

ARTICLE

# Variation in adjunct islands: The case of Norwegian

Ingrid Bondevik<sup>1</sup>, Dave Kush<sup>1</sup> and Terje Lohndal<sup>1,2</sup>

<sup>1</sup>Department of Language and Literature, NTNU, 7491 Trondheim, Norway and <sup>2</sup>UiT The Arctic University of Norway

Email for correspondence: [ingrid.bondevik@ntnu.no](mailto:ingrid.bondevik@ntnu.no), [dave.kush@ntnu.no](mailto:dave.kush@ntnu.no), [terje.lohndal@ntnu.no](mailto:terje.lohndal@ntnu.no)

(Received 29 June 2020; revised 10 October 2020; accepted 11 October 2020)

Finite adjunct clauses are often assumed to be among the strongest islands for filler–gap dependency creation cross-linguistically, but Kush, Lohndal & Sprouse (2019) found experimental evidence suggesting that finite conditional *om*-adjunct clauses are not islands for topicalization in Norwegian. To investigate the generality of these findings, we ran three acceptability judgment experiments testing topicalization out of three adjunct clause types: *om* ‘if’, *når* ‘when’ and *fordi* ‘because’ in Norwegian. Largely replicating Kush et al. (2019), we find evidence for the absence of strong island effects with topicalization from *om*-adjuncts in all three experiments. We find island effects for *når*- and *fordi*-adjuncts, but the size of the effects and the underlying judgment distributions that produce those effects differ greatly by island type. Our results suggest that the syntactic category ‘adjunct’ may not constitute a suitably fine-grained grouping to explain variation in island effects.

**Keywords:** adjunct complementizers; adjunct islands; acceptability judgments; contrastive topicalization; filler–gap dependency; islands; Norwegian; variation

## 1. Introduction

A common trait for natural languages is the ability to establish filler–gap dependencies between two elements across a distance in a sentence. For example, in (1), the *wh*-words *what/hva* ‘what’ are interpreted as the object of the verbs *fix/fikse* ‘fix’ in the English and Norwegian sentences.

- (1) a. What did Andreas think that Ole said that he probably could not fix \_?  
b. Hva trodde Andreas at Ole sa at han mest  
*what thought Andreas that Ole said that he most*  
*sannsynlig ikke kunne fikse \_?*  
*likely not could fix*  
‘What did Andreas think that Ole said that he probably could not fix?’

Filler–gap dependencies are unbounded, but there are constraints that limit the establishment of a dependency across certain domains. These domains are often

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

referred to as ISLANDS (Ross 1967). Many researchers hold that island constraints are unlearnable from input alone, and, thus, they theorize that islands somehow arise from innate principles (either constraints or learning biases) and are therefore part of Universal Grammar (UG; Chomsky 1964, 1973, 1986; Ross 1967; Huang 1982; Rizzi 1990; Lasnik & Saito 1992; Manzini 1992; Phillips 2013a:107).<sup>1</sup>

Adjuncts were first identified as islands by Huang (1982). In the examples in (2), trying to link a *wh*-filler to a gap inside an adjunct clause renders the sentences unacceptable:

- (2) a. \*Who did Mary cry [after John hit \_]?  
(Huang 1982:503)
- b. ?\* Which bottle of wine was Mick annoyed [because Keith drank \_]?  
(Roberts 1997:217)

Huang (1982:505) posited the CONDITION ON EXTRACTION DOMAINS (CED) such that both subjects and adjuncts would be considered islands for extraction:

- (3) *Condition on Extraction Domains* (CED)  
A phrase A may be extracted out of a domain B only if B is properly governed.

Although the notion of proper government has been abandoned in recent theoretical frameworks, the notion that adjuncts, as a general structural class, are islands remains pervasive.

While certain non-finite adjuncts have been acknowledged to be exceptions to the CED,<sup>2</sup> in addition to certain complex subject clauses (Stepanov 2007, Abeillé et al. 2020), finite adjuncts are often considered among the strongest islands cross-linguistically (Huang 1982, Stepanov 2007, Truswell 2011, Sprouse & Hornstein 2013a). However, anecdotal evidence suggests that Mainland Scandinavian (MSc) languages allow filler-gap dependencies to be formed into a tensed adjunct clause (Bermingrud 1979, Anward 1982, Maling & Zaenen 1982, Faarlund 1992). The sentences in (4) provide examples of reportedly acceptable filler-gap dependencies into tensed adjunct clauses in MSc languages.

- (4) *Reportedly acceptable filler-gap dependencies into tensed adjunct clauses in MSc languages*
- a. Det blir han sint [når eg seier \_].  
*that becomes he angry when I say*  
'That he becomes angry when I say.'
- b. Den saka ventar vi her [mens de ordnar \_].  
*that case.DEF wait we here while they fix*  
'That case we wait here while they fix.'  
(Norwegian; Faarlund 1992:117)
- c. 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.'  
(Swedish; Anward 1982:74)

In (4a), the pronoun *det* 'that' appears to have been topicalized from the direct object position of the adjunct-internal verb *seier* 'say'. In (4b), the definite DP *den saka* 'that

case' has been topicalized from the object position of the adjunct-internal simple verb *ordnar* 'fix'. Similarly, in the Swedish example in (4c), the definite DP *sportspeglen* 'the sports program' appears to have been topicalized from the object position of the adjunct-internal verb *ser* 'see'.

Recent experimental evidence provides some support for the observations about MSc (e.g. Nyvad, Christensen & Vikner 2017; Kush et al. 2018, 2019; C. Müller 2019). In several studies, the acceptability of island extraction in MSc languages has been investigated by way of formal experiments. We focus on two studies (using the factorial design developed by Sprouse 2007; see Section 2.1.1 below for details) that investigated Norwegian: (i) Kush et al. (2018), which tested the acceptability of *wh*-extraction from five islands types: 'whether', complex NP, subject, (conditional) adjunct, and relative clause, and (ii) Kush et al. (2019), which tested the acceptability of contrastive topicalization from the same five island types.

Kush et al. (2018) found clear evidence of subject, adjunct, complex NP, and relative clause-island effects on *wh*-extraction with simple (e.g. *hva* 'what') and complex (e.g. *hvilken bok* 'which book') *wh*-phrases.<sup>3</sup> The authors failed to find reliable 'whether'-island effects, which reflected significant inter-individual variation in whether participants accepted *wh*-extraction from embedded polar questions. Notably, many participants did not exhibit any sensitivity to 'whether'-island violations at all. The authors reasoned that the absence of statistically reliable 'whether'-island effects and variability in the underlying distribution of judgments of 'whether'-island violations was inconsistent with the conclusion that embedded questions were syntactic islands in Norwegian.

Following up on these findings, Kush et al. (2019) investigated the island-sensitivity of contrastive topicalization. Many of the reported naturally-occurring examples of island violations in MSc involve topicalization. As a type of A'-movement, topicalization is expected to respect the same syntactic locality conditions as *wh*-movement under traditional syntactic accounts (see e.g. den Dikken & Lahne 2013; Phillips 2013a:68). However, topicalization is subject to different semantic and discourse-pragmatic factors. Thus, insofar as the island effects observed in Kush et al. (2018) reflect syntactic constraint violations, similar effects should obtain with topicalization. However, if any of the island effects observed for *wh*-extraction were semantic or discourse-pragmatic in origin, then a different pattern might be found for topicalization.

Kush et al. (2019) replicated large island effects for subjects and complex NPs, and once again failed to find a reliable 'whether'-island effect. Relevant for our purposes, the authors unexpectedly found no island effect for dependencies like (5) in their second experiment, where an object has been topicalized from a finite conditional adjunct clause introduced by the complementizer *om* 'if'.

- (5) Bakdøren blir han nervøs [om de lar stå ulåst].  
*back.door.DEF gets he nervous if they leave stand unlocked*  
 'The backdoor he gets nervous if they leave unlocked.'

Judgments of topicalizations from adjuncts were variable: participants rejected the dependencies on some trials, but accepted on others. On balance, participants were more likely to accept topicalizations from *om*-adjuncts than to reject them.<sup>4</sup> Tellingly, the probability of accepting topicalization from a conditional adjunct

was comparable to the probability of accepting long-distance topicalization from a non-island embedded declarative clause.

The findings suggest that conditional adjuncts are not categorical islands for A'-movement in Norwegian and that the type of dependency has a significant impact on acceptability of A'-dependencies into certain islands (see also Sprouse et al. 2016). However, given the potentially large theoretical consequences of revising our standard understanding of the islandhood of adjuncts, we should be sure that the such findings can be replicated with a larger sample. A further question concerns the generality of the findings. Kush et al. (2019) only investigated conditional adjunct clauses. Many syntactic accounts of extraction from adjuncts predict that adjuncts should behave as a coherent class with respect to their island status (Huang 1982; Lasnik & Saito 1992; Uriagereka 1999, 2012; Boeckx 2003, 2012; Stepanov 2007; G. Müller 2011; Hunter 2015). We therefore ask whether similar island-insensitivity would be observed with other finite adjuncts in Norwegian. It is also possible that island effects might vary by adjunct type (a possibility hinted at in Truswell 2007, 2011, and C. Müller 2019). Insofar as we observe variability in island-sensitivity across adjuncts, this variability might provide clues about a finer-grained set of features governing adjunct islandhood beyond the coarse cut made by conditions like the CED.

## 2. Experiments

To investigate these questions, we ran two acceptability judgment experiments testing the acceptability of topicalization dependencies into three different types of finite adjunct clauses, partly using the same material as in Kush et al. (2019).

### 2.1 Experimental design

#### 2.1.1 The factorial definition of island effects

We describe common design characteristics of our experiments before discussing the specifics of each experiment individually. Our experiments adopted the general factorial definition of islands, introduced by Sprouse (2007) and used in much recent work (Sprouse et al. 2011, Sprouse, Wagers & Phillips 2012, Sprouse et al. 2016). In a standard design, participants judge multi-clausal sentences with a filler-gap dependency. The two factors, *Distance* and *Structure*, determine the properties of the sentences. *Distance* determines whether the filler is linked to a gap in the matrix clause (*Short-distance*) or the embedded clause (*Long-distance*). *Structure* determines whether the embedded clause is a *non-Island* or (contains) an *Island*. *Island* is here used as a label for conditions that simply contain domains characterized as islands (both (6c) and (6d) in example (6) below). The factorial design crosses these factors, creating conditions that correspond to combinations of the factors' levels, as shown in Table 1.

The factorial design is illustrated with a test item that uses a 'whether'-island below. *Short-distance* is realized as the movement of the *wh*-word from subject position in the matrix clause in (6a) and (6c). *Long-distance* is realized as the movement of the *wh*-word from object position of verb in the embedded clause in (6b) and (6d). In *no-Island* sentences the embedded clause is a declarative complement clause. In *Island* sentences, the embedded clause is a 'whether'-clause in (6c) and (6d).

**Table 1.** A schematic of a 2 × 2 factorial design for testing for island effects.

		Structure	
		no-Island	Island
Distance	Short-distance	<i>Short-distance, no-Island</i> (6a)	<i>Short-distance, Island</i> (6c)
	Long-distance	<i>Long-distance, no-Island</i> (6b)	<i>Long-distance, Island</i> (6d)

- (6) ‘Whether’-island example from Sprouse (2007:56)
- Who \_ thinks [that you wrote the letter]?
  - What \_ do you think [that you wrote \_]?
  - Who \_ wonders [whether you wrote the letter]?
  - What \_ do you wonder [whether you wrote \_]?

The factorial design proceeds from the assumption that linear distance and structural complexity may have effects on sentence acceptability. For example, participants might like longer dependencies less than shorter dependencies or prefer simpler structures to more complex structures due to processing burden. Such effects are, however, orthogonal to the question of whether there is an island effect. The strength of the factorial design is that it allows for the main effects that distance and complexity might have on acceptability to be isolated, so that the independent island effect (if there is one) can be isolated. The factorial definition treats island effects as the super-additive interaction of the two independent factors (*Distance* and *Structure*), independent of the main effects.

