Evidence for a Discourse Account of Manner-of-Speaking Islands

September 14, 2024

Abstract

Sentences with syntactic movement out of sentential complements of manner-of-speaking (MoS) verbs (e.g., *whisper*, *shout*) are degraded in acceptability, an effect called the "manner-of-speaking (MoS) island effect". Accounts variably attribute the MoS island effect to the violation of the subjacency condition, to the low frequency of MoS verbs taking sentential complements, or to a general information structural constraint that discourse-backgrounded constituents cannot be extracted. In five acceptability judgment experiments, we find that the MoS island effect can be modulated by foregrounding or backgrounding the extracted constituent, suggesting a causal relationship between discourse backgroundedness and the MoS island effect. Our findings challenge syntactic and frequency accounts of the MoS island effect.

1 Introduction

The degradedness of sentences like those in (1) has traditionally been attributed to the inability of syntactic movement to cross certain structural domains, a phenomenon called "island effects" (Ross, 1967). Examples of such structural domains include complex noun phrases, clausal adjuncts, and conjuncts.

(1) a. Extraction from a complex noun phrase
*What$_i$ does John know the fact that Mary ate t_i?

b. Extraction from a clausal adjunct
*Who$_i$ did John have lunch after he talked to t_i?

c. Extraction from a conjunct
*Who$_i$ does John like Mary and hate t_i?

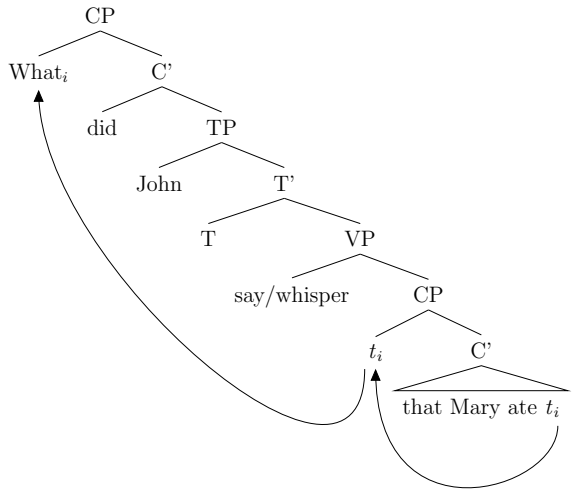
Traditionally, island effects have been considered to arise due to the constraints that particular structures impose on syntactic dependencies (Bresnan, 1976; Chomsky, 1977, 1986; Ross, 1967; Sag, 1976), or the cyclic phase-by-phase manner in which syntactic derivations proceed (Fox and Pesetsky, 2005; Nunes and Uriagereka, 2000). Syntactic approaches to

island effects face a challenge in explaining a particular type of island, namely those that are sentential complement clauses. Sentential complement clauses are generally permeable to movement, as shown in example (2a). However, complements of certain verbs behave like islands. For example, the complements of manner-of-speaking (MoS) verbs (e.g., *whisper*) resist movement from within (Ambridge and Goldberg, 2008; Erteschik-Shir, 2006, 2007; Goldberg, 2013; Richter and Chaves, 2020; Snyder, 1992, *inter alia*), as shown in example (2b). The degraded status of sentence (2b) is called the "manner-of-speaking (MoS) island effect."

(2) a. What$_i$ did John say that Mary ate t_i?
b. *What$_i$ did John whisper that Mary ate t_i?

On the surface, the two sentences in (2) are identical except for the matrix verb. If we assume that (2a) and (2b) have the same canonical sentential complement structure, shown in (3), the contrast in their acceptability cannot be attributed to a structural difference, thereby posing a challenge to syntactic accounts of MoS island effects.

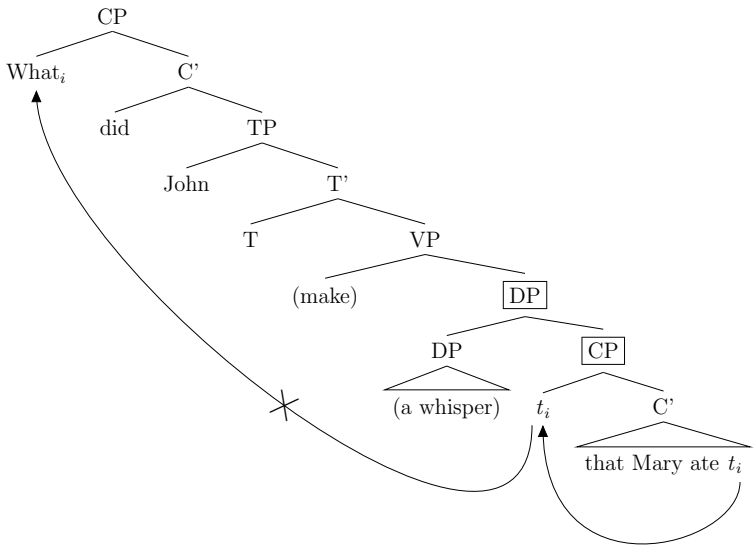
(3) Sentential complement structure of (2)



Several proposals have been made to explain this MoS island puzzle. The "Subjacency Account" proposes that (2a) and (2b) are in fact structurally distinct, and (2b) involves violation of a syntactic subjacency condition (Snyder, 1992; Stowell, 1981). According to this account, a sentence like (2b) has the syntactic structure in (4). The matrix verb *make* takes a DP complement *a whisper*, and the sequence *make a whisper* has the option of being realized as *whisper* in PF. The embedded clause is adjoined to the DP node, creating a complex-NP structure with two bounding nodes (CP and DP, boxed in (4)) intervening on the wh-movement step.[1] The resulting sentence is thus ungrammatical because it violates the subjacency condition, which states that no movement step can cross more than one bounding node.

[1]In the structure in (4), we adopt the DP hypothesis, which is the dominant analysis of nominal structures in generative syntax (Abney, 1987). Therefore, the island structure in (4) should be more accurately described as a complex-DP, rather than a complex-NP. However, we follow the notational tradition in the literature on island effects, and label such structures as complex-NP structures.

(4) Reducing MoS island to subjacency structure of (2b)



Other accounts of the MoS island effect maintain the syntactic structure in (3) and propose a non-syntactic source of unacceptability. Under the "Verb-Frame Frequency Account," the MoS island effect is due to the low frequency with which MoS verbs take complement clauses Kothari (2008); Liu et al. (2019, 2022b). When comprehenders read a sentence with an MoS verb that takes a clausal complement, the low expectedness (high surprisal) of the complement clause is taken to lead to higher processing cost and hence lower acceptability ratings, giving rise to the MoS island effect. Evidence for this account comes from an observed negative correlation between the corpus frequency of MoS verbs taking complement clauses and the acceptability rating of sentences showing the MoS island effect (Kothari, 2008; Liu et al., 2019, 2022b) [2]

Under another non-syntactic account of the MoS island effect, the "Backgroundedness Account," extraction from within MoS verb complements is unacceptable because elements contained inside the complement clauses of MoS verbs are by default backgrounded rather than foregrounded in the discourse (Ambridge and Goldberg, 2008; Goldberg, 2013). The term "foreground" (Hopper, 1979; Jones and Jones, 1979; Tomlin, 1985) refers to a similar concept as that of "dominance" (Erteschik-Shir, 1973, 1979),"focus" (Chomsky, 1969, 1972; Roberts, 1996, 2012), "topic" (Kuno, 1976, 1987), and "at-issue" (Potts, 2005, 2007). For consistency, we adopt the term "foregrounded" to describe constituents that are the topic, dominant, focused, or at-issue, and the term "backgrounded" to describe constituents that are not. We adopt the following working definition for "foregrounded/backgrounded" that draws on the concepts of Question Under Discussion (QUD) from Roberts (1996, 2012), following previous literature (Beaver et al., 2017; Simons et al., 2010; Tonhauser et al., 2018). There are other ways of formalizing backgroundedness that are not QUD-based (see, e.g., Erteschik-Shir, 2007). The QUD-based definition is adopted simply for concreteness of discussion, while other definitions proposed in the literature may also be compatible with the arguments and the interpretation of the experimental results to be presented in this study.

[2]But see Richter and Chaves (2020) for a similar study that did not find such a correlation.

Definition 1. *A constituent C is* ***foregrounded*** *iff the alternative set relative to C is a subset of Q-alternatives determined by the QUD.*

We further adopt the following definitions for *alternative set* (adapted from the concept of *focus alternative set*) and *Q-alternatives* from Roberts (1996) and Roberts (2012).

Definition 2. *The* ***alternative set*** *of an utterance U relative to a constituent C is the set of all interpretations obtained by replacing C with variables, and then interpreting the result relative to each member of the set of all assignment functions which vary at most in the values they assign to those variables.*

Definition 3. *The* ***Q-alternatives*** *determined by a question A is the set of all complete answers to A.*

In sentences with MoS verbs taking embedded clauses, the backgroundedness of embedded elements is taken to be the result of MoS verbs' lexical properties. We follow Erteschik-Shir (2007) and assume that MoS verbs are manner verbs, and can be lexically decomposed into a "light verb" *say* and a manner component.[3] Given an utterance like (5a), there is a salient alternative set relative to the manner component of the MoS verb "whisper" (i.e., the set of propositions of the form "John said in x manner that Mary was in the courtyard"). Therefore, comprehenders are likely to consider the QUD to be "In what manner did John say that Mary was in the courtyard?", and thus the manner of speaking component of the MoS verb to be foregrounded.

In contrast, the verb *say*, which do not restrict syntactic movement and is thus considered a "bridge verb", is lexically lighter than the MoS verbs since it is otherwise semantically equivalent to a MoS verb except that it lacks a manner component Erteschik-Shir (2007). Therefore, *say* is less likely to activate a salient alternative set: when encountering an utterance like (5b), it is unlikely that the comprehenders would take the QUD to be "What is the two-part relation between John and the proposition 'Mary was in the courtyard'?", or "How did John communicate the proposition 'Mary was in the courtyard'?". As a result, the verb "say" in (5b) is less likely to be interpreted as foregrounded compared to "whisper" in (5a).

(5) a. John whispered that Mary was in the courtyard.
    b. John said that Mary was in the courtyard.

Since movement has a discourse foregrounding function (moved elements are focused) (Abeillé et al., 2020; Cuneo and Goldberg, 2023; Erteschik-Shir, 1973; Kuno, 1976, 1987; Takami, 1988; Van Valin and LaPolla, 1997), any movement of backgrounded wh-elements would result in an information structural clash (Ambridge and Goldberg, 2008; Erteschik-Shir, 1973, 1979, 2007). In a clausal embedding structure with MoS verbs, the matrix MoS verb is foregrounded, and thus the elements inside the complement clause are more likely to be interpreted as backgrounded. Syntactic movement from within the embedded clause

[3]Some other studies assume that MoS verbs are verbs of implicit creation, and a MoS verb can be decomposed into a light verb *make* and a nominal cognate (Snyder, 1992; Stoica, 2020). We will return to this alternative in the General Discussion.

is thus degraded due to the aforementioned information structural clash. In contrast, when the matrix verb is the lexically light bridge verb "say" that is by default backgrounded, the embedded contents are more likely foregrounded. The foregrounding of the embedded elements is compatible with the foregrounding associated with syntactic movement, and movement of embedded elements therefore does not lead to an information structural clash and associated unacceptability. The Backgroundedness Account thus captures the contrast in (2). [4]

One way in which backgroundedness has been operationalized is via the "negation test", which is motivated by the generalization that only foregrounded contents are affected by sentential negation (Erteschik-Shir, 1979; Takami, 1988; Van Valin, 1998). The negation test captures that backgrounded constituents are less likely to be interpreted as negated by a matrix sentential negation than foregrounded constituents (Ambridge and Goldberg, 2008; Cuneo and Goldberg, 2023; Goldberg, 2006; Namboodiripad et al., 2022). For example, in sentence (6a), comprehenders are more likely to rate the embedded clause *Mary was in the courtyard* as true (i.e., unaffected by the matrix negation) compared to sentence (6b), suggesting that the embedded clause is more likely interpreted as backgrounded in (6a) than in (6b).

(6) a. John didn't whisper that Mary was in the courtyard.
b. John didn't say that Mary was in the courtyard.

Using the negation test, previous experimental studies have shown a positive correlation between the degree of backgroundedness of MoS complements and the magnitude of sentence acceptability degradation due to extraction from MoS complements (Ambridge and Goldberg, 2008; Goldberg, 2013).[5] These results suggest that backgroundedness predicts MoS islandhood, and provide support for the Backgroundedness Account of MoS islands.[6]

But there is a caveat to these correlational studies: correlation does not imply causation. A proponent of the Subjacency Account might argue that it is the structural difference between (3) and (4) that gives rise to the difference in acceptability and—independently—to the difference in discourse backgroundedness. The syntactic account is thus not ruled out by the observations made by Ambridge and Goldberg (2008) and Goldberg (2013). In this study, we test whether the relationship between discourse backgroundedness and the MoS island effect is causal. If such a causal relation exists, MoS islandhood should change in response to changes in discourse backgroundedness of the extracted constituent through non-syntactic means (e.g., prosodic manipulations), while holding the syntactic structure and the lexical items of the sentence constant. Neither the Subjacency Account

[4]The account builds on the "Dominance Condition on Extraction" (Erteschik-Shir, 1973), which states that movements out of constituents that are considered dominant in the context are degraded. Generalizations equivalent to Erteschik-Shir (1973)'s Dominance Condition on Extraction have also been proposed under the name "Topichood Condition on Extraction" (Kuno, 1987), and the "Backgroundedness Constituents are Islands Generalization"(Ambridge and Goldberg, 2008; Goldberg, 2006, 2013).

[5]Backgroundedness can also be measured by other tests in addition to the negation test. For example, Cuneo and Goldberg (2022) and Cuneo and Goldberg (2023) used the "appropriate answer discourse task" to operationalize backgroundedness.

[6]For similar correlational results for a wider range of island effects, see Goldberg (2013), Cuneo and Goldberg (2022), Namboodiripad et al. (2022), and Cuneo and Goldberg (2023).

nor the Verb-Frame Frequency Account predicts a change in MoS islandhood in response to such a manipulation, because MoS islandhood is considered causally dependent on the syntactic structure or the choice of the complement-taking verbs, respectively, but not on the discourse backgroundedness of the extracted constituents.

In Experiment 1 we test whether the MoS island effect can be ameliorated by discourse foregrounding the MoS complement. In Experiments 2a, we test whether extraction from the complements of *say* can be made unacceptable by discourse foregrounding *say*, an effect predicted by a causal relationship between backgroundedness and islandhood. In Experiment 2b, we test whether the MoS verbs are by default more likely to be interpreted as foregrounded by comprehenders than the bridge verb *say*. Finally, in Experiments 3a and 3b, we test whether the MoS island effect can be replicated in sentences with the bridge verb *say* by adding manner adverbs to the matrix predicate. These experiments directly test whether the backgrounded status of MoS complements gives rise to the MoS island effect.

