



Bondevik, Ingrid & Lohndal, Terje. 2023. Extraction from finite adjunct clauses: an investigation of relative clause dependencies in Norwegian. *Glossa: a journal of general linguistics* 8(1). pp. 1–41. DOI: <https://doi.org/10.16995/glossa.9033>



Open Library of Humanities

Extraction from finite adjunct clauses: an investigation of relative clause dependencies in Norwegian

Ingrid Bondevik, NTNU Norwegian University of Science and Technology, Norway, ingrid.bondevik@ntnu.no

Terje Lohndal, NTNU Norwegian University of Science and Technology, UiT the Arctic University of Norway, Norway, terje.lohndal@ntnu.no

Recent experiments have confirmed earlier informal evidence that finite adjuncts are not islands categorically. Specifically, it has been shown that adjuncts are not necessarily islands for all dependency types (Sprouse et al. 2016), and that the island status of an adjunct depends on the type of the adjunct clause in question (Kush et al. 2019; Müller 2019; Bondevik et al. 2021; Nyvad et al. 2022). The current study further explores these questions by testing three different adjunct clause types: Clauses introduced by *om* ‘if’, *fordi* ‘because’ and *når* ‘when’, in a relative clause (*rc*) dependency in Norwegian. We find that forming an *rc*-dependency into a finite adjunct in Norwegian overall causes island effects, but that there are fine-grained differences within the category ‘adjunct’. Specifically, we find that *fordi* ‘because’ and *når* ‘when’ yield large island effects, while *om* ‘if’, on a par with Kobzeva et al. (2022) and Nyvad et al. (2022), yields intermediate results. Rather than relying on binary distinctions only, we argue that any theory that is to explain the empirical landscape must be sufficiently fine-grained and allow for gradient distinctions.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction

Natural languages allow dependencies to be formed across a distance. This means that in (1) *the book* is interpreted as the object of the verb *buy*.

- (1) Forming a long-distance dependency
They discussed *the book* that Mary had recommended that John should buy.

There are, however, a number of domains that seem to block such dependency formation. These domains, given the metaphorical name *islands*, were first explored in detail in Ross (1967).

Since Huang (1982), there has been, and to some degree still is, a consensus that *finite* adjunct clauses are islands (see Bode 2020 for an overview; Truswell 2007; 2011; Stepanov 2007).

- (2) Finite adjunct clauses are islands
**Who* did John meet Bill before he phoned __ ? (Bode 2020: 120)

There is some experimental evidence to support this view (Sprouse et al. 2016; Kush et al. 2018), but there is also a growing body of evidence that finite adjunct clauses are not always islands. The empirical evidence to date has revealed that adjunct island violations are allowed under certain conditions. Specifically, it has been shown that dependency types might differ in their island sensitivity. A general pattern that has emerged is that finite adjunct clauses are islands for *wh*-dependencies (Sprouse et al. 2012; Sprouse et al. 2016; Kush et al. 2018; Kohrt et al. 2020, though see Kobzeva et al. 2022 and Chaves & Putnam 2020 (on satiation effects)) but might not be so for relative clause (*rc*-) dependencies in English (Sprouse et al. 2016) or topicalization (*top*-) dependencies in Norwegian (Kush et al. 2019; Bondevik et al. 2021), Swedish (Müller 2019) or Chinese (Zenker & Schwartz 2017). Furthermore, several studies have found that acceptability of adjunct island violations depends on the type of adjunct clause from which extraction takes place (Müller 2019; Chaves & Putnam 2020; Bondevik et al. 2021; Nyvad et al. 2022).

This paper investigates *rc*-dependencies into three different finite adjunct clause types in Norwegian: clauses introduced by *om* ‘if’, *når* ‘when’ and *fordi* ‘because’.

- (3) Examples of adjunct clause types in Norwegian
- a) *Om* ‘if’
De diskuterer båten som Jon blir glad om foreldrene kjøper.
they discuss boat.DEF that/which John gets happy if parents.DEF buy
‘They discuss the boat that John will be happy if his parents buy.’¹
- b) *Når* ‘when’
Nils unngår spillet som han blir frustrert når han taper.
Nils avoids game.DEF that/which he gets frustrated when he loses
‘Nils avoids the game that he gets frustrated when he loses.’

¹ The idiomatic translations into English show island violations, and so may not be grammatical. We have chosen to do this to make the relevant dependency clear.

c) *Fordi* ‘because’

Samtalen handler om tv-serien som mange blir redde
 conversation.DEF revolves about tv-serie.DEF that/which many become scared
fordi de ser.
 because they watch.

‘The conversation is about the tv-series that many get scared because they watch.’

The purpose of the study is to investigate the uniformity of adjunct island effects: Do *rc*-dependencies formed into finite adjunct clauses yield island effects in the same way as *top*-dependencies, or in the same way as *wh*-dependencies, or neither? And do different finite adjunct clauses yield uniform island effects or not in *rc*-dependencies?² In a broad sense, the goal is to contribute to determining how fine-grained theories of adjunct islands must be in order to account for the observed extraction patterns. Foreshadowing slightly, we find that *rc*-dependencies in Norwegian yield similar island effects for finite adjunct islands as *top*-dependencies do, and that for both types of dependencies adjunct clauses are not islands uniformly.

In the following section we give an overview of previous research on islands, specifically adjunct islands, and variation. Sections 3 and 4 provide an overview of the methodology and results of the first and second acceptability judgment experiments respectively, before our findings are discussed in Section 5. Section 6 concludes the paper.

2 Adjuncts as islands

2.1 Previous findings

When islands were first characterized and described in detail in Ross (1967) and later by Chomsky (1973; 1977; 1986), islands were explained in terms of syntactic principles. The claim was that islands arose from innate, universal, syntactic constraints on general movement operations. The traditional syntactic accounts such as the Subjacency Condition (Chomsky 1973; 1977) and *Barriers* (Chomsky 1986) alongside *Phases* (e.g., Chomsky 2000) predict that there will be minimal variation between island domains and between languages, and that any variation observed must be due to independent syntactic differences (see e.g. Rizzi 1982). Much research has, however, questioned this clear set of predictions, both within and across languages.

According to many researchers, particularly within traditional syntactic approaches to islands, adjunct clauses have maintained their status as strong and universal islands (see e.g., Stepanov 2007, and the overview in Bode 2020). Thus, the empirical predictions that follow are (i) adjunct islands should have universal validity, unless there is (preferably independently observable) evidence of relevant structural differences between languages; (ii) the acceptability

² By *island effect* we mean the observable “reaction” that speakers have to a structure where a filler must be posited in an illicit gap position, and where there are no other syntactic reasons why this gap position should be illicit (i.e., binding conditions, argument structure etc.).

of adjunct island violations should be categorically low (though see Chomsky 1986: 28). Some formal investigations find exactly this. Both Sprouse et al. (2012; 2016) and Kush et al. (2018) find large island effects of forming a *wh*-dependency into finite adjunct clauses in English and Norwegian, respectively.

Despite the claimed universal validity of the Adjunct Island Condition, much variation has also been uncovered for this island type. Sprouse et al. (2016) find no island effect for finite adjunct clauses in an *rc*-dependency in English, and Goldberg (2006) and Chaves (2021), among others, provide examples of acceptable extraction from finite adjunct clauses in English.

Norwegian and Swedish have figured prominently in the literature as languages with exceptions to the universal validity of island constraints. The papers collected in Engdahl & Ejerhed (1982) demonstrate a range of variation in MSc languages, among them examples of licit extractions from finite adjunct islands in Norwegian and Swedish (see also e.g., Teleman et al. 1999; Faarlund 1992; Bermingrud 1979 etc.).³

(4) Examples of licit extractions from finite adjunct islands

a. *Norwegian*

“Krig og fred” husker jeg ikke når kom ut
 “War and peace” remember I not when came out
 “War and peace”, I don’t remember when was published’ (Engdahl 1982: 167)

b. *Swedish*

Sportspegeln somnar jag om/när jag ser
 sports-program.DEF fall asleep I if/when I see
 ‘The sports program, I fall asleep if/when I see’ (Anward 1982: 74)

Engdahl & Ejerhed claim that such data challenges the “proposed universal principles of rule application” (1982: 9). Nevertheless, Engdahl (1982) maintains that long-distance dependencies are not unbounded in Norwegian and Swedish as there are several examples of illicit extraction provided alongside licit examples (see e.g. Bermingrud 1979; Faarlund 1992; Teleman et al. 1999).

More recent formal investigations corroborate that there are both licit and illicit extractions from adjunct clauses in Norwegian. Kush et al. (2018) find island effects for finite adjunct clauses in Norwegian in a *wh*-dependency. In a second series of experiments, Kush et al. (2019) find island effects for topicalization out of finite adjunct clauses, but no island effects for finite adjunct clauses when a context sentence facilitating contrastive topicalization is presented alongside the test sentence. An example of their test material is provided in (5).

³ For an overview of finite adjunct clauses in Danish see Poulsen (2008).

- (5) Example test sentence from Kush et al. (2019)

Preamble:

Kollegaene bryr seg ikke om at advokaten antageligvis vil glemme
 colleagues.DEF care themselves not about that lawyer.DEF probably will forget
kofferten sin,
 suitcase.DEF his
 ‘The colleagues do not care that the lawyer probably will forget his suitcase,’

Test sentence:

... men mappene blir de sinte om han glemmer igjen på kontoret.
 ... but files.DEF get they upset if he forgets again at office.DEF
 ... ‘but the files, they will be upset if he leaves at the office.’

Kush et al. (2019: 406) report that contrastive topicalization from a finite adjunct clause with context, on average, was rated to be almost as acceptable as topicalization from embedded declarative clauses. In addition, they find that judgments varied between and within participants. Kush et al. (2019) conclude that conditional adjuncts are not islands for topicalization in Norwegian.

Bondevik et al. (2021) further investigate Kush et al.’s (2019) findings for finite adjunct clauses in a contrastive topicalization dependency with context. Bondevik et al. (2021) test three different adjunct clauses – conditional *om* ‘if’-clauses, habitual *når* ‘when’-clauses and causal *fordi* ‘because’-clauses. Overall, they replicate Kush et al.’s (2019) findings for *om* ‘if’ showing that *om* ‘if’ is not treated as an island in Norwegian. However, they find large island effects for *fordi* ‘because’-clauses, and variable effects for *når* ‘when’-clauses. They conclude that with regards to islandhood, “adjunct” does not behave as a uniform class in the manner predicted by traditional syntactic approaches.⁴ Additionally, Bondevik et al. (2021) find a wide distribution underlying the average judgments for *om* ‘if’, much like Kush et al. (2019). They also see this for *når* ‘when’. They find no predictor which reliably explains differences between participants, nor are they able to identify any syntactic, semantic, or pragmatic factors that reliably predict differences between items which could explain the wide distribution of ratings.

Two recent studies have investigated *demonstrative rc*-dependencies (*dem rcs*)⁵ into finite adjunct clauses. Nyvad et al. (2022) investigated English *dem rcs* into the same three finite

⁴ Müller (2019, on Swedish) and Dal Farra (2020, on Italian) also argue that adjunct clauses must be distinguished.

⁵ Both Kobzeva et al. (2022) and Nyvad et al. (2022) use the term *rc*-dependency to refer to the dependency type tested in their studies. Kobzeva et al. (2022) provide the term *demonstrative rc* as an explanation of the type. Nyvad et al. (2022) use the same type of dependency in their study. Because there are substantial differences between the constructions which the *rc*-dependencies tested in Sprouse et al. (2016) and the *rc*-dependencies tested in Kobzeva et al. (2022) and Nyvad et al. (2022) appear in, we think it is important to separate the two as they clearly have different properties. The following test based on McCawley (1981) shows that different syntactic operations can apply to these types of *rcs*. It is not unlikely that this might carry over to island phenomena, but this needs to be tested carefully.

adjunct clauses that were tested in Bondevik et al. (2021) – *if*, *because* and *when*. Despite the widely held assumption that all finite adjunct clauses are strong islands in English, the authors find non-uniformity between the different adjunct clause types. As Bondevik et al. (2021) found for Norwegian, they find that forming an A'-dependency into finite *if*-adjuncts in English is rated much higher than A'-dependencies formed into finite *because*- and *when*-clauses. It is worth noting that the same proportional relationship between adjunct clause types replicates across languages (Norwegian vs. English) and across a different dependency type as well (*top* vs. *dem rc*). Unlike Bondevik et al. (2021), Nyvad et al. (2022) find that *when*- and *because*-adjuncts yield intermediate⁶ island effects. Thus, they argue that their results indicate that all finite adjunct clause types require a gradient theory of adjunct islands.

Kobzeva et al. (2022) do not find a strong island effect for *dem rcs* in Norwegian conditional *om* 'if'-adjuncts. They find a null effect, similar to Sprouse et al.'s (2016) findings for *rc*-dependencies in English, and average judgments on the “long, island” condition to be just below the acceptable range. In comparing conditional *om* 'if'-adjuncts on *dem rcs* and *wh*-dependencies, Kobzeva et al. (2022) find that *dem rcs* yield lower acceptability ratings compared to *wh*-dependencies, contrary to previous findings in Kush et al. (2018) that both simple and complex *wh*-dependencies yield large island effects in Norwegian *om* 'if'-adjuncts. Kobzeva et al. (2022) suggest that differences between studies might be related to the predicate types used in the different experiments.

2.2 Dealing with variation

Above, we have seen that the traditional claim that all adjuncts are islands cross-linguistically is disputed by more recent evidence of cross-linguistic variation (Sprouse et al. 2016), variation between dependency types (Kush et al. 2018; 2019; Kobzeva et al. 2022), and even variation between and within adjunct clause types (Müller 2019; Bondevik et al. 2021; Nyvad et al. 2022). Variation poses a problem for traditional syntactic accounts, and variation in adjunct islands particularly so. On these approaches, adjunct clauses are constrained by general principles that restrict all adjuncts categorically. For instance, within Huang's (1982: 505) Condition

-
- (i) Sprouse et al.'s (2016) test item:
- a) I called the client who the secretary thought that the lawyer insulted __.
 - b) ?I called the client, as you know, who the secretary thought that the lawyer insulted.

