands. In Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman (eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics*, pages 38–45, Somerville, MA: Cascadilla Press.

Creel, Sarah C., Richard N. Aslin, and Michael K. Tanenhaus. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106(2): 633-664.

Davies, Mark. 2008. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

DeLong, Katherine A., Thomas P. Urbach, and Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8): 1117–1121.

Dempsey, Jack, Qiawen Liu, and Kiel Christianson. 2020. Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46(10): 1906-1921.

Do, Monica L. and Elsi Kaiser. 2017. The relationship between syntactic satiation and syntactic priming: A first look. *Frontiers in Psychology* 8: 1851.

Evans, Kelly E. and Elizabeth Hampson. 2015. Sex-dependent effects on tasks assessing reinforcement learning and interference inhibition. *Frontiers Psychology* 6: 1044.

Farmer, Thomas A., Alex B. Fine, Shaorong Yan, Spyridoula Cheimariou, and Florian T. Jaeger. 2014. Error-driven adaptation of higher-level expectations during natural reading. In Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati (eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 2181-2186. Austin, TX: Cognitive Science Society.

Ferreira, Fernanda and John M. Henderson. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word selfpaced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(4): 555-568.

Ferreira, Fernanda and Henderson, John M. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 31: 725–745.

Fine, Alex and Florian T. Jaeger. 2016. The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(9): 1362-1376.

Fine, Alex B., Florian T. Jaeger, Thomas A. Farmer, and Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8(10): e77661.

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

Fine, Alex B., Ting Qian, Florian T. Jaeger, and Robert A. Jacobs. 2010. Syntactic adaptation in language comprehension. In John T. Hale (ed.), *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics,* 18-26.

Francom, Jerid. 2009. *Experimental syntax: exploring the effect of repeated exposure to anomalous syntactic structure - Evidence from rating and reading tasks.* PhD Dissertation. University of Arizona.

Frazier, Lyn and Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14(2): 178-210.

Gahl, Susanne, Dan Jurafsky, and Douglas Roland. 2004. Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, and Computers* 36(3): 432–443.

Gibson, Edward. 2006. The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language* 54(3): 363-338.

Gigerenzer, Gerd, Peter M. Todd, and the ABC research group. 1999. *Simple heuristics that make us smart.* Oxford: Oxford University Press.

Goldman, Patricia S., Howard T. Crawford, Linton P. Stokes, Thelma W. Galkin, and H. Enger Rosvold. 1974. Sex-dependent behavioral effects of cerebral cortical lesions in the developing rhesus monkey. *Science* 186: 540–542.

Goodall, Grant. 2011. Syntactic satiation and the inversion effect in English and Spanish wh-questions. *Syntax* 14: 29–47.

Grissom, Nicola M. and Teresa M. Reyes. 2019. Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology* 44: 86–96.

Haegeman, Liliane, Ángel Luis Jiménez-Fernández, and Andrew Radford. 2014. Deconstructing the Subject Condition: Cumulative constraint violation and tolerance thresholds. *The Linguistic Review* 31(1): 73-150.

Hiramatsu, Kazuko. 2000. *Accessing linguistic competence: Evidence from children's and Adults's acceptability judgments.* PhD Dissertation. University of Connecticut.

Hofmeister, Philip. 2015. Experience-driven acceptability effects. Ms. University of Essex.

Hofmeister, Philip and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86(2): 366-415.

Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammar.* PhD Dissertation. Massachusetts Institute of Technology.

Jacob, Gunnar and Claudia Felser. 2016. Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology* 69(5): 907-925.

Just, Marcel Adam, Patricia A. Carpenter, and Jacqueline D. Wooley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2): 228-238.

Kamide, Yuki and Don C. Mitchell. 1997. Relative clause attachment: No determinism in Japanese parsing. *Journal of Psycholinguistic Research* 26: 247-254.

Kayne, Richard. 1981. ECP Extensions. *Linguistic Inquiry* 12: 93-133.

Kluender, Robert. 2004. Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher, and A. Rodriguez (eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics,* 101-125, Somerville, MA: Cascadilla Press.

Kutas, Marta and Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 5947(307): 161–163.

Lasnik, Howard and Mamoru Saito. 1992. *Move α: conditions on its application and output.* Cambridge, MA: The MIT Press.

Lau, Ellen, Clare Stroud, Silke Plesch, and Colin Phillips. 2006. The role of structural prediction in rapid syntactic analysis. *Brain and Language* 1(98): 74-88.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 3(106): 1126-1177.

Levy, Roger, Evelina Fedorenko, Mara Breen, and Ted Gibson. 2012. The processing of extraposed structures in English. *Cognition* 12(1): 12-36.

Levy, Roger and Frank Keller. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 2(68): 199-222.