2 Experiment 1: Discourse Effects on MoS Islands

This acceptability judgment experiment tested whether the MoS island effect is modulated by discourse backgroundedness. Backgroundedness of MoS complements was manipulated prosodically. This manipulation is crucially based on the observation that prosodically focused elements in English are in most cases interpreted as information structurally foregrounded and addressing the current QUD (Roberts, 1996, 2012). The Backgroundedness Account predicts that extraction from inside the MoS complement should be rated as more acceptable when the complement is prosodically focused compared to when it is not. In contrast, neither the Subjacency Account nor the Verb-Frame Frequency Account predicts an effect of prosodic foregrounding on acceptability.

2.1 Methods

Participants.

We recruited 100 participants through the online crowd-sourcing platform Prolific to complete the experiment. Based on the preregistered exclusion criteria, responses from 6 participants were excluded, either because they did not self-report as native monolingual English speakers or because their ratings for unacceptable fillers were on average higher than those for acceptable fillers.[7]

Materials. Example critical and filler items are shown in (7). Each item consisted of a two-sentence written dialog. On critical trials, the first utterance was a declarative sentence with an MoS verb taking a complement clause. The second utterance was a wh-interrogative with the embedded object extracted from within the MoS complement. There were two focus conditions: in the Verb Focus condition, the matrix verb of the first utterance was capitalized and bolded, representing a prosodic focus that foregrounds

[7]The pre-registration is available at `https://osf.io/rsza5`. All data, materials, and analysis scripts for each experiment can be accessed at `https://github.com/pennydy/MOS_Island`.

the matrix verb and hence backgrounds the embedded constituents. In the Embedded Focus condition, the embedded object of the first utterance was capitalized and bolded, representing a prosodic focus that foregrounds the embedded object. The second utterance was identical across focus conditions.

(7) Example critical items in Experiment 1

a. Verb Focus condition
Hanako said: John didn't **WHISPER** that Mary met with the lawyer.
Scott said: Then who did John whisper that Mary met with?

b. Embedded Focus condition
Hanako said: John didn't whisper that Mary met with the **LAWYER**.
Scott said: Then who did John whisper that Mary met with?

(8) Example filler items in Experiment 1

a. Grammatical filler
Hanako said: **RONALD** didn't imply that Jacy rented the truck.
Scott said: Then who implied that Jacy rented the truck?

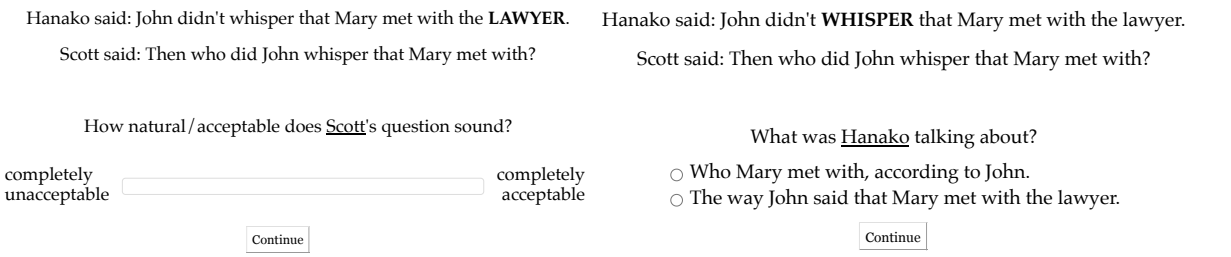
b. Ungrammatical filler
Hanako said: Tony didn't suggest that Frank and **LISA** were in the office.
Scott said: Then who did Tony suggest that Frank and were in the office?

Each participant completed 12 critical trials and 24 filler trials. For each participant, half of the 12 critical items were randomly assigned to the Verb Focus condition and the other half to the Embedded Focus condition. No participant saw an item in both the Verb and Embedded Focus condition. All participants were presented with the same set of 24 filler items, half of which were grammatical and the other half ungrammatical.

Procedures. After reading each dialog, participants completed either an acceptability judgment task (main task) or a backgroundedness task (manipulation check). On acceptability judgment trials, they rated the acceptability of the second utterance (Scott's utterance in Fig. 1) on a sliding scale with endpoints labeled "completely unacceptable" (coded as 0 for analysis) and "completely acceptable" (coded as 1 for analysis). Fig. 1a shows an example trial that probes the acceptability judgment.

On backgroundedness trials, participants answered a forced-choice question that probed the backgroundedness of the embedded object. An example of the backgroundedness task is shown in Fig. 1b. The first option represents an interpretation of the polar interrogative under which the embedded object (i.e., "the lawyer") is foregrounded, and the second option represents one under which the embedded object is backgrounded and the MoS verb (i.e., "whisper") is foregrounded.[8] The order of the two choices was randomized across trials.

[8]We deviate from Ambridge and Goldberg (2008) and Goldberg (2013) in our choice of backgroundedness diagnostic for two reasons. First, the task we used directly probes whether the embedded object (i.e., the DP that would be extracted in the second utterance) is backgrounded, as opposed to the negation test that probes whether the whole complement clause is backgrounded. Therefore, this task serves as a better manipulation check for our design. Second, the negation test is arguably a test of projection (i.e., whether certain content is presupposed). Discourse backgroundedness and projection, albeit correlated, are not the same concept (Beaver et al., 2017; Tonhauser et al., 2018). Therefore, our backgroundedness task is a more direct measurement of discourse backgroundedness than the negation test.



(a) Acceptability judgment task.

(b) Backgroundedness task.

Figure 1: Examples of the two tasks in the experiment.

For each participant, the 36 trials were divided into 6 blocks. Each block contained 2 randomly sampled critical items (one in each focus condition), 2 randomly sampled acceptable fillers, and 2 randomly sampled unacceptable fillers, in random order. In addition, regardless of the focus condition, half of the trials within each block occurred with the comprehension task and half with the rating task. Task order was randomized.

There were 4 practice trials at the beginning of the experiment, 2 acceptable and 2 unacceptable practice items, none of which contained MoS verbs. One of each was followed by an acceptability judgment task, and the other by a forced-choice comprehension question, just as on experimental trials.

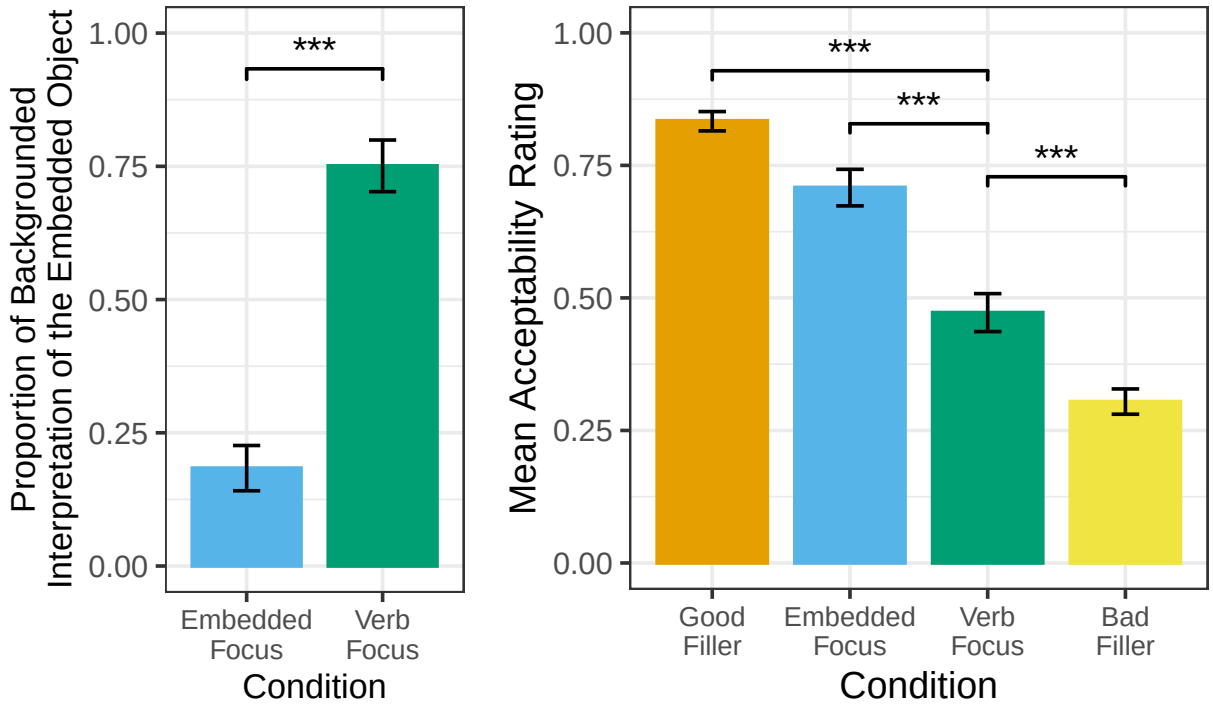
2.2 Results

The results of Experiment 1 are visualized in Fig. 2. We begin by reporting the manipulation check (Fig. 2a), followed by the effect of focus on acceptability ratings (Fig. 2b).

If the manipulation of embedded object backgroundedness via capitalization (prosodic focus) of verb or object succeeded, then the embedded object in the Embedded Focus condition should have had fewer backgrounded interpretations than in the Verb Focus condition. To test this, we fit a logistic mixed-effects regression model predicting backgrounded interpretation from a sum-coded fixed effect of focus condition (reference level: Embedded Focus), as well as by-item and by-participant random intercepts and slopes for the fixed effect.[9] There were fewer backgrounded responses in the Embedded Focus condition ($\beta = -2.46, SE = 0.40, z = -6.14$,$p < 0.001$), suggesting that the focus manipulation changed the backgroundedness of the embedded object in the expected way.

To test our main question of interest, whether embedded focus leads to higher acceptability, we fit a linear mixed-effects regression predicting acceptability from a dummy-coded fixed effect of focus condition (reference level: Verb Focus) and the maximal random effects structure that allowed the model to converge: by-item and by-participant intercepts and by-participant random slopes for the fixed effect of condition. There was a significant main effect of focus condition, such that sentence acceptability was higher in the Embedded Focus condition than in the Verb Focus condition ($\beta = 0.23, SE = 0.03, t = 7.10, p < 0.001$).

[9] All models were run using the `lmer4` package (Bates et al., 2015) in R (R Core Team, 2022). All p-values were obtained using Satterthwaite's method using the `lmerTest` package (Kuznetsova et al., 2017).



(a) Backgroundedness responses.

(b) Mean acceptability ratings.

Figure 2: Results of Experiment 1 across focus condition. Error bars indicate 95% bootstrapped confidence intervals.

In addition, in comparison to sentences in the Verb Focus condition, grammatical fillers were rated to be more acceptable ($\beta = 0.37, SE = 0.04, t = 9.46, p < 0.001$) and ungrammatical fillers were less acceptable ($\beta = -0.16, SE = 0.04, t = -3.62, p < 0.001$). This suggests that foregrounding the embedded object via prosodic focus in the context sentence attenuates the MoS island effect.

In addition, we conducted two post-hoc analyses to assess the Frequency Account of the MoS island effect. Under the Frequency Account, MoS island sentences should be more acceptable with verbs that are more likely to take a sentential complement (SC). We considered two different SC frequency measures from the previous literature.

First, Liu et al. (2019, 2022a) measure the verb-frame frequency using the joint probability of the verb taking an SC with the complementizer *that*, as defined in Equation (1). The estimates they used were collected from the Google books corpus since the year 2000 (Davies, 2011).

$$P(\text{verb}_{lemma}, \text{SC}) = P(\text{verb}_{lemma}) * P(\text{SC} \mid \text{verb}_{lemma}) \quad (1)$$

Fig. 3 shows the acceptability ratings in the two conditions against the log-transformed verb-frame frequency scores as reported in Liu et al. (2022a).[10] We added the mean-

[10]Since we could not replicate the values reported by Liu et al. (2022a) using the methods provided

centered verb-frame frequency predictor and its interaction with focus condition to the above-reported linear mixed effects model for the main analysis. The effect of focus condition remained robust ($\beta = 0.12, SE = 0.02, t = 6.71, p < 0.001$), but neither the main effect of verb-frame frequency ($\beta = -0.003, SE = 0.02, t = -0.16, p = 0.874$) nor the interaction between frequency and focus condition ($\beta = -0.004, SE = 0.01, t = -0.26, p = 0.796$) reached significance.

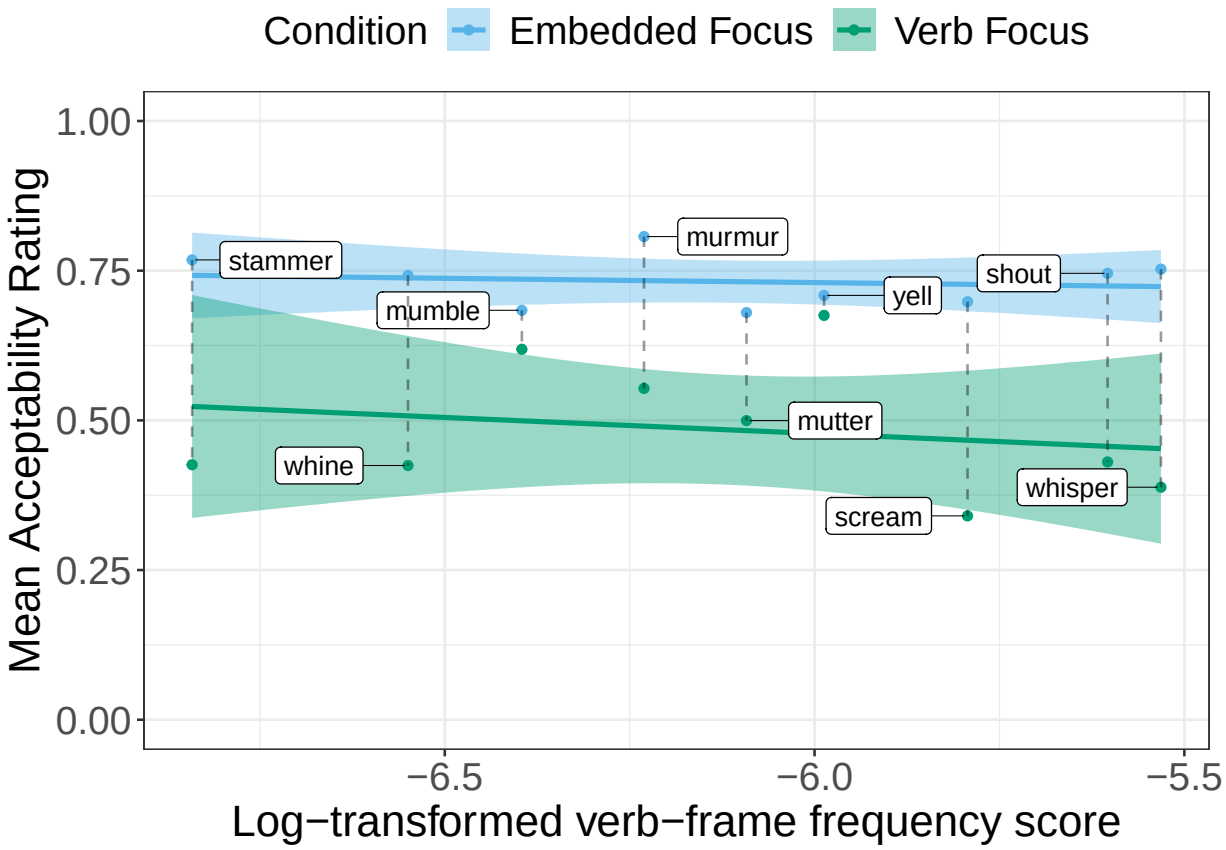


Figure 3: Mean acceptability ratings in the Verb Focus condition and the Embedded Focus condition against the log-transformed verb-frame frequency. The values are directly taken from Liu et al. (2022a). The shaded ribbons represent bootstrapped 95% confidence intervals.