- (ii) Nyvad et al.'s (2022) test item:
- a) This is the exercise that I was surprised that she actually completed __.
 - b) This is the exercise, as you know, that I was surprised that she actually completed.

⁶ They define the results for *when*- and *because*-adjuncts as “intermediate” by showing that the effect sizes are below a threshold set in Kush et al. (2019) for the normal range for typical island effect sizes. As Kush et al. (2019) set this threshold at 0.75, and Nyvad et al. (2022) report an effect size of 0.74 for *because* and 0.63 for *when*, these clause types are numerically below the threshold for a “typical island effect”, but exceedingly close to the boundary.

on Extraction Domain (CED), all adjuncts are islands based on the claim that no adjuncts are (properly) governed.

(6) Condition on Extraction Domain (CED):

A phrase A may be extracted out of a domain B only if B is properly governed.

The notion of proper government has been abandoned in recent theoretical frameworks, but the idea remains that adjuncts are islands namely because adjuncts provide a special type of constituent that is less closely integrated with the matrix clause (see e.g., Bode 2020 for an overview). This is implemented in different ways in Minimalism (see e.g., Chomsky 2000; Stepanov 2007; Hornstein & Nunes 2008). Consequently, all adjuncts are islands simply because they are adjuncts. Thus, traditional syntactic approaches generally do not allow fine-grained variation between and within adjuncts.

Sprouse et al. (2016) review several syntactic approaches to islands looking at how each of these can account for variation in dependency type between languages. For each of the syntax-based approaches that they review, they find that their results are difficult to accommodate. This indicates that none of the syntax-based approaches can easily handle variability. However, they discuss the possibility that *Relativized Minimality* might have the power to account for differences in dependency types, but they do not provide an explicit analysis of differences between *rc*-dependencies and *wh*-dependencies into *if*-adjuncts in English. Nyvad et al. (2022) come to a similar conclusion as Sprouse et al. (2016) regarding syntax-based approaches. Bondevik et al. (2021) and Nyvad et al. (2022) also review some extra-syntactic approaches, but find that these struggle to readily handle the differences between adjunct clause types.

2.3 Research questions, predictions, and hypotheses

It seems clear that adjuncts are not categorical islands for all A'-dependencies as predicted by traditional syntactic accounts, but that there are some factors that facilitate variation across constructions, languages, and adjunct types. Our main aim is to map the empirical landscape of finite adjunct clauses in Norwegian. Finite adjunct clauses have been tested in a *wh*-dependency, a *top*-dependency and a *dem rc*-dependency in Norwegian. There is evidence of cross-dependency variation for finite adjunct clauses in Norwegian, such that *top*-dependencies and *dem rcs* are less sensitive to finite adjunct island effects compared to *wh*-dependencies (though see Kobzeva et al.'s findings for *wh*-dependencies). Sprouse et al.'s (2016) findings for English and Italian point in different directions as to whether or not *rc*-dependencies are sensitive to adjunct islands constraints. We therefore want to test different finite adjunct clause types in an *rc*-dependency in Norwegian.

In addition, Norwegian finite adjunct clauses provide an interesting case study for investigating the island sensitivity of *rc*-dependencies. Previous research documents systematic differences between adjunct clauses in Norwegian (Bondevik et al. 2021). It is therefore possible

to test (i) whether *rc*-dependencies are sensitive to adjunct island effects in general, and (ii) whether *rc*-dependencies are sensitive to all adjunct clause types equally. These are important for two reasons – firstly, it is an empirical problem that we do not know the descriptive adequacy of these constructions, and secondly, by studying these two phenomena in tandem we can begin to create better models for capturing variation displayed by adjunct clauses. Specifically, our research questions are:

- i. Are adjuncts islands for relativization in Norwegian?
- ii. Do different types of adjunct clauses behave like a uniform group for relativization?

The rest of the paper is organized as follows. In Section 3, we give a detailed overview of the experimental design employed in Experiments 1 and 2 and provide a detailed overview of methodology and results for Experiment 1. The second experiment is presented in Section 4. Section 5 provides a discussion of our research questions in view of both experiments. Finally, Section 6 concludes the paper.⁷

3 Experiment 1

3.1 Experimental design

To investigate our research questions, we ran an acceptability judgment study following the 2×2 factorial design popularized by Sprouse and colleagues (Sprouse 2007; Sprouse et al. 2016).⁸ This allows for a direct comparison with previous findings for adjunct clauses in *rc*-dependencies in English (Sprouse et al. 2016), in *dem rcs* in Norwegian (Kobzeva et al. 2022) and in *top*-dependencies in Norwegian (Kush et al. 2019; Bondevik et al. 2021). The goal is to isolate any effects of an island violation that goes beyond potential processing difficulties involved with complex sentences. The design controls for two confounds that potentially put a strain on processing, and subsequently lower acceptability: (i) the length of time that a filler must be maintained in working memory before the gap is encountered (*short* vs. *long*), and; (ii) the complexity of the domain present in the sentence (*no-island* (declaratives) vs. *island* (domains claimed to be islands)). The idea is that domains claimed to be islands (e.g. adjunct clauses), irrespective of extraction, might be more complex to process than declaratives.

⁷ All test materials and data analyses are made available in the following OSF repository: https://osf.io/d6wfe/?view_only=344f4132528b432593808e05d622d9bd.

⁸ To read more on the advantages of this design see Sprouse & Villata (2021) and references therein. Since Sprouse (2007), many experiments using this design have been conducted in several different languages to assess the inventory of islands in different languages and dependency types (see e.g., Sprouse et al. 2011; Sprouse et al. 2012; Kush et al. 2018, 2019; Keshev & Meltzer-Asscher 2019; Pañeda & Kush 2022; Kobzeva et al. 2022).

The 2×2 design crosses the two factors, each with two layers: *Structure* : *no island* vs. *island* \times *Distance* : *short* vs. *long*. This yields four test conditions which together make up one test item. An example is provided in (7).

(7) Example of test item

- | | |
|--|---------------------------------|
| a. Who _ believed that Monica bought a house? | <i>short</i> <i>no-island</i> |
| b. What did Rachel believe that Monica bought _ ? | <i>long</i> <i>no-island</i> |
| c. Who _ was sad because Monica bought a house? | <i>short</i> <i>island</i> |
| d. What was Rachel sad because Monica had bought _ ? | <i>long</i> <i>island</i> |

If the main effects illustrate linear additivity (i.e., no interaction effect), we will see that the decrease in acceptability is constant between the short and long conditions, and equally, that it is constant between the no-island and island conditions. This is illustrated in the interaction plot in **Figure 1** under “No island effect”. Here, the lowered acceptability on the “long, island” condition can be explained by the linear sum of the processing costs.

If, however, the main effects illustrate a super-additive interaction, the effect of forming a filler-gap dependency into an island domain is larger than the sum of processing costs. This is termed an *island effect* and is illustrated in **Figure 1** under “Island effect”. Here, the additional decrease in acceptability on the island violating sentence indicates that there is something outside of processing costs that causes an “unexpected” decrease in acceptability. Importantly, the effect is predicted to be directional such that the “long, island”-condition is rated as least acceptable.

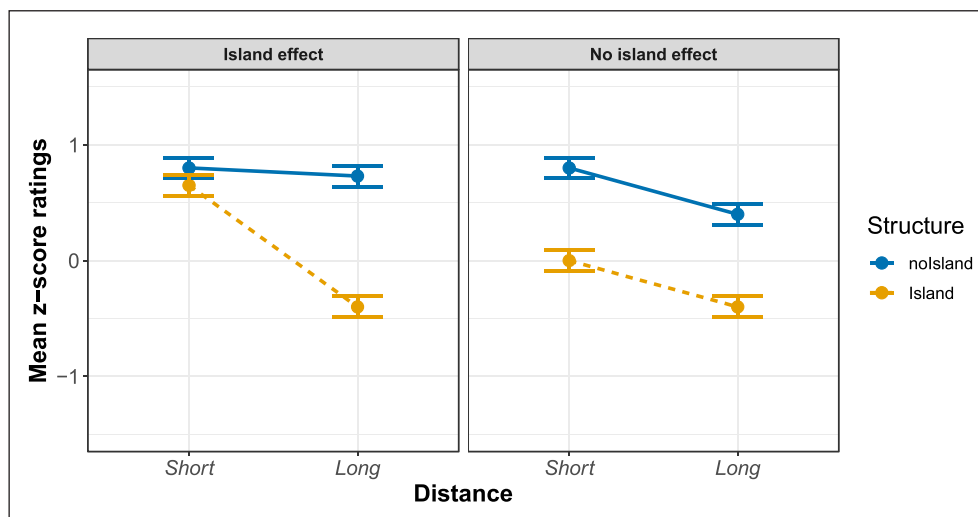


Figure 1: Examples of interaction patterns.

3.2 Test material

We tested three different adjunct clause types in an *rc*-dependency, each introduced by a different complementizer *om* – conditional ‘if’, *fordi* – causal ‘because’ and *når* – habitual ‘when’. In addition, we included two control clause types for baseline comparisons: complex subjects, which have been shown to yield large and robust island effects in Norwegian, and complement *om* ‘whether’ clauses, which have been shown to yield small or no island effects in Norwegian in a *top*-dependency (Kush et al. 2019; Bondevik et al. 2021).

A relative clause is a clause in which the nominal phrase is associated with a position both in the matrix and the subordinate clause. Unlike Kobzeva et al. (2022) and Nyvad et al. (2022), we tested restrictive relative clauses in which the head noun is the object of the matrix verb. The most common type of restrictive relative clauses in Norwegian are *som*-relatives, introduced by the complementizer *som* (Åfarli 1994: 82).

(8) Example of relative clauses in Norwegian

a. Subject relative clause

*Han *(som) kjøpte skoene*
 He *(SOM) bought shoes.DEF
 ‘He/The man who bought shoes’

b. Object relative clause

Skoene (som) han kjøpte
 Shoes.DEF (SOM) he bought
 ‘The shoes that he bought’ (Åfarli 1994: 82)

(8) shows that the relative complementizer is obligatory in subject relative clauses, but not in object relative clauses. All target items were created with *rc*-dependencies forming restrictive relative clauses. For the object relative clauses, the complementizer *som* ‘who/which/that’ was included to maintain as much of the structure as identical as possible across subject and object relative clauses.

The test items were modelled on previous experiments with this design (Sprouse et al. 2016; Kush et al. 2018; 2019; Bondevik et al. 2021). Specifically, the items followed the structure in Sprouse et al. (2016) for testing island violations in an *rc*-dependency, where there are three clauses – a matrix clause, a relative clause modifying the object in the matrix clause and finally a finite adjunct clause embedded under the relative clause. The finite verb in each clause will henceforth be referred to as *V*matrix, *V*rel and *V*adjunct, respectively. An example item for *om* ‘if’ is provided in (9).⁹

⁹ For examples of test sentences for all island types tested see Supplementary file.

(9) **Adjunct *om* ‘if-clauses**a. *No island, short*

De erter fotballspilleren som __ misliker at de nevner selvmålet.
 they tease football-player.DEF who __ dislikes that they mention own-goal.DEF
 ‘They tease the football player who dislikes that they mention the own goal.’

b. *No island, long*

De diskuterer selvmålet som fotballspilleren misliker at de nevner __.
 they discuss own-goal.DEF that football-player.DEF dislikes that they
 mention __.
 ‘They discuss the own goal that the football player dislikes that they mention.’

c. *Island, short*

De erter fotballspilleren som __ blir flau om de nevner selvmålet.
 they tease football-player.DEF who __ gets embarrassed if they mention
 own-goal.DEF.
 ‘They tease the football player who gets embarrassed if they mention the own goal.’

d. *Island, long*

De diskuterer selvmålet som fotballspilleren blir flau om de nevner __.
 they discuss own-goal.DEF that football-player.DEF gets embarrassed if they
 mention __.
 ‘They discuss the own goal that the football player will be embarrassed if they mention.’

The items are matched on several syntactic and semantic parameters that might influence acceptability. Every verb phrase is in the present tense, none of the embedded clauses are negated (Szabolcsi & Lohndal 2017), and every relative clause head is a definite DP. Finally, all adjunct clauses can be classified as Central Adverbial Clauses in the sense of Haegeman (2012) (see also Müller 2019). There are minor differences between items such as type of subject (e.g., indefinite determiners *noen* ‘someone’, full NPs *studentene* ‘the students’, general 3rd person pronouns *de* ‘they’) in Vmatrix, Vrel or Vadjunct. This means that items are not minimally distinct, but items are matched on the features that have been suggested in the literature to be relevant for judgments of islandhood.

As pointed out to us by an anonymous reviewer, the conditions are not minimally different on two important aspects which potentially confound the results: (i) there are different lexicalizations across the different conditions, and (ii) on the short conditions, the gap is in the subject position

of the relative clause. Regarding the first point, we believe that this fact is not detrimental since the different lexicalizations are the same for two and two conditions within each item. Thus, any effects of word choice will subtract (see Sprouse & Villata 2021). Turning to (ii), this means that there is a subject gap in the relative clause in the short conditions and an object gap in the clause embedded within the relative clause in the long conditions. Thus, the DISTANCE factor controls both whether there is a subject gap or an object gap *and* whether the filler-gap dependency is short or long. Given the subtractive logic of the 2×2 factorial design (see e.g., Sprouse 2016: 314), the main effect of DISTANCE can be attributed to the difference in length *or* the difference in argument structure properties. This design will not be able to distinguish between these two possibilities.