Identifying the presence or absence of an island effect within the paradigm can be done visually by plotting the acceptability of each of the four conditions with an interaction plot. If there is no island effect, we expect that the unacceptability of the *Long-distance, Island* condition should be equal to the linear sum of the costs of *Distance* and *Structure*. Such a state of affairs would correspond to the plot in Figure 1A. If, on the other hand, there is an island effect, we expect the unacceptability of the *Long-distance, Island* condition to be greater than the sum of the linear costs of *Distance* and *Structure*, we expect a super-additive interaction like Figure 1B.

The size of the *Distance* × *Structure* interaction, and hence the island effect can be quantified using a *Differences-in-Differences* (DD)<sup>5</sup> score (Maxwell & Delaney 2003). This allows (mean) effect sizes to be compared across islands and experiments.

### 2.1.2 Materials<sup>6</sup>

Our experiments tested extraction from five different clause types: three adjunct clauses – *om* ‘if’, *når* ‘when’, and *fordi* ‘because’ – and two control islands – subject islands and ‘whether’-islands. The subject- and ‘whether’-island sub-experiments were included as baselines for comparison. Kush et al. (2018, 2019) found very large island effects for subject islands in Norwegian, making the subject island a good baseline for a large island effect. In comparison, they found small and unreliable effects for extraction from an embedded ‘whether’-question in Norwegian. Moreover, the authors identified the variability in judgments observed with

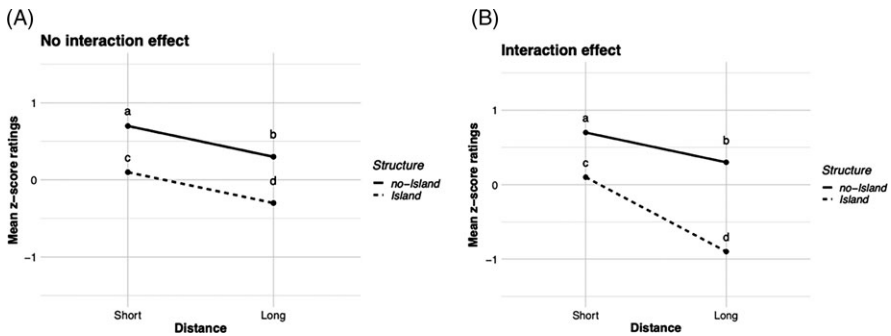


Figure 1. Example interaction plots illustrating the absence of a *Distance* × *Structure* island effect (A) or the presence of a *Distance* × *Structure* island effect (B).

extraction from embedded ‘whether’-questions as characteristic of ‘extra-syntactic’ effects on acceptability. Thus, other island effects that exhibit similar variability might be argued to be similarly ‘extra-syntactic’ in nature.

Since we were interested in testing whether Kush et al.’s (2019) results can be replicated, we used the design for their test items for all our items. Each test item contained four test sentences that were different realizations of *Distance* × *Structure*. Each test sentence was preceded by a preamble that facilitated topicalization in the test sentence. Context was included because Kush and colleagues found that participants rejected indisputably grammatical contrastive topicalization dependencies presented *in vacuo* without supporting context at surprisingly high rates. The context sentence introduced felicitous context for topicalization. Below are example items for all the islands tested. The example items for *om* ‘conditional if’, ‘whether’- and subject islands are from Kush et al. (2019), while the items for *når* ‘(temporal) when’ and *fordi* ‘causal because’ adjunct clauses were created for the current study.

(7) *Conditional om* ‘if’ item set (item number exp1: 38; exp2a and 2b: 54)

Preamble:

Moren var glad for at brudeparet husket  
*mother.DEF was glad for that bridal.couple.DEF remembered*  
 å sende ut invitasjoner i tide, ...  
*to send out invitations in time*

‘The mother was happy that the bride and the groom remembered to send out the invitations in time, ...’

a. men hun forventer at de kommer til å glemme å sende ut  
*but she expects that they come to to forget to send out*  
 takkekortene med en gang.  
*thank.you.cards.DEF with one time*

‘but she expects that they will forget to send out the thank you cards right away.’

b. men takkekortene forventer hun at de kommer  
*but thank.you.cards.DEF expects she that they come*  
 til å glemme å sende ut med en gang.  
*to to forget to send out with one time*

‘but the thank you cards expects she that they will forget to send out right away.’

- c. men hun blir skuffet om de glemmer å sende  
*but she becomes disappointed if they forget to send*  
 ut takkekortene med en gang.  
*out thank.you.cards.DEF with one time*  
 'but she will be disappointed if they forget to send out the thank you cards right away.'
- d. men takkekortene blir hun skuffet om de glemmer  
*but thank.you.cards.DEF becomes she disappointed if they forget*  
 å sende ut med en gang.  
*to send out with one time*  
 'but the thank you cards she will be disappointed if they forget to send out right away.'

## (8) Temporal når 'when' item set (item number exp1: 9; exp2b: 25)

Preamble:

John føler seg i god form når han drikker øl, ...  
*John feels himself in good shape when he drinks beer*  
 'John feels fine when he is drinking beer, ...'

- a. men han blir ofte dårlig av å drikke whisky.  
*but he becomes often sick of to drink whisky*  
 'but he often feels sick from drinking whisky.'
- b. men whisky blir han ofte dårlig av å drikke.  
*but whisky becomes he often sick of to drink*  
 'but whisky he often feels sick from drinking.'
- c. men han blir dårlig når han drikker whisky.  
*but he becomes sick when he drinks whisky*  
 'but he feels sick when he drinks whisky.'
- d. men whisky blir han dårlig når han drikker.  
*but whisky becomes he sick when he drinks*  
 'but whisky he feels sick when he drinks.'

## (9) Causal fordi 'because' item set (item number exp1: 2; exp2a: 2)

Preamble:

Mette er ikke fornøyd med sommertemperaturene i Nord Norge, ...  
*Mette is not satisfied with summer.temperatures.DEF in North Norway*  
 'Mette is not happy with the summer temperatures in Northern Norway, ...'

- a. men hun sier at hun liker vintertemperaturene.  
*but she says that she likes winter.temperatures.DEF*  
 'but she says that she likes the winter temperatures.'
- b. men vintertemperaturene sier hun at hun liker.  
*but winter.temperatures.DEF says she that she likes*  
 'but the winter temperatures she says that she likes.'
- c. men hun blir boende fordi hun liker vintertemperaturene.  
*but she becomes living because she likes winter.temperatures.DEF*  
 'but she stays there because she likes the winter temperatures.'
- d. men vintertemperaturene blir hun boende fordi hun liker.  
*but winter.temperatures.DEF becomes she living because she likes*  
 'but the winter temperatures she stays there because she likes.'

(10) *Subject-island item set*

Preamble:

Vitenskapsmannen tror ikke at den gamle behandlingen er god, ...  
*scientist.DEF think not that the old.DEF treatment.DEF is good*

'The scientist does not think that the old treatment is good, ...'

- a. men han synes den nye behandlingen fortjener Nobelprisen.  
*but he thinks the new.DEF treatment.DEF deserves Nobel.Prize.DEF*  
 'but he thinks the new treatment deserves the Nobel Prize.'
- b. men den nye behandlingen synes han fortjener Nobelprisen.  
*but the new.DEF treatment.DEF thinks he deserves Nobel.Prize.DEF*  
 'but the new treatment thinks he deserves the Nobel Prize.'
- c. men han synes den nye behandlingen mot kreft fortjener Nobelprisen.  
*but he thinks the new.DEF treatment.DEF against cancer deserves Nobel.Prize.DEF*  
 'but he thinks the new treatment against cancer deserves the Nobel Prize.'
- d. men kreft synes han den nye behandlingen mot fortjener Nobelprisen.  
*but cancer thinks he the new.DEF treatment.DEF against deserves Nobel.Prize.DEF*  
 'but cancer he thinks the new treatment against deserves the Nobel Prize.'

(11) *'Whether'-island item set*

Preamble:

Servitøren antok at Christina ville nekte å drikke Farris, ...  
*waiter.DEF assumed that Christina would refuse to drink Farris*

'The waiter assumed that Christina would refuse to drink Farris, ...'

- a. men han trodde at hun ville drikke Bris stedet.  
*but he thought that she would drink Bris instead*  
 'but he thought that she would drink Bris instead.'
- b. men Bris trodde han at hun ville drikke stedet.  
*but Bris thought he that she would drink instead*  
 'but Bris he thought that she would drink instead.'
- c. men han lurte på om hun ville drikke Bris stedet.  
*but he wondered on if she would drink Bris instead*  
 'but he wondered whether she would drink Bris instead.'
- d. men Bris lurte han på om hun ville drikke stedet.  
*but Bris wondered he on if she would drink instead*  
 'but Bris he wondered whether she would drink instead.'

2.1.3 *Procedure and analysis*

Test items were distributed online on IbexFarm (Drummond 2012). Participants were instructed to rate the test sentences between 1 and 7, with 1 given as *dårlig* 'bad' and 7 as *god* 'good' and to imagine that the sentences were uttered in a conversation. All test items contained a context sentence in italics followed by the test



sentence. Participants were instructed to base their ratings on the acceptability of the second sentence.

Before analysis, participant ratings were  $z$ -score transformed by participant to control for scale bias (e.g. Sprouse et al. 2016).<sup>7</sup> Analysis was conducted using linear mixed effects models using the *lme4* (Bates et al. 2015) and *lmerTest* (Kuznetsova, Brockhoff & Christensen 2017) packages in R (R Core Team 2019). Separate models for each island type with *Distance*, *Structure* and their interaction (*Distance*  $\times$  *Structure*) as the fixed effects were constructed with simple difference coding. The model included random intercepts for subject and items as well as by-subject random slopes for the fixed effects and their interaction. In the few cases when a model did not converge, the random effects structure was simplified. The Satterthwaite approximation was used to calculate  $p$ -values in the *lmerTest* package. We only report the size of the *Distance*  $\times$  *Structure* interaction effect, as main effects are orthogonal to our questions of interest. All plots were constructed with *ggplot2* (Wickham 2016).

## 2.2 Experiment 1

### 2.2.1 Participants

One hundred and five self-reported native Norwegian-speaking volunteers took part in Experiment 1 (66 females, mean age = 43.5 years). Participants were recruited via announcements on social media sites. Four participants were excluded for reporting a different native language than Norwegian. All speakers self-identified as native speakers of Norwegian.

### 2.2.2 Materials

Eight item sets were constructed for each of the five island types. The test sentences were distributed across four lists in a Latin-Square fashion, such that each participant encountered 40 test sentences – two items per condition per island. The 40 test sentences were pseudo-randomly mixed with 46 fillers, 15 acceptable fillers and 31 unacceptable. Only 10 of the 40 encountered test sentences were unacceptable sentences (i.e. sentences testing the *Long-distance*, *Island* condition). In order to balance the experiment between unacceptable and acceptable test sentences, we included 31 unacceptable fillers. In effect, participants encountered 86 test sentences, out of which, 45 could be considered acceptable and 41 unacceptable. The order of the test items differed for each participant.