As opposed to the joint probability of a verb lemma and the SC in Liu et al. (2019, 2022a), Richter and Chaves (2020) used the conditional probability of an SC given a verb lemma as the measure for verb-frame frequency, which they referred to as the *Sentence Complement Ratio* (SCR). They obtained the estimates from a random sample of the Corpus of Contemporary American English (COCA)(Davies, 2008) using the formula shown

by the original paper, we directly used the reported verb-frame frequency estimates provided at `https://osf.io/g38rj` for analysis. Three verbs that were not included in Liu et al. (2022b) ("shriek", "moan", and "groan") were therefore excluded from this analysis.

in 2.

$$\text{SCR}_{lemma} = \frac{\#\text{ (verb lemma used with SC)}}{\#\text{(verb lemma)}} \tag{2}$$

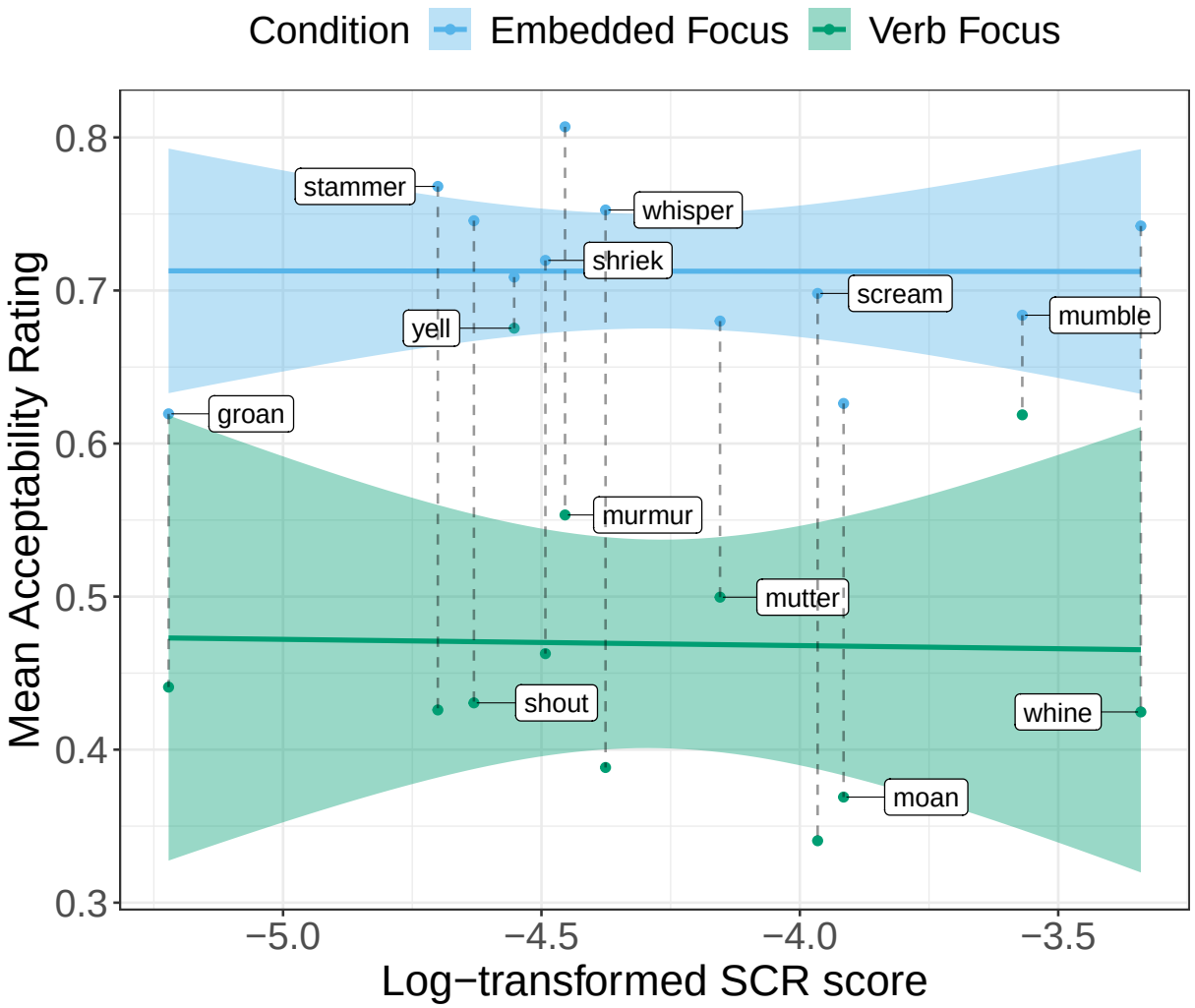


Figure 4: Mean acceptability ratings in the Verb Focus condition and the Embedded Focus condition against the log-transformed sentence complement ratio (SCR) scores. The shaded ribbons represent bootstrapped 95% confidence intervals.

Fig. 4 shows the acceptability ratings in the two conditions against the log-transformed SCR scores provided in Richter and Chaves (2020).[11] Using the same statistical model, we tested the effect of frequency as measured by the SCR scores. The effect of focus condition remained robust ($\beta = 0.12, SE = 0.02, t = 6.92, p < 0.001$), but neither the main effect

[11]The verb "groan" was not included in Richter and Chaves (2020)'s analysis. We calculated the SCR score for "groan" based on the method provided by Richter and Chaves (2020), and included it for analysis.

of SCR ($\beta = -0.0002, SE = 0.02, t = -0.02, p = 0.987$) nor the interaction between SCR and condition ($\beta = 0.008, SE = 0.01, t = 0.83, p = 0.408$) reached significance. [12]

2.3 Discussion

Our results show that foregrounding the embedded constituent ameliorates the degradedness of extracting that constituent from an MoS island, suggesting that there is a causal relationship between discourse backgroundedness and the magnitude of the MoS island effect. This provides support for the Backgroundedness Account, which attributes the degradedness of MoS island violations to a clash between the discourse backgroundedness of constituents contained in an MoS verb complement and the foregrounding wh-movement operation (Ambridge and Goldberg, 2008; Goldberg, 2013).

We did not find evidence in support of the Verb-Frame Frequency Account. Verb-frame frequency does not correlate with the acceptability of the wh-interrogatives containing an MoS island violation. More importantly, the Verb-Frame Frequency Account cannot capture the observed contrast between the focus conditions, since the comparison is between sentences with the same matrix verbs. Similarly, the Subjacency Account is not supported, either: whether the subjacency condition is violated is entirely syntactically determined and focus-independent. Therefore, subjacency cannot account for the contrast between the two focus conditions.

The findings of Experiment 1 lead to an important novel prediction: if the MoS island effect is truly the result of foregrounding the MoS verb, one should be able to create island effects using the same prosodic foregrounding manipulation as in Experiment 1 for verbs that don't typically exhibit island effects, like the bridge verb *say*. The reason that MoS complements are considered islands while *say* complements are not is that MoS verbs are by default more likely to be interpreted as foregrounded than *say*. In Experiments 2a and 2b, we put these predictions to test.

3 Experiment 2

In Experiment 1, we showed that the MoS island effect can be attenuated by discourse foregrounding the embedded object with a context sentence. In Experiment 2a, we extend the manipulation to the bridge verb *say*, and show that we can *create* a "*say* island effect" by foregrounding the bridge verb *say* using the same prosodic manipulation as in Experiment 1. In Experiment 2b, we explore why only the complements of MoS verbs but not the complements of *say* are considered islands in the literature, despite that they behave similarly in Experiment 2a. Specifically, we examine whether *say* is by default more likely to be interpreted as backgrounded than MoS verbs, and whether such interpretive contrast

[12] An anonymous reviewer suggested that as shown in Figures 3 and 4, the acceptability contrast between the two focus conditions varies across different MoS verbs. The variation is not captured by verb-frame frequency, since there is no significant interaction between SCR/VFF and condition; the variation also does not seem to be captured by any lexical properties of the verbs, or reflect any fine-grained taxonomy of MoS verbs. We leave this puzzle for future research.

maps to a difference in the extraction acceptability of the embedded contents when no prosodic manipulation is present.

3.1 Experiment 2a: Discourse Effects on MoS verbs and *say*

3.1.1 Methods

Participants. A total of 100 participants who did not participate in Experiment 1 were recruited through Prolific to complete the experiment. With the same pre-registered exclusion criteria as the previous experiment, responses from 3 participants were excluded because their ratings for unacceptable fillers were on average higher than those for acceptable fillers.

Materials. To replicate the finding in Experiment 1, we used the same 12 critical items with MoS verbs from Experiment 1 and followed the same experimental design. Moreover, we also included 12 additional critical items with *say* as the matrix predicate, resulting in a total of 24 critical items. Example critical items are shown in 9. As in Experiment 1, there were two focus conditions: in the Verb Focus condition, the matrix verb (either *say* or an MoS verb) in the first utterance was capitalized and bolded. In the Embedded Focus condition, the embedded object in the first utterance was capitalized and bolded. The second utterance was the same in both conditions. In addition, all participants saw the same set of 12 fillers, with 6 grammatical fillers and 6 ungrammatical fillers taken from Experiment 1.

(9) Example critical items in Experiment 2a

a. Verb Focus condition
Hanako said: John didn't **WHISPER/SAY** that Mary met with the lawyer.
Scott said: Then who did John whisper/say that Mary met with?

b. Embedded Focus condition
Hanako said: John didn't whisper/say that Mary met with the **LAWYER**.
Scott said: Then who did John whisper/say that Mary met with?

Taken together, there were 24 critical items, including 12 items with MoS verbs and 12 items with *say*, as well as 12 filler items. Each participant viewed each of the 12 MoS verbs once, with 6 items in the Verb Focus condition and 6 in the Embedded Focus condition. Likewise, they viewed each of the 12 *say* critical items once, with 6 in the Verb Focus condition and 6 in the Embedded Focus condition.

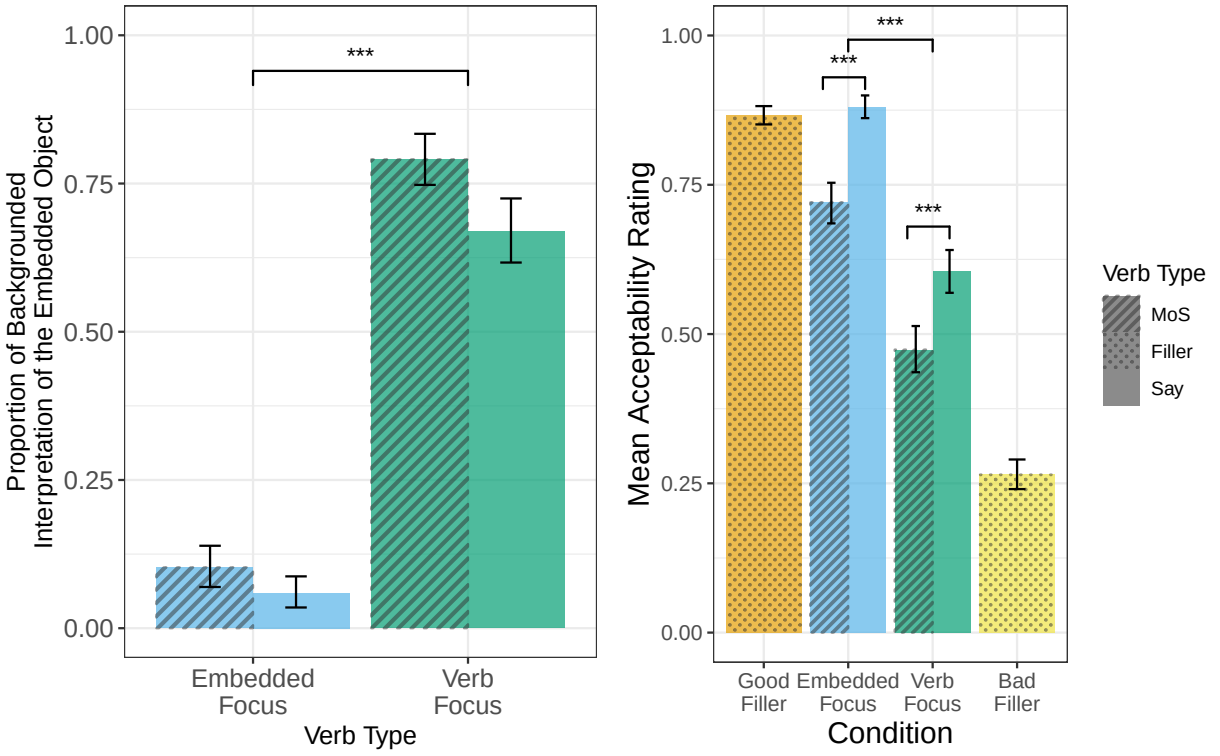
Procedure. Participants were either asked to rate the acceptability of the second utterance (acceptability judgment task) or answer a two-alternative forced-choice question that probed the backgroundedness of the embedded object (backgroundedness task). Example questions for each task are shown in Fig. 1. The conditions and tasks were randomized using the same block design as Experiment 1. For each participant, the 24 critical items and 12 fillers were divided into 6 blocks. Each block contained 2 randomly selected critical items with MoS verbs (one from each focus condition), 2 randomly selected critical items

with *say* (one from each focus condition), 1 randomly selected acceptable filler, and 1 randomly selected unacceptable filler. The order of items in each block was randomized. For critical items, half of the critical trials in each block were paired with the comprehension task and half with the rating task, regardless of the focus condition. All filler items were paired with the rating tasks. The task order was also randomized.

At the beginning of the experiment, participants saw 4 practice trials with 2 acceptable and 2 unacceptable practice items, none of which contained the critical matrix verbs. One of the two items of each type was paired with an acceptability judgment task, while the other with a two-alternative forced-choice comprehension task. All practice items included focus manipulation, and the participants were informed that the highlight and capitalization were intended to mimic the emphasis on the word or phrase after they provided their response in the comprehension task.

3.1.2 Results

The results of Experiment 2a are shown in Fig. 5. Following the analyses in Experiment 1, we first analyzed the results of the comprehension task, followed by the results of the acceptability rating task.