3.3 Participants

100 Participants were recruited through Prolific and offered 7 GBP for participation. The study was made available to all participants who registered “Norway” as their nationality on Prolific. A background survey collected data on language history and demographics. Participants were asked to briefly describe how to get to their closest bus stop. Here three participants were excluded for providing a written reply that did not comply with Norwegian written standards. Next, 14 participants who self-reported being ‘bilingual’ were excluded.¹⁰ In addition, among the 14 participants who reported living outside of Norway, we excluded five participants who reported having lived abroad for a long period of time and/or who reported rarely speaking Norwegian. Participants were rewarded regardless of their responses. Finally, we excluded three participants for having >5 responses with <1000 ms. reaction times. We consider <1000 ms. insufficient time to read and judge any of our test sentences. After the exclusion criteria were applied, 76 participants were included in our data set.

Out of 76 participants, 30 reported being in the 18–24 age group, 30 between 25–34, 12 between 35–44, 3 between 45–54 and one older than 65. Participants were also asked to report dialectal background. Dialects were grouped into 10 larger dialectal areas based on Mæhlum & Røyneland’s (2012: 179) map of dialectal areas in Norwegian. In addition, *bergensk* ‘Bergendialect’ and *ingen av disse* ‘none of these’ were added as possible responses. All dialectal areas were represented in the study, the most frequent response being *østlandsk* ‘Eastern Norwegian’ (40 responses).

¹⁰ Kush & Dahl (2020) find evidence of transfer of functional structure allowing Norwegian speakers to accept island violating sentences in L2 English that have been shown to be acceptable in Norwegian. Such findings emphasize the importance of excluding multilanguage influence.

3.4 Procedure

16 items were tested for each adjunct clause types (16 items \times 3 clause types = 48 adjunct items), while 8 items were tested for each of the control clause types (8 items \times 2 clause types = 16 control items). Items were distributed across 4 lists in a Latin Square procedure, such that participants only saw one condition per item. This left 64 test sentences in each list. Under the assumption that every island violating sentence is unacceptable, the ratio between acceptable and unacceptable sentences was 3:1 for the target test sentences within each list.

The experiment was designed to be balanced both with regards to the ratio of target to filler sentences, and acceptable and unacceptable sentences. The experiment included 64 fillers, of which 48 were created to be unacceptable fillers. The bad fillers included syntactic, semantic, and orthographic violations. The good fillers included relative clauses and finite adjunct clauses that differed from target sentences, e.g., non-restrictive relative clauses, other adjunct clause types. All fillers were used across all four lists. Test sentences and fillers were pseudo-randomized by list for every individual participant by condition.

The experiment was distributed via Prolific and run on JATOS with JsPsych (de Leeuw 2015). Following previous experiments using this design, the experiment was designed as an acceptability judgment task where each test sentence was presented alone. Judgments were given on a labelled 1–7 Likert Scale with end points given as 7 *god* ‘good’ and 1 *dårlig* ‘bad’ (i.e., a full Likert Scale as defined in Marty et al. 2020).¹¹

Inside the experiment, the background survey was presented first. Next, task instructions were given. Specifically, participants were instructed to imagine a context in which the sentence was uttered by someone in their own dialect. Moreover, the instructions specified that long sentences are not necessarily unacceptable and short sentences are not necessarily acceptable. An example of a grammatical, but long sentence was shown and rated 7, and an example of a short, but ungrammatical sentence, rated 1.

Two unmarked practice items initiated the experimentation phase: one was clearly grammatical, the second ungrammatical.

3.5 Data analysis

The data was analyzed using similar procedures as previous experiments following this design (e.g. Sprouse et al. 2016). The raw responses were z-score transformed by participant prior to analysis. Following Sprouse et al. (2016), there are three procedures for identifying island effects within this design: (i) a visual inspection of the relationship between conditions: a superadditive pattern vs. a linear additive pattern; (ii) a numerical identification process of calculating

¹¹ Marty et al. (2020) show that a full Likert Scale with singular presentation provides higher effect detection rates than a non-labelled scale.

differences-in-differences scores (DD-scores) (see e.g., Sprouse & Villata 2021: 230 for a detailed explanation of the DD-score): a score above 0 is indicative of an island effect, while a score below 0 is characterized by Sprouse et al. (2011) as a *reverse island effect*, and; (iii) a statistical procedure fitting linear mixed effects models.

Data visualizations for visual inspection were created with `ggplot2` (Wickham 2016). The size of the island effect for each island type was calculated with a DD-score.¹² Linear mixed effects models were fitted with `lmer()` from the `lme4` package (Bates et al. 2015) in R (R Core Team 2021). An omnibus model was fit with a three-way interaction term crossing the main effects *island type*, *distance*, and *structure*. We included the three-way interaction term as we predict that the interaction of the main effects will differ by island type. By-item and by-participant varying slopes and intercepts were estimated as random effects. The model was simplified in a stepwise fashion to arrive at a model that converged without warning messages (though see Winter 2020: 266–267 for problems with such an approach). The categorical predictors were contrast coded -1 and 1 . The omnibus model returns the results for the reference level (which is alphabetically set to *fordi* ‘because’) and the rest of the model must be interpreted in relation to the reference level. To measure the island effect for each specific island clause type, we also fit separate models for each island type with a two-way interaction term crossing the main effects *distance* and *structure*.

We also checked to see if there was satiation of judgments. Satiation is a term used to describe the “perception of acceptability after repeated exposures to the same sentence or the same structure” (Sprouse & Villata 2021: 242). Several studies on English have found that there are no satiation effects for adjunct islands (see overview in Sprouse & Villata 2021). Chaves & Putnam (2020), however, found satiation effects with 24 exposures to the same adjunct island structure. Moreover, they found that conditional adjunct clauses satiated at a higher rate than causal and temporal adjunct clauses. Given that participants were only exposed to 4 test sentences of the same structure in Experiment 1, we predict that we will not see any satiation effects for either adjunct clause type. Nevertheless, we want to exclude this as a potential source of variation. We looked for this in two ways: (i) we checked if the results in Experiment 1 replicated when only the first two responses to each condition were included in a partial data set.¹³ As participants were only presented with two test sentences per control clause type in the full data set, the control clause types are the same for partial and full data sets. (ii) Following Chaves & Putnam (2020), we fit linear mixed effects models for each of the adjunct island’s “long, island” condition crossing z-scores and trial index as main effects and fitting by-subject and by-item varying intercepts.

¹² The DD-scores were calculated with the following formula based on Sprouse et al. (2012): (“long, no-island” – “long, island”) – (“short, no-island” – “short, island”).

¹³ Thanks to an anonymous reviewer for suggesting this approach.

3.6 Results

The bad fillers received an average rating of $z = -0.834$, while the good fillers received an average rating of $z = 0.859$, both yielding narrow distributions of scores. **Table 1** provides an overview of the main results of the omnibus model.

Main effects	Estimate	SE	<i>t</i>	<i>p</i>
distance: short	-0.353	0.019	-18.369	<0.0001
structure: no-island	-0.253	0.017	-13.690	<0.0001
Interaction				
<i>Forði</i> ‘because’*	-0.254	0.017	-14.718	<0.0001
<i>Når</i> ‘when’	-0.033	0.024	1.385	0.166
<i>Om</i> ‘if’	0.140	0.024	5.747	<0.0001
Subject	-0.064	0.029	-2.148	0.0317
<i>Whether</i>	0.265	0.029	8.853	<0.0001

Table 1: Results of omnibus model. See the Supplementary file for the full model output.

The omnibus model returned a significant interaction effect between the three main effects – *island type*, *distance*, and *structure*. In addition, there was a main effect of distance and structure. On the interaction term, the model did not distinguish between *forði* ‘because’-adjunct clauses (= the alphabetically determined reference level), the *når* ‘when’-adjunct clauses and the subject-islands. There were, however, significant differences between the *om* ‘if’-adjunct clauses and the *forði* ‘because’-clauses, and similarly between the ‘whether’-clauses and the *forði* ‘because’-clauses. This indicates that the interaction of distance and structure is statistically significantly different between *forði* ‘because’- and *når* ‘when’-adjunct clauses on the one hand, and *om* ‘if’-adjunct clauses on the other.

Looking at each island type separately, we ran separate linear mixed effects models for each island type and calculated DD-scores. We found significant island effects for all island types except for the control ‘whether’-clauses. For the ‘whether’-clauses only the main effect of distance was significant. The subject-island, the other control condition, yielded significant island effects and the largest effect size of all clause types. See **Table 2** reports the results for the control clause types.

All three adjunct clause types yielded significant interaction effects. However, as the omnibus model indicated, there are differences between adjunct clause types: *Forði* ‘because’ and *når* ‘when’ on the one hand show large DD-scores, while *om* ‘if’ shows a much smaller score. **Table 3** provides an overview of the model output for each target clause type, while the interaction plot in **Figure 2** visualizes the island effect and the effect size for each island type.

	Estimate	<i>t</i>	<i>p</i>	<i>DD</i>	Avg. z-score: <i>isl.cond.</i>
Subject				1.226	-0.605
<i>intercept</i>	0.377	7.581	<0.0001		
<i>distance</i>	-0.328	-14.283	<0.0001		
<i>structure</i>	-0.348	-15.137	<0.0001		
<i>distance</i> × <i>structure</i>	-0.308	-13.406	<0.0001		
‘whether’				-0.086	0.458
<i>intercept</i>	0.600	13.520	<0.0001		
<i>distance</i>	-0.117	-5.105	<0.0001		
<i>structure</i>	-0.040	-1.760	0.079		
<i>distance</i> × <i>structure</i>	0.014	0.614	0.539		

Table 2: Main results of the linear models by control clause type and calculated DD-scores, Experiment 1.

	Estimate	<i>t</i>	<i>p</i>	<i>DD</i>	Avg. z-score: <i>isl.cond.</i>
<i>Fordi</i> ‘because’				1.006	-0.568
<i>intercept</i>	0.278	6.078	<0.0001		
<i>distance</i>	-0.352	-19.084	<0.0001		
<i>structure</i>	-0.254	-13.337	<0.0001		
<i>distance</i> × <i>structure</i>	-0.252	-13.619	<0.0001		
<i>Når</i> ‘when’				0.876	-0.342
<i>intercept</i>	0.386	6.709	<0.0001		
<i>distance</i>	-0.294	-16.548	<0.0001		
<i>structure</i>	-0.218	-12.223	<0.0001		
<i>distance</i> × <i>structure</i>	-0.221	-12.422	<0.0001		
<i>Om</i> ‘if’				0.469	0.082
<i>intercept</i>	0.489	10.211	<0.0001		
<i>distance</i>	-0.109	-6.384	<0.0001		
<i>structure</i>	-0.183	-10.627	<0.0001		
<i>distance</i> × <i>structure</i>	-0.118	-6.870	<0.0001		

Table 3: Main results of the linear models by island type and calculated DD-scores, Experiment 1.

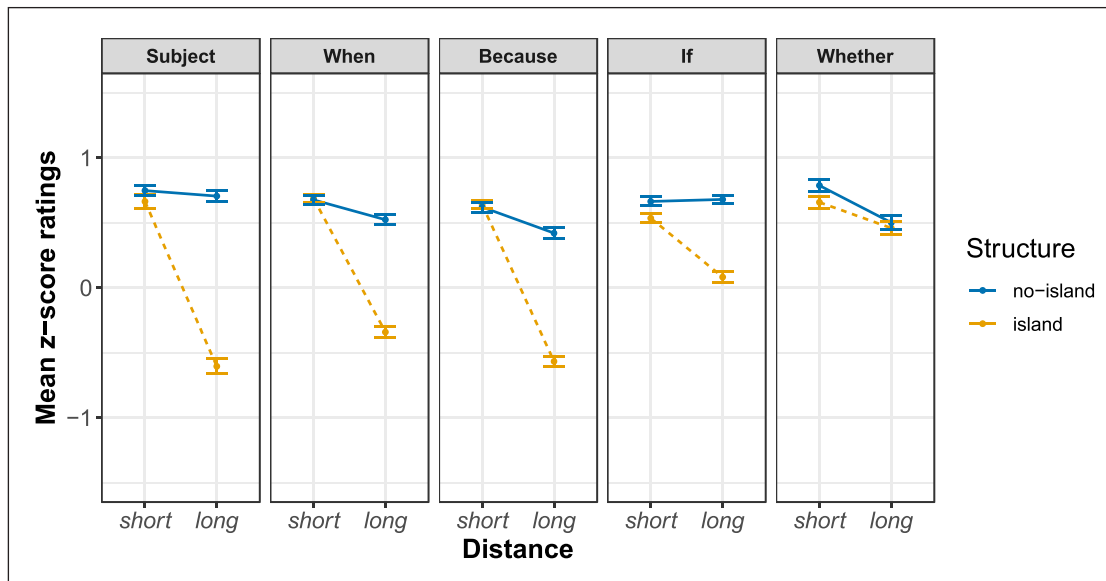


Figure 2: Interaction plot for all island types, Experiment 1 – average ratings on every condition for each clause.

We see that the average z-score for the “long, island” condition varies between island type, while the average ratings for the three non-island violating conditions are relatively stable across clause types. The average z-score on the “long, island” condition for the subject-island is low and for the ‘whether’ island it is high. Again, *fordi* ‘because’ and *når* ‘when’ pattern together with average ratings well below 0, while the island condition in the *om* ‘if’-items received average ratings just above 0.

Following findings for topicalization, we expect to see inter-trial variation, especially for *om* ‘if’ (Kush et al. 2019; Bondevik et al. 2021) and partly for *når* ‘when’ (Bondevik et al. 2021). We therefore investigated the distribution of z-scored ratings for each condition for each island type. In **Figure 3**, the distribution of z-scored ratings for each condition for each clause type is plotted.

We see a unimodal and quite narrow distribution for the ‘whether’ island condition. The distribution of scores for the “long, island” condition largely overlaps with the distribution for the “long, no-island” condition, where scores predominantly fall well above 0.¹⁴ On the “long, no-island” condition, there is a mostly unimodal distribution around -1 for the subject island.

¹⁴ A Kolmogorov-Smirnov (KS) test returns a significant difference between the distribution of the “long, no-island” condition and the “long, island” condition for ‘whether’ ($p = 0.0195$). The KS-test was run with `ks.test()` from the `dgof`-package (Arnold & Emerson 2011).