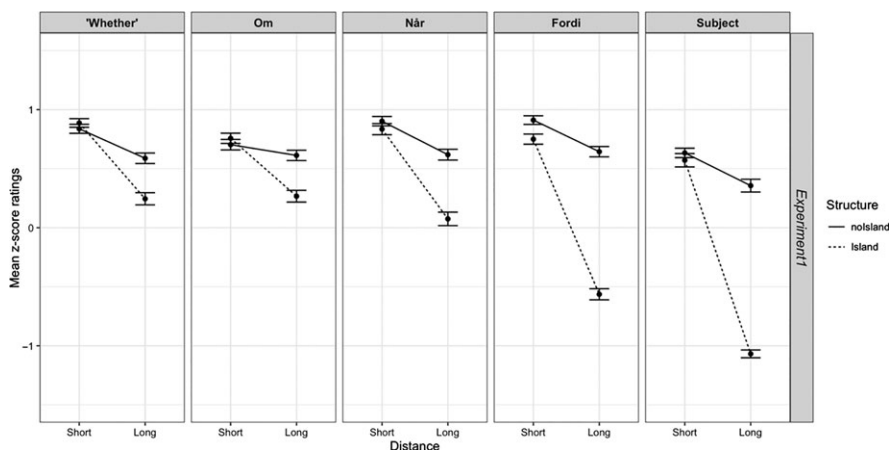
### 2.2.3 Results

The unacceptable fillers received a mean score of  $z = -0.84$ , whereas the good fillers received a mean of  $z = 0.63$ . Interaction plots displaying the average rating by condition and island type are presented in Figure 2. Table 2 provides a statistical summary of the *Distance*  $\times$  *Structure* interaction effects for each island. As can be seen, superadditive interaction effects were observed for all islands tested ( $p < .001$ ).

The size of the interaction effects varies by island: subject-island effects were large ( $DD = 1.375$ ), while ‘whether’-island effects were considerably smaller ( $DD = 0.375$ ). This replicates previous findings for these island types (Kush et al. 2018, 2019). The

**Table 2.** Statistical summary of the *Structure* × *Distance* interaction effects for each island type in Experiment 1.

Island type	Experiment 1		
	<i>p</i> -value	<i>t</i> -value	DD-score
'Whether'	< .001	−4.211	0.375
<i>Om</i> 'if'	< .001	−4.358	0.397
<i>Når</i> 'when'	< .001	−5.036	0.485
<i>Fordi</i> 'because'	< .001	−11.803	1.032
Subject	< .001	−15.017	1.375



**Figure 2.** Interaction plots for Experiment 1. Error bars indicate standard error.

adjunct island effect sizes also vary: the *om*-, *når*- and *fordi*-islands have DD scores of 0.397, 0.485, and 1.032, respectively.

What is also evident from Figure 2 is that the mean acceptability of the island-violating sentence differs for each adjunct. On average, participants rated extraction from *om*-adjuncts around  $z = 0.25$  (similar to their judgments for 'whether'-islands) and from *når*-adjuncts around  $z = 0$ , but extraction from *fordi*-adjuncts was rated much lower: closer to  $z = -0.75$ . Kush et al. (2018, 2019) showed that average acceptability scores that fall in the acceptable or intermediate range can conceal rather variable judgments of island-violations. To investigate the judgment pattern underlying the mean scores, we inspected the distribution of ratings by condition.

Distributions in Figure 3 show the density of ratings for each  $z$ -score by island type and by condition. If a sentence is always rated as acceptable we should see a unimodal distribution around +1, which we can see for the *Short-distance, no-Island* conditions. The distributions for the *Short-distance, Island* conditions are also

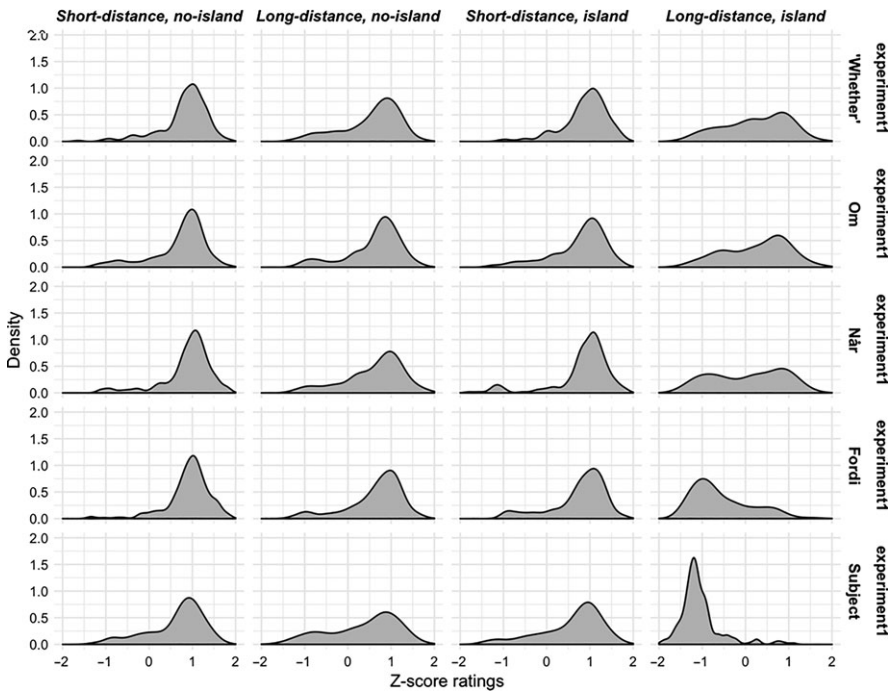
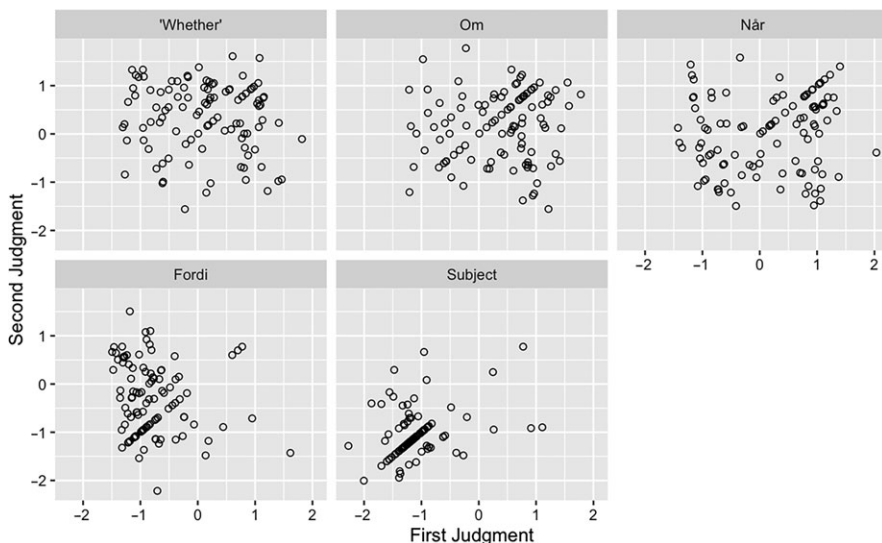


Figure 3. Distribution of z-scores for each island type tested and for each condition.

unimodally distributed around +1. The distributions for the *Long-distance, no-Island* conditions provide a point of comparison for how ratings of acceptable long-distance topicalization pattern. Here we see a mode at or close to +1, but also a longer leftward tail. This indicates that the items in this condition are not always accepted unequivocally and are perhaps rejected at a slightly higher rate than the short conditions.

Turning to the distributions for the *Long-distance, Island* condition, we see great differences between island types. The two control-island types show, as expected, very different behavior: judgments of the subject island are narrowly and unimodally distributed around  $z = -1.5$ . This means that topicalization from a complex subject is always rejected. Judgments of topicalization from embedded ‘whether’-clauses largely fall, as in Kush et al. (2019), above  $z = 0$ . The distribution for ‘whether’ exhibits a longer, fatter left tail than seen in the corresponding *Short-distance, Island* condition. This left tail indicates that participants judged topicalization from a ‘whether’-embedded question as either less acceptable or wholly unacceptable on a subset of trials.

The distribution of *Long-distance, Island* ratings differed considerably across all three adjuncts.<sup>8</sup> Ratings of topicalization from a conditional adjunct, show a distribution similar to the ‘whether’-clauses, again consistent with Kush et al. (2019). The distribution is roughly bimodal: the majority of judgments cluster around  $z = 1$ , but there is a smaller group of judgments that cluster around  $z = -1$ . This entails that



**Figure 4.** Each participant's judgments split by island type in Experiment 1. Each dot represents one participant, with their first judgment ( $x$ -axis) plotted against their second judgment ( $y$ -axis) on the *Long-distance, Island* condition.

extraction from this adjunct is more often accepted than it receives intermediate or poor ratings. The *fordi*-island exhibits unimodal distribution on the *Long-distance, Island* condition, however, unlike *om*, the distribution patterns well below 0 around  $z = -0.75$ . *Fordi*-extractions pattern more like the subject island, indicating relatively consistent rejection, though there does appear to be a small number of trials where topicalization was accepted. For the temporal *når*-island, we see clear bimodality. Bimodal distributions entail EITHER-OR-JUDGMENT, sometimes the condition is accepted, sometimes it is rejected, but it is less often given an intermediate rating. Accordingly, the *når*-adjunct does not pattern like any of the other conditions, with clustering around  $z = -1$  and  $z = 1$ .

Figure 3 above shows that there is variability in judgments, but does not allow us to distinguish between different origins of variability. Does the variability reflect inter-subject, inter-item differences, or both? We first investigate inter-subject differences using a visualization method from Kush et al. (2018, 2019); see also Kush & Dahl (published online on 15 September 2020). Figure 4 provides scatterplots of each participant's first and second judgment for each island type on the *Long-distance, Island* condition. When dots cluster in the bottom left quadrant, participants are consistently rejecting the island violating condition. Dots that lie in the top right quadrant indicate that participants are consistently accepting this condition. Dots that fall in the lower right or upper left quadrant correspond to INCONSISTENT RATERS, who accepted on one trial and rejected on another.

Almost all participants consistently rejected subject island violations, as evidenced by the preponderance of dots in the lower left quadrant for subject islands in Figure 4. Many participants consistently accepted 'whether'-island

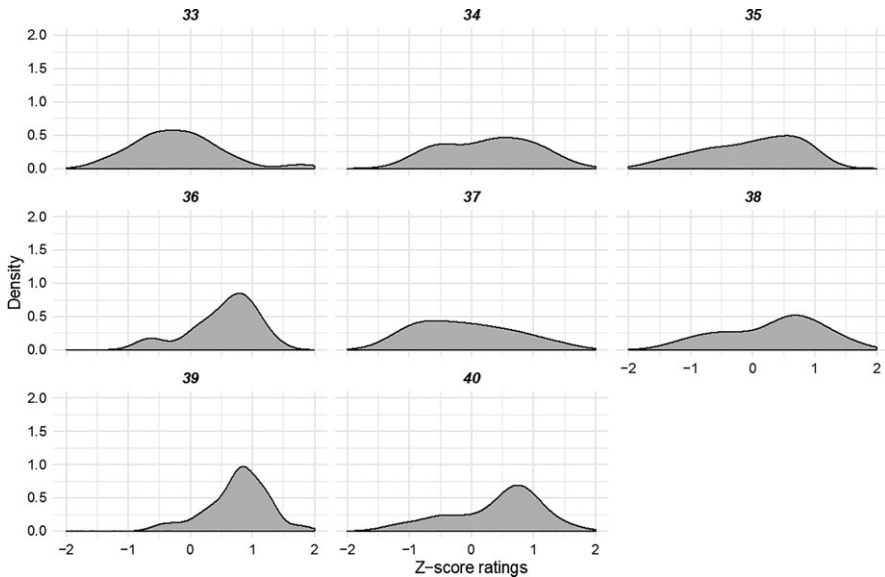


Figure 5. Distribution of z-scores for the *Long-distance, Island* condition for *om*-items tested. Item numbers are provided for cross-reference in the materials list.

violations, though there were also many inconsistent raters. For *om*-adjunct violations, a substantial portion of participants were consistent accepters, judging both trials above  $z = 0$ , as seen by the large number of dots in the upper right quadrant in Figure 4 (in line with the findings of Kush et al. 2019). A few participants consistently rejected topicalization from *om*, but most of the participants judged inconsistently: appearing to accept one trial and reject another.