(a) Backgroundedness responses.

(b) Mean acceptability ratings.

Figure 5: Results of Experiment 2a across focus condition. Error bars indicate 95% bootstrapped confidence intervals.

To test the effects of the context sentence and verb type on the backgroundedness of the embedded object, we fit a logistic mixed-effects regression model. To predict the backgroundedness interpretation of the embedded object, we included the sum-coded condition (reference level: Embedded Focus) and verb types (reference level: MoS verbs) as the main effects as well as their interaction, along with the maximal random effect structure that allows the model to converge (Barr et al., 2013), which includes the by-participant random intercept and slope for the fixed effect of condition. There were fewer backgrounded responses in the Embedded Focus condition than in the Verb Focus condition ($\beta = -5.21, SE = 0.90, z = -5.82, p < 0.001$), suggesting that as expected, having the focus manipulation on the embedded object made it less likely to be interpreted to be discourse backgrounded, similar to the results in Experiment 1. In addition, there was a main effect of verb type, such that the embedded object of *say* had fewer backgrounded interpretations than when it was embedded under the MoS verbs ($\beta = 0.59, SE = 0.14, z = 4.27, p < 0.001$). The interaction between the verb type and focus condition was not significant ($\beta = -0.07, SE = 0.14, z = -0.49, p = 0.623$).

For the main analysis, we fit a linear mixed-effects regression predicting the acceptability rating from the sum-coded fixed effects of condition (reference level: Embedded Focus) and verb type (reference level: MoS verbs) and the by-item and by-participant random intercepts and the by-item slope for the focus condition and the by-participant random slopes for both focus condition and verb type as well as their interaction.[13] There was a significant effect of focus condition, such that the acceptability ratings in the Embedded Focus condition were higher than those in the Verb Focus condition ($\beta = 0.14, SE = 0.02, t = 9.14, p < 0.001$), similar to the results of focus manipulation to MoS verbs in Experiment 1. In addition, there was a main effect of verb type, such that the acceptability ratings for sentences with MoS verbs were lower, in comparison to those for sentences with *say* ($\beta = -0.08, SE = 0.01, t = -5.49, p < 0.001$). On the other hand, the interaction between focus condition and verb type was not significant ($\beta = 0.00, SE = 0.01, t = 0.67, p = 0.509$), suggesting that there is no evidence that the focus manipulation affected the acceptability ratings differently for *say* and MoS verbs.

[13]Only the data from *say* and MoS verbs were included in the analysis since filler items do not have the full focus condition and verb type combination.

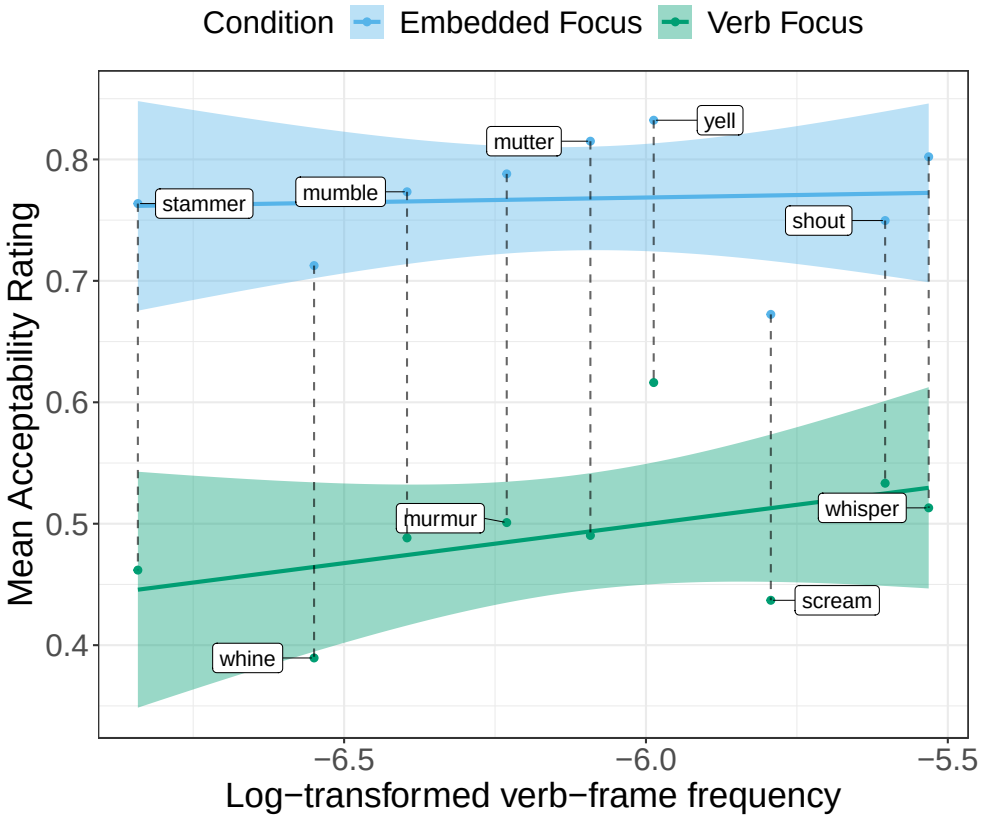


Figure 6: Mean acceptability ratings in the Verb Focus condition and the Embedded Focus condition against the log-transformed verb-frame frequency. The values are directly taken from Liu et al. (2022a). The shaded ribbons represent bootstrapped 95% confidence intervals.

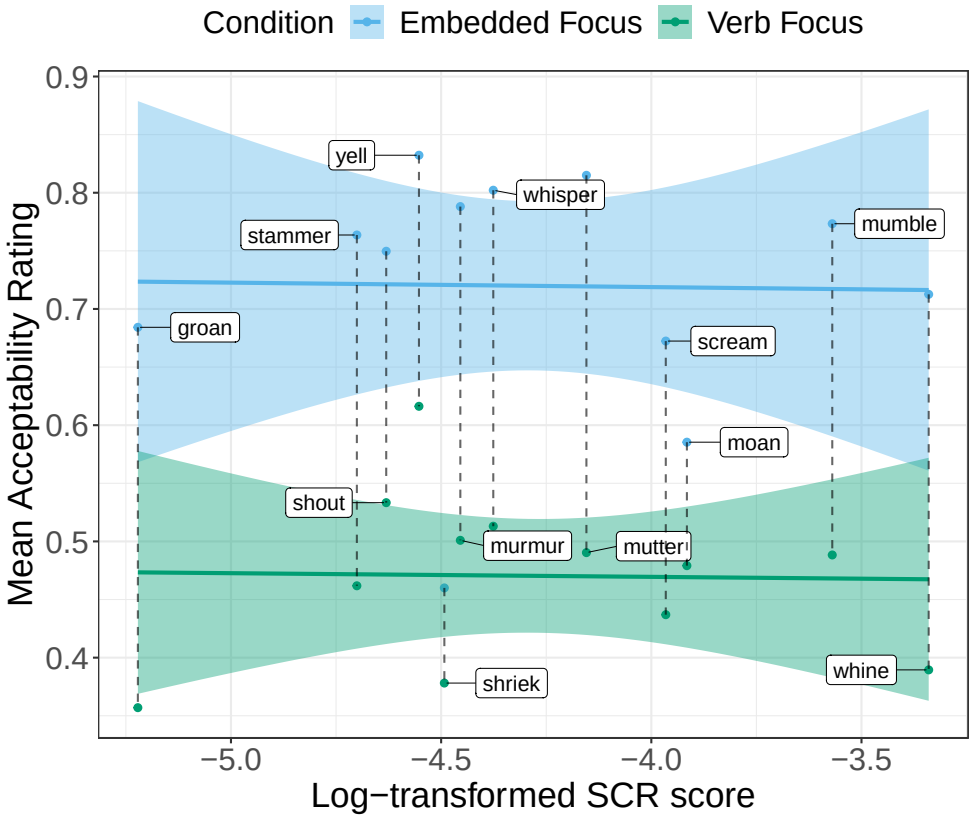


Figure 7: Mean acceptability ratings in the Verb Focus condition and the Embedded Focus condition against the log-transformed sentence complement ratio (SCR) scores. The shaded ribbons represent bootstrapped 95% confidence intervals.

Moreover, we also tested the effect of frequency on the acceptability ratings for each MoS verb with both the verb-frame frequency measure and sentential ratio (SCR) score, and the results are shown in Fig. 6 and Fig. 7 respectively. We fit a mixed-effect linear regression model predicting the acceptability ratings from sum-coded fixed effects of condition (reference level: Embedded Focus) and the log-transformed and mean-centered verb-frame frequency values as well as the maximal random effect structure justified by the design, which includes the by-item random intercept and random slope for the effects of condition and the by-participant random intercept and random slopes for the effects of condition, verb-frame frequency, and their interaction. The effect of condition remained robust ($\beta = 0.14, SE = 0.02, t = 8.11, p < 0.001$), but crucially, neither the effect of verb-frame frequency ($\beta = 0.01, SE = 0.02, t = 0.71, p = 0.498$) nor its interaction with condition ($\beta = -0.004, SE = 0.01, t = -0.32, p = 0.753$) was significant. The same model was used for the SCR frequency measure and yielded the same pattern of results, where there was a significant effect of condition ($\beta = 0.13, SE = 0.02, t = 7.74, p < 0.001$), yet neither the SCR ($\beta = 0.01, SE = 0.03, t = 0.27, p = 0.791$) nor its interaction with condition reached significance ($\beta = -0.004, SE = 0.01, t = -0.35, p = 0.728$).

3.1.3 Discussion

First, foregrounding the embedded objects via a prosodically manipulated context sentence attenuated the MoS island effect, replicating the results of Experiment 1. This is in line with the predictions of the Backgroundedness Account, which attributes the MoS island effect to the discourse status of the embedded object: the more likely the embedded object is to be interpreted as backgrounded, the less acceptable it is to extract it.

In addition, replicating the results in Experiment 1, we did not find evidence supporting the Verb-Frame Frequency Account. With either the verb-frame frequency measure or the SCR score, the frequency of the MoS verb cannot predict the acceptability of sentences that contain extraction of the embedded content. Neither can frequency account for the contrast between the two focus conditions. A similar argument also applies to the Subjacency Account: since the target wh-interrogatives containing an MoS island violation across the two focus conditions have the same MoS verbs as the matrix verb, the two sentences should have the same syntactic structure and thus would not have different acceptability ratings between the two conditions.

Finally, extraction of the embedded object from the complement of *say* is less acceptable when the verb *say* is foregrounded than when the embedded object is foregrounded. This confirms our prediction that we can create a "*say* island effect" if we force comprehenders to interpret the verb *say* as foregrounded. This finding provides further support for the causal relationship between discourse backgroundedness and islandhood.

This finding raises an important question: if the complements of both MoS verbs and *say* behave like islands when the matrix verb is foregrounded, then why are only MoS complements labeled as islands? One possible hypothesis is that MoS verbs, due to the extra manner component in their semantics, are by default more likely to be interpreted as foregrounded than *say* when no overt prosodic cue is provided. We tested this hypothesis in Experiment 2b.

3.2 Experiment 2b: Default Backgroundedness Interpretation and Island Effects of *Say* and MoS Verbs

3.2.1 Methods

Participants. A total of 101 participants who did not participate in the previous experiments were recruited through Prolific. Following the same preregistered exclusion criteria, responses from 7 participants were excluded, either because they did not self-report as native monolingual English speakers or because their ratings for ungrammatical fillers were on average higher than their ratings for grammatical fillers.

Materials. The design was similar to the previous experiments but did not include a focus manipulation. Each participant saw all 24 critical items, including 12 items with MoS verbs and 12 with *say*, and 24 filler items, including 12 grammatical items and 12 ungrammatical items. Each item was in the form of a two-sentence dialog presented in text. Example stimuli are shown in 10. The first sentence was a declarative sentence with either an MoS verb (MoS condition) or *say* (Say condition). The second utterance was

a wh-question involving extracting the embedded object from the embedded complement clause. Each participant viewed each of the critical items once.

(10) Example critical items in Experiment 2b

a. MoS condition
Hanako said: John didn't whisper that Mary met with the lawyer.
Scott said: Then who did John whisper that Mary met with?

b. Say condition
Hanako said: John didn't say that Mary met with the lawyer.
Scott said: Then who did John whisper/say that Mary met with?

(11) Example fillers from Experiment 2b

a. Grammatical filler
Hanako said: Ronald didn't imply that Jacy rented the truck.
Scott said: Then who implied that Jacy rented the truck?

b. Ungrammatical filler
Hanako said: Tony didn't suggest that Frank and Lisa were in the office.
Scott said: Then who did Tony suggest that Frank and were in the office?

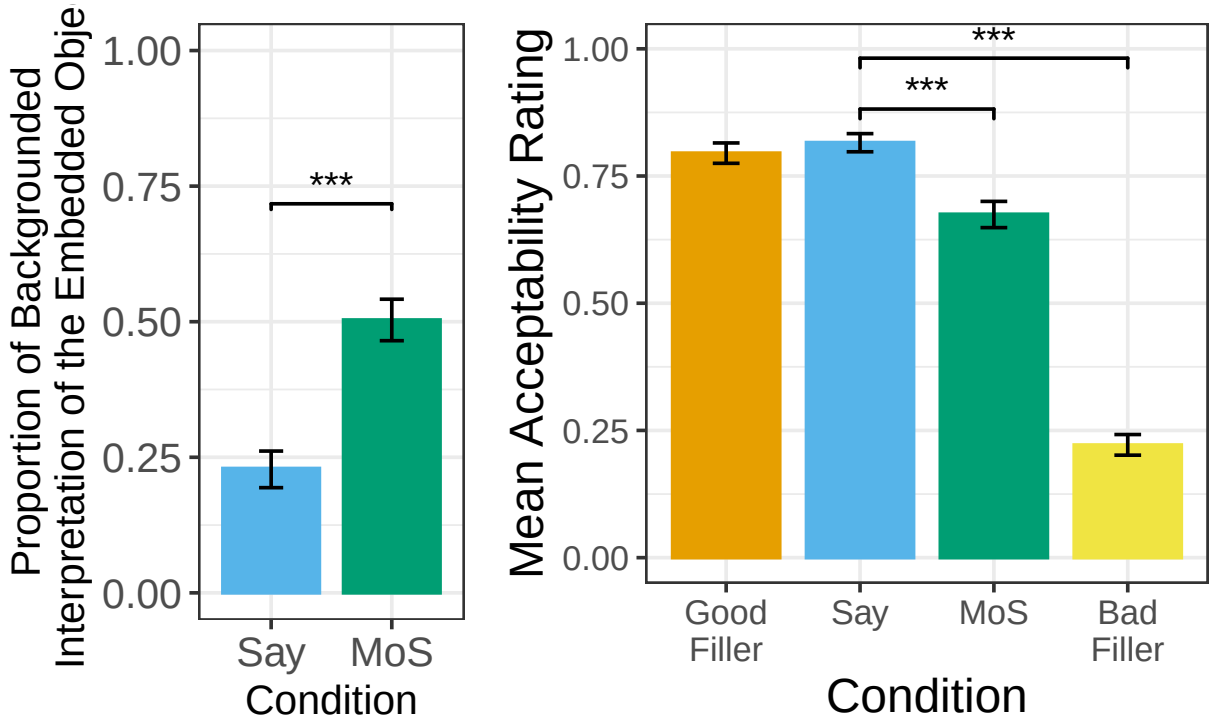
Procedure. Just as in the previous experiments, participants either rated the acceptability of the second utterance (acceptability judgment task) or answered a two-alternative forced-choice question about the backgroundedness of the embedded object (backgroundedness task). See Fig. 1 for example questions of the two tasks. To randomize the task and condition, we divided the 24 critical items and 24 fillers into 12 blocks. Each block contained 2 critical items, one from each condition, 1 grammatical filler, and 1 ungrammatical filler. The task was randomly chosen for each trial, such that in each block, half of the trials were paired with the acceptability rating task and the half with the backgroundedness task. We randomly selected items from each condition, and both the block order and the order of the items within each block were also randomized.