Lewis, Carolin A., Melina Grahlow, Anne K"Nuhnel, Birgit Derntl, and Nils B. Kroemer. 2023. Women compared with men work harder for small rewards. *Scientific Reports* 13(5456).

Lu, Jiayi, Michael Frank and Judith Degen. 2024. A meta-analysis of syntactic satiation in extraction from islands. *Glossa Psycholinguistics* 3(1): 1-33.

Lu, Jiayi, Daniel Lassiter, and Judith Degen. 2021. Syntactic satiation is driven by speaker-specific adaptation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 43, 1493-1499.

MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4): 67-703.

Metzing, Charles and Susan E. Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 49(2): 201-213.

Nagata, Hiroshi. 1987a. Change in the modulus of judgmental scale: An inadequate explanation for the repetition effect in judgments of grammaticality. *Perceptual and Motor Skills* 65(3): 907–910.

Nagata, Hiroshi. 1987b. Long-term effect of repetition on judgments of grammaticality. *Perceptual and Motor Skills* 65(1), 295–299.

Nagata, Hiroshi. 1988. The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. *Journal of Psycholinguistic Research* 17(1): 1-17.

Newell, Allen and Herbert A. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Noppeney, Uta and Cathy J. Price. 2004. An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience* 16: 702-713.

Overman, William H., Jocelyne Bachevalier, E. Schuhmann, and P. Ryan. 1996. Cognitive gender differences in very young children parallel biologically based cognitive gender differences in monkeys. *Behavioral Neuroscience* 110: 673-684.

Prasad, Grusha and Tal Linzen. 2019. Reassessing the evidence for syntactic adaptation from self-paced reading studies. A poster presented at the *32nd CUNY Conference on Human Sentence Processing*. Boulder, Colorado. March 29-31.

Prasad, Grusha and Tal Linzen. 2021. Rapid syntactic adaptation in selfpaced reading: detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory and Cognition* 47(7): 1156-1172.

Snyder, William. 1994. A psycholinguistic investigation of weak crossover, islands, and syntactic satiation effects: implications for distinguishing competence from performance. A poster

presented at *the 7th CUNY Conference of human sentence processing*. New York, NY. March 17-19.

Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31: 575-582.

Snyder, William. 2021. Satiation. In Grant Goodall (ed.), *The Cambridge handbook of experimental syntax,* Cambridge: Cambridge University Press.

Sprouse, Jon, Carson T. Schütze and Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001-2010. *Lingua* 134, 219–248.

Stack, Caoimhe M. Harrington, Ariel N. James, and Duane G. Watson. 2018. A failure to replicate rapid syntactic adaptation in comprehension. *Memory & cognition* 46(6): 864-877.

Staub, Adrian. 2007. The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(3): 550-569.

Staub, Adrian and Charles Clifton Jr. 2006. Syntactic prediction in language comprehension: Evidence from *either... or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(2): 425-436.

Tabor, Whitney, Cornell Juliano, and Michael K. Tanenhaus. 1997. Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes* 2-3(12): 211-271.

Traxler, Matthew J. 2015. Priming of early closure: Evidence for the lexical boost during. *Language, Cognition and Neuroscince* 30(4): 478-490.

Traxler, Matthew J. and Martin J. Pickering. 2005. Syntactic priming in comprehension: Evidence from eye-movements. *Proceedings of the 18th annual CUNY Conference on Human Sentence Processing*. Tucson, AZ. May 16-18.

Wells, Justine, Morten H. Christiansen, David S. Race, Daniel J. Acheson, and Maryellen C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58: 250-271.

Xu, Fei and Tamar Kushnir. 2013. Infants are rational constructivist learners. *Current Directions in Psychological Science* 22(1): 28-32.

Yano, Masataka, Shugo Suwazono, Hiroshi Arao, Daichi Yasunaga, and Hiroaki Oishi. 2021. Selective adaptation in sentence comprehension: Evidence from event-related brain potentials. *Quarterly Journal of Experimental Psychology* 74(4): 645-668.

Zehr, Jeremy and Florian Schwarz. 2018. PennController for internet Bbased experiments (IBEX). Ms. Availab https://doi.org/10.17605/OSF.IO/MD832.

Zervakis, Jennifer, Seiko Mazuka. 2013 Effect of repeated evaluation and repeated exposure on acceptability ratings of sentences. *Journal of Psycholinguistic Research* 42: 505-525.

Rui P. Chaves
Professor
Department of Linguistics
University of Buffalo, The State University of New York
609 Baldy Hall University at Buffalo, North Campus
Buffalo, NY 14260, USA
E-mail: rchaves@buffalo.edu

Received: 2025. 01. 29.
Revised: 2025. 02. 14.
Accepted: 2025. 02. 14.