Greater inter-participant variability is found with judgments of topicalization from *når*. A number of participants appear to consistently accept topicalization from *når*, somewhat similar to *om*, but there are more participants who consistently rejected *når* test sentences compared to *om*. This matches the bimodal distribution found for *når* in Figure 3. There are also a number of inconsistent raters. The majority of the *fordi*-adjunct ratings lie in the bottom left quadrant, indicating generally consistent rejection. Three participants appear to have consistently accepted the sentences, and a few more participants exhibited inconsistency.

We also inspected inter-item variability, by comparing distributions of judgments for different items separately by island type.

The plots in Figures 5–7 reveal that there is also variation between items within each adjunct type. For *om*, most items have ratings centered around  $z = 0.75$ . Three items show a clear single mode close to  $z = 1$  (36, 39, 40), and three others show a bimodal or left-skewed distribution slightly favoring positive scores (34, 35, 36). Only one item (33) appears to have consistently received a negative z-score. For *når*-items, judgments were either clustered around  $z = 1$  (items 10, 12), or exhibited bimodal distributions. Only one item seems to have received mostly negative

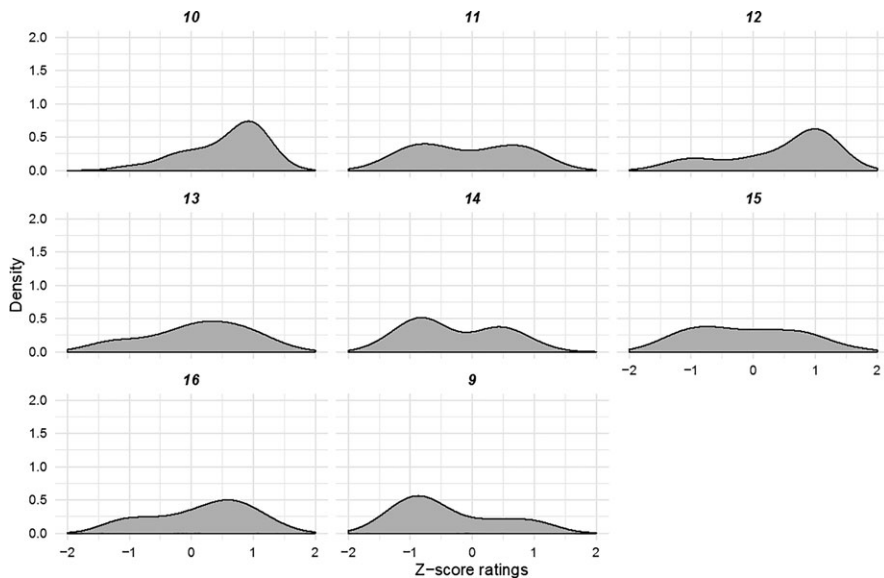


Figure 6. Distribution of z-scores for the *Long-distance, Island* condition for *når*-items tested. Item numbers are provided for cross-reference in the materials list.

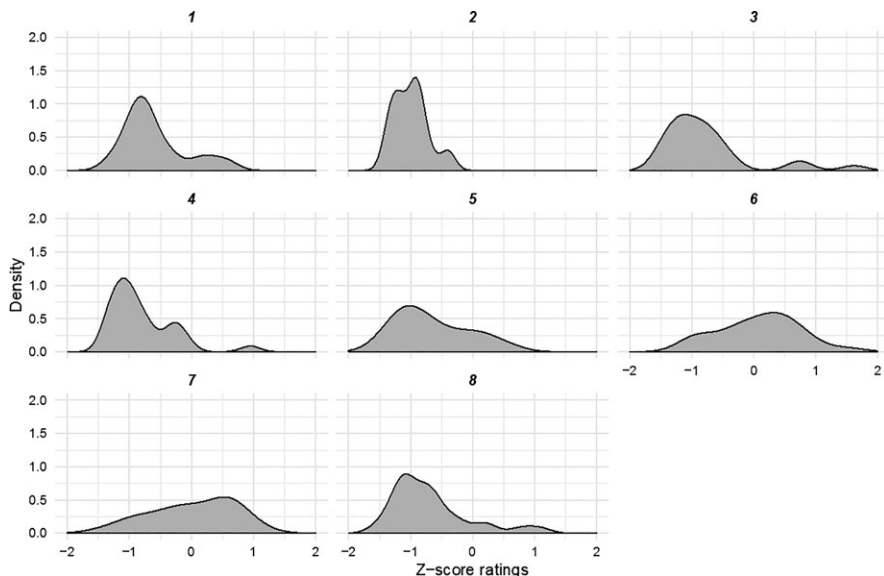


Figure 7. Distribution of z-scores for the *Long-distance, Island* condition for *fordi*-items tested. Item numbers are provided for cross-reference in the materials list.

*z*-scores. In contrast to *om*, six of eight *fordi*-items show relatively consistent ratings centered around  $z = -1$ . Two items (6, 7) have ratings centered around  $z = 0.5$ .

In order to determine whether there were any features that reliably contribute to acceptable topicalization or correlate with it, we coded each item for a number of surface features, which have been proposed to affect acceptability of extraction (e.g. Truswell 2011, Dal Farra 2020): tense in the matrix and embedded clauses, agentivity of the matrix and embedded predicates, aspectual class of the matrix clause, telicity of the matrix VP, spatiotemporal overlap between matrix and embedded clause, direct causation between matrix and embedded clause and type of matrix verb. We also checked the definiteness of the moved constituent (Szabolcsi & Lohndal 2017) and, the number of words between the filler and the gap (i.e. processing difficulty, Hofmeister, Casanto & Sag 2013). We then compared ratings of the *Long-distance, Island* condition by items grouped across shared features through visual inspection of plotted ratings to investigate whether any of the *om*-, *nâr*-, or *fordi*-items that were disproportionately accepted shared any features with one another to the exclusion of the items that were rejected. We could not find any surface features that could explain the variation between items for any of the islands.

#### 2.2.4 Discussion

The experiment roughly replicates Kush et al.'s (2019) findings for extraction from subject, 'whether'- and *om*-clauses. Subject island effects were large, while island effects for 'whether'-clauses and conditional *om*-adjuncts were considerably smaller. Though there were small differences in the significance of the interaction effect, these can be attributed to a lower sample size in Kush et al.'s (2019) experiment compared to this experiment, 36 versus 105, respectively. We also found that average judgments of topicalization from 'whether'- and conditional *om*-islands fell in the range of 'acceptable' sentences ( $z > 0$ ) and were roughly comparable to long-distance extractions from non-islands. Moreover, judgments of topicalization from both 'whether'-clauses and *om*-adjuncts were highly variable, just as Kush et al. (2019) found.

Next, we turn to the two new adjunct types we investigated. The island effect size of extraction from *nâr*-adjuncts ( $DD = 0.485$ ) was smaller than for subject islands ( $DD = 1.375$ ), but larger than for 'whether'-islands ( $DD = 0.375$ ). Judgments of topicalization from *nâr*-adjuncts were bimodally distributed, indicating significant variation. Bimodality can partly be explained as inter-participant variation: we see some consistent accepters, some consistent rejecters and some inconsistent participants.<sup>9</sup> The bimodal distribution of *z*-scores for the *Long-distance, Island* condition is also partly due to variation between items.

Contrary to the pattern found for *nâr*, we found a large *fordi*-island effect similar in size to subject islands. Topicalization from a *fordi*-adjunct was almost always rejected. However, the judgments for *fordi* are nevertheless more variable than the subject-island judgments. *Fordi*-island sentences were less often categorically rejected than subject-island sentences. Still, *fordi* is much less accepted than *nâr*.

The variation seen within each adjunct type, as well as between the different adjuncts, is surprising. We could not find any surface features that could



straightforwardly explain the variation between items or the variation between island types. We observed a large number of inconsistent participants, as in Kush et al.'s (2019) study, and some participants who were consistent rejectors. Inter- and intra-participant inconsistency could be explained in a number of ways. For example, observed differences could reflect meaningful differences at the population level, or could be attributed to noise. With the current design, it is difficult to tease apart various hypotheses due to lack of power at the individual participant level, given that each participant has only encountered two *Long-distance, Island* items per island type. To better understand the source of inconsistent ratings we ran an experiment with more observations per participant.

### 2.3 Experiments 2a and 2b

To better investigate the variation seen in Experiment 1, Experiments 2a and 2b were conducted. We increased the number of observations per participant per condition in the *om-*, *når-* and *fordi-*islands to five per participant (20 items in total). We also increased the number of subject islands to four per participant (16 items in total). To avoid participant fatigue, island types were distributed into two different experiments: Experiment 2a included items of *om-*, *fordi-*, as well as the control islands; 'whether'- and subject islands. Experiment 2b included items of *om-*, *når-*, and the same control items as in Experiment 2a.

#### 2.3.1 Participants

In Experiment 2a there were 28 participants (20 female, mean age = 25 years), three participants were excluded for having reported a different native language than Norwegian. In Experiment 2b there were 37 participants (27 female, mean age = 26 years); one participant was excluded for reporting a different native language than Norwegian. All speakers were self-identified native speakers of Norwegian. Participants were recruited through various social media sites or through virtual learning environments for various courses. We were careful to distribute the link for Experiment 2a and the link for Experiment 2b to different channels. In the instructions, we also added that participants who knew that they participated in Experiment 1 should not participate in Experiment 2a or 2b.

#### 2.3.2 Materials

In Experiment 2a, participants saw 64 test sentences across all four test conditions – 5 *om*-adjunct items, 5 *fordi*-adjunct items, 4 subject island items, 2 'whether'-adjunct items. In Experiment 2b, participants saw 64 test sentences across all four test conditions – 5 *om*-adjunct items, 5 *når*-adjunct items, 4 subject island items, 2 'whether'-adjunct items. Test items in Experiments 2a and 2b were pseudo-randomly intermixed among 40 unacceptable fillers, out of which 31 were the same as in Experiment 1.<sup>10</sup> In addition we added four acceptable fillers featuring local topicalization to have a rough baseline of acceptability for topicalization across a single clause.



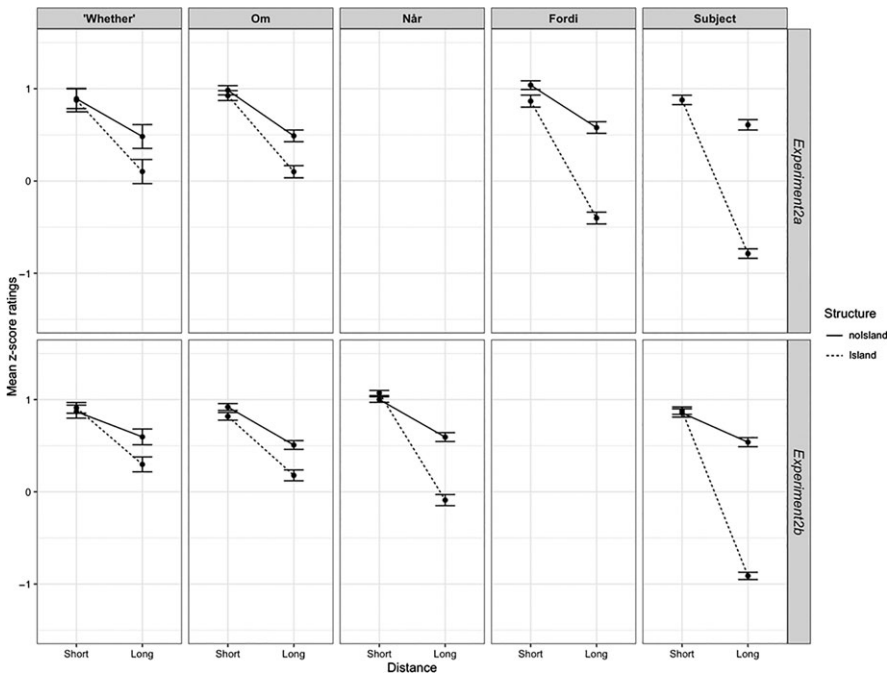


Figure 8. Interaction plots for Experiment 2a and 2b. Error bars indicate standard error.