At the start of the experiment, participants saw 4 practice trials with 2 acceptable and 2 unacceptable practice items. Similar to the experimental trials, half of the items were followed by an acceptability judgment task, and the other half by a forced-choice comprehension question.

3.2.2 Results.

Fig. 8 shows the results of the backgroundedness and the acceptability judgment tasks.

As shown in Fig. 8a, the embedded objects of *say* received fewer backgrounded responses than those of MoS verbs. This result is borne out statistically. We fit a logistic mixed-effects regression model predicting backgroundeness from a sum-coded fixed effect of verb type (reference level: MoS verbs) along with the by-item random intercept and by-participant random intercept and slope for verb type. There was a main effect of verb type ($\beta = 0.96, SE = 0.16, z = 6.06, p < 0.001$), suggesting that the embedded object of MoS verbs



(a) Backgroundedness responses.

(b) Mean acceptability ratings.

Figure 8: Results of Experiment 2b across conditions. Error bars indicate 95% bootstrapped confidence intervals.

was more likely to be interpreted as being backgrounded than when it was embedded under *say*.

Fig. 8b shows the mean acceptability ratings in the four conditions. We fit a linear mix-effects regression model predicting the acceptability ratings from a dummy-coded fixed effect of verb type (reference level: Say), with the maximal random effects structure that allowed the model to converge, which includes the by-participant random intercept and random slope of verb type. There was a main effect of verb type, such that the acceptability ratings in the Say condition were higher than those in the MoS condition ($\beta = -0.14, SE = 0.02, t = -9.26, p < 0.001$) and those of the ungrammatical fillers ($\beta = -0.59, SE = 0.03, t = -22.45, p < 0.001$). The ratings were not significantly different between the Say condition and the grammatical fillers ($\beta = -0.02, SE = 0.01, t = -1.83, p = 0.067$).

For the frequency analysis, we used the two frequency measures of each MoS verb as in Experiment 1 with the verb-frame frequency values from Liu et al. (2022a) and the SCR scores from Richter and Chaves (2020). The acceptability ratings against the verb-frame frequency values and SCR scores are shown in Fig. 9 and in Fig. 10, respectively. To test the effect of frequency on acceptability for MoS verbs, we fit a linear mixed-effects regression model predicting acceptability ratings in the MoS condition from a fixed effect of the log-transformed and mean-centered verb-frame frequency value or the SCR score and the maximal random effect structure, which includes the by-item random intercept and

the by-participant random intercept and the random slope for the fixed effect. There was no effect of frequency, neither with the verb-frame frequency measure ($\beta = -0.001, SE = 0.02, t = -0.03, p = 0.981$) nor with the SCR ($\beta = 0.008, SE = 0.03, t = 0.32, p = 0.754$).

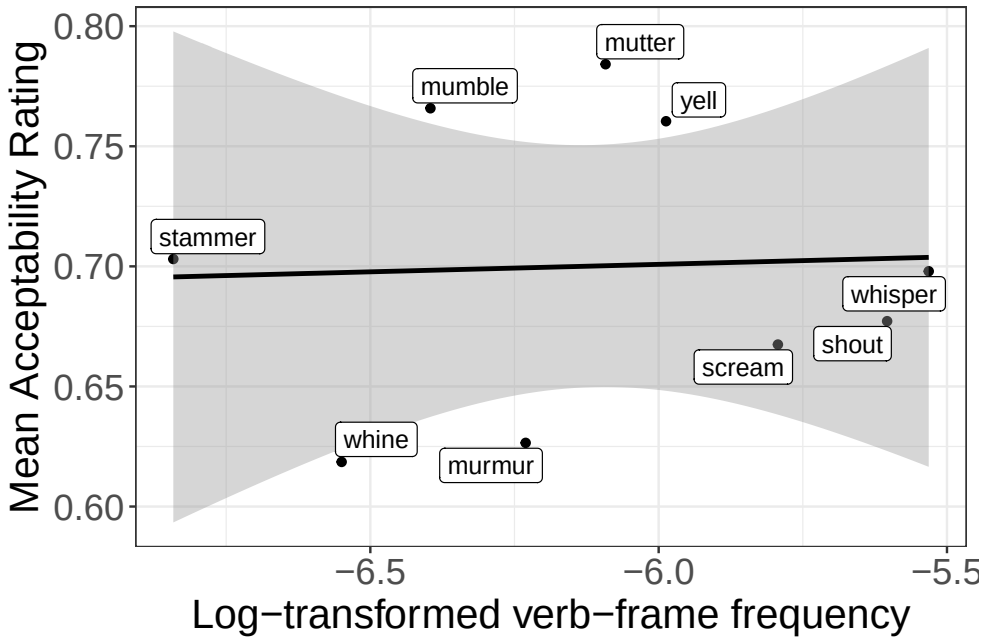


Figure 9: Mean acceptability ratings in the MoS condition against the log-transformed verb-frame frequency. The values are directly taken from Liu et al. (2022a). The shaded ribbons represent bootstrapped 95% confidence intervals.

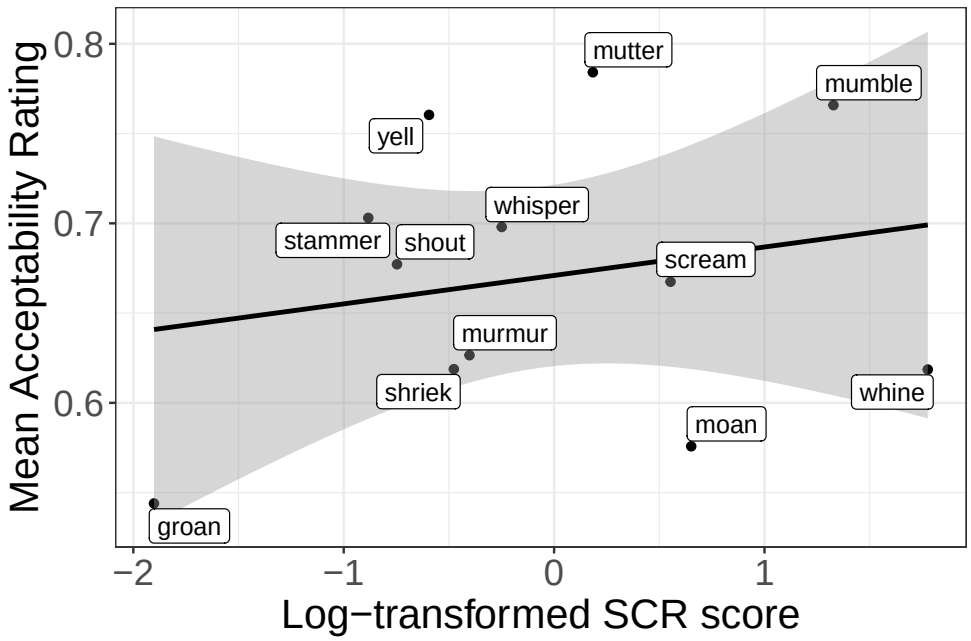


Figure 10: Mean acceptability ratings in the MoS condition against the log-transformed sentence complement ration (SCR) scores. The shaded ribbons represent bootstrapped 95% confidence intervals.

3.2.3 Discussion.

In Experiment 2b, we investigated the default backgroundedness status of the embedded object in the complement of *say* and MoS verbs, without introducing any focus manipulation, as well as the acceptability of extracting that embedded object from the embedded clause. The results of the backgroundedness task show that the embedded object of *say* is by default more likely to be interpreted as discourse backgrounded than the embedded object of MoS verbs. This confirms the assumption of the Backgroundedness Account, which suggests that elements in the complement clauses of MoS verbs are default discourse backgrounded (Ambridge and Goldberg, 2008; Goldberg, 2013).

In addition, the acceptability task confirms that without any focus manipulation in place, extracting the embedded object from the complement clause of MoS verbs is degraded. On the other hand, *say* as a light verb does not show the island effect.

Moreover, among the attested MoS verbs, frequency is not a significant predictor of the acceptability of wh-interrogatives containing MoS-island violation, even when the discourse status of the embedded clause was not controlled for. Hence, there is no evidence in support of the Frequency Account.

4 Experiment 3

In Experiments 2a and 2b, we saw that MoS verbs are by default more likely to be interpreted as foregrounded than *say* is. Moreover, this interpretive contrast is associated with an acceptability contrast in extraction from MoS complements versus *say* complements. What is the reason for this greater baseline probability of MoS verb foregroundedness? Following (Erteschik-Shir, 2007), we hypothesize that MoS verbs are lexically composed of the light verb *say* and a manner component. Due to the extra manner component, MoS verbs are more likely to activate a salient set of alternative verbs than simple *say*, and are thus more likely to be interpreted as addressing the QUD. As a result, MoS verbs tend to be interpreted as foregrounded, which, under the common assumption that there is typically only one QUD active at a time (Beaver et al., 2017; Büring, 2003; Roberts, 1996, 2012, *inter alia*), results in their complements being backgrounded. One natural prediction of this hypothesis is that by modifying the light verb *say* with manner adverbs, we should be able to replicate the MoS island effect even without using MoS verbs. We tested this prediction in Experiment 3. In Experiment 3a, we tested whether adding manner adverbs to *say* results in the complement clause being backgrounded, and hence, turned into an island. In Experiment 3b, we tested whether the MoS-like island effect observed in Experiment 3a behaves like the regular MoS island effect in showing sensitivity to prosodic discourse manipulations.

4.1 Experiment 3a: Replicating MoS Islands using Manner Adverbs

4.1.1 Methods

Participants. A total of 100 participants who did not participate in in the previous experiments were recruited through Prolific to complete the experiment. Based on the preregistered exclusion criteria, responses from 7 participants were excluded, either because they did not self-report as native monolingual English speakers or because their ratings for unacceptable fillers were on average higher than those for acceptable fillers.

Materials. Similar to the design of Experiment 1, each participant was presented with 12 critical items and 24 filler items. Each item was in the form of a two-sentence dialog presented in text. Example stimuli are shown in (12). The first utterance was a declarative sentence with either *say* (Say condition) or *say* followed by an adverb (Say + Adverb condition) taking a complement clause. The second utterance was a wh-question with the embedded object moved from within the embedded complement. Each participant viewed each critical item once, with 6 items in the Say condition and 6 in the Say + Adverb condition. The 6 adverbs in the Say + Adverb condition were distinct and randomly selected from a total set of 12 adverbs. In addition, all participants were presented with the same set of 24 filler items, half of which are grammatical, and the other half ungrammatical, similar to those used in Experiment 2b shown in (11).

(12) Example stimuli from Experiment 3a

a. Say condition
   Hanako said: John didn't say that Mary met with the lawyer.
   Scott said: Then who did John say that Mary met with?

b. Say + Adverb condition
   Hanako said: John didn't say softly that Mary met with the lawyer.
   Scott said: Then who did John say softly that Mary met with?

Procedures. After reading each dialog, participants were asked to rate the acceptability of the second utterance on a sliding scale with endpoints labeled "completely unacceptable" (coded as 0 for analysis) and "completely acceptable" (coded as 1 for analysis). The pairing between the embedded clause and the matrix verb was randomized for both critical and filler trials. We divided 12 critical items and 24 fillers into 6 blocks. Each block contained 2 critical items, one from each condition, 2 acceptable fillers, and 2 unacceptable fillers. We randomly selected items from each condition, and both the block order and the order of items within each block were randomized.

4.1.2 Results

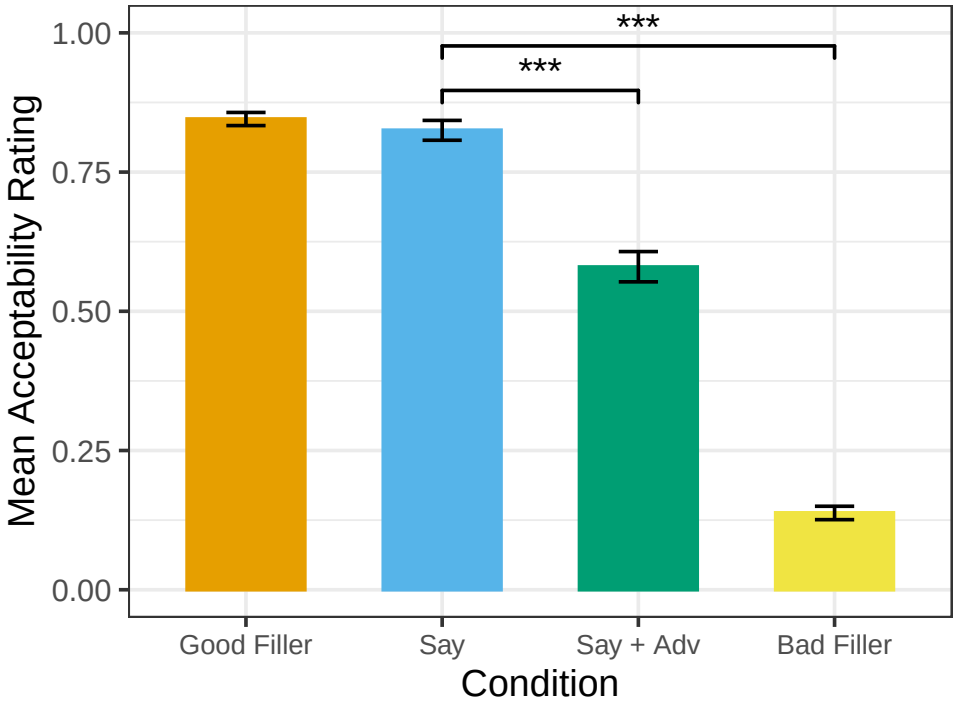


Figure 11: Mean acceptability ratings in different conditions. Error bars indicate 95% bootstrapped confidence intervals.