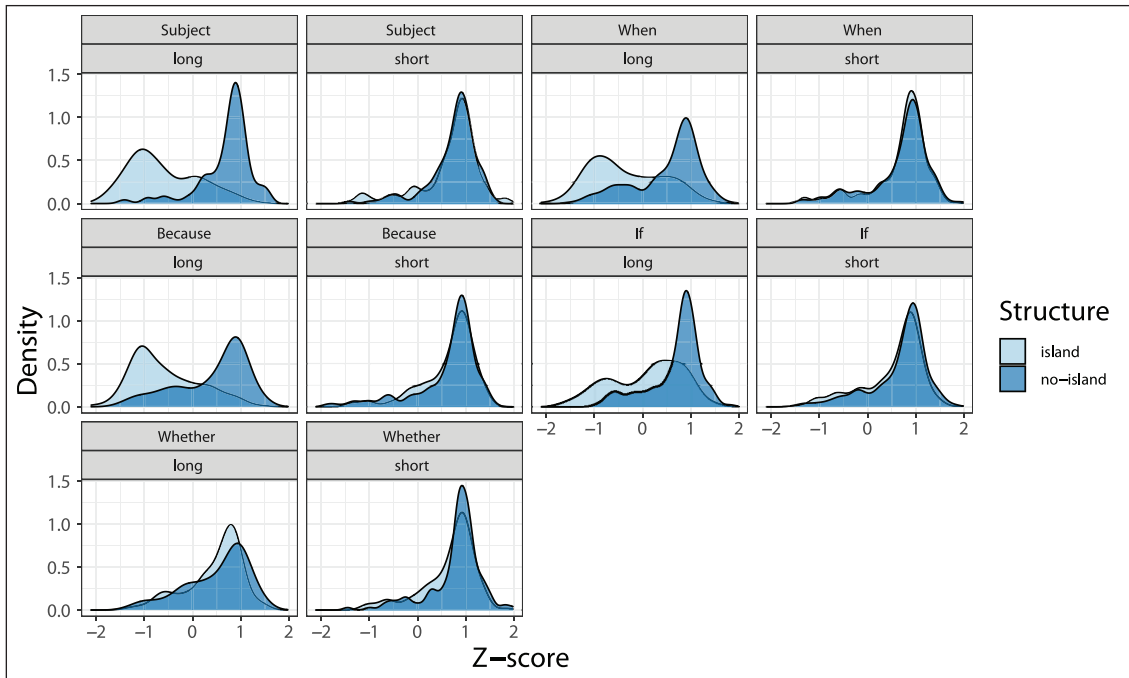


Figure 3: Comparing the distribution of z-scores on the no-island and the island conditions for the long and short conditions separately, Experiment 1.

Again, the distribution of scores is similar between *fordi* ‘because’ and *når* ‘when’ on the “long, island” condition, such that the majority of scores fall below 0. However, the leftward tail for *når* ‘when’ is wider than for *fordi* ‘because’, indicating that there is some variation between trials for *når* ‘when’ that is not observed for *fordi* ‘because’.¹⁵

The ratings for *om* ‘if’ have a wide, bimodal distribution: the biggest cluster of scores falls above 0, and a smaller cluster of scores below 0. The distribution of scores on the “long, island” condition resembles the distribution of scores on the “long, no-island” condition, but there is more variation for the “long, island” condition.¹⁶

Investigating the raw scores, we see the same pattern that we do for the z-scored ratings. In **Figure 4** we see that *om* ‘if’ is different from the two other adjunct clause types – while *fordi* ‘because’ and *når* ‘when’ resemble the subject clause type, *om* ‘if’ resembles the ‘whether’-clauses.

¹⁵ A KS test yielded significant differences between the distribution of ratings on the “long, island” condition of the two island types ($p = 0.0060$).

¹⁶ KS tests show that the distributions for each of the long conditions for *om* ‘if’ are significantly different ($p < 0.0001$), and that the “long, island” condition for *om* ‘if’ is different from the “long, island” conditions for *fordi* ‘because’ ($p < 0.0001$) and *når* ‘when’ ($p < 0.0001$).

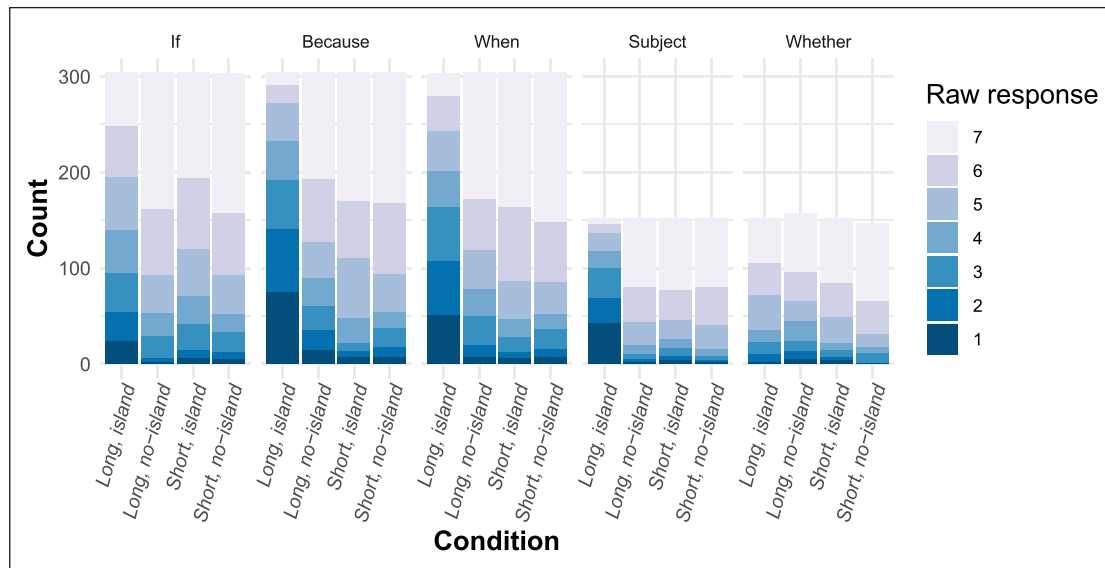


Figure 4: Barplots displaying the count of raw responses per condition for Experiment 2.

Checking for satiation effects, we find the exact same pattern for the partial data set that we find for the full data set (see Supplementary file). The omnibus model returns a significant interaction effect, main effects of distance and structure. The model finds *om* ‘if’- and ‘whether’-clauses to be significantly different from the reference level (*fordi* ‘because’). Running a linear mixed effects model modelling z-score on the island violating condition by trial index for each adjunct clause type reveals that there is a significant effect of z-score by trial index, but that as Sprouse & Villata (2021) point out, it is very small across adjunct clause type, see model output in **Table 4**.

Adjunct type	Intercept	Estimate	SE	<i>t</i>	<i>p</i>
<i>Fordi</i> ‘because’	−0.878	0.005	0.0001	4.928	<0.0001
<i>Når</i> ‘when’	−0.542	0.003	0.0010	3.021	0.0027
<i>Om</i> ‘if’	0.068	0.002	0.0001	2.128	0.0342

Table 4: Output of linear mixed effects model investigating z-score by trial index for each adjunct clause type.

This means that for each repetition, the z-score is predicted to rise by > 0.005 for each of the island conditions. As we presented participants with 4 repetitions of the same structure, we exclude satiation as having any effect on ratings.

In the plots in **Figure 5** (based on Chaves & Putnam 2020), we see judgments for items by block for each of the adjunct clause types. Block 1 contains the first two responses given to a

certain condition, and block 2 the last two. We do see differences between blocks, such that some items show an increase in acceptability from block 1 to block 2. However, we also see instances of a decrease in acceptability between blocks. We understand this to mean that overall there is a slight increase in acceptability as the experiment proceeds, but as the model demonstrates, the increase is very small.

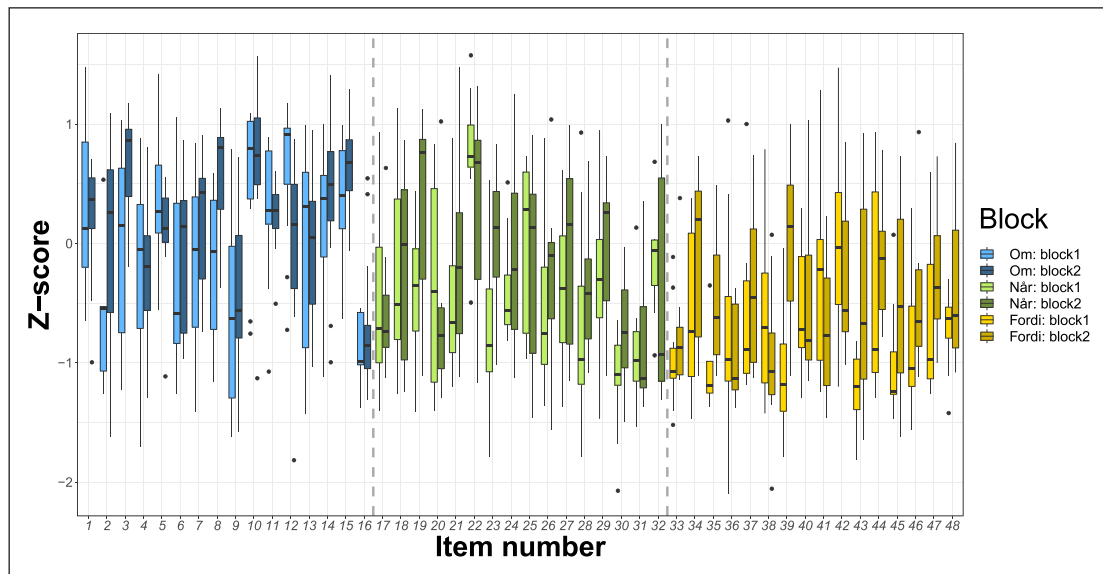


Figure 5: Boxplot illustrating average judgments on the “long, island” condition by item for each island type. The dashed line highlights the border between adjunct clause types. The plot legend provides the explanation of the colors.

3.7 Intermediate summary

Experiment 1 reveals that *rc*-dependencies *are* sensitive to island constraints in Norwegian. Collapsing across island types, we find island effects of forming a relative clause dependency into these domains. Fitting separate models for each island type, we find statistically significant island effects for all adjunct clauses and for the subject island, while the ‘whether’-island did not yield any significant interaction effects. As such, findings for the control island types replicate previous findings for *top*-dependencies in Norwegian (Kush et al. 2019; Bondevik et al. 2021).

Though we find island effects across the three adjunct clause types, we see clear indications that *fordi* ‘because’, *når* ‘when’ and *om* ‘if’ do not behave like a group in *rc*-dependencies. We find statistically significant differences between *fordi* ‘because’ and *når* ‘when’ on the one hand, and *om* ‘if’ on the other. While *om* ‘if’ shows a small island effect size, z-scored ratings clustering above 0 and a distribution of scores indicating variation between trials, *fordi* ‘because’ and *når* ‘when’ show large island effect sizes and z-scored ratings clustering well below 0. Thus, our

findings substantiate Bondevik et al.'s (2021) and Nyvad et al.'s (2022) findings: adjuncts do not behave like a uniform group with regard to islandhood.

As previously discussed, many theories of islands predict that there will be a categorical split between islands and non-islands, such that islands should be clearly unacceptable, while non-islands should be clearly acceptable. To that end, the intermediate island effect that we see for *om* 'if' is problematic for these theories. *Om* 'if' seems to fall in an intermediate position between acceptable (null effects) and unacceptable (large island effects). Thus, we need some way of accounting for *om* 'if'.

One possible interpretation of the intermediate effect size is that intermediacy is caused by averaging over variable results. The other studies testing *om* 'if' in Norwegian report substantial variation between trials. We see indications of this too in the distribution of scores for *om* 'if' on the "long, island" condition. Kush et al. (2019) suggest that the variation might be caused by inconsistent raters, i.e., either between- or within-speaker variation. Another option implied by Bondevik et al. (2021) is that there is variation between items. However, Bondevik et al. (2021) fail to find any factor across items that can explain said variation. If *om* 'if' sporadically induces island effects depending on certain factors (that we have yet to identify), which yield intermediate effects when averaged over, *om* 'if' is an adjunct type that variably causes large or small-to-nonexistent island effects. Such an interpretation predicts groupings of judgments on either side of the scale.

Another possibility is that the intermediate result we uncovered for *om* 'if' is a true representation of the acceptability of extraction from *om* 'if'. This means that extraction from *om* 'if' is systematically judged to be less acceptable than extraction from embedded complements ('whether' and declarative-clauses) and systematically more acceptable than extraction from *fordi* 'because' and *når* 'when' clauses. If this is true, we predict that there will be normal distribution around an intermediate score, i.e., variation between trials will be within the expected range.

In order to classify *om* 'if' with regard to islandhood, it is important to understand the source of the intermediate effects. Experiment 1 does not reveal much about the source of the intermediate effect. Thus, we carried out a follow-up experiment where we controlled for between- and within-speaker and -item variation.

4 Experiment 2

We ran a follow-up experiment to investigate the source of the on average intermediate effect seen for *om* 'if' in Experiment 1. We hypothesized that there would be no difference between judgments in Experiments 1 and 2 such that the intermediate effect size would replicate. We were interested in investigating three plausible sources of the intermediate effect size: (i) participant variation and/or; (ii) item variation; or (iii) order effects.

4.1. Test material

In Experiment 2, only *om* ‘if’ was tested with the same exact 16 items as were tested in Experiment 1. We also re-used the fillers.

4.2 Participants

100 participants completed the study. The exclusion criteria applied in Experiment 1 were also applied in Experiment 2. Six participants were excluded for reporting being bilingual. One participant was excluded for failing to report being a native Norwegian speaker. 37 participants were excluded for having >5 responses below 1000 ms. We characterized these respondents as “false respondents” as they typically had >50 responses below 1000 ms.