### 2.3.3 Results

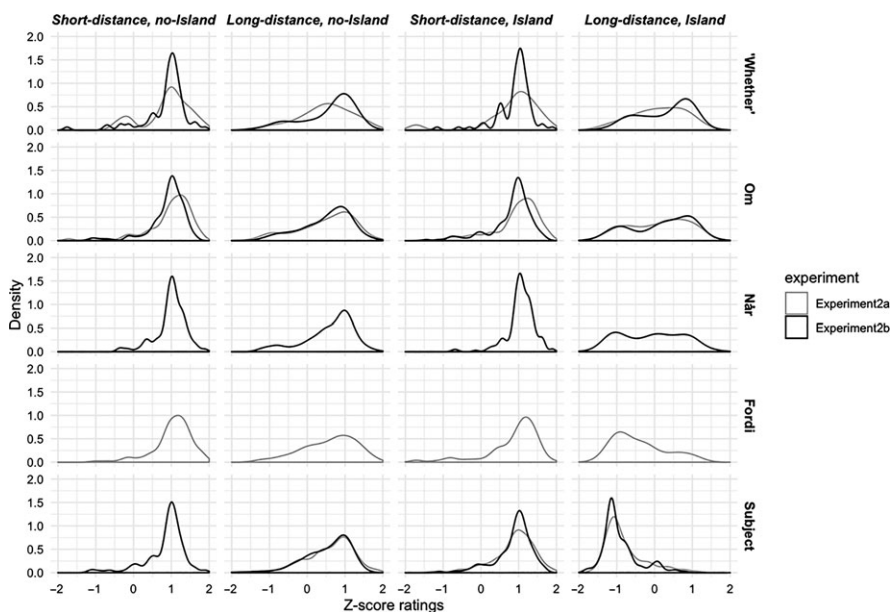
In Experiment 2a, unacceptable fillers received a mean score of  $z = -0.79$  and the local topicalization fillers a mean of  $z = 0.00$ . The average ratings of fillers in Experiment 2b were similar: unacceptable fillers  $z = -0.84$ ; acceptable local topicalization  $z = -0.05$ . Interaction plots displaying the average rating by condition and island type are presented in Figure 8. Table 3 provides a statistical summary of the interaction effects for each island. The findings in Experiments 2a and 2b are similar to the findings in Experiment 1. Significant super-additive interaction effects were found for all clause types tested. The effect sizes (DD) are also comparable to Experiment 1.

As in Experiment 1, judgments and effect sizes differ across adjunct types. Similarly, distributions of  $z$ -scores in each condition and island for Experiments 2a and 2b are comparable to what was observed in Experiment 1. This can be seen in Figure 9. Judgments of *om*- and *når*-island violations both exhibit bimodality, with a greater proportion of acceptances of extraction from *om*- than *når*-clauses. Judgments of *fordi*-adjunct violations cluster unimodally around  $z = -1$ , seemingly showing agreement across participants.

Once again, we inspected the results for inter-subject variation. Figures 10 and 11 provide overviews of individual participant ratings on the *Long-distance, no-Island condition* in each adjunct island sub-experiment. Each column represents an individual participant. The box reports the median (black line inside the box) and the range within which 50% of the ratings lie. The top and bottom ‘whiskers’ (thin lines)

**Table 3.** Statistical summary of the *Distance* × *Structure* interaction effect for each island type for each experiment.

Island type	Experiment 2a			Experiment 2b		
	<i>p</i> -value	<i>t</i> -value	DD-score	<i>p</i> -value	<i>t</i> -value	DD-score
'Whether'	.007	−2.803	0.534	.027	−2.295	0.355
<i>Om</i> 'if'	.008	−0.728	0.310	.029	−2.220	0.214
<i>Når</i> 'when'	—	—	—	< .001	−6.792	0.746
<i>Fordi</i> 'because'	< .001	−6.455	0.857	—	—	—
Subject	< .001	−14.119	1.337	< .001	−15.093	1.464

**Figure 9.** Distribution of z-scores for each condition in adjunct island comparisons in Experiments 2a and 2b.

report the range within which 25% of the lowest and highest ratings lie. Finally, dots represent outliers. Great variance between a participant's ratings on the same condition can be seen in the plots as a long box and long whiskers.

Participants' judgments of extraction from *om*-adjuncts vary in both Experiments 2a and 2b. Nearly all participants exhibit a degree of inconsistency, but 30/37 participants in Experiment 2b exhibit a median rating above  $z = 0$ . Since we see similar variation across experiments, it is likely that some of the variability of judgments for *om*-adjuncts is not caused by BETWEEN-PARTICIPANT variation. Instead, some of the variability must be attributed to BETWEEN-ITEM or WITHIN-PARTICIPANT variation. Figure 11 reveals that participants were not consistent in their judgments of *når*-adjunct island violations, though some speakers show

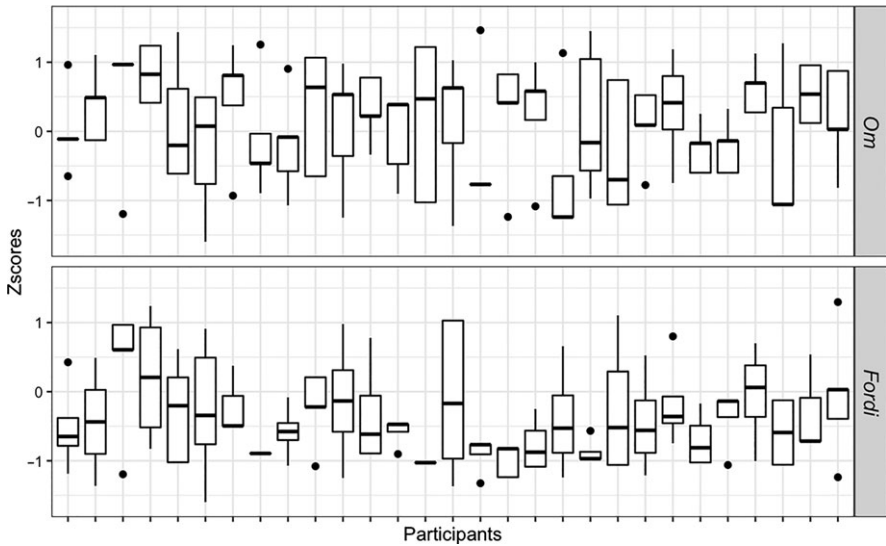


Figure 10. Overview of participant ratings of *om*- and *fordi*-adjunct items in Experiment 2a on the *Long-distance, no-Island* condition.

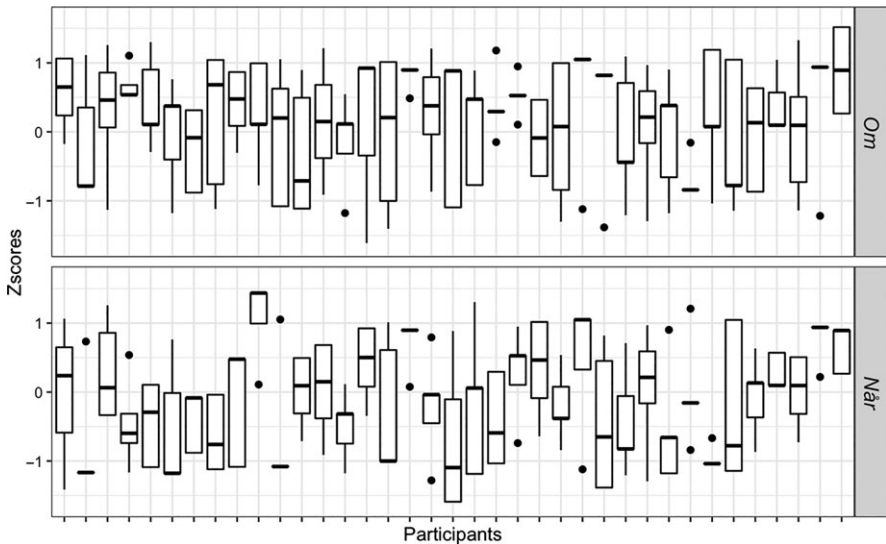
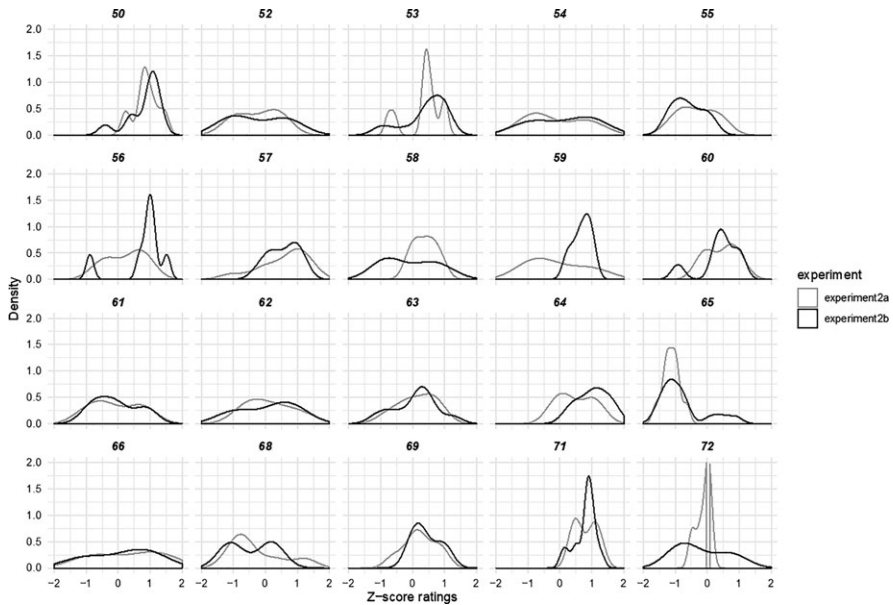


Figure 11. Overview of participant ratings of *om*- and *når*-adjunct items in Experiment 2b on the *Long-distance, no-Island* condition.

greater consistency than others. Here, 17/37 participants had median ratings above  $z = 0$ . As in Experiment 1, most participants (27/28) consistently rejected topicalization from *fordi*-adjuncts showing median ratings below  $z = 0$ , however, there were a few consistent accepters and inconsistent raters.



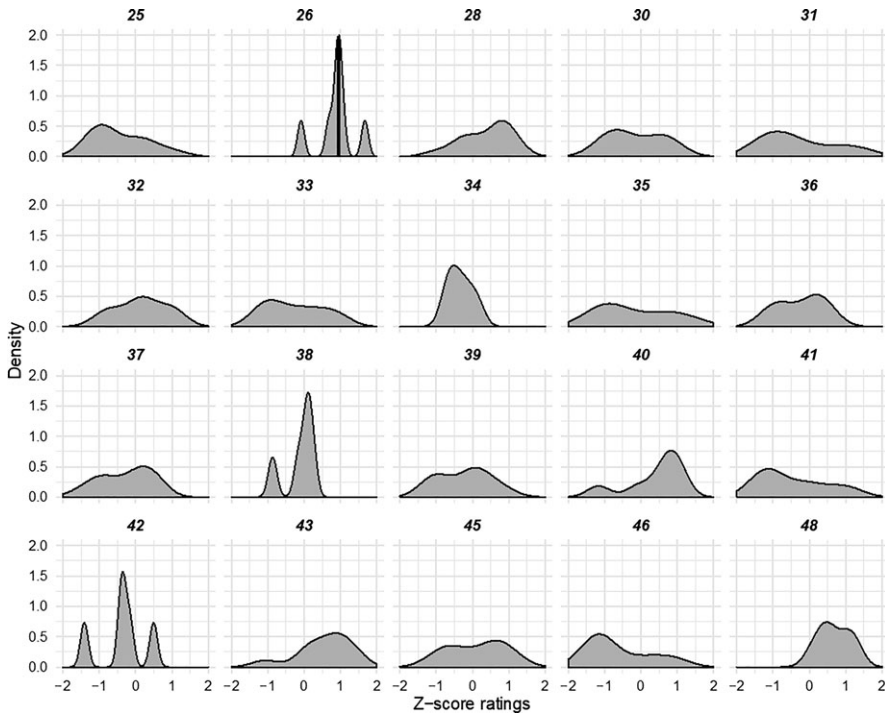
**Figure 12.** Distribution of *z*-scores in the *Long-distance, Island* condition for *om*-items tested in Experiments 2a and 2b. Item numbers are provided for cross-reference in the materials list.