Fig. 11 shows the mean acceptability ratings in the four conditions. To test whether the MoS island effect can be replicated in the "*say* + adverb" construction, we fit a linear mixed-effects regression predicting acceptability from a dummy-coded fixed effect of condition

(reference level: Say) with the maximal random effects structure that allowed the model to converge: by-item and by-participant random intercepts as well as by-participant random slopes for the fixed effect of condition. There was a significant main effect of condition, such that the acceptability ratings in the Say condition were higher than those in the Say + Adv condition ($\beta = -0.24, SE = 0.02, t = -12.4, p < 0.001$), and than those of the ungrammatical fillers ($\beta = -0.69, SE = 0.02, t = -28.9, p < 0.001$). In addition, the ratings were not significantly different between the Say condition and the grammatical fillers ($\beta = 0.02, SE = 0.02, t = 1.09, p = 0.279$). These results indicate that extracting the embedded constituent of a "*say* + adverb" construction was more degraded than extracting the complement of *say*, replicating the MoS island effect.

For the frequency measure, we first used COCA (Davies, 2008) to obtain two raw counts of each "*say* + adverb" construction: the total number of occurrences of the lemmatized *say* followed by the adverb and the number of occurrences of that construction with a sentential complement preceded by the complementizer *that* (excluding direct quotes).[14] We use the obtained frequency as the predicate-frame frequency measure and computed the SCR score of each construction as formulated in Equation (3).

$$\text{SCR}_{say+adverb} = \frac{\#\ (\text{say}_{lemma}\ \text{adverb, SC})}{\#(\text{say}_{lemma}\ \text{adverb})} \tag{3}$$

The acceptability ratings against the predicate-frame frequency and the SCR score are shown in Fig. 12 and Fig. 13, respectively. To test the effect of frequency on acceptability, we fit a linear mixed-effects regression model predicting acceptability ratings in the Say + Adverb condition from a main effect of the log-transformed and mean-centered predicate-frame frequency obtained in the corpus study and the maximal random effects structure justified by the design: by-participant and by-item random intercepts and random slopes for the fixed effect. There was no significant main effect of frequency ($\beta = -0.005, SE = 0.01, t = -0.44, p = 0.664$). Likewise, we used the same statistical model with the SCR measure, and there was no significant main effect of SCR ($\beta = -0.003, SE = 0.01, t = -0.26, p = 0.793$). These results indicate that the acceptability of extracting the embedded object from a "*say* + adverb" construction does not vary by the relative frequency of that "*say* + adverb" construction taking a sentential complement.

4.1.3 Discussion

The results of Experiment 3a show that when the matrix verb is the bridge verb *say*, extraction from the complement clause is more degraded when *say* is modified by MoS adverbs (e.g., loudly, softly) compared to when no adverb is present. This suggests that the MoS island effect can be replicated even in sentences without MoS verbs by adding manner adverbs to the matrix predicate. This observation is predicted by the Backgroundedness

[14] We did not include cases where the sentential complement was not preceded by the complementizer *that*. This is theoretically motivated since the complement clauses of MoS verbs resist complementizer dropping (Pesetsky, 1995; Snyder, 1992; Stoica, 2016; Stowell, 1981; Zwicky, 1971). In addition, all context and target sentences presented in the current study contain the complementizer, and thus we only include occurrences where the SC is preceded by the complementizer *that*.

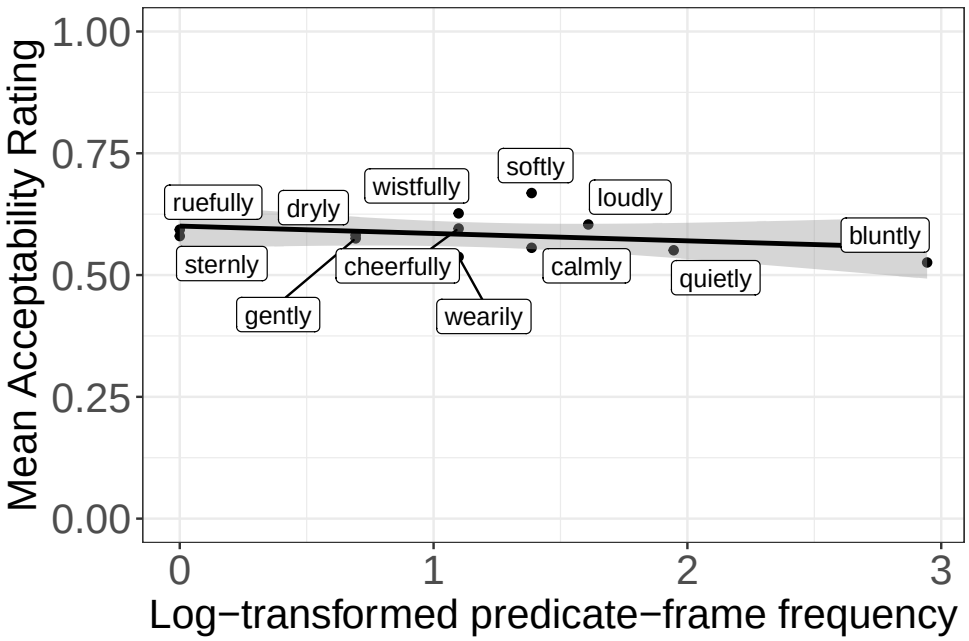


Figure 12: Mean acceptability ratings in the Say + Adverb condition against the log-transformed frequency of the predicate-frame. The shaded ribbon represents bootstrapped 95% confidence intervals.

Account: the added manner adverbs have the same discourse backgrounding effect on the complement clause as MoS verbs do, resulting in an island effect.

On the other hand, predicate-frame frequency did not predict the acceptability ratings in the "*say* + adverb" construction, which suggests a lack of evidence for the Frequency Account. In addition, since sentences in the two conditions both have *say* as their matrix verb and only differ in terms of the presence or absence of the following adverb, none of them contained the complex-NP structure that would lead to the violation of the subjacency condition. Hence, the Subjacency Account is also not supported.

One caveat of the current experiment is that the critical contrast is between two sentence types of different length and structural complexity: the Say + Adverb condition sentences are longer and more complex than the Say condition sentences due to the extra manner adverb. It is possible that the length and structural complexity difference could have contributed to the acceptability contrast. In Experiment 3b, we further verify that the MoS island-like effect we observed in Experiment 3a is discourse-driven. Specifically, we tested whether using context sentences to discourse foreground the embedded object in the "*say* + adverb" construction would ameliorate the penalty on extracting the embedded object, analogous to the results of Experiment 1. [15]

[15]An anonymous reviewer suggested an alternative way of testing whether the result of Experiment 3a is a length/structural complexity effect or a discourse effect: if the degradedness of the Say + Adverb condition is due to length/structural complexity, the unacceptability should persist even when the adverb is a non-manner adverb that cannot help foreground the predicate (e.g., *actually*). We leave this potential follow-up to future work.

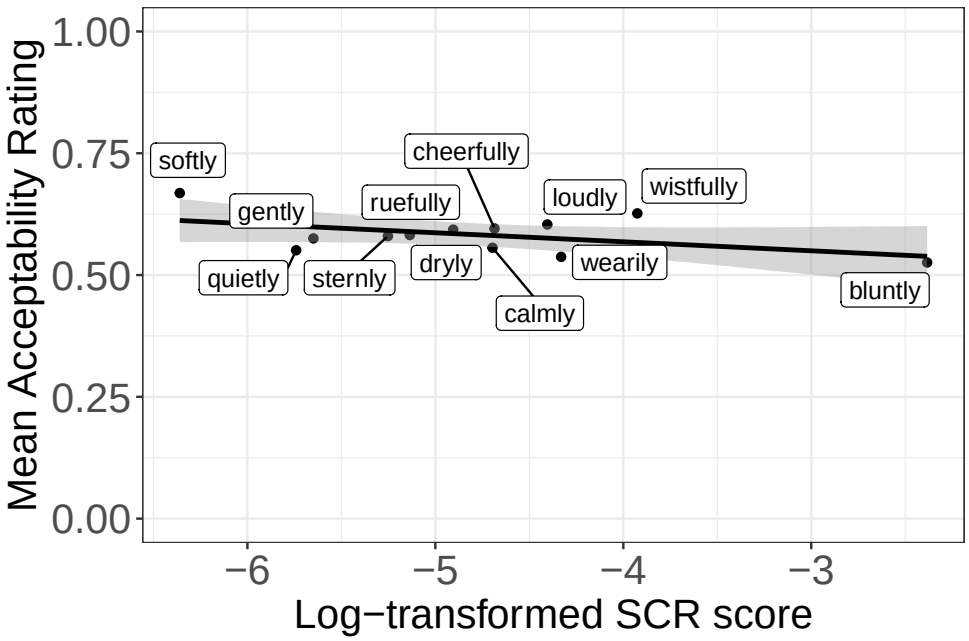


Figure 13: Mean acceptability ratings in the Say + Adverb condition against the log-transformed sentence SCR scores. The shaded ribbon represents bootstrapped 95% confidence intervals.

4.2 Experiment 3b: Discourse Effect on Islands Induced by Manner Adverbs

In Experiment 3a, we observed the MoS island effect in the "*say* + adverb" construction. In this experiment, we tested the discourse effect on MoS islands with "*say* + adv" construction to verify that the MoS island-like effect observed in Experiment 3a behaves like the regular MoS island effect.

As in Experiment 1, we manipulated the backgroundedness of the constituent in the complement clause of MoS verbs using context sentences that correspond to different prosodic patterns. We hypothesized that when the complement clause selected by a "*say* + adverb" construction is more foregrounded (i.e., in the Embedded Focus condition), extracting from that clause should be more acceptable than when it is discourse backgrounded.

4.2.1 Methods

Participants. A total of 100 participants who did not participate in the previous two experiments were recruited through Prolific to complete the experiment. Based on the preregistered exclusion criteria, responses from 6 participants were excluded, either because they did not self-report as native monolingual English speakers or because their ratings for unacceptable fillers were on average higher than those for acceptable fillers.

Materials. The experimental materials and procedure were identical to Experiment 1,

except that all MoS verbs in the stimuli were replaced by "*say* + adverb" combinations. Example stimuli are shown in (13). There were two experimental conditions: the Adverb Focus condition, in which the adverb following *say* in the first utterance was capitalized and bolded, and the Embedded Focus condition, in which the embedded object in the first utterance was capitalized and bolded. The second utterance was identical in both focus conditions. All participants were presented with the same set of 24 filler items, which were identical to those used in Experiment 1 (examples shown in 8).

(13) Example stimuli from Experiment 3b

a. Adverb Focus condition
Hanako said: John didn't say **SOFTLY** that Mary met with the lawyer.
Scott said: Then who did John say softly that Mary met with?

b. Embedded Focus condition
Hanako said: John didn't say softly that Mary met with the **LAWYER**.
Scott said: Then who did John say softly that Mary met with?

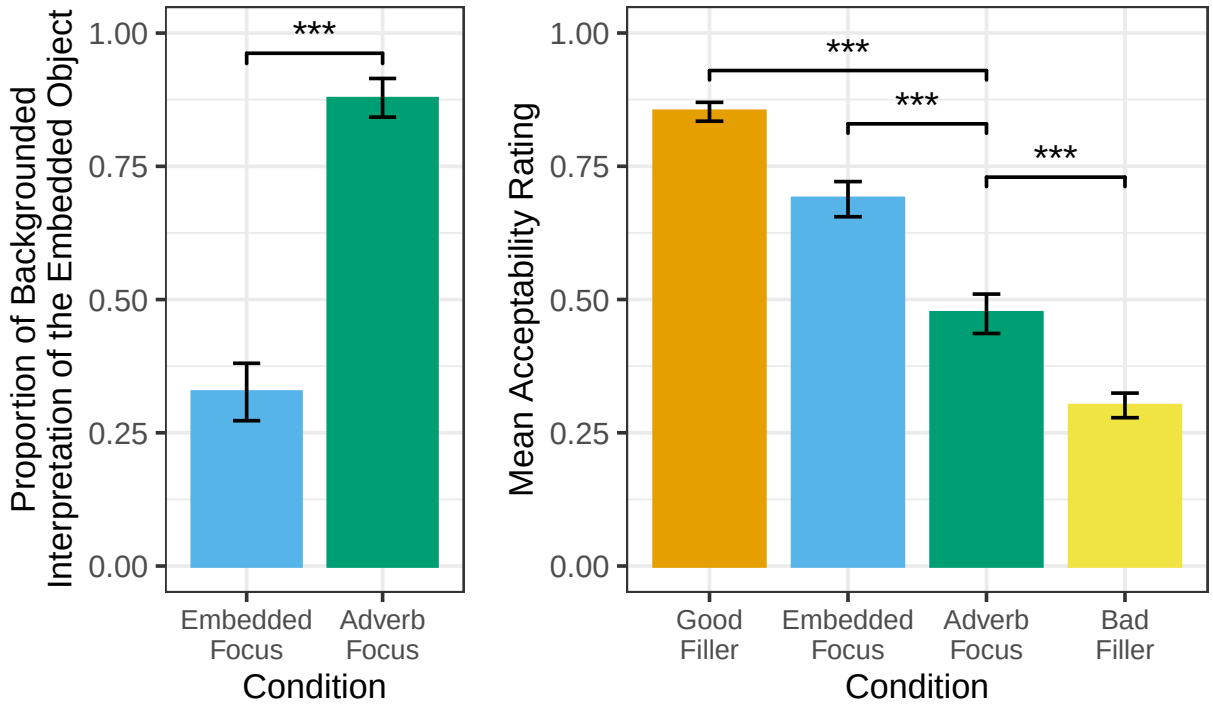
Procedures. The same rating tasks and comprehension tasks as in Experiment 1 were used. See Fig. 1b for example questions of the comprehension task used to probe the backgroundedness of the extracted content.

4.2.2 Results

Similar to the analyses in Experiment 1, we first tested the effect of the context sentence on the backgroundedness of the embedded object. Fig. 14a shows the proportion of responses that received the backgrounded interpretation of the embedded object in the comprehension task. We fit a logistic mixed-effects regression model predicting backgroundedness from a sum-coded fixed effect of focus condition (reference level: Embedded Focus) along with by-item and by-participant random intercepts and slopes for the fixed effect. In comparison to the Adverb Focus condition, there were fewer backgrounded responses in the Embedded Focus condition ($\beta = -3.99, SE = 0.74, z = -5.42, p < 0.001$), suggesting that the focus manipulation changed the backgroundedness of the embedded content in the expected way.

For the main analysis, Fig. 14b shows the mean acceptability ratings in the four conditions. We fit a linear mixed-effects regression predicting acceptability from a dummy-coded fixed effect of focus condition (reference level: Adverb Focus) and the maximal random effects structure allowing model convergence: by-participant random intercepts and slopes for the fixed effect of condition. There was a significant main effect of focus condition, such that sentence acceptability was higher in the Embedded Focus condition than in the Adverb Focus condition ($\beta = 0.21, SE = 0.03, t = 6.90, p < 0.001$). Additionally, in comparison to sentences in the Adverb Focus condition, the grammatical fillers were ($\beta = 0.39, SE = 0.03, t = 13.38, p < 0.001$), whereas ungrammatical fillers were less acceptable ($\beta = -0.16, SE = 0.04, t = -3.62, p < 0.001$). This increase in acceptability when the adverb received focus suggests that foregrounding the embedded content via the context sentence attenuates the island effect.