In total, 56 participants were included in the data material. Out of 56, 49 participants reported being aged between 18–24. All dialect groups were represented, with the most frequent reply being *østlandsk* ‘Eastern Norwegian’ (14 responses).

4.3 Procedure

The study followed the same procedure as Experiment 1, with two exceptions. First, items were not distributed across different lists. The Latin Square distribution of test sentences in Experiment 1 makes it impossible to distinguish participant variation from item variation. To control for this, every participant was presented with all test-sentences in experiment 2 in the exact same randomized order. Such a design allows us to control for (i) participant effects, which will be the same across items, (ii) item effects, which will be the same across participants, and finally (iii) potential ordering effects, which will be the same across items and participants. Participants saw 64 (16×4) test sentences for *om* ‘if’, 64 fillers (48 bad, 16 good) and 2 unmarked practice sentences.

Second, participants were recruited through NTNUs internal student platforms and one external student’s social media platforms. We think it is highly unlikely for someone to have participated in both Experiments 1 and 2. Participants received monetary reward for completing the study (150 NOK).

4.4 Data analysis

Data analysis was conducted as for Experiment 1. A linear mixed-effects model was fit with a two-way interaction term crossing the main effects *distance* and *structure*. We also fit a linear mixed effects model that included *item* as a fixed effect in an interaction with *distance* and *structure*. Here the model makes item 1 the reference level, and the model outputs must be read in relation to this reference level. We calculated by-participant DD-scores aggregated over all items and by-item

DD-scores aggregated over all participants. As we did for Experiment 1, we checked for satiation effects. As test sentences were given in the same order across participants, satiation effects are conflated with potential item effects. Thus, we will not rely too heavily on any results of these analyses here. We ran a model for target conditions, modelling z-scores by trial index, with by-subject and by-item varying intercepts. We also ran separate models for each target condition and for bad fillers, checking whether trial index co-varied with z-scores for each condition. Here, we also fit by-subject and by-item intercepts. Based on the evidence in Chaves & Putnam (2020) for conditional clauses, since participants were exposed to 16 island violating conditions we hypothesized that we would see some evidence of satiation for the “long, island”-condition.

4.5 Results

4.5.1 Overall results

The bad fillers received low ratings, and the good fillers received high ratings. **Table 5** provides an overview of average ratings for each condition included in Experiment 2.

<i>Condition</i>	<i>Mean z-score</i>	<i>SD</i>
<i>Bad fillers</i>	-0.896	0.726
<i>Good fillers</i>	0.761	0.666
<i>Short, no-island</i>	0.736	0.569
<i>Long, no-island</i>	0.675	0.581
<i>Short, island</i>	0.484	0.673
<i>Long, island</i>	0.031	0.762

Table 5: Overview of average ratings (z-scored) and standard deviations for every condition, Experiment 2.

The linear mixed effects model with a two-way interaction between *distance* and *structure* returned a significant interaction effect, in addition to significant main effects of distance and structure (see **Table 6**). The model indicates, through the size of *t*, that the main effect of *structure* is greater than the effect of *distance*. We also see an intermediate effect size, and an average z-scored rating of the “long, island” condition just above 0. This implies that *om* ‘if’ yields intermediate island effects in Norwegian, as can be visually confirmed in **Figure 6**. As such, Experiment 2 replicates Experiment 1.

We also investigated the distribution of z-scores on the four conditions, which shows that there is more variation on the “long, island” condition compared to the three baseline conditions.

For the three baseline conditions there is a narrow distribution around $z = 1$, with a thin rightward tail indicating some variation. For the “long, island” condition, however, we see a wide distribution.

	Estimate	SE	t	p
<i>Intercept</i>	0.030	0.095	0.308	0.791
<i>distance: short</i>	0.454	0.105	4.320	<0.001
<i>structure: no-island</i>	0.645	0.085	7.558	<0.0001
<i>distance × structure</i>	-0.392	0.106	-3.680	0.002

Table 6: Main results of the linear mixed effects model, Experiment 2.

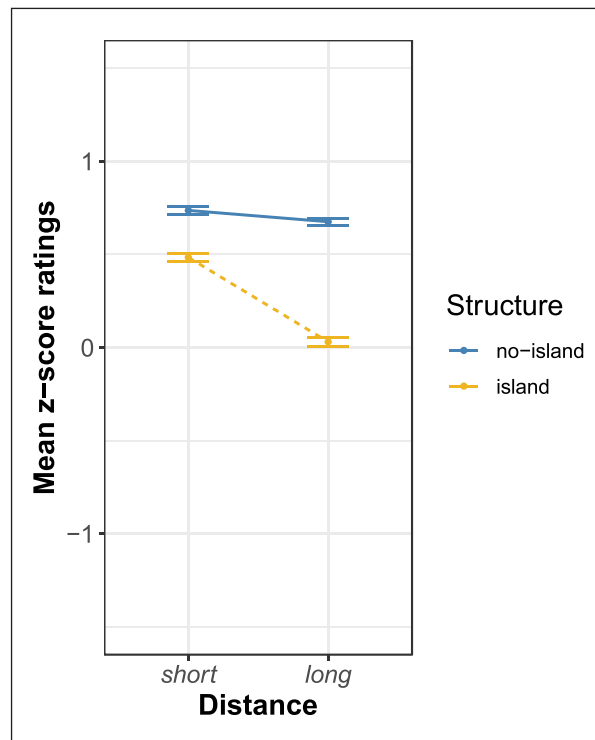


Figure 6: Interaction plot for *om* ‘if’, Experiment 2.

The density plot in **Figure 7** shows that a portion of scores on the “long, island” condition overlaps with the “long, no-island” condition, indicating that for some portion of the trials, the “long, island” condition is indistinguishable from the “long, no-island” condition. An analysis with `overlap()` from the `overlapping-package` in R (Pastore 2018) shows that these distributions are 44% different (following the procedure detailed in Pastore & Calcagni 2019). This means that the distributions of the scores for the “long, no-island” and the “long, island” conditions are more similar than they are different.

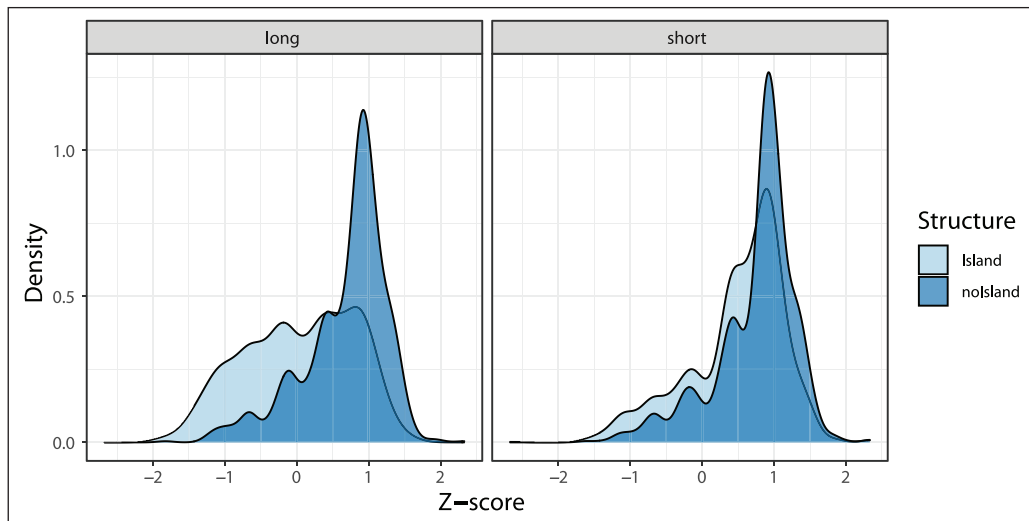


Figure 7: Comparing the distribution of z-scores for *om* 'if' on the no-island and the island conditions for the long and short conditions separately, Experiment 2.

Comparing **Figure 7** to the distribution of scores on the bad and good fillers, we see the way in which scores are distributed for two conditions that are consistently distinguished by participants.¹⁷ **Figure 8** shows that there is only marginal overlap between z-scores for the filler conditions, meaning that the fillers were consistently distinguished across trials. An overlap analysis finds that they are 85% different.

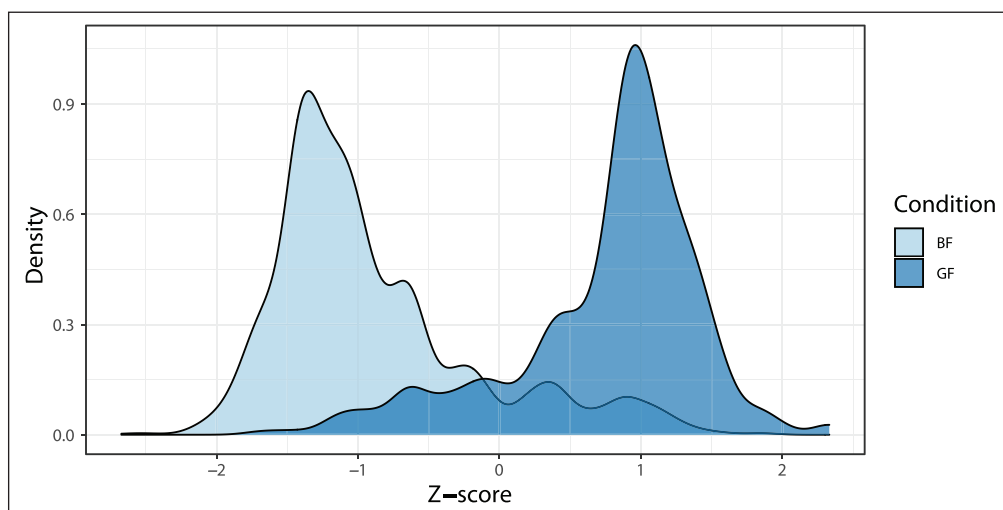


Figure 8: Distribution of z-scores for the fillers, Experiment 2. The bad fillers show a narrow distribution around -1.5 . The good fillers show a narrow distribution around 1 .

¹⁷ Figure (8) also shows that participants understood the task and executed it according to instructions.

We see here that for participants *om* ‘if’-adjuncts are not unacceptable in the same way as the bad fillers, nor acceptable in same way as the good fillers.

Looking at satiation effects, we ran a similar linear mixed effects model investigating the effect of trial index on z-score as we did for the “long, island”-conditions in Experiment 1 (see **Table 4**). An overview of model outputs is provided in **Table 7**. We see an overall satiation effect across conditions, but the estimate is very low. With an estimate of 0.0025, each new test sentence will see a very small increase in rating (across all conditions), which means that after being exposed to 64 test sentences a z-score of e.g. $z = 0.2$ will increase to $z = 0.34$. Fitting models for each condition separately, we do not see a significant increase in rating as the experiment proceeded.

	<i>Number of data points</i>	<i>Estimate</i>	<i>t</i>	<i>p</i>
Overall	3570	0.0025	6.440	< 0.0001
Short, no-island	892	0.0005	0.289	0.776
Long, no-island	891	0.0020	1.348	0.1989
Short, island	896	0.0007	0.340	0.7388
Long, island	891	0.0007	0.286	0.779
Bad fillers	2677	0.0024	1.352	0.183

Table 7: Overview of results from linear mixed effects models testing for satiation, Experiment 2.

Separating the responses on the “long, island”-condition into four blocks (the first four responses in block 1, etc.), we see the same pattern that we see in the model for this condition, i.e., no indication that late blocks are rated better than earlier blocks. This is illustrated in **Figure 9** below.

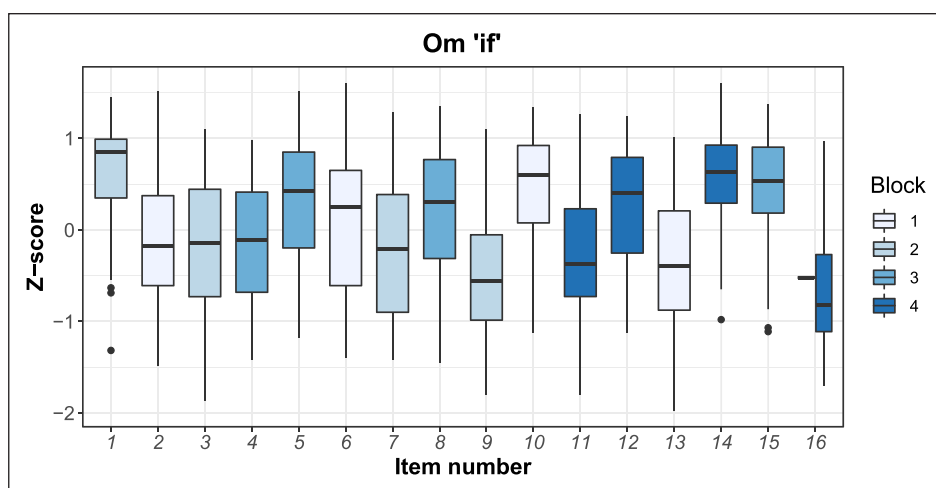


Figure 9: Boxplot illustrating the average judgments on the “long, island” condition by item. The different shades of blue indicate the different block numbers.

4.5.2 Results – variation

The average results for *om* ‘if’ are in the intermediate range. However, in the distribution of scores we see variation between trials. The distribution of scores is wider than the distribution for any of the filler and other target conditions. Therefore, we want to investigate this variation more closely to see if there are any meaningful patterns either between participants or between items. If so, we expect to see grouping of participants and/or items.

First, we looked at variation between items. We fit a linear mixed effects model on our data in a three-way interaction between *item*, *distance*, and *structure*. The model did not return a significant interaction effect and found only a significant effect of *structure*. The model returned significant differences between item 1 (reference level) and several items, but there were also several items that were found not be distinguishable from item 1 (see Supplementary file).

Visually inspecting the items in an interaction plot in **Figure 10**, it is clear why the model did not return a significant interaction effect, nor significant main effects when item 1 was set as the reference level. For item 1, there is only minimal linear additivity between conditions, which is reflected in a DD-score of -0.09 . Linear additivity and DD-scores close to 0 are the common denominators for items that the model did not distinguish from item 1. In comparison, the items that were found to be distinct from item 1 show super-additivity. These also have DD-scores well above 0. There is, however, variation in the size of the DD-score between the items that the model distinguished from item 1.