To further address the source of the variation, we also examined the distribution of *z*-scores on the *Long-distance, Island* condition for each item of the adjunct clause types in Experiments 2a and 2b.

The distributions across adjunct types are similar to distributions across adjunct types in Experiment 1. As in Experiment 1, we also see significant variation between items within each adjunct type. Interestingly, for the items that were tested in Experiment 1 and Experiment 2a and/or 2b, we see similar variation across experiments, suggesting that the differences between items in Experiment 1 were not due to just random noise.

For *om*-adjuncts (see Figure 12 above), nine items in Experiment 2a and 10 in Experiment 2b show a mostly unimodal distribution around a positive *z*-score. Eight items in each of the two experiments have bimodal ratings or highly variable ratings across the full range. Only two items in Experiment 2a and two in 2b show a unimodal distribution around  $z = -1$ . Examining *om*-items based on the same surface features as in Experiment 1 (see results section in Section 2.2.3 for the list of features), we did not find any similarities across items.

The *når*-adjuncts (see Figure 13 above) show a large degree of variation between items: four items show a unimodal, narrow distribution around  $z = 0.5$ – $0.75$  and five items have a bimodal distribution. Many of the items with a bimodal distribution have a larger mode below  $z = 0$ , in contrast to *om*-adjuncts. Again, we could not find any shared features between items that show similar behavior.



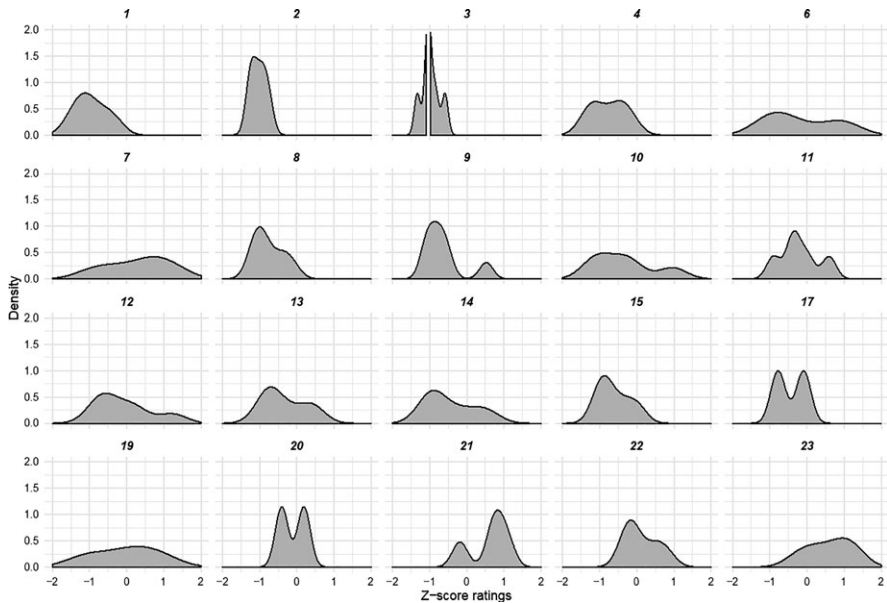
**Figure 13.** Distribution of z-scores in the *Long-distance, Island* condition for *nâr*-items tested in Experiment 2b. Item numbers are provided for cross-reference in the materials list.

Finally, the majority of the *fordi*-items (11 out of 20; see Figure 12 above) show a quite narrow unimodal distribution of z-scores centering around  $z = -0.75$ . Seven items received inconsistent ratings. Two *fordi*-items show ratings clustering around a positive z-score resembling the distributions of some *om*-items. These items do not share any surface features or feature combinations that accepted items do not have.

### 2.3.4 Discussion

Experiments 2a and 2b roughly replicated the findings from Experiment 1 and Kush et al. (2019). Island effects for topicalization from conditional *om*-adjuncts were comparable in size to ‘whether’-island effects, as were the average absolute judgments of such island violations. Intermediate judgments of *om*- and ‘whether’-island violations reflected highly variable underlying judgment distributions, in which a large number of trials represent ‘acceptable’ judgments.

As in Experiment 1, island effects were slightly larger for topicalization from *nâr*-adjuncts than *om*-adjuncts, but judgments of topicalization from *nâr*-adjuncts were bimodally distributed. Thus, the slightly larger island effects reflect a higher probability of rejecting topicalization from *nâr*-adjuncts than *om*-adjuncts. The island effects do not, however, appear to indicate that topicalization is always unacceptable



**Figure 14.** Distribution of z-scores in the *Long-distance, Island* condition for *fordi*-items tested in Experiment 2a. Item numbers are provided for cross-reference in the materials list.

from *når*-adjuncts (as it appears to be from subject phrases). For *fordi*-adjuncts, the same distribution in Experiment 1 was also seen in Experiment 2a. Topicalization from *fordi*-adjuncts was mostly rejected across trials, though there was a small subset of trials where such dependencies were accepted.

The fact that we observed a similar degree of variation as in Experiment 1 indicates that inconsistent judgments at an individual participant-level should not be attributed to noise. Further, the differences between the types of adjuncts were replicated across more items, indicating reliable differences between adjunct types.

### 3. Discussion

We investigated the acceptability of (contrastive) topicalization from three types of finite adjunct clauses *om* ‘if’, *når* ‘when’ and *fordi* ‘because’, in Norwegian. Our goal was to replicate Kush et al.’s (2019) findings of the absence of island effects with *om*-adjuncts and to determine whether the absence of island effects extended to other adjuncts in Norwegian. We compared the ratings of adjunct island violations to similar topicalizations from subject islands and ‘whether’-islands, as ‘anchor points’ for interpretation.

The most significant finding is the great amount of cross-trial variability in ratings both between and within adjunct types. Such variability is unexpected under most accounts of adjunct islands and has not previously been observed in formal investigations of adjunct islands. As we discuss below, this finding is at odds with

established accounts of adjunct islands, which predict relatively uniform unacceptability across sentences containing the same ‘island violation’.

Before going into the variation in more detail, we point out that across the variable ratings all three adjunct clauses show super-additive interaction effects. Following the factorial definition of an island effect, all three adjunct clauses can be defined as ISLANDS for the formation of filler–gap dependencies. This entails that SOMETHING causes filler–gap dependencies into these adjuncts to be judged less acceptable than might be expected based on simple considerations of distance and structural complexity alone. The mere presence of island effects alone does not tell us what the underlying cause of those effects is.

Our study shows that the TYPE of adjunct clause impacts the acceptability of extraction to a large extent. We observed considerable variation between adjunct clauses in (i) the size of the island effect; (ii) the mean *z*-score rating of the *Long-distance, Island* condition; and (iii) the distribution of *z*-scores on the *Long-distance, Island* condition. Similarly to Kush et al. (2019), we found that contrastive topicalization from *om*-adjuncts resulted in relatively small island effects (in comparison to subject-island effects, but similar to ‘whether’-island effects), mean judgments of island violations fell in the range of acceptability (e.g.  $z > 0$ ), and that judgments of such topicalizations exhibited a bimodal distribution, though the majority of judgments fell above  $z = 0$ . Topicalization from *når*-adjuncts also resulted in smaller island effects, higher average acceptability scores, and a bimodal rating distribution. *Fordi*-islands differed in that effect sizes were reliably larger and test sentences were almost consistently rejected.

Kush et al. (2018, 2019) argued that judgment distributions could inform the theoretical interpretation of different island effects and, in particular, where to apportion responsibility for island effects. The authors argued that a high degree of variability in judgments was inconsistent with the conclusion that A'-movement was (syntactically) prohibited from that domain *tout court*. More specifically, Kush et al. (2019) suggest that small or inconsistent island effects paired with bimodal judgment distributions should be taken as evidence that a particular domain was not a syntactic island, under the assumption that syntactic islands should categorically block A'-dependency formation. Under this interpretation our results (and theirs) imply at the very least that *om*-adjuncts are not syntactic islands in Norwegian. The variability observed with *når*-adjuncts could also be interpreted as evidence against *når*-adjuncts being syntactic islands.

### 3.1 Implications for syntactic approaches to adjunct islands

Neither the fact that extraction is ever judged acceptable from any adjuncts we tested or that there is substantial variation across adjunct types is predicted under any of the syntactic theories on adjunct islands that treat adjuncts as one uniform class of island domains (e.g. Huang's 1982 Condition on Extraction Domains, Chomsky's 1986 Barriers; Rizzi's 1990, 2004 Relativized Minimality, or the spell-out based approach of Uriagereka 1999, Nunes & Uriagereka 2000). If all of the adjuncts share the same structural feature (e.g. adjuncthood) that determines opacity for A'-dependencies, then differences are not predicted. To account for our findings within these frameworks would require a number of stipulations which have



little independent justification and which would weaken their appeal, which lies in their generality. For example, to be treated as non-islands, *om*- and *når*-adjuncts would have to be properly governed, or merged in such a way to avoid early spell-out, while *fordi*-adjuncts should not. Furthermore, to account for the variability, proper government or evading late spell-out would have to be optionally available for *om*- and *når*-adjuncts. It is not at all clear how such optionality could be formally implemented in a principled way.

Traditional approaches to adjunct clauses appear to be too coarse in their classification to account for our data. Syntactic analyses that allow for finer-grained distinctions could, in principle, fare better. If, for example, different adjunct interpretations corresponded to different attachment heights (e.g. Ernst 2002), a correlation between position and extractability might be tenable. Recently, C. Müller (2019) proposed an analysis of extraction from adjuncts in Swedish where the height of an adjunct's merge position determines its opacity to A'-movement (see also Truswell 2011). C. Müller adopts Haegeman's (2012) distinction between central and peripheral adjunct clauses and postulates that extraction is only allowed from central adjunct clauses that are adjoined low in the structure, at TP or  $\nu$ P (C. Müller 2019:42). The adjunct clauses we tested in our experiments are classified as central adjunct clauses according to Haegeman's (2012) and C. Müller's (2019) definitions: *om*-, *når*- and *fordi*-clauses can have both a central and a peripheral reading, but they are considered central adjunct clauses when they provide information about the condition for, the time of and the cause of the event expressed in the matrix clause, respectively (Haegeman 2012:161–164). The items in (12) below provide prototypical examples of items with respect to the classification of the type of adjunct clause:

(12) *Items as presented in (7)–(9) above, repeated*

a. *Om* 'if' test sentence

men takkekortene blir hun skuffet om de  
*but thank.you.cards.DEF becomes she disappointed if they*  
 glemmer å sende ut med en gang.  
*forget to send out with one time*  
 'but she will be disappointed if they forget to send out the thank you cards  
 right away.'

b. *Når* 'when' test sentence

men whisky blir han dårlig når han drikker.  
*but whisky becomes he sick when he drinks*  
 'but he feels sick when he drinks whisky.'

c. *Fordi* 'because' test sentence

Men vintertemperaturene blir hun boende fordi hun liker.  
*but winter.temperatures.DEF becomes she living because she likes*  
 'but she stays there because she likes the winter temperatures.'