Finally, we used the two frequency measures of each "*say* + adverb" construction from the corpus analysis reported in Experiment 2 to test the effect of predicate-frame frequency



(a) Backgroundedness responses.

(b) Mean acceptability ratings.

Figure 14: Results of Experiment 3 across focus condition. Error bars indicate 95% bootstrapped confidence intervals.

on acceptability. Fig. 15 and Fig. 16 show acceptability ratings in the two conditions against the predicate-frame frequency values and the SCR scores, respectively. Using the same statistical models for the frequency analyses in Experiment 1, we first fit a linear mixed-effects regression predicting acceptability rating from a sum-coded main effects of condition (reference level: Embedded Focus), log-transformed and mean-centered predicate-frame frequency, and their interaction. The model also included the maximal random effects structure that allowed the model to converge: by-participant random intercepts and random slopes for the effects of condition, predicate-frame frequency, and their interaction; as well as by-item random intercepts and random slopes for the fixed effect of condition. There was a significant main effect of focus condition ($\beta = 0.11, SE = 0.02, t = 6.99, p < 0.001$), but neither the effect of predicate-frame frequency ($\beta = 0.01, SE = 0.02, t = 0.38, p = 0.712$) nor its interaction with condition ($\beta = -0.01, SE = 0.01, t = -0.89, p = 0.375$) was significant.

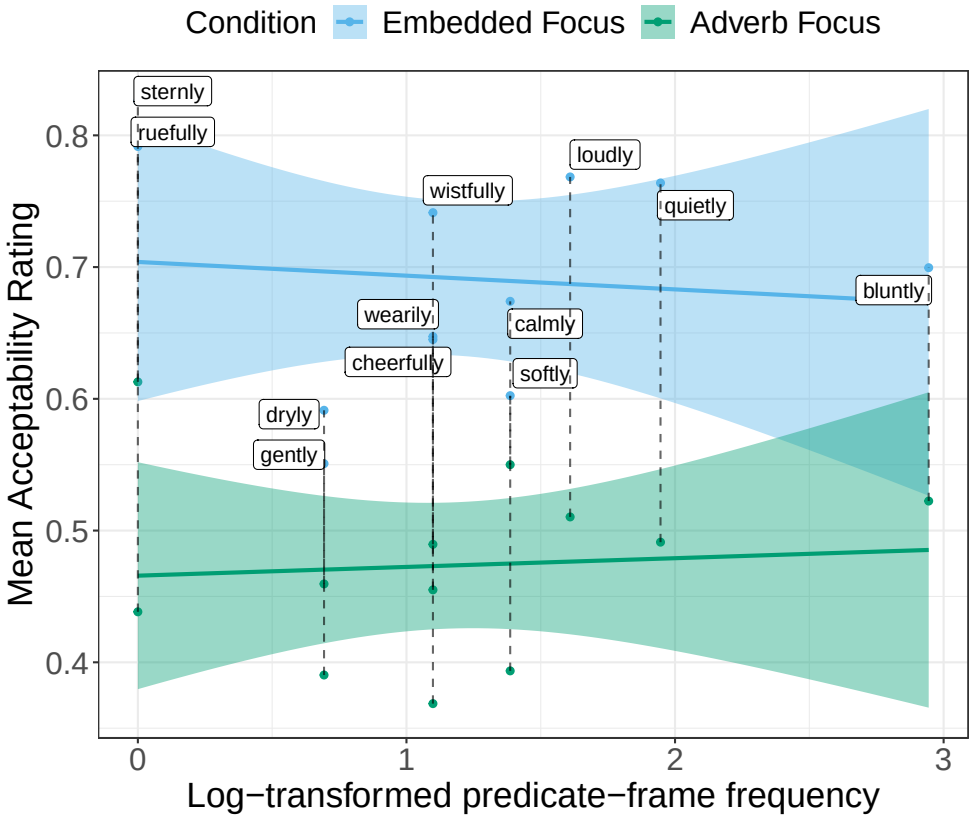


Figure 15: Mean acceptability ratings in the Adverb Focus condition and the Embedded Focus condition against the log-transformed predicate-frame frequency. The shaded ribbons represent 95% bootstrapped confidence intervals.

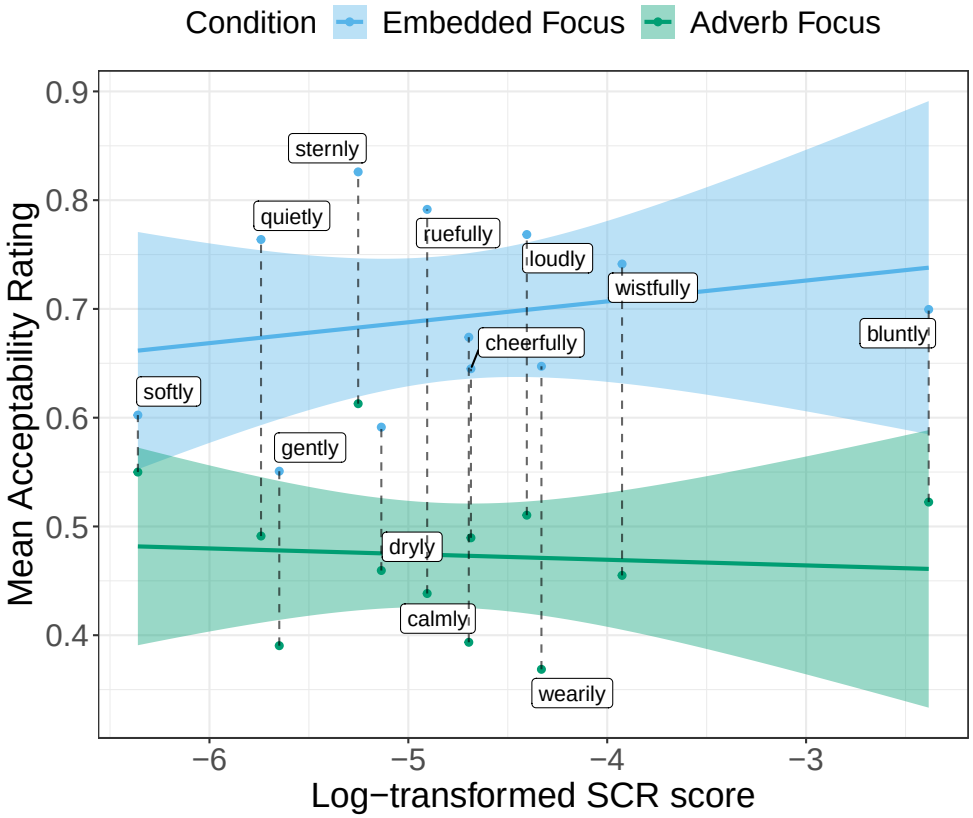


Figure 16: Mean acceptability ratings in the Adverb Focus condition and the Embedded Focus condition against the log-transformed Sentential Complement Ratio (SCR) scores. The shaded ribbons represent 95% bootstrapped confidence intervals.

The same model run with the SCR frequency score yielded the same results: a significant main effect of focus condition ($\beta = 0.11, SE = 0.01, t = 7.27, p < 0.001$), but neither a significant effect of SCR ($\beta = 0.01, SE = 0.02, t = 0.56, p = 0.587$) nor of its interaction with condition ($\beta = 0.01, SE = 0.01, t = 0.70, p = 0.484$).

4.2.3 Discussion

In combination with the results from Experiment 3a, where we replicated the MoS island effect using manner-of-speaker adverbs, the results of the current experiment show that foregrounding the embedded constituent in the "*say* + manner adverb" construction ameliorates the island effect. This is in line with findings in Experiment 1, which revealed a discourse effect on MoS islands. Therefore, the degradedness of extracting the embedded constituent contained in the complement clause of an MoS verb or *say* modified by an MoS adverb can be attributed to the discourse backgrounded status of the embedded constituent, further supporting the Backgroundedness Account. The results of Experiment 3b also shows that the "island effect" induced by the manner adverbs in Experiment 3a is discourse sensitive, and cannot be fully reduced to a processing-level effect of sentence

length or structural complexity. Moreover, as in Experiment 3a, the results of Experiment 3b support neither the Frequency Account nor the Subjacency Account.

5 General Discussion

In this paper, we presented three acceptability judgment experiments. In Experiment 1, we showed that the MoS island effect can be ameliorated through discourse manipulation. In Experiment 2, we showed that the same discourse manipulation can be used to create island effects with the light verb *say* (Experiment 2a). Moreover, by default the embedded object is more likely to be discourse backgrounded when it is in the embedded clause of MoS verbs than of *say* (Experiment 2b). In Experiment 3, we showed that the MoS island effect can be replicated even in sentences with the bridge verb *say* taking complement clauses by increasing the lexical weight of the matrix predicate using manner adverbs, such as loudly, softly (Experiment 3a), and this observed MoS island effect in the "*say* + adverb" construction behaves exactly like the MoS island effect observed in sentences with MoS verbs: it is also sensitive to prosodic discourse manipulations (Experiment 3b).

These experimental results bear out two predictions of the Backgroundedness Account of the MoS island effect. First, the Backgroundedness Account implicates the discourse backgroundedness of the extracted elements rather than structural properties of the sentence as the source of island effects. Therefore, without altering the syntactic structure of a sentence, the island effect should be ameliorated by foregrounding an extracted element that is initially discourse backgrounded. This is exactly what we found in Experiment 1: DPs embedded in MoS complements are by default backgrounded and resist extraction; but when they receive prosodic focus and are thus discourse foregrounded, extracting them becomes more acceptable.

Second, one should be able to conjure up an island effect by backgrounding an extracted element that is initially in the foreground. This is supported by the results from Experiments 2 and 3. In Experiment 2, we show that DPs embedded in the complement clause of *say* are by default not backgrounded, and thus can be extracted; however, they can be backgrounded by foregrounding the matrix predicate, either by prosodically focusing the matrix verb as in Experiment 2a, or by adding lexical weight to the matrix predicate using manner adverbs as in Experiment 3a, resulting in less acceptable extraction. In Experiment 3b, we further confirmed that the island effect that surfaced in Experiment 3a behaves just like the MoS island effect observed in Experiment 1, in that it is sensitive to information structural manipulations.

These results cannot be explained under either the Subjacency Account or the Verb-Frame Frequency Account. Under the Subjacency Account, the MoS island effect is attributed to the underlying complex NP structure of sentences with MoS verbs (i.e., the structure in (4)). Syntactic rather than information structural properties are considered the direct cause of the MoS island effect under this account. Therefore, the Subjacency Account predicts that foregrounding the extracted DP by changing the prosodic pattern as in Experiment 1 should not alter the syntactic structure of the critical sentences and thus should not affect the magnitude of the MoS island effect, contrary to the findings of Exper-

iment 1. The Subjacency Account also does not predict the effect observed in Experiment 3: in both the *say* and *say* + adverb conditions, the critical sentences involved a matrix verb *say* taking a complement clause, which do not have the complex NP structure as in (4). Therefore, the Subjacency Account does not predict the contrast between the *say* and *say* + adverb conditions in Experiment 3.

One way to potentially reconcile the results of Experiment 3 with the Subjacency Account is to propose a structural difference between the sentences of the *say* and the *say* + adverb conditions. Some may argue that the adverb modifying *say* initially originates in a position following the embedded complement clause, and the complement clause needs to undergo rightward movement (akin to extraposition and heavy NP shift) to achieve the word order where the adverb immediately follows *say*. This rightward movement of the embedded CP is shown in 14.

(14) ??Who did John say t softly [that Mary met with]?

Following this analysis, wh-movement from within the embedded clause would be prohibited due to the Freezing Effect (Culicover and Wexler, 1977, 1980; Ross, 1967): no movement can occur from within an extraposed constituent.[16] Following Johnson (1986), the Freezing Effect can be accounted for using the Subjacency Condition.[17] Therefore, we can still maintain a purely syntactic explanation for the degradedness of the *say* + adverb condition observed in Experiment 3a. However, in Experiment 3b, we observed that the degradedness of sentences like (14) can be ameliorated by information structural manipulations, an effect that cannot be explained if subjacency is the only reason that the *say* + adverb condition is degraded in Experiment 3a.

The results also do not support the Verb-Frame Frequency Account. In Experiment 1, sentences in the two critical conditions (Verb and Embedded focus conditions) involved the exact same set of MoS verbs, and thus the observed acceptability contrast cannot be attributed to a difference in verb-frame frequency. Follow-up analyses further showed that in Experiment 1, there was no correlation between the acceptability of the MoS island sentences and the verb-frame frequency of each MoS verb taking complement clauses, further confirming that verb-frame frequency is not the source of the MoS island effect. In Experiments 3a and 3b, all critical sentences included the same matrix verb *say*. Therefore, verb-frame frequency cannot account for any contrasts observed in the critical comparisons in Experiments 3a and 3b. One may argue that the most relevant frequency measure in both experiments should take into account the whole matrix predicate (i.e., the frequency

[16]The empirical evidence for the Freezing Effect is mixed. Culicover and Wexler (1980) noted certain acceptable examples of extraction from extraposed clauses like (1):

(1) What was it clear that Mary had bought?

Other exceptions to the Freezing Effect were also noted in Rizzi (2007), Gallego (2009, 2010), and Müller (2010). See Corver (2017) for a comprehensive review.

[17]It is rather controversial whether the Freezing Effect should have a syntactic explanation. Past studies have also proposed processing-level (Hofmeister et al., 2015; Konietzko et al., 2018; Winkler et al., 2016) and pragmatic (Martens, 2021) accounts of the Freezing Effect.

of *say* + different manner adverbs followed by complement clauses), rather than just the frequency of verbs taking complement clauses. However, the predicate-frame frequency measures calculated based on corpus counts of *say* + different manner adverb followed by complement clauses did not correlate with the acceptability of the *say* + adverb wh-questions in Experiment 3. As a result, the Verb-Frame Frequency Account falls short of accounting for the results of the current study. [18]

In sum, the current study provides support for the Backgroundedness Account of the MoS island effect, and challenges the Subjacency Account and the Verb-Frame Frequency Account. This does not mean that we ruled out the possibility of syntactic and frequency factors playing a role in the MoS island effect. Below we present several possible ways that syntactic and frequency factors may interact with the MoS island effect.

It is possible that the complex-NP structure as shown in (4) is in fact the correct underlying structure for sentences showing the MoS island effect, but the direct source of the island effect is the discourse backgrounded nature of appositive CPs rather than a violation of the subjacency condition. In this case, syntactic properties (the complex-NP syntactic structure) determine the discourse backgroundedness of constituents, which in turn gives rise to island effects. Note that this analysis diverges from our proposal, which states that the default backgroundedness of MoS complements is due to the heavier lexical weight of MoS verbs than the bridge verb *say* (due to the extra manner component). We favor our proposal for two reasons. First, we showed in Experiment 3 that island effect can be created by adding manner-of-speaking adverbs to matrix predicates, which serves as a proof of concept that the extra manner component can affect discourse backgroundedness. In contrast, it is less clear whether the appositive CP structure gives rise to discourse backgroundedness.