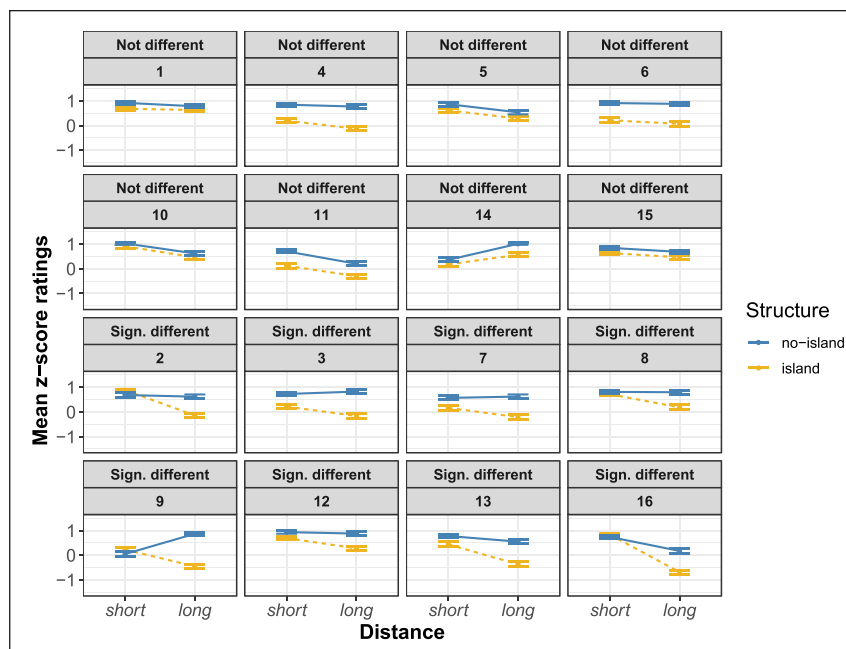


Figure 10: Interaction plot by item for *om* ‘if’, Experiment 2. The items that the model did not distinguish from item 1 are labelled “Not different”, while the items that the model did distinguish from item 1, “Sign. different”.

Looking at each item separately in this manner, we see that there are differences between items. As participant- and ordering-effects are kept constant across the experiment, the variation can in fact be attributed to item variation. Nevertheless, investigating the distribution of the DD-scores by items aggregated over participants in **Figure 11b**, we see that there is in fact normal distribution (with a positive skew) around an intermediate score. In other words, we do not see indications of item grouping. This suggests that the variation we see in **Figure 10** might be random variation that we can expect to see by chance.

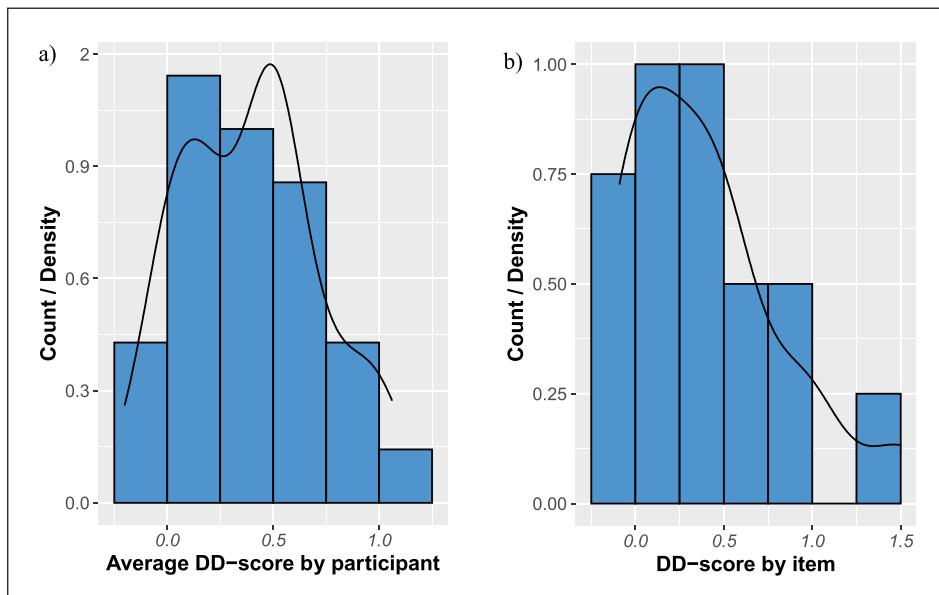


Figure 11: DD-scores calculated by participant across 16 items (a) and DD-scores calculated by item across 56 participants (b), Experiment 2. Histograms are plotted with `geom_histogram()`, `boundary = 0`, `binwidth = 0.25`.¹⁸

This does not exclude the possibility that there is variation at the participant level. The design allows us to calculate DD-scores for each participant aggregated over the same 16 items. This means that we have a large sample of items that make up the average DD-scores per participant. As such, if we see differences between DD-scores we will assume that these reflect real differences between participants. Investigating the range of DD-scores in a histogram we see that there is a wide range of DD-scores ranging from an average score below 0 to an average score above 1. However, the histogram in **Figure 11a** shows that participants' DD-scores are widely,

¹⁸ Plot specifications are set following suggestions from Jon Sprouse (p.c.). The absolute split between an island effect and a reverse island effect (see Sprouse et al. 2011) is 0. Thus, setting the boundary at 0 allows us to visually inspect the number of DD-scores above and below this point. As seen in previous experiments, the relative effect size that can be set as a distinction between an island effect and a null effect is close to 0.25. Thus, setting the binwidth to 0.25 allows us to see the number of DD-scores that fall within this range.

but normally distributed around the average DD-score ($DD = 0.39$). In other words, we do not see signs of participant grouping. Accordingly, we do not see indications in the variation between items or between participants that the intermediate effect is caused by aggregating over variable judgments.

Importantly, intermediate scores are also represented in the raw ratings.¹⁹ Looking at the raw ratings by condition in **Figure 12**, we can recognize the pattern of the z-scored ratings. They tell us that participants use the full range of the scale, but that the most frequent responses are in the intermediate range. We also see that there is a large portion of ratings on the “long, island” condition at 7, i.e., the highest score possible. In terms of absolute ratings of an island violating sentence, this tells us that for some items some participants did not find these island violations to be unacceptable. Comparing the raw scores of the test sentences to the fillers in **Figure 12**, we see that there is a larger proportion of intermediate ratings for the “long, island” condition than for bad fillers and fewer high ratings than for good fillers.

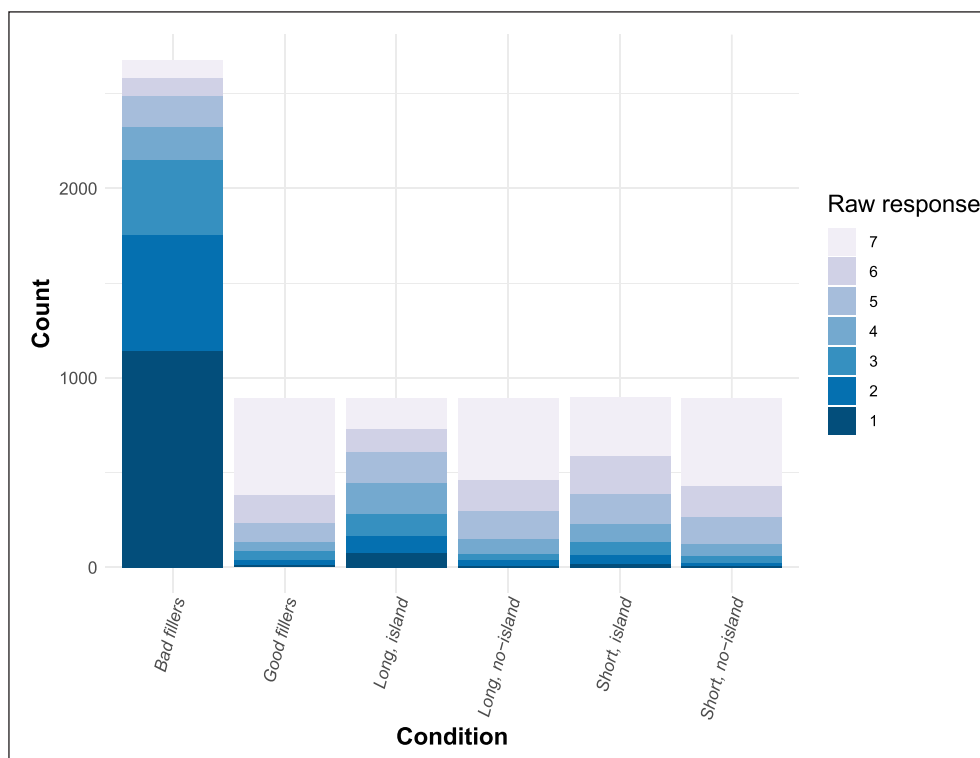


Figure 12: Barplots displaying the count of raw responses per condition, Experiment 2.

¹⁹ We also ran ordinal logistic regressions with the raw data for Experiment 2. As the results were the same as with the linear mixed effects models with z-scores, we will not report the ordinal logistic regressions here but see Supplementary file.

4.6 Intermediate summary

In Experiment 2 the on average intermediate island effects of forming an *rc*-dependency into *om* ‘if’-adjuncts in Norwegian were replicated. The results replicate in a design where participants see every test sentence, as opposed to distributing items in a Latin Square Design. Thus, it seems that the number of exposures to lexicalizations of the same test conditions does not influence acceptability. Overall, we find a significant interaction effect, a super-additive judgment pattern (see **Figure 4**), a DD-score of 0.39 and an average rating of the “long, island” condition just above 0. Though we see variation both at the item and participant level, there is normal distribution around an intermediate effect size. Such a distribution of DD-scores indicates that the average intermediate results for *om* ‘if’ do not conceal meaningful variation between items and/or participants or order of exposure. Thus, it seems that the intermediate effect is not caused by (the most obvious) extra-grammatical factors. Accordingly, the intermediate results for *om* ‘if’ seem to reflect the accurate underlying acceptability pattern for this adjunct clause type.

5 Discussion

The present study investigates adjunct clauses in *rc*-dependencies in Norwegian. The goal of the study is to conduct a formal investigation of the empirical landscape and map out general patterns. Specifically, we ask whether adjunct clauses are islands for relativization in Norwegian and whether adjunct clauses behave like a uniform group for relativization. The current section is organized around these questions. Experiment 1 reveals consistent variation between adjunct clause types. For that reason, we will first discuss the second research question before turning to the first.

5.1 Do adjunct clauses behave like a uniform group for relativization?

Following up on Bondevik et al.’s (2021) results where finite adjunct clauses introduced by *fordi* ‘because’, *når* ‘when’ and *om* ‘if’ did not behave as a uniform group for *top*-dependencies, the present study finds that these three adjunct clauses do not behave like a uniform group for *rc*-dependencies either. First, the linear mixed effects model did not distinguish between *fordi* ‘because’ and *når* ‘when’, but distinguished *fordi* ‘because’ and *om* ‘if’. This indicates that *fordi* ‘because’ and *når* ‘when’ received judgments that, on average, were similar enough to accept the null hypothesis that these behave alike in *rc*-dependencies. In addition, judgments on the “long, island” condition are similarly distributed around a negative z-score across the two adjunct clause types. There is slightly more variation in the scores for *når* ‘when’ than is seen in *fordi* ‘because’. *Om* ‘if’, on the other hand, yields smaller DD-scores across experiments 1 and 2, compared to *fordi* ‘because’ and *når* ‘when’. In addition, ratings of the “long, island” condition fall above 0. In other words, *om* ‘if’ yields intermediate results.

The data clearly shows that there are systematic differences between the adjunct clause types. accordingly, we need our theory of adjunct constraints to explain these differences. However, there are no traditional syntactic approaches to islands that can readily accommodate the distinction between adjunct clause types which our, Bondevik et al.'s (2021) and Nyvad et al.'s (2022) results necessitate. Comparing the recent findings, we find systematicity in the extraction patterns: Causal clauses (tested with *because*) and habitual clauses (tested with *when*) yield low ratings and island effects across three dependency types and across two languages (Norwegian: *top*, *rc*; English: *dem rc*, simple and complex *wh*). Conditional clauses (tested with *if*) yield high ratings, more closely resembling declarative clauses than the other adjunct clause types across dependency types and languages (Norwegian: *top*, *rc*, *dem rc*²⁰; English: *dem rc*, and no island effect in *rc*). We believe that the systematicity in the recent findings indicate that there is some identifiable and general constraint that governs extraction from adjunct clauses. However, none of the current theories provide a ready-made solution to this puzzle. Thus, we must explore more fine-grained versions of the current theoretical approaches.²¹

Investigating the semantics of each complementizer, we see clear differences between adjunct clause type. In (10) the same sentence is presented with the three different complementizers to allow for an easy comparison of the meaning of each complementizer.

- (10) *Nils snakker med kunstsamlere som jubler om / når / fordi noen kjøper maleriet av Van Gogh.*²²
 Nils talks with art dealer.DEF who celebrates if / when / because someone buys painting.DEF by Van Gogh
Om 'if': 'Nils is talking to the art dealer who will celebrate if someone buys the painting by Van Gogh.'
Når 'when': 'Nils is talking to the art dealer who will celebrate when someone buys the painting by Van Gogh.'
Fordi 'because': 'Nils is talking to the art dealer who is celebrating because someone is buying the painting by Van Gogh.'

Om 'if' introduces a conditional clause. The *om 'if'*-clause specifies a condition, and the clause that it modifies conjectures an outcome of the fulfillment of the condition (Hornstein 1990: 74).

²⁰ High ratings of *wh*-extraction from finite *om 'if'* in Norwegian in Kobzeva et al. (2022), but low ratings in same dependency type in Kush et al. (2018).

²¹ Another possibility recently explored in Abeillé et al. (2020) is that islands are constrained by general discourse factors. However, very recent research has found several indications that Abeillé et al.'s (2020) *focus-background conflict* constraint makes the incorrect predictions (see Kobzeva et al. 2022; Šimík et al. 2022; Nyvad et al. 2022). Therefore, we do not pursue this approach further.

²² Norwegian present tense covers the meaning of both simple present and present progressive in English, in addition to having a broader future-oriented use than English present tense.

A causal relationship between the conditional and relative clause is implied – if the condition is not satisfied, the event in Vrel (finite verb in relative clause)²³ might still occur, but not for the reason expressed in the conditional clause. *Når* ‘when’ introduces a habitual clause. The meaning of *når* ‘when’ in this use is “every time the event expressed in Vadjunct (finite verb in the adjunct clause) occurs, the event in Vrel also occurs”. It is presupposed that both the event in the relative clause and in the ‘when’-clause have minimally occurred once. A causal relationship between the adjunct and the relative clause is implied. *Fordi* ‘because’ introduces a causal clause that explicitly expresses the cause of the event in Vrel.

As each complementizer contributes different meanings to the sentence, it is possible that each complementizer conditions the adjunct clause’s opacity differently. As semantic conditions have been shown to govern extraction from non-finite adjunct clauses (Truswell 2011; Ernst 2022), it is not improbable that there are semantic conditions on finite adjunct clauses as well (see also Abrusán 2014 on semantic conditions in weak islands). Truswell (2007; 2011) proposes the semantic condition on adjunct islands that *wh*-extraction is only possible if two events can be construed as one event (2011: 157). This is captured in the *Single Event Grouping Condition* (SEGC). For events described in different clauses to be construed as a single event (i) the events described in the two clauses must have spatiotemporal overlap; and (ii) there can be maximally one agentive verb. Spatiotemporal overlap means that the event grouping must “[...] happen in a single place, as well as in a single time” (Truswell 2011: 48). Accordingly, Truswell’s SEGC can make distinctions between adjuncts that are structurally the same, but differ in their semantics. Truswell (2011) does not explicitly extend the SEGC to non *wh*-dependency types.

Truswell (2011) specifically shows that his condition does not apply to finite adjunct clauses, arguing that a finite operator blocks extraction from finite adjuncts regardless of the SEGC (2011: 118). Extending Truswell’s approach, Ernst (2022) relaxes the complete ban on extraction from finite adjunct clauses.²⁴ Ernst (2022) assumes that non-finite clauses also include a tense operator, and thus, he rejects that a tense operator in and of itself blocks movement from finite adjunct clauses.

Ernst’s (2022) extension of Truswell’s (2011) SEGC has the potential to explain the difference between adjunct clause types without additional machinery: The three different sentences, on their most natural readings, imply different temporal relations between Vrel and Vadjunct. While temporal location of the event time in either the relative or the conditional clause is possible with *om* ‘if’, temporal location of two separate event times is possible for both *fordi* ‘because’ and *når* ‘when’. Thus, although there is grammatical tense in the *om* ‘if’-clauses, the lack of temporal interpretation means that there are not two “independently determined” times that the clauses

²³ See description of test sentences in Section 3.2.

²⁴ Müller (2019) finds that finiteness does not matter for Swedish adjunct clauses in the same way as Truswell (2011) argues for English. Bondevik et al.’s (2021) and Kush et al.’s (2019) results also strongly suggest that finiteness should not matter for topicalization in Norwegian either.

can be associated with, in fact the event times are undetermined. It is possible, following Ernst (2022), that extraction is facilitated as there is only one determined time (Vmatrix) in the *om* ‘if’-items. It would be interesting to test extraction from *om* ‘if’-adjuncts where both the event in Vrel and Vadjunct can be temporally located.²⁵

There is, however, nothing within this theory that can explain the robustness of the intermediate results for extraction from conditional clauses.

5.2 Are adjuncts islands for relativization in Norwegian?

5.2.1 Classic island effects: *Fordi* ‘because’- and *når* ‘when’-clauses

We find evidence that *fordi* ‘because’ and *når* ‘when’ both yield classic, super-additive island effects in *rc*-dependencies in Norwegian. Our results for *fordi* ‘because’ and *når* ‘when’ are in alignment with the traditional syntactic view of adjunct clauses as islands. The results for *fordi* ‘because’ and *når* ‘when’ provide initial evidence that *top*- and *rc*-dependencies behave similarly with respect to islandhood in Norwegian. *Fordi* ‘because’ and *når* ‘when’ yield classic island effects both in a contrastive *top*-dependency (Bondevik et al. 2021) and *rc*-dependency in Norwegian. We see a similar pattern for English – *because* and *when* yield much lower ratings in a *dem rc* than *if* (Nyvad et al. 2022).

The pattern for *fordi* ‘because’ and *når* ‘when’ differs from previous findings for adjunct clauses in an *rc*-dependency (Sprouse et al. 2016). Sprouse et al. (2016) conclude that there is only evidence of a processing constraint of forming an *rc*-dependency into finite adjunct clauses in English. In other words, there is not evidence of a grammatical constraint. Given that we find that adjuncts do not behave like a uniform group for *rc*-dependencies in Norwegian and for *dem rcs* in English, it would be interesting to see how our and Sprouse et al.’s (2016) results compare to judgments for *because*- and *when*-adjunct clauses in *rc*-dependencies in English, and how Kobzeva et al.’s (2022) and Nyvad et al.’s (2022) results for *dem rcs* compare to results for *fordi* ‘because’ and *når* ‘when’ in *dem rcs* in Norwegian. We think that there are important differences between *dem rcs* and the *rc*-dependencies that we and Sprouse et al. (2016) tested, and that the two should not be collapsed. Thus, we do not necessarily expect the same results across these dependency types. For one, the ratings of the “long, no-island” condition for adjunct-items in Kobzeva et al. (2022) are much lower than ratings for the same condition in the current experiments. We also see this pattern for the subject island. The lowered ratings on a non-island violating condition is an indication that the dependency types are not directly comparable, irrespective of islandhood. An additional confound, which Kobzeva et al. (2022) point out, is that *dem rcs* have the same surface structure as clefts in Norwegian, and their test sentences are subsequently ambiguous between a cleft and a *dem rc* reading.

²⁵ Such an interpretation would be possible in Norwegian if both clauses were in the past tense or, as an anonymous reviewer suggests, with a temporal adverbial clause like *i morgen* ‘tomorrow’.

5.2.2 Island undecided: the special case of *om* ‘if’

While *fordi* ‘because’ and *når* ‘when’ display classic island effects, *om* ‘if’ does not in Norwegian. In both experiments in the current study, *om* ‘if’ (i) yields island effects which are smaller than the island effects of extracting from the other two adjunct clauses, and (ii) causes a larger decrease in acceptability compared to extraction from declarative clauses. That is, on average, forming an *rc*-dependency into *om* ‘if’ yields a judgment pattern that fits the description of an intermediate effect. Experiments 1 and 2 provide evidence that *om* ‘if’ consistently yields intermediate acceptability judgments. As such, we can say that *om* ‘if’ causes less breakdown of acceptability compared to the superadditive islands *fordi* ‘because’ and *når* ‘when’.

The problem, however, is how to interpret such intermediate results. The data clearly suggests a theory that can accommodate intermediate effects. Since the results show that the intermediate judgments are not caused by aggregating over variable results, we need to explain how such intermediate ratings arise. We believe the results indicate that we are dealing with a gradient island effect. This means that we are not looking at a binary division between “island” and “no-island”, but instead we see that *om* ‘if’ consistently falls somewhere in between. This is also the conclusion that Nyvad et al. (2022) reach. We draw this conclusion somewhat reluctantly as postulating gradience in island effects has wide-spread theoretical implications. Traditionally, a gradient judgment pattern is impossible to entertain without assuming that the intermediate judgments reflect gradience in acceptability, as opposed to gradience in grammaticality. Thus, we will begin by exploring one way in which gradience in acceptability can be implemented.

One possibility is to assume that *om* ‘if’ is a *subliminal island* (Almeida 2014). Subliminal island effects are defined as cases where “measurable island sensitivity effects are observed, and yet do not lead to gross sentence unacceptability” (2014: 87). Thus, like we see for *om* ‘if’, a subliminal island will be more acceptable than an island yielding traditional island effects and less acceptable than a non-island. Interpreting the intermediate effect for *om* ‘if’ as a subliminal island, we would have to assume that a grammatical constraint applies to *om* ‘if’, *fordi* ‘because’ and *når* ‘when’ in the same way, but that there is something that causes *om* ‘if’ to be perceived as more acceptable. Almeida (2014) theorizes that subliminal effects occur when speakers perceive the island violation to be in the acceptable range, but a subconscious island constraint causes a decrease in acceptability. This allows us to sustain a theory that does not distinguish between adjunct clause types syntactically. However, we do not favor this interpretation. First, it is difficult to imagine a scenario where the acceptability of a categorically ill-formed sentence can improve, unless there is a grammaticality illusion at play. It is unlikely that grammatical factors such as plausibility, semantic felicitousness etc. can ameliorate syntactic/pragmatic/semantic violations (see e.g., Juzek & Häussler 2019 for an experimental investigation of this issue). Second, defining conditional clauses as “subliminal islands” does not provide an explanation for *why* the constraint is subliminal with *om* ‘if’-clauses and not with *fordi* ‘because’ and *når* ‘when’-clauses.

As the explanation of *om* ‘if’ as a subliminal island does not seem to provide a satisfying account of the intermediate ratings, we find that the data leads us to explore options that allow a more direct mapping between acceptability and grammaticality. It is possible to conceptualize a non-binary theory of grammar that can account for the relevant differences.

Within a *Barriers*-like system, it is possible that relativizing a DP from inside an *om* ‘if’-clause crosses one barrier, while relativizing from *fordi* ‘because’ and *når* ‘when’ crosses at least two. Chomsky (1986: 28) assumes gradience in the acceptability of long-distance movement: “[...] movement should become “worse” as more barriers are crossed, the best case being the crossing of zero barriers”. We might rely on differences in the height of adjunction or level of integration with the relative clause to distinguish between clause types and the number of barriers that are crossed. For instance, ‘whether’-clauses are complement clauses, and subsequently properly governed in a *Barriers*-system. Thus, the mover will not cross any barriers on its way to the matrix Spec-CP. Adjunct clauses, however, are not properly governed, and like all adjunct clauses, cause the mover to cross two barriers on its way to the matrix Spec-CP. It is possible that the place of adjunction or the internal structure of the conditional adjunct clause is different from other types of adjunct clauses such that the mover will only cross one barrier leading to an intermediate decrease in acceptability. Irrespective of the specific implementations of the *Barriers*-system, the main point remains: There does not have to be just one main hurdle that the dependency must cross in order for the filler-gap-dependency to be established, there might be several smaller hurdles that must be crossed. The best case being that no hurdles are crossed, the worst being many. A theory of what the relevant hurdles might be for adjunct islands remains to be extensively investigated within a post-Government-and-Binding approach (though see Villata et al. 2016 and Beljon et al. 2021 for approaches using featural Relativized Minimality to explain gradience in *wh*-islands).²⁶

Villata & Tabor (2022) provide a model in which gradience in islandhood is the outcome of *coercion*. When the parser finds no other outcome, a new interpretation of the structure is forced (coerced) such that a possible parse is available. For instance, Villata & Tabor (2022) argue that in cases of the Complex Noun Phrase Constraint (CNPC) – the parser cannot integrate the encountered filler into a gap position as the only available gap is inside a complex NP. Thus, in cases where the complex NP is similar enough, it will coerce the interpretation to be VP + CP. For instance, “hear the rumor that” is coerced as “hear that”. Such coercion means that the sentence will be perceived as less acceptable, even if the coerced structure is grammatical. They argue that this provides an explanation of weak islands. In cases of strong islands, the parser cannot find a gap position for the filler, but coercion is not available in such a way that the outcome will be somewhat grammatical.

²⁶ There are also other possible implementations. For instance, Kathol (2001) provides a lexicalist account for the difference in the distribution of island effects based on lexical properties of complementizers introducing parasitic gaps in German. Space does not allow us to discuss such implementations further.

Villata & Tabor's (2022) model provides an explanation of how certain grammatical features are relevant for the parser. If we apply Villata & Tabor's (2022) model to our Norwegian data, the parser is able to coerce the *om* 'if'-clause into a complement, but unable to do so successfully with *fordi* 'because' and *når* 'when'-clauses. The semantics of *om* 'if' is very similar to the semantics of *om* 'whether', and perhaps it is possible that this similarity is utilized by the parser (at the expense of acceptability) such that a gap site inside the *om* 'if'-clause is found, just as it is in *om* 'whether'-clauses. Furthermore, it is possible that the availability of a SEG reading is necessary for the adjunct clause to be coerced as a complement clause.

We believe that there is something to be gained from the approaches that assume a (semi-) direct mapping between acceptability and grammaticality. However, in order for them to satisfactorily solve the problem that non-uniformity in adjunct island constraints poses, further investigations of the syntax and semantics of different adjunct clauses are required. A concrete proposal for which features, structures or interpretations that are responsible for the difference between *om* 'if' on the one hand, and *fordi* 'because' and *når* 'when' on the other, is beyond the scope of this paper.

6. Conclusion

We have shown that, overall, adjunct clauses yield significant island effects in an *rc*-dependency in Norwegian. However, the three adjunct clause types we tested – conditional *om* 'if', habitual *når* 'when' and causal *fordi* 'because' – do not behave as a uniform class for *rc*-dependencies. Instead, *fordi* 'because' and *når* 'when' pattern together yielding classic island effects, while *om* 'if' yields a judgment pattern that is best described as intermediate. Thus, our findings replicate previous findings showing that adjuncts are not a uniform class for islands (Müller 2019; Bondevik et al. 2021; Nyvad et al. 2022). Our second experiment provides evidence that the intermediate island effects seen for *om* 'if' replicate across a new sample of participants and with increased exposure to target conditions. Additionally, as we see normal distribution around an intermediate result, the experiments presented in this paper contribute strong evidence that the intermediate effect for *om* 'if' reflects the true underlying pattern. We believe that the origin of the intermediate effect size, which is not discernable from our, Bondevik et al.'s (2021), Kobzeva et al.'s (2022) or Nyvad et al.'s (2022) data, should be further investigated in future work as this is central to our understanding of islandhood. Our study, together with previous findings, provide evidence that we need fine-grained theories of islands and adjunct clauses. We need further research in order to build a theory or to extend already existing theories of gradience to fit the empirical landscape of this phenomenon, which is integral to our understanding of the constraints that govern language.

Data availability

All materials and analyses are provided in an osf: https://osf.io/d6wfe/?view_only=344f4132528b432593808e05d622d9bd. A README file is provided for ease of reproducibility. The test materials are provided in a csv file and the data analyses are provided in a knitted html file. In addition, a Supplementary file [DOI: <https://doi.org/10.16995/glossa.9033.s1>] provides easy access to test materials and model outputs in a pdf. One example item per clause type tested in Experiment 1 and only the model outputs of the models referenced in the paper are provided in the Supplementary file.

Ethics and consent

The studies were conducted in accordance with the Declaration of Helsinki. We obtained a written confirmation from the Norwegian Centre for Research Data (NSD) that the acceptability judgments studies including the background surveys did not collect personal data that allowed participants directly or indirectly to be identified, and accordingly, they assessed that the study did not need to be approved by NSD. All participants were informed about this before consenting to participate in the studies, including the fact that this level of anonymization meant that they could not withdraw their participation after completing the study.

Acknowledgements

We thank Dave Kush for ideas and inspiration at the early stage of this study. We also thank there anonymous reviewers, audiences at AcqVA seminars at NTNU, Jon Sprouse, Andrew Weir, Anastasia Kobzeva and Kristin Klubbo Brodahl for helpful comments and ideas. We would also like to thank Myrte Vos for helping us to set up the first experiment and Natalia Mitrofanova, Anna Giskes, and Sunniva Briså Strætkvern for looking at early versions of the paper. Any faults or weaknesses remain our own.

Competing interests

The authors have no competing interests to declare.

References

- Abeillé, Anne & Hemforth, Barbara & Winckel, Elodie & Gibson, Edward. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204. 104293. DOI: <https://doi.org/10.1016/j.cognition.2020.104293>
- Abrusán, Márta. 2014. *Weak island semantics*. Oxford Academic. DOI: <https://doi.org/10.1093/acprof:oso/9780199639380.001.0001>
- Áfarli, Tor A. 1994. A promotion analysis of restrictive relative clauses. *The Linguistic Review* 11. 81–100. DOI: <https://doi.org/10.1515/tlir.1994.11.2.81>

- Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABralin* 3(2). 55–93. DOI: <https://doi.org/10.5380/rabl.v13i2.39611>
- Anward, Jan. 1982. Basic Swedish. In Engdahl, Elisabet & Ejerhed, Eva (eds.), *Readings on unbounded dependencies in Scandinavian languages*. Stockholm: Almqvist & Wiksell, 47–75.
- Arnold, Taylor B. & Emerson, John. W. 2011. Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3(2). URL: http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf. DOI: <https://doi.org/10.32614/RJ-2011-016>
- Bates, Douglas & Maechler, Mächler & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Beljon, Maud & Joosen, Dennis & Koeneman, Olaf & Ploum, Bram & Sommer, Nöelle & de Swart, Peter & Wilms, Veerle. 2021. The effect of filler complexity and context on the acceptability of wh-island violations in Dutch. *Linguistics in the Netherlands* 38. 4–20. DOI: <https://doi.org/10.1075/avt.00047.bel>
- Bermingrud, Knut O. 1979. Setningsknutekonstruksjonen: en analyse av konstruksjonens grammatikalitet i moderne norsk. Oslo: University of Oslo. (MA thesis.)
- Bode, Stefanie. 2020. *Casting a minimalist eye on adjuncts*. New York: Routledge. DOI: <https://doi.org/10.4324/9780367822613>
- Bondevik, Ingrid, & Kush, Dave & Lohndal, Terje. 2021. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics* 44(3). 1–32. DOI: <https://doi.org/10.1017/S0332586520000207>
- Chaves, Rui. 2021. Island phenomena and related matters. In Müller, Stefan & Abeillé, Anne, Borsley & Robert D. & Koenig, Jean-Pierre. (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 633–687. Berlin: Language Science Press.
- Chaves, Rui & Putnam, Mike. 2020. *Unbounded dependency construction: Theoretical and experimental perspectives*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198784999.001.0001>
- Chomsky, Noam. 1973. Conditions on transformations. In Halle, Morris & Anderson, Stephen R. & Kiparsky, Paul (Eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart and Winston.
- Chomsky, Noam. 1977. On wh-movement. In Peter Culicover & Thomas Wasow & Adrian Akmajian (Eds.), *Formal syntax*, 71–132. Academic Press.
- Chomsky, Noam. 1986. *Barriers* (Vol. 13). Cambridge, Mass: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Martin, Roger & Michaels, David & Uriagereka, Juan & Keyser, Samuel J. (Eds.), *step by step—essays on minimalist syntax in honor of Howard Lasnik*, 89–155. MIT.
- Dal Farra, Chiara. 2020. To be or not to be an island – the status of adjuncts. Venice: Ca' Foscari University of Venice. (Doctoral dissertation.)
- de Leeuw, Joshua R. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods* 47(1). 1–12. DOI: <https://doi.org/10.3758/s13428-014-0458-y>

- Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. In Engdahl, Elisabet & Ejerhed, Eva (Eds.), *Readings on unbounded dependencies in Scandinavian languages*, 151–174. Umeå: Almqvist & Wiksell.
- Engdahl, Elisabet & Ejerhed, Eva. 1982. *Readings on unbounded dependencies in Scandinavian languages*. Umeå: Almqvist & Wiksell.
- Ernst, Thomas. 2022. The adjunct condition and the nature of adjuncts. *The Linguistic Review*. DOI: <https://doi.org/10.1515/tlr-2021-2082>
- Faarlund, Jan T. 1992. *Norsk syntaks i et funksjonelt perspektiv*. Oslo: Universitetsforlaget.
- Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and composition of the left periphery*. New York: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199858774.001.0001>
- Hornstein, Norberg. 1990. *As time goes by. Tense and Universal Grammar*. Cambridge: MIT Press.
- Hornstein, Norbert & Nunes, Jairo. 2008. Adjunction, labeling, and bare phrase structure. *Biolinguistics* 2(1). 57–86. DOI: <https://doi.org/10.5964/bioling.8621>
- Huang, C-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: MIT. (Doctoral dissertation.)
- Juzek, Tom S. & Häußler, Jana. 2019. Semantic influences on syntactic acceptability ratings. In Gattnar, Anja, Hörnig, Robin, Störzer, Melanie & Featherston, Sam (Eds.), *Proceedings of linguistic evidence 2018: Experimental data drives linguistic theory*. Tübingen: University of Tübingen.
- Kathol, Andreas. 2001. On the nonexistence of true parasitic gaps in Standard German. In Culicover, Peter W. & Postal, Paul M. (Eds.), *Parasitic Gaps*, 315–338. Cambridge, MA: MIT Press.
- Keshev, Maayan, & Meltzer-Asscher, Aya. 2019. A processing-based account of subliminal wh-island effects. *Natural Language & Linguist Theory* 37. 621–657. DOI: <https://doi.org/10.1007/s11049-018-9416-1>
- Kobzeva, Anastasia & Sant, Charlotte & Robbins, Parker T. & Vos, Myrte & Lohndal, Terje & Kush, Dave. 2022. Comparing island effects for different dependency types in Norwegian. *Languages* 7(3). 197. DOI: <https://doi.org/10.3390/languages7030197>
- Kohrt, Annika & Sorensen, Trey & O'Neill, Peter & Chacón, Dustin. 2020. Inactive gap formation: An ERP study on the processing of extraction from adjunct clauses. *Proceedings of the Linguistic Society of America* 5(1). 15. DOI: <https://doi.org/10.3765/plsa.v5i1.4775>
- Kush, Dave & Dahl, Anne. (2020). L2 Transfer of L1 Island-insensitivity: The case of Norwegian. *Second Language Research* 38(2). 315–346. DOI: <https://doi.org/10.1177/0267658320956704>
- Kush, Dave & Lohndal, Terje, & Sprouse, Jon. 2018. Investigating variation in island effects. *Natural Language & Linguistic Theory* 36. 743–779. DOI: <https://doi.org/10.1007/s11049-017-9390-z>
- Kush, Dave & Lohndal, Terje, & Sprouse, Jon. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95. 393–420. DOI: <https://doi.org/10.1353/lan.0.0237>

- Marty, Paul & Chemla, Emmanuel & Sprouse, Jon. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: a journal of general linguistics* 5(1). 72. DOI: <https://doi.org/10.5334/gjgl.980>
- McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53(2–3). 99–149. DOI: [https://doi.org/10.1016/0024-3841\(81\)90014-0](https://doi.org/10.1016/0024-3841(81)90014-0)
- Müller, Christiane. 2019. *Permeable islands*. Lund: Lund University. (Doctoral dissertation.)
- Mæhlum, Brit & Røyneland, Unn. 2012. *Det norske dialektlandskapet: Innføringer i studiet av dialekter*. Oslo: Cappelen Damm akademisk.
- Nyvad, Anne Mette & Müller, Christiane & Christensen, Ken Ramshøj. 2022. Too true to be good? The non-uniformity of extraction from adjunct clauses in English. *Languages* 7(4). 244. DOI: <https://doi.org/10.3390/languages7040244>
- Pañeda, Claudia & Kush, Dave. 2022. Spanish embedded question island effects revisited: an experimental study. *Linguistics* 60(2). 463–504. DOI: <https://doi.org/10.1515/ling-2020-0110>
- Pastore, Massimiliano. 2018. Overlapping: a R package for estimating overlapping in Empirical distributions. *The Journal of Open Source Software* 3(32). 1023. DOI: <https://doi.org/10.21105/joss.01023>
- Pastore, Massimiliano & Calcagni, Antonio. 2019. Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in Psychology* 10. 1089. DOI: <https://doi.org/10.3389/fpsyg.2019.01089>
- Poulsen, Mads. 2008. Acceptability and processing of long-distance dependencies in Danish. *Nordic Journal of Linguistics* 31(1). 73–107. DOI: <https://doi.org/10.1017/S0332586508001832>
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.r-project.org/>.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Berlin: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110883718>
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT. (Doctoral dissertation.)
- Šimík, Radek & Biskup, Petr & Bartasová, Kateřina & Dančová, Markéta & Dostálková, Eliška & Hrdinková, Kateřina & Kosková, Gabriela & Kozák, Jaromír & Lupoměská, Klára & Maršík, Albert & Schejbalová, Edita & Yekimov, Illia. 2022. Extraction from clausal adjuncts in Czech: A rating experiment. Preprint published on researchgate, May 2022. Accessed from: https://www.researchgate.net/publication/360973028_Extraction_from_clausal_adjuncts_in_Czech_A_rating_experiment
- Sprouse, Jon. 2007. *A program for experimental syntax*. College Park: University of Maryland. (Doctoral dissertation.)
- Sprouse, Jon & Caponigro, Ivano & Greco, Ciro, & Cecchetto, Carlo. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34. 307–344. DOI: <https://doi.org/10.1007/s11049-015-9286-8>
- Sprouse, Jon & Fukuda, Shin & Ono, Hajime & Kluender, Robert. 2011. Reverse island effects and the backward search for a licenser in multiple wh-questions. *Syntax* 14(2). 179–203. DOI: <https://doi.org/10.1111/j.1467-9612.2011.00153.x>

- Sprouse, Jon & Villata, Sandra. 2021. Island effects. In Goodall, Grant (Ed.), *The Cambridge Handbook of Experimental Syntax*, 227–257. Padstow: Cambridge University Press. DOI: <https://doi.org/10.1017/9781108569620.010>
- Sprouse, Jon & Wagers, Matt & Phillips, Colin. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. DOI: <https://doi.org/10.1353/lan.2012.0004>
- Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1). 80–126. DOI: <https://doi.org/10.1111/j.1467-9612.2007.00094.x>
- Szabolcsi, Anna & Lohndal, Terje. 2017. Strong vs. weak islands. In Everaert, Martin & Riemsdijk, Henk van (Eds.), 1–51. *The Wiley Blackwell companion to syntax*, 8. John Wiley and Sons, Inc. DOI: <https://doi.org/10.1002/9781118358733.wbsyncom008>
- Teleman, Ulf & Hellberg, Staffan & Andersson, Erik. 1999. *Svenska Akademiens grammatik—4 Satsor och meningar*. Stockholm: Svenska Akademien.
- Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117. 1355–1377. DOI: <https://doi.org/10.1016/j.lingua.2006.06.003>
- Truswell, Robert. 2011. *Events, phrases, and questions*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199577774.001.0001>
- Villata, Sandra & Rizzi, Luigi & Franck, Julie. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179. 76–96. DOI: <https://doi.org/10.1016/j.lingua.2016.03.004>
- Villata, Sandra & Tabor, Whitney. 2022. A self-organized sentence processing theory of gradience: The case of islands. *Cognition* 222. 104943. DOI: <https://doi.org/10.1016/j.cognition.2021.104943>
- Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer Verlag, 2016. Retrieved from: <https://ggplot2.tidyverse.org>. DOI: <https://doi.org/10.1007/978-3-319-24277-4>
- Winter, Bodo. 2020. *Statistics for linguists*. New York: Routledge.
- Zenker, Fred & Schwartz, Bonnie D. 2017. Topicalization from adjuncts in English vs. Chinese vs. Chinese-English interlanguage. In LaMendola, Maria & Scott, Jennifer (Eds.), *Proceedings of the 41st annual Boston University Conference on language development* (806–819). Somerville: Cascadilla Press.