The embedded *om*-clause in (12a) provides the condition for why the disappointment occurs. In (12b) the adjunct clause provides the time of the event expressed in the matrix VP. In (12c) the cause of 'the staying' is expressed by the *fordi*-clause. Insofar as they are all central adjuncts, the central versus peripheral distinction cannot be the ONLY relevant distinction for determining acceptability (if it is relevant at all).<sup>11</sup>



More generally, any proposal that automatically maps particular adjunct types to rigid attachment positions and uses attachment position as the sole determinant of acceptability of extraction would be hard-pressed to explain the inter- and intra-participant variation we see within individual adjunct types. Whatever the ultimate explanation for adjunct island effects is, it must account for variability by presumably allowing the precondition(s) for acceptable extraction to be variably assigned within an experimental setting.

### 3.2 *Extra-syntactic explanations*

We suspect that an account of adjunct island effects will have to take seriously semantic and discourse-pragmatic factors in order to provide an explanation of the fine-grained differences that we observe. Interpretive differences between the semantics of the different adjunct types (conditional, temporal, causal) could, for example, provide a foundation for differences between adjunct types. However, once again, semantic accounts would have to provide room for inter-trial variation, so the lexical semantics of the different complementizers cannot be the only factor determining acceptability of extraction. It seems more likely that the individual lexical semantics of the complementizers interact with semantic or pragmatic properties of the larger sentence. Under some frameworks, islandhood is tied to pragmatic focus or the foreground/background distinction (e.g. Erteschik-Shir 1973, Erteschik-Shir & Lappin 1979, Ambridge & Goldberg 2008). Within these frameworks, adjuncts would be non-islands insofar as they constitute the ‘main focus’, ‘informational center’, or insofar as their content was foregrounded. This status would be influenced by a number of different factors within the clause and interactions between various features would be expected. For example, differences in how often topicalization out of different adjunct types was accepted might reflect how easy the lexical semantics of the individual complementizers make it to adopt a pragmatically central/relevant reading of the adjunct.

Moreover, the differences that we observe between dependency types might also reflect differences in how easy it is to meet the relevant information structural conditions for extraction given the discourse function of different dependency types (see also Abeillé et al. 2020 for a similar idea). Kush et al. (2018, 2019) found that topicalization is more often judged acceptable than *wh*-movement from adjuncts: this could reflect that the (yet-to-be determined) conditions on acceptable extraction are harder to meet with *wh*-movement than with topicalization. We note that, insofar as pragmatic conditions are not expected to vary across languages, we would expect differences in adjunct island effects to vary by dependency type across languages. To some extent, this prediction is borne out: Sprouse et al. (2016) found a conditional adjunct island effect in a *wh*-dependency in English, but did not find one in a relative clause dependency.<sup>12</sup>

Erteschik-Shir & Lappin (1979) also propose that stress pattern and particularly relevant for our data, contrastive stress pattern, also influence the pragmatic focus of the sentence. They argue that extraction of an element is licit if it is contrastively paired and marked with a contrastive stress pattern with another element outside the embedded clause. Erteschik-Shir & Lappin’s (1979) account could provide an

explanation for why topicalization dependencies have been found to be accepted more often than *wh*-dependencies in Norwegian (see Kush et al. 2018, 2019). Applied to our data, all our test sentences in the *Long-distance* condition have contrastive topicalization, which means that the stress pattern must, in order for this account to work, interact with other features to allow extraction in some test sentences and not in others. It could perhaps also be the case that some of our items more felicitously than others encourage a contrastive reading between the preamble and the test sentence. We have not been able to identify any conditions or features that allow a contrastive reading to a larger or lesser extent in our test sentences. However, given the difference in judgments between the two *Long-distance* conditions, it is clear that the type of embedded clause influences acceptability to a greater extent than a contrastive stress pattern.

Truswell (2011) proposes a semantic condition in which extraction is possible if the event denoted by the embedded adjunct clause and the matrix clause can be construed as a single event grouping in the SINGLE EVENT GROUPING CONDITION:

(13) *The Single Event Grouping Condition* (SEGC)

An instance of *wh*-movement is legitimate only if the minimal constituent containing the head and the foot of the chain can be construed as describing a single event grouping.

(Truswell 2011:157)

A core assumption for this condition is that it only applies to non-finite adjunct clauses (Truswell 2011:118), as tensed adjunct clauses will force a two-event reading.<sup>13</sup> Nevertheless, we will dispose of this premise to consider whether the SEGC can account for some of the patterns in our data with finite adjunct clauses.

Truswell (2011:157) identifies the following conditions for a single event grouping (SEG):

- (i) spatiotemporal overlap between events denoted by matrix and embedded clause
- (ii) a maximum of one (maximal) event is agentive

Under this account, we would expect the distribution of SEG-items to roughly mirror the distribution of accepted items across adjunct type, such that *om* with the largest proportion of accepted items also would have the largest proportion of items with an SEG-reading. In fact, we do see slightly more items that, with the exception of tense, meet the criteria for being construed as an SEG in *når*- and *om*-items, compared to *fordi*-items. However, the proportion of SEG-items with *fordi* is much larger than the acceptability ratings for this adjunct type would predict.

Turning to the between-items variation, we see instances of accepted topicalization from both SEG items and non-SEG items within the same adjunct type. For example, in (14) we have one item with a single event grouping reading (14a)

and one where the most natural interpretation is arguably consistent with a multiple events reading (14b) (though see endnote 10).

(14) Når *test sentences*, Long-distance, Island condition

a. Single event grouping (item 10/26)

Preamble:

Håndballtreneren interesserer seg ikke spesielt i  
*handball.coach.DEF interest himself not particularly in*  
 fotballkampene på NRK  
*football-matches.DEF on NRK*

'The handball coach is not particularly interested in the football matches on NRK.'

Test sentence:

men håndballkampene på TV2 blir han ivrig når  
*but hand.ball.matches.DEF on TV2 becomes he eager when*  
 han ser.  
*he sees*

'but the handball matches on TV2 he becomes eager when he watches.'

b. Multiple events (item 48)

Preamble:

Sondre blir sur når de han bor med arrangerer fester, ...  
*Sondre becomes mad when they he lives with organize parties*

'Sondre gets mad when the people he lives with organizes parties, ...'

Test sentence:

men spillekvelder blir han glad når de arrangerer.  
*but gamenights becomes he glad when they organize*

'but gamenights he becomes happy when they organize.'

Both items received similar ratings (14a: mean rating  $z = 0.87$ , percentage of  $z > 0 = 85$ ; 14b: mean rating  $z = 0.68$ , percentage of  $z > 0 = 100$ ). The matrix and embedded clause in (14a) can be construed as a single event grouping as (i) the events overlap spatiotemporally – the activity of watching is occurring in the same space and at the same time as his interest rises; and (ii) only the embedded clause is agentive – the handball coach is deliberately watching the game, but not deliberately becoming interested in it. The reading of (14b) is ambiguous with regard to spatiotemporal overlap. The most obvious reading, when also taking into account the reading of the preamble, is one in which the item does not constitute a single event grouping as the events do not overlap spatiotemporally: the accomplishment ARRANGERE 'organize' does not occur at the same time as the change in mood.<sup>14</sup>

This implies that the patterns in our data do not match perfectly with what is predicted by the SEGC. Nevertheless, we do see that the majority of accepted items are SEG-items, particularly when we also consider items that are ambiguous with regard to spatiotemporal overlap as SEG-items. However, there is still a substantial number of SEG-items that are not accepted and a significant number of non-SEG-items that are accepted. This does not exclude the possibility that the SEGC is a precondition for extraction, but it implies that other features also interact with acceptability of extraction. Of the surface features we tracked, we could

not find any shared features/combination of features between the unaccepted SEG-items.

Truswell (2011:44) furthermore proposes that causation between the matrix and the embedded clause enables extractability, as it facilitates a single event reading. C. Müller (2019) supports this. It is interesting to note that there is a potential causative relationship between the matrix and embedded clause in all items that are accepted in our study, across adjunct type. However, this relationship alone is not enough to guarantee extraction as most items that are rejected also have a causation link between the matrix and embedded clause. Thus, it might be the case that causation is ONE prerequisite for extraction, but not the only one.

If relations like causation or SEG are interpretive preconditions on extraction, but those interpretations were not FORCED by our materials, then some variability in our data could be explained as a result of participants failing to adopt the appropriate interpretation on a given trial. Individual surface level features (e.g. tense, verb choice, plausibility, lexical semantics of individual complementizers or matrix predicates) – or their interactions – might also conspire to lead towards or away from causation readings or single event construal (Truswell 2011, Dal Farra 2020). As Truswell (2011:124) notes, participants may differ in the probability that they will construe events into a single event grouping depending on world-knowledge and creative ability to perceive a link between two events.

#### 4. Conclusion

Our experiments investigated the acceptability of contrastive topicalization dependencies from three adjunct types in Norwegian – *om* ‘if’, *når* ‘when’, and *fordi* ‘because’. Our results suggest that *om*-adjuncts are not categorical islands for A'-movement (replicating the findings of Kush et al. 2019). We found island effects for *når*-adjuncts, but we reasoned, on the basis of judgment distributions, that these effects were also incompatible with a strict ban on movement from structural adjuncts. Participants largely rejected topicalization from *fordi*-adjuncts, suggesting variation in island effects between adjunct type. The large variation within each adjunct type implies that ‘adjunct’ is not a uniform group in relation to island extraction, as it has previously been treated. We also uncovered great inter-item variation, which we think implies that there are extra-syntactic conditions that govern the extraction from these adjunct clauses, as no known syntactic account can explain the variation seen in our experiments. Current extra-syntactic explanations for extraction from adjunct clauses can not, however, straightforwardly explain the pattern found for extraction from Norwegian adjunct clauses and should be addressed in future work.

**Acknowledgements.** The data presented in this paper was collected for and first published in Bondevik (MA thesis, NTNU, 2018). An early version of the paper was presented at MONS, 2019. We want to thank three anonymous *NJL* reviewers for feedback, comments and questions, as well as valuable insights for future work.

## Notes

1 Though see Pearl & Sprouse (2013) and Phillips (2013b) for a relatively recent discussion of this issue.  
 2 Infinitival adjuncts provide an apparent exception to the CED. Truswell (2007) and Szabolcsi & Lohndal (2017) show that untensed, gerundive adjunct clauses allow extraction in English in certain contexts:

(i) a. Which topic did you leave [without talking about \_]?

(Szabolcsi & Lohndal 2017:4)

b. What did John arrive [whistling \_]?

c. What did John drive Mary crazy [trying to fix \_]?

(Truswell 2007:1356)

3 Interestingly, participants rated extraction out of complex NPs and relative clauses as just as unacceptable as extraction out of subjects and adjuncts, despite the fact that these constituents have been argued not to be islands in MSc (e.g. Allwood 1982, Maling & Zaenen 1982, Engdahl 1997). We return to this point later.

4 C. Müller (2019) also reports variable acceptability of extraction from finite adjunct clauses in Swedish. The same is reported for Italian in Dal Farra (2020).

5 The DD-score is a measure of the size of the island effect. It measures the residual acceptability difference between the baseline condition (*Short-distance, no-Island*) and 'island violation' condition (*Long-distance, Island*), after the main effects of structure and distance have been subtracted away. The DD-score can be calculated in three steps: (i) calculating the difference between the long conditions [ $D1 = \text{Long-distance, no-Island} - \text{Long-distance, Island}$ ]; (ii) calculating the difference between the short conditions [ $D2 = \text{Short-distance, no-Island} - \text{Short-distance, Island}$ ]; and (iii) calculating the difference between the difference scores [ $D1 - D2$ ]. For example, the size of the island effect in Figure 1B is  $DD = 0.5 ((0.3 - (-0.9)) - (0.8 - 0.1) = 0.5)$ . Larger (positive) DD-scores indicate that an island violation incurs a larger acceptability cost. When there is no interaction effect, the size of the island effect is (close to) zero. See e.g. Sprouse et al. (2012) and Sprouse et al. (2016) for a more detailed explanation of DD-scores in relation to the factorial definition of islands.