Second, our proposal is more parsimonious in that it maintains the same syntactic structure as in (3) for both bridge verbs and MoS verbs taking embedded complement clauses, and does not posit any silent structure. Nevertheless, it is an empirical question whether the complex-NP syntactic structure is the indirect cause of the MoS island effect, which we leave for future studies.

It is possible that the magnitude of the MoS island effect is affected by certain frequency measures other than the verb-frame frequency as defined in Liu et al. (2019, 2022a). For example, the acceptability of MoS island sentences could be predicted by the frequency of MoS verbs taking complement clauses that contain extraction gaps, or the frequency of wh-questions containing MoS verbs. Future corpus studies are needed to test these possibilities. However, one should be cautious in using a correlation between acceptability ratings and certain frequency measurements to argue for a frequency-based account of an acceptability

[18] Richter and Chaves (2020) suggested a possible explanation for why the effect of verb-frame frequency reported in Liu et al. (2019, 2022a) may fail to replicate in other studies: it is possible that the correlation between verb-frame frequency and acceptability that Liu et al. (2019, 2022a) observed was mainly driven by a verb-class effect. Bridge verbs in general have a higher frequency of taking complement clauses than MoS verbs, and extracting from bridge verb complements is more acceptable than extracting from MoS complements. This difference between verb classes gives rise to a false picture of correlation between verb-frame frequency and extraction acceptability. However, when analyzing only MoS verbs, verb-frame frequency may not correlate with extraction acceptability.

effect. If such a correlation is found, it is still unclear whether the low frequency of use leads to degraded acceptability or whether the degraded acceptability leads to the structure being less frequently used.

The current study has implications for the study of island effects and the experimental syntax enterprise in general. First, it supports the claim that the wide range of island effects identified in human language cannot be reduced to only syntactic factors, and that some island effects may have a multivariate explanation(Abrusán, 2011; Chaves and Putnam, 2020; Francis, 2022; Huang et al., 2023; Szabolcsi and Zwarts, 1993; Villata and Sprouse, 2023, *inter alia*). We demonstrated that at least for the MoS island effect, information structural factors contribute to degraded acceptability. Future work should carefully explore the possibility of non-syntactic accounts of islands before considering island effects only the domain of syntax. Second, we showed that sentence acceptability ratings can be affected by information structural factors. This adds to the long list of observations that extra-grammatical factors affect acceptability ratings (Chaves and Putnam, 2020; Francis, 2022; Schutze, 1996, *inter alia*).

Finally, several questions are left unanswered. First, it has been observed that the complement clauses of MoS verbs cannot be headed by null complementizers, whereas the complement clauses of bridge verbs can (Pesetsky, 1995; Snyder, 1992; Stoica, 2016; Stowell, 1981; Zwicky, 1971). This contrast is shown below in (15).

(15) a. John said (that) Mary is in the courtyard.
b. John whispered *(that) Mary is in the courtyard

This contrast naturally follows from the Subjacency Account, where the two sentences in (15) involve two different syntactic structures, and null complementizer is not possible when the embedded clause is not under syntactic selection by the matrix verb (i.e., in an appositive structure, as is the case with the MoS verb). This contrast in (15) could also follow from the Verb-Frame Frequency Account: it has been shown that the use of the overt complementizer *that* is preferred over its null counterpart when the embedded clause that it heads is less predictable given the prior context (Jaeger, 2010; Wasow et al., 2011). Therefore, the low frequency (and hence low predictability) of complement clauses following MoS verbs gives rise to a preference of using overt complementizers with MoS verbs. By contrast, the Backgroundedness Account makes no predictions about the effect shown in (15), and hence has less explanatory power compared to its alternatives. It is possible that there is a causal link between the backgroundedness of a complement clause and the overtness of its complementizer, a possibility that should be investigated by future studies.

Another open question concerns the decomposition of the MoS verbs. MoS verbs have often been analyzed as a manner verb composed from the light verb *say* and a manner component (e.g., *whisper* = say in a whispering manner), or a creation verb composed from the light verb *make* and a nominal component (e.g., *whisper* = make a whisper). In the current study, we assumed the manner verb analysis following Erteschik-Shir (2007). However, it is evident that a creation verb interpretation of MoS verbs is possible, especially considering the following example of an MoS verb that does not entail a *saying* action:

(16) She howled something at me, but she wasn't saying anything. (Zwicky, 1971)

As discussed earlier, the Subjacency Account assumes not only that MoS verbs can be lexically decomposed as creation verbs (*make* + a nominal component), but also that such decomposition is reflected syntactically, giving rise to the appositive structure in (4). In the current study, we provided ample evidence against such a syntactic analysis. However, it is still possible that the creation verb decomposition is an available interpretation of the MoS verbs to comprehenders, but such decomposition does not give rise to a special syntactic structure, unlike what is suggested by the Subjacency Account. Both the manner verb decomposition and the creation verb decomposition make the same prediction that MoS verbs are lexically complex and thus can bear focus, and therefore their complement clauses can be discourse backgrounded. Under the Backgroundedness Account, both ways of analyzing the MoS verbs predict that MoS complements are islands. In the current study, we do not have direct evidence that participants in Experiment 1 interpret the MoS verbs as manner verbs as opposed to creation verbs. Future studies can investigate which interpretation(s) of the MoS verbs participants arrive at when reading MoS island sentences, and test whether the interpretation interacts with the MoS island effect.

6 Conclusion

In the current study, we demonstrated that the MoS island effect can be ameliorated by foregrounding the extracted DP. While the embedded object of the bridge verb *say* is inherently discourse foregrounded by default, the complements of *say* behave like MoS islands when its embedded object is backgrounded, either by adding prosodic focus or by increasing its lexical weight. These results support a causal relationship between discourse backgroundedness and the MoS island effect, and challenge accounts that attribute the MoS island effect to syntactic or frequency factors.

References

Abeillé, A., Hemforth, B., Winckel, E., and Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*, 204:104293.

Abney, S. P. (1987). *The English noun phrase in its sentential aspect*. PhD thesis, Massachusetts Institute of Technology.

Abrusán, M. (2011). Presuppositional and negative islands: A semantic account. *Natural Language Semantics*, 19:257–321.

Ambridge, B. and Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics*, 19(3):357–389.

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

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Beaver, D. I., Roberts, C., Simons, M., and Tonhauser, J. (2017). Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics*, 3:265–284.

Bresnan, J. W. (1976). On the form and functioning of transformations. *Linguistic Inquiry*, 7(1):3–40.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and philosophy*, 26:511–545.

Chaves, R. P. and Putnam, M. T. (2020). *Unbounded dependency constructions: Theoretical and experimental perspectives*, volume 10. Oxford University Press, USA.

Chomsky, N. (1969). Deep structure, surface structure, and semantic interpretation. In Jakobson, R. and Kawamoto, S., editors, *Studies in General and Oriental Linguistics Presented to Shiro Hattori on the Occasion of His Sixtieth Birthday*. Tokyo: TEC Co. Ltd.

Chomsky, N. (1972). *Studies on semantics in generative grammar*. De Gruyter Mouton.

Chomsky, N. (1977). On wh-movement. In Culicover, P. W., Wasow, T., and Akmajian, A., editors, *Formal Syntax*, pages 71–132.

Chomsky, N. (1986). *Barriers*. MIT press.

Corver, N. (2017). Freezing effects. *The Wiley Blackwell Companion to Syntax, Second Edition*, pages 1–33.

Culicover, P. W. and Wexler, K. (1977). Some syntactic implications of a theory of language learnability. In Culicover, P. W., Wasow, T., and Akmajian, A., editors, *Formal Syntax*.

Culicover, P. W. and Wexler, K. (1980). *Formal principles of language acquisition*. The MIT Press.

Cuneo, N. and Goldberg, A. E. (2022). Islands effects without extraction: the discourse functions of constructions predicts island status. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, volume 44.

Cuneo, N. and Goldberg, A. E. (2023). The discourse functions of grammatical constructions explain an enduring syntactic puzzle. *PsyArXiv*.

Davies, M. (2008). Word frequency data from *the Corpus of Contemporary American English (COCA)*. Data available at https://www.wordfrequency.info.

Davies, M. (2011). Google books (american english) corpus (155 billion words, 1810-2009). Available online at http://googlebooks.byu.edu/.

Erteschik-Shir, N. (1973). *On the nature of island constraints.* PhD thesis, Massachusetts Institute of Technology.

Erteschik-Shir, N. (1979). Discourse constraints on dative movement. In *Discourse and syntax*, pages 441–467. Brill.

Erteschik-Shir, N. (2006). Bridge phenomena. *The Blackwell companion to syntax*, pages 284–294.

Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface*, volume 3. OUP Oxford.

Fox, D. and Pesetsky, D. (2005). Cyclic linearization of syntactic structure.

Francis, E. (2022). *Gradient acceptability and linguistic theory*, volume 11. Oxford University Press.

Gallego, Á. J. (2009). On freezing effects. *Iberia: An International Journal of Theoretical Linguistics*.

Gallego, Á. J. (2010). *Phase theory*. John Benjamins Publishing Company.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Goldberg, A. E. (2013). Backgrounded constituents cannot be "extracted". *Experimental syntax and island effects*, 221.

Hofmeister, P., Culicover, P., and Winkler, S. (2015). Effects of processing on the acceptability of "frozen" extraposed constituents. *Syntax*, 18(4):464–483.

Hopper, P. J. (1979). Aspect and foregrounding in discourse. In *Discourse and syntax*, pages 211–241. Brill.

Huang, N., Almeida, D., and Sprouse, J. (2023). A nearly-exhaustive experimental investigation of bridge effects in english. ms.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1):23–62.

Johnson, K. B. (1986). *A case for movement*. PhD thesis, Massachusetts Institute of Technology.

Jones, L. B. and Jones, L. K. (1979). Multiple levels of information in discourse. In Johnes, L. K., editor, *Discourse Studies in Mesoamerican Languages*, pages 3–27. The Summer Institute of Linguistics and The University of Texas at Arlington.

Konietzko, A., Winkler, S., and Culicover, P. W. (2018). Heavy NP shift does not cause freezing. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 63(3):454–464.

Kothari, A. (2008). Frequency-based expectations and context influence bridge quality. unpublished ms.

Kuno, S. (1976). Subject, theme, and the speaker's empathy: A reexamination of relativization phenomena in subject and topic. In Li, C., editor, *Subject and topic*, pages 417–444. Academic Press New York, NY.

Kuno, S. (1987). *Functional syntax: Anaphora, discourse and empathy*. University of Chicago Press.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.

Liu, Y., Ryskin, R., Futrell, R., and Gibson, E. (2019). Verb frequency explains the unacceptability of factive and manner-of-speaking islands in english. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pages 685–691.

Liu, Y., Ryskin, R., Futrell, R., and Gibson, E. (2022a). A verb-frame frequency account of constraints on long-distance dependencies in english. *Cognition*, 222:104902.

Liu, Y., Winckel, E., Abeillé, A., Hemforth, B., and Gibson, E. (2022b). Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics*, 8:495–525.

Martens, G. (2021). Frozen by context: Pragmatic factors of syntactic freezing. *Glossa: a journal of general linguistics*, 6(1).

Müller, G. (2010). On deriving ced effects from the pic. *Linguistic inquiry*, 41(1):35–82.

Namboodiripad, S., Cuneo, N., Kramer, M. A., Sedarous, Y., Bisnath, F., Sugimoto, Y., and Goldberg, A. (2022). Backgroundedness predicts island status of non-finite adjuncts in english. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, volume 44.

Nunes, J. and Uriagereka, J. (2000). Cyclicity and extraction domains. *Syntax*, 3(1):20–43.

Pesetsky, D. (1995). *Zero syntax: Experiencers and cascades*. The MIT press.

Potts, C. (2005). *The logic of conventional implicatures*. OUP Oxford.

Potts, C. (2007). Into the conventional-implicature dimension. *Philosophy compass*, 2(4):665–679.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richter, S. and Chaves, R. (2020). Investigating the role of verb frequency in factive and manner-of-speaking islands. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, pages 1771–1777.

Rizzi, L. (2007). On some properties of criterial freezing. *CISCL Working Papers on Language and Cognition*, 1:145–158.

Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. unpublished ms.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics*, 5:1–69.

Ross, J. R. (1967). *Constraints on variables in syntax*. PhD thesis, Massachusetts Institute of Technology.

Sag, I. A. (1976). *Deletion and logical form*. PhD thesis, Massachusetts Institute of Technology.

Schutze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. University of Chicago Press.

Simons, M., Tonhauser, J., Beaver, D., and Roberts, C. (2010). What projects and why. In *Proseedings of SALT 20*, pages 309–327.

Snyder, W. (1992). Wh-extraction and the lexical representation of verbs. unpublished ms.

Stoica, I. (2016). Island effects and complementizer omission: the view from manner of speaking verbs. *Constructions of Identity*, pages 191–200.

Stoica, I. (2020). Are manner of speaking verbs truly manner? *Bucharest Working Papers in Linguistics*, 22(1):29–48.

Stowell, T. A. (1981). *Origins of phrase structure*. PhD thesis, Massachusetts Institute of Technology.

Szabolcsi, A. and Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Natural language semantics*, 1(3):235–284.

Takami, K. (1988). Preposition stranding: Arguments against syntactic analyses and an alternative functional explanation. *Lingua*, 76(4):299–335.

Tomlin, R. S. (1985). Foreground-background information and the syntax of subordination. *Text-Interdisciplinary Journal for the Study of Discourse*, 5(1-2):85–122.

Tonhauser, J., Beaver, D. I., and Degen, J. (2018). How projective is projective content? gradience in projectivity and at-issueness. *Journal of Semantics*, 35(3):495–542.

Van Valin, R. D. (1998). The acquisition of wh-questions and the mechanisms of language acquisition. In Tomasello, M., editor, *The new psychology of language*, pages 221–249. Routledge.

Van Valin, R. D. and LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*. Cambridge University Press.

Villata, S. and Sprouse, J. (2023). A large-scale test of "d-linking" as a diagnostic of the strong/weak island distinction in english. Talk presented at the Penn Locality Workshop.

Wasow, T., Jaeger, T. F., and Orr, D. M. (2011). Lexical variation in relativizer frequency. In Heike, W. and Horst, S., editors, *Expecting the unexpected: Exceptions in grammar*, pages 175–96.

Winkler, S., Radó, J., and Gutscher, M. (2016). What determines 'freezing'effects in was-für split constructions. *Quantitative Approaches to Grammar and Grammatical Change: Perspectives from Germanic*, 290:207.

Zwicky, A. M. (1971). In a manner of speaking. *Linguistic Inquiry*, 2(2):223–233.