6 A full overview of all our test material can be found at our OSF project site at <https://osf.io/6tx3n/>.

7 Raw ratings were z-score transformed by participant to standardize the response variable and control for scale bias across participants (Sprouse et al. 2011; Sprouse et al. 2012; Sprouse et al. 2016; Kush et al. 2018, 2019). By-participant z-scores are calculated as follows: each rating value is centered by subtracting the participant's average rating. Each rating is then divided by the participant's standard deviation. The resulting z-score is a standardized score that quantifies how many standard deviations from a participant's mean a given rating is. Z-scoring enables us to compare relative differences across participants on a standardized scale.

8 A Friedman test comparing all three distributions showed a significant difference ( $p < .001$ ). Post-hoc Wilcoxon signed-rank tests showed that all pairwise differences were significantly different between adjunct types ( $p < .05$ ).

9 Interestingly, we could not find any patterns in the self-reported dialectal background data that could readily account for the differences between participants.

10 Forty unacceptable fillers + 16 unacceptable test sentences (5 *fordi*, 5 *om*, 4 subject and 2 'whether') = 56 unacceptable test sentences. Four acceptable fillers + 48 acceptable test sentences (15 *fordi*, 15 *om*, 12 subject, 6 'whether') = 52 acceptable sentences. Given our results in Experiment 1 counting extractions from 'whether'-clauses as acceptable leads to a total of 54 acceptable sentences in the experiment.

11 C. Müller (2019) encounters the same problem in her Swedish data and proposes that central adjunct clauses can have the internal syntax of peripheral clauses and thus disallow extraction. Haegeman (2012:182) separates between central clauses with a peripheral or central internal syntax by showing that the former do not allow *it*-clefting. All items in our experiment with extraction from the mostly rejected *fordi*-adjuncts pass the tests as central adjunct clauses both internally and externally. Thus, this proposal cannot account for the variation between adjunct types in our data.

12 All items in our study have contrastive topicalization. An interesting pattern seen in our data is that items where there is a very close semantic relationship between the objects of contrast, are more often accepted. Two *fordi*-items with *glemme* 'forget' as the embedded verb where, with the exception of definiteness of the moved constituent, all other variables were the same (agentivity, event grouping, tense, number of words between filler and gap), received very different ratings: item 7 contrasting *te* 'tea' and *kaffe* 'coffee'

was accepted by approximately 60% of participants in Experiment 2a, whereas item 9 contrasting *ananas* ‘pineapple’ and *eple* ‘apple’ was accepted only by approximately 14% of participants in the same experiment. Similarly, an *om*-item with *se* ‘see’ as the embedded verb contrasting *filmer* ‘movies’ and *dokumentarer* ‘documentaries’ was mostly rejected by participants, whereas a *når*-item with the same embedded verb *se* ‘see’ contrasting  *fotballkamper på NRK* ‘football matches on NRK’ and  *håndballkamper på TV2* ‘handball matches on TV2’ was mainly accepted by participants.

13 The explanation being that the tense operator *Op* will block extraction from tensed adjunct clauses as it will force a two-event reading.

14 A different reading of (14b) can be a scenario in which Sondre sees his roommates sitting in the living room planning a game night, and this is the event that makes him happy. In this reading, both events overlap spatiotemporally and as only one clause is agentive, it can be construed as a Single Event Grouping. For most, though, the event of organizing a gamenight is preceded by the actual happening. And if you are familiar with the differences between parties (loud and disorderly) and gamenights (calmer and well-organized), the most likely reading is that Sondre becomes happy when he comes home to find that his roommates have already organized the gamenight and are now (relatively) quietly having fun in the living room. As we did not control for interpretation of events, we cannot be certain how this sentence was interpreted and we cannot exclude the possibility that participants understood these events to overlap spatiotemporally.

## References

- Abeillé, Anne, Barbara Hemforth, Elodie Winckel & Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204, <https://doi.org/10.1016/j.cognition.2020.104293>.
- Allwood, Jens. 1982. The Complex NP Constraint in Swedish. In Engdahl & Ejerhed (eds.), 15–32.
- Ambridge, Ben & Adele E. Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19(3), 349–381.
- Anward, Jan. 1982. Basic Swedish. In Engdahl & Ejerhed (eds.), 47–75.
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. doi: 10.18637/jss.v067.i01.
- Bermingrud, Knut Otto. 1979. Setningsknutekonstruksjonen: en analyse av konstruksjonens grammatikalitet i moderne norsk [The island construction: An analysis of the grammaticality of the construction in Modern Norwegian]. Masters thesis, University of Oslo.
- Boeckx, Cedric. 2003. *Islands and Chains*. Amsterdam: John Benjamins.
- Boeckx, Cedric. 2012. *Syntactic Islands*. Cambridge: Cambridge University Press.
- Chomsky, Noam. 1964. Current issues in linguistic theory. In Jerold J. Katz & Jerry A. Fodor (eds.), *The Structure of Language: Readings in the Philosophy of Language*, 50–118. Englewood Cliffs, NJ: Prentice-Hall.
- Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson, Paul Kiparsky & Morris Halle (eds.), *A Festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart and Winston.
- Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monographs 13). Cambridge, MA: MIT Press.
- Dal Farra, Chiara. 2020. *To Be or not to Be an Island: The Status of Adjuncts*. Ph.D. dissertation, Ca’ Foscari University of Venice.
- Dikken, Marcel den & Antje Lahne. 2013. The locality of syntactic dependencies. In Marcel den Dikken (ed.), *The Cambridge Handbook of Generative Syntax*, 655–698. Cambridge: Cambridge University Press.
- Drummond, Alex. 2012. IbxFarm. Retrieved from <http://spellout.net/ibxfarm/> (February 2018).
- Engdahl, Elisabet. 1997. Relative clause extractions in context. *Working Papers in Scandinavian Syntax* 60, 51–79.
- Engdahl, Elisabet & Eva Ejerhed (eds.). 1982. *Readings on Unbounded Dependencies in Scandinavian Languages*. Stockholm: Almqvist & Wiksell.
- Ernst, Thomas. 2002. *The Syntax of Adjuncts*. Cambridge: Cambridge University Press.
- Erteschik-Shir, Nomi. 1973. *On the Nature of Island Constraints*. Ph.D. dissertation, MIT.
- Erteschik-Shir, Nomi & Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6(1–3), 41–86.
- Faarlund, Jan T. 1992. *Norsk syntaks i et funksjonelt perspektiv* [Norwegian syntax from a functional perspective]. Oslo: Universitetsforlaget.

- Haegeman, Liliane.** 2012. *Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery*. New York: Oxford University Press.
- Hofmeister, Philip, Laura S. Casanto & Ivan A. Sag.** 2013. Islands in the grammar? Standards of evidence. In Sprouse & Hornstein (eds.), 64–108.
- Huang, C.-T. James.** 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. dissertation, MIT.
- Hunter, Tim.** 2015. Deconstructing Merge and Move to make room for adjunction. *Syntax* 18, 266–319.
- Kush, Dave & Anne Dahl.** L2 Transfer of L1 Island-insensitivity: The case of Norwegian. *Second Language Research*, doi:10.1177/0267658320956704. Published online on 15 September 2020 by Sage.
- Kush, Dave, Terje Lohndal & Jon Sprouse.** 2018. Investigating variation in island effects. *Natural Language & Linguistic Theory* 36, 743–779.
- Kush, Dave, Terje Lohndal & Jon Sprouse.** 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95, 393–420.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen.** 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13), doi: 10.18637/jss.v082.i13.
- Lasnik, Howard & Mamoru Saito.** 1992. *Move Alpha: Conditions on its Application and Output*. Cambridge, MA: MIT Press.
- Maling, Joan & Annie Zaenen.** 1982. A phrase structure account of Scandinavian extraction phenomena. In Pauline Jacobson & Geoffrey K. Pullum (eds.), *The Nature of Syntactic Representation. Synthese Language Library* (Texts and Studies in Linguistics and Philosophy 15), 229–282. Dordrecht: Reidel.
- Manzini, Rita M.** 1992. *Locality: A Theory and some of its Empirical Consequences* (Linguistic Inquiry Monographs 19). Cambridge, MA: MIT Press.
- Maxwell, Scott E. & Harold D. Delaney.** 2003. *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. London: Routledge.
- Müller, Christiane.** 2019. *Permeable Islands*. Ph.D. dissertation, Lund University.
- Müller, Gereon.** 2011. *Constraints on Displacement*. Amsterdam: John Benjamins.
- Nunes, Jairo & Juan Uriagereka.** 2000. Cyclicity and extraction domains. *Syntax* 3(1), 20–43.
- Nyvad, Anne M., Ken R. Christensen & Sten Vikner.** 2017. CP-recursion in Danish: A cP/CP-analysis. *The Linguistic Review* 34(3), 449–477.
- Pearl, Lisa. & Jon Sprouse.** 2013. Computational models of acquisition for islands. In Sprouse & Hornstein (eds.), 109–131.
- Phillips, Colin.** 2013a. On the nature of island constraints I: Language processing and reductionist accounts. In Sprouse & Hornstein (eds.), 64–108.
- Phillips, Colin.** 2013b. On the nature of island constraints II: Language learning and innateness. In Sprouse & Hornstein (eds.), 132–158.
- R Core Team.** 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> (7 November 2020).
- Rizzi, Luigi.** 1990. *Relativized Minimality* (Linguistic Inquiry Monographs 16). Cambridge, MA: MIT Press.
- Rizzi, Luigi.** 2004. Locality and left periphery. In Adriana Belletti (ed.), *Structures and Beyond*, 223–251. Oxford: Oxford University Press.
- Roberts, Ian G.** 1997. *Comparative Syntax*. London: Arnold.
- Ross, John R.** 1967. *Constraints on Variables in Syntax*. Ph.D. dissertation, MIT.
- Sprouse, Jon.** 2007. *A Program for Experimental Syntax*. Ph.D. dissertation, University of Maryland.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco & Carlo Cecchetto.** 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34, 307–344.
- Sprouse, Jon, Shin Fukuda, Hajime Ono & Robert Kluender.** 2011. Reverse Island Effects and the Backward Search for a Licensor in multiple *wh*-questions. *Syntax* 14(2), 179–203.
- Sprouse, Jon & Norbert Hornstein.** 2013a. Experimental syntax and island effects: Toward a comprehensive theory of islands. In Sprouse & Hornstein (eds.), 1–18.
- Sprouse, Jon & Norbert Hornstein** (eds.). 2013b. *Experimental Syntax and Island Effects*. Cambridge: Cambridge University Press.
- Sprouse, Jon, Matt Wagers & Colin Phillips.** 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1), 82–123.
- Stepanov, Arthur.** 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1), 80–126.



- Szabolcsi, Anna & Terje Lohndal.** 2017. Strong vs. weak islands. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax*, 2nd edn. 1–51. Malden, MA: Wiley Blackwell.
- Truswell, Robert.** 2007. Extraction from adjuncts and the structure of events. *Lingua* 117, 1355–1377.
- Truswell, Robert.** 2011. *Events, Phrases, and Questions*. Oxford: Oxford University Press.
- Uriagereka, Juan.** 1999. Multiple spell-out. In Samuel D. Epstein & Norbert Hornstein (eds.), *Working Minimalism*, 251–282. Cambridge, MA: MIT Press.
- Uriagereka, Juan.** 2012. *Spell-Out and the Minimalist Program*. Oxford: Oxford University Press.
- Wickham, Hadley.** 